DON'T SHOOT THE MESSENGER

Non-coding RNAs and RNA-binding proteins in control of gene expression

Marieke van Kouwenhove

The research described in this thesis was performed at the division of Gene Regulation of the Netherlands Cancer Institute – Antoni van Leeuwenhoekzoekenhuis (NKI-AVL) in Amsterdam, The Netherlands, with financial support from the Dutch Cancer foundation – Koningin Wilhelmina Fonds (KWF), European Research Council (ERC), Horizon – NWO (Nederlandse organisatie voor Wetenschappelijk Onderzoek) to R. Agami.

Cover: M.C. Escher's "Place Filling II" © 2012 The M.C. Escher Company B.V. - Baarn - Holland. All rights reserved. www.mcescher.com. Author's note: the place in this image can be interpreted as the surface of the earth, on which mutual adaptation and adjustment may have resulted in the survival of individual species. Also, the form of optical illusion demonstrates that while seeing everything in front of you, it can be difficult to see the complete picture.

Lay-out and printing: Off Page, www.offpage.nl

ISBN: 978-94-6182-139-3

The printing of this book was financially supported by the Netherlands Cancer Institute, KWF, and Erasmus University Rotterdam.

DON'T SHOOT THE MESSENGER

Non-coding RNAs and RNA-binding proteins in control of gene expression

Niet-coderende RNAs en RNA-bindende eiwitten beheersen genexpressie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.G. Schmidt
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 26 oktober 2012 om 11.30 uur.

door

Marieke van Kouwenhove
geboren te Stadskanaal

Promotiecommissie

Promotor:	Prof.dr. R. Agami
Overige leden:	Prof.dr. L.H.J. Looijenga Prof.dr. R. Fodde Prof.dr. F.G. Grosveld

CONTENTS

ABBREVIATIONS

aa	amino acid
bp	basepair
CH	choriocarcinoma
ChIP	chromatin immunoprecipitation
CHX	cycloheximide
chr	chromosome
CIS	carcinoma *in situ*
CpG	-C-phosphate-G-
dpc	days post coitus
ds	double stranded
EC	embryonal carcinoma
EMSA	electromobility shift assay
ESC	embryonic stem cell
FL	full length
fw	forward
GCT	germ cell tumour
Hsa	*Homo sapiens*
IPA	intronically polyadenylated
iPS	induced pluripotent stem
kb	kilobase
kDa	kilodalton
miR	microRNA
Mmu	*Mus musculus*
NS	non-seminoma
nt	nucleotide
ORF	open reading frame
PAS	polyadenylation signal
PGC	primordial germ cell
PRE	Pumilio recognition element
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RBD	RNA-binding domain
RBP	RNA-binding protein
RISC	RNA-induced silencing complex
RPA	RNase protection assay
RRM	RNA recognition motif
rv	reverse
SAM	significance analysis of microarrays
SD	standard deviation
SE	seminoma
SEM	standard error of the mean
SG	stress granules
ss	single stranded
TC	teratocarcinoma
TE	normal testis

TF	transcription factor
TSS	transcription start site
URR	uridine-rich region
UTR	untranslated region
WB	Western blot
XX	female sex chromosomes
XY	male sex chromosomes
YS	yolk sac tumour

chapter ONE

GENERAL INTRODUCTION AND OUTLINE OF THE THESIS

INTRODUCTION

Every human life starts with the fusion of an egg cell and a sperm cell, whereby the DNA content of both is combined into one cell and the blueprint of the human body is established. The DNA of this cell contains all information necessary to form a complete organism. During the development of the embryo, every newly formed cell will either continue to divide thereby contributing to expansion, or will specialize to form certain tissues or to carry out particular functions. The latter process is termed differentiation and underlies the development of all different cell types and tissue types present in the body. Together, the cells in a multicellular organism collaborate to support the germ cells: the cells that will form eggs or sperm to generate a next generation being the ultimate goal. In principle, a human body is a system consisting of many (10^{14}, or one hundred thousand billion) cells that are grouped in specialized tissues and organs. Still, each somatic cell contains the complete DNA package, which is replicated during every cell cycle and is passed on to both daughter cells at cell division.

Cell division is central to renewal of cell types with a limited life span such as blood and skin cells. It is estimated that 200 billion (10^9) new cells are generated in a human body every day, which equals approximately 2,500,000 new cells per second. However, most cells do not multiply, but are held in an inactive state called replicative senescence which is irreversible under regular conditions. Cells can also enter a state of reversible growth arrest referred to as quiescence, when the cell cycle is interrupted due to the absence of growth factors that stimulate proliferation. The cell cycle is a tightly controlled process; it pauses to check for damage to a cell's DNA content. One can imagine that certain damage arises every day, most of which is detected, repaired and rendered harmless. However, cells can experience damage to their genetic content that enables evasion from senescence and these cells continue or resume cell division. As a result, the stability of the genome is further compromised by accumulation of DNA damage, allowing unrepaired cells to proliferate in an uncontrolled manner.

At the G1/S checkpoint, it is decided whether a cell continues to cycle. In case of DNA damage, the cycle can be interrupted for repair, or the cell transiently or permanently enters are sting stage, or proceeds to apoptosis (cell death) in case of extensive damage. Generally, the tumour suppressor protein p53 plays the most critical role in this decision. In approximately half of solid cancers, p53 is mutated and functions aberrantly as a result. In many other tumours with intact p53, components downstream of p53 are mutated or dysfunctional thereby neutralizing the p53 pathway without affecting the p53 gene itself.

The DNA content is virtually identical in every cell, however, a cell's specific function is orchestrated by simultaneous expression of individual genes. The expression of genes is tightly coordinated in time and space and involves a variety of genetic and epigenetic regulators. One can imagine that small deviations occur from cell to cell, but a gain or loss of one regulator will affect multiple target genes. In addition, the expression of a gene can be manipulated not only at the level of transcription, but also thereafter by mRNA splicing, polyadenylation, translational control or by other processes. Furthermore, post-translational modifications can

also direct the function of a protein. In a healthy environment, these processes are orchestrated such that an organism can survive and reproduce. However, genetic alterations or mutations that disturb normal cell proliferation or homeostasis can lead to the progressive conversion of normal cells into cancer cells. In their latest paper[1], Weinberg and Hanahan added two emerging hallmarks to the earlier defined ones for cancer cells: crucial alterations in cellular physiology that are characteristic for malignancy. These include 1) insensitivity to anti-growth signals, 2) self-sufficiency in growth signals, 3) limitless replicative potential, 4) escape from apoptosis, 5) blood vessel formation for oxygen and nutrient supply (angiogenesis), 6) invasion of nearby tissue or distant metastasis, supplemented with 7) adaptation of energy metabolism to support continue growth and 8) evasion from attack by immune cells. Acquisition of these capabilities enables malignant transformation and allows cancer cells to proliferate, metastasize and invade other tissues.

RNA AND ITS ROLE IN LIFE

The number of defined protein-coding genes in the human genome is estimated to be between 20,000 and 25,000, which in total make up only 1-2% of the human genome. Since it became possible to sequence complete transcriptomes by relatively modern sequencing techniques, we know that more than 90% of the human genome is actually expressed in RNA[2,3] and that much more sequence is evolutionarily conserved across species than can be explained by protein-coding regions[4,5]. Non-protein-coding DNA not just includes introns or untranslated regions (UTRs) related to protein-coding genes, but also independently transcribed regions. Based on these notions, non-protein-coding DNA was no longer considered "junk" DNA but became recognized as active part of the genome. Moreover, whereas the absolute amount of protein-coding DNA does not substantially increase with complexity of a species, the amount of non-protein-coding DNA, or non-coding RNA, does[6]. Also, the average length of mRNA 3′ UTRs increases with developmental complexity[7]. These discoveries, just a decade ago, laid the basis for the discovery of new RNA species and encouraged novel conceptual thinking about the role of RNA in life.

RNA molecules have a wide range of functions beyond their role as messengers from DNA to protein. RNA is known to function in amino acid transfer (tRNA), peptide bond formation (rRNA) and catalyse a variety of enzymatic reactions (snRNA)[8]. RNA is also proficient in catalysing RNA replication[9]. RNA can both be catalytically active as so-called ribozymes, and carry and mobilise genetic information; the combination of these capacities makes this molecule unique. Moreover, this notion supports the possibility for pre-cellular evolution at the RNA level in the absence of proteins or DNA; a functional biomolecule at the origin of life[10,11].

RNA AND ITS STRUCTURE

RNA is the transportable copy of DNA present in the nucleus and also in cytoplasm. The primary structure of RNA is expressed in ribonucleotide sequence from the 5′ to the 3′ end, with lengths varying from around twenty to thousands of

nucleotides. Single-stranded RNA copies are less stable than the corresponding double-stranded DNA, due to the protective nature of both strandedness and to chemical properties of the different phosphate backbones with regard to protection against hydrolysis. RNA molecules were found to form base pairs in a similar way as assessed for the double-stranded helical structure of DNA[12]. A decade later, the occurrence of energetically weaker Wobble pairs was demonstrated[13], allowing guanine (G) to pair with uracil (U) and leading to the possible C:G, A:U and G:U base pairs. In single-stranded molecules, base-pairing implies that the molecule folds complementary regions to establish intramolecular interactions. RNA transcripts adapt thermodynamically stable conformations for function, plasticity, or protection[14]. Of the many possible conformations, only few are presumed to be functional and non-toxic[15,16]. The two-dimensional RNA structure can dynamically change over time or upon binding of regulative factors with or without function conservation[17]. Only little is known about the three-dimensional (tertiary) structure of RNA and its contribution to functionality. It appears that RNA dimensional structure is more conserved than RNA sequence; this fact is useful for comparative RNA structure prediction in a way similar to that is done for proteins[18]. Another important feature of RNA is that most species are part of RNA-protein (RNP) complexes wherein the protein components assist RNA folding and function[19,20].

RNA AND ITS SMALL SPECIES

In addition to the already mentioned functional class of mRNA, several additional ones have been characterized over the last years. One important class is formed by microRNAs (miRNAs), the smallest RNA species in length that has been described so far[21]. MiRNAs fine-tune gene expression by strong or subtle inhibition of mRNA translation through binding to specific sites in the 3′ UTR of the target mRNA. The biogenesis pathway of miRNAs involves the principles of gene transcription mediated by RNA polymerase II, nuclear export, cytoplasmic processing and RNA degradation. The process of miRNA gene transcription is dependent on principal mechanisms of transcription activation, such as transcription factor recruitment. Whereas active transcription is imperative for the presence of mature miRNA in a cell, the levels of mature miRNA and their actual effect on mRNA translation cannot directly be deduced from their transcription rate. It has become apparent that miRNA regulators themselves are controlled by RNA-binding proteins (RBPs) (chapter 2). First, processing of a miRNA transcript into a mature miRNA is modulated by RBPs, either generally or in a sequence-specific or structure-specific manner affecting a particular subset of miRNAs. Second, the function of mature miRNAs is modulated by RBPs that either interact with the miRNA incorporated in the RNA-induced silencing complex (miRISC) or bind the 3′ UTR of the target mRNA to enhance or alleviate miRISC interaction with its target. Many factors participating in miRNA biogenesis or in miRNA-mediated gene repression have been identified so far, some of which with a link to cancer.

RNA AND ITS APPEARANCE IN CANCER

A general phenomenon in cancer cells is the global down regulation of miRNA levels. This is possibly due to aberrant functioning of a universal regulator of miRNA biogenesis or function. Also, deregulated expression of specific clusters or families of miRNAs is found in certain tumour types. A causal relationship has not been proved for all of these miRNAs, but some have been demonstrated to induce or favour cancer development as so-called oncomiRs[22]. For other miRNAs with tumour-suppressive effects, their expression or effect is repressed in cancer. Some miRNAs can have either an oncogenic or tumour-suppressive effect dependent on the cellular context. The nature of miRNAs results from the elicited effect upon translational repression of their mRNA targets. Also, mRNA expression patterns differ globally between normal cells and tumours; this knowledge has been applied more and more for clinical purposes.

ALTERNATIVE POLYADENYLATION

Following the production of mRNA by transcription, multiple events must occur before functional proteins can be formed by translation. Initially, most pre-mRNAs are subjected to splicing to remove introns. Addition of the poly(A) tail upon binding of the polyadenylation and cleavage machinery is essential for proper mRNA processing and facilitates efficient mRNA export, translation, and stability[23]. The resulting mRNA is then exported from the nucleus into the cytoplasm for translation. In the cytoplasm, mRNA abundance is regulated by stability, which is mostly influenced by binding of post-transcriptional regulators to the 5′ and 3′ UTRs. About half of the mammalian genes contain, in addition to a polyadenylation signal (PAS) at the distal end of the 3′ UTR, alternative PASs in the 3′ UTR or in the coding region[24]. Usage of alternative PASs in the 3′ UTR, as observed in proliferating and less differentiated cells, generally increases levels of the target mRNA and protein[25]. An escape from 3′ UTR-mediated destabilization by miRNAs and RBPs could underlie this outcome. Alternative polyadenylation in the coding region interferes with assembly of the full-length protein, but functional consequences of this phenomenon have only been sparsely demonstrated[26].

MIRNAS IN GERM CELL TUMOURS

During human embryonic development, embryonic stem cells (ESCs) start to differentiate into primordial germ cells (PGCs); both these cell types are characterized by the expression of specific pluripotency markers and of miRNAs such as the miR-302-367 and miR-371-373 clusters. ESCs and PGCs can give rise to germ cell tumours (GCTs) of type I. Once the PGCs have migrated to the genital ridge, i.e. the future gonads, these cells are referred to as gonocytes. Late PGCs or early gonocytes can arrest in differentiation, thereby forming a carcinoma *in situ* (CIS) or intratubular germ cell neoplasia unclassified (IGCNU) lesion, which is a typical precursor of type II GCT. CIS lesions can occur in the first few years of life and progress after a latent state

to type II GCTs, which are subdivided in seminomas and nonseminomas. According to the current understanding, proliferation of CIS will result in seminoma whereas progression of CIS into embryonic carcinoma (EC) cells, i.e. the malignant counterpart of PGCs, characterises development of nonseminoma[27]. Non-seminomas can contain EC cells mixed with various components of more differentiated carcinoma tissue. Corresponding to the migration path of PGCs, type I and II GCTs can originate along the total body axis. Eventually, gonocytes differentiate to sustain gametogenesis, either through oogenesis or spermatogenesis. More differentiated primary spermatocytes can give rise to type III GCT in elderly men. Testicular GCTs are the most common cancer in males aged 15-35 years. Options for treatment are dependent on tumour type and stage, but are overall very effective with an average cure rate of more than 80%[28]. Testicular GCTs are correlated with cryptorchidism and reduced fertility, but underlying genetic defects or environmental factors have not been clearly linked.

REPROGRAMMING

Due to their potential to generate various types of specialized tissue, human ES cells are of great interest for clinical applications. Nonetheless, ethical issues surround the derivation of human ES cells from *in vitro* fertilized blastocysts. A recently developed alternative approach is cellular reprogramming, which is the process of transforming fully differentiated, specialized cells into a less differentiated, pluripotent cell type. Cellular reprogramming has been achieved using a variety of methods, including somatic cell nuclear transfer, cell-cell fusion and, most recently, through the introduction of three or more transcription factors (Chapter 3, table S1). Retroviral expression of the transcription factors Sox2, Oct3/4 and Klf4 has been shown to be minimally required for induction of pluripotent stem (iPS) cells from mouse and human fibroblast cells. Since this involves induction of a less differentiated cellular program, there might be links to dedifferentiation process in germ cell tumorigenesis.

OUTLINE OF THE THESIS

In eukaryotes, multiple variables are involved in the expression of genes into functional proteins. Clearly, transcription of DNA and translation of mRNA are important steps, but regulatory processes that direct mRNA processing, localization or stability are equally important determinants of outcome. We aim to unravel post-transcriptional mechanisms controlling the abundance of cellular proteins, predominantly by investigating the role of specific non-coding RNAs and RBPs towards the mRNA 3' UTR.

In **chapter 2**, we outline the biogenesis pathway of miRNAs, and we explain how deregulation of RBPs with a central role in miRNA processing can contribute to cancer. In the second part of the chapter, the mechanism of miRNA-mediated repression is summarized. Also, we illustrate various ways of interplay between miRNAs and RBPs that determine post-transcriptional regulation of target mRNAs in cellular processes with a link to cancer. In **chapter 4**, the specific mechanism of interplay between RBP Pumilio and miRNAs 221/222 is demonstrated. By doing so, we highlight in

this chapter the importance of mRNA secondary structure in post-transcriptional regulation, focusing on cell cycle control. Also, we provide an example of regulation of RBP activity with regard to target binding under specific cellular conditions.

In **chapter 3 and 5**, we focus on post-transcriptional control of gene expression in the germ line. The germ line originates from PGCs, which are vital for continuity of a species. During embryonic development, hampered migration and differentiation of PGCs may result in a pre-malignant lesion. In **chapter 3**, we investigate regulation of oncogenic miRNAs 371-373 which are found expressed in clinical germ cell tumours. Both genetic and epigenetic factors that sustain transcription of these miRNAs can be relevant in the context of germ cell tumorigenesis. Further, we discuss parallels between embryonic development and cellular reprogramming with regard to the role of pluripotency-related genes. In **chapter 5**, we examine the functionality of a truncated protein form of Dnd1, an RBP important for proper gene expression in the germ line. We investigate whether RBP expression can be modulated at the level of mRNA processing and whether this phenomenon is applicable to clinical GCTs. Further, we present the effect of a RBP with changed conformation on mRNA target regulation.

Chapter 6 is a general discussion on the findings presented in this thesis.

REFERENCES

1. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
2. Mattick, J.S. & Gagen, M.J. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Molecular biology and evolution* **18**, 1611-30 (2001).
3. Bertone, P. et al. Global identification of human transcribed sequences with genome tiling arrays. *Science (New York, NY)* **306**, 2242-6 (2004).
4. Venter, J.C. et al. The sequence of the human genome. *Science (New York, NY)* **291**, 1304-51 (2001).
5. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science (New York, NY)* **304**, 1321-5 (2004).
6. Mattick, J.S. RNA regulation: a new genetics? *Nature reviews Genetics* **5**, 316-23 (2004).
7. Frith, M.C., Pheasant, M. & Mattick, J.S. The amazing complexity of the human transcriptome. *European journal of human genetics : EJHG* **13**, 894-7 (2005).
8. Noller, H.F., Hoffarth, V. & Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science (New York, NY)* **256**, 1416-9 (1992).
9. Cech, T.R. A model for the RNA-catalyzed replication of RNA. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 4360-3 (1986).
10. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618-618 (1986).
11. Joyce, G.F. RNA evolution and the origins of life. *Nature* **338**, 217-24 (1989).
12. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
13. Holley, R.W. et al. STRUCTURE OF A RIBONUCLEIC ACID. *Science (New York, NY)* **147**, 1462-5 (1965).
14. Hofacker, I.L. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie - Chemical Monthly* **125**, 167-188 (2010).
15. Shapiro, B.A. An algorithm for comparing multiple RNA secondary structures. *Computer applications in the biosciences : CABIOS* **4**, 387-93 (1988).
16. Herschlag, D. RNA chaperones and the RNA folding problem. *The Journal of biological chemistry* **270**, 20871-4 (1995).

17. Al-Hashimi, H.M. & Walter, N.G. RNA dynamics: it is about time. *Current opinion in structural biology* **18**, 321-9 (2008).
18. Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. & Chang, H.Y. Understanding the transcriptome through RNA structure. *Nature reviews Genetics* **12**, 641-55 (2011).
19. Lorsch, J.R. RNA chaperones exist and DEAD box proteins get a life. *Cell* **109**, 797-800 (2002).
20. Mansfield, K.D. & Keene, J.D. The ribonome: a dominant force in co-ordinating gene expression. *Biology of the cell / under the auspices of the European Cell Biology Organization* **101**, 169-81 (2009).
21. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
22. Esquela-Kerscher, A. & Slack, F.J. Oncomirs - microRNAs with a role in cancer. *Nature reviews Cancer* **6**, 259-69 (2006).
23. Licatalosi, D.D. & Darnell, R.B. RNA processing and its regulation: global insights into biological networks. *Nature reviews Genetics* **11**, 75-87 (2010).
24. Tian, B., Hu, J., Zhang, H. & Lutz, C.S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic acids research* **33**, 201-12 (2005).
25. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. & Burge, C.B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science (New York, NY)* **320**, 1643-7 (2008).
26. Yao, P. et al. Coding Region Polyadenylation Generates a Truncated tRNA Synthetase that Counters Translation Repression. *Cell* **149**, 88-100 (2012).
27. Honecker, F. et al. New insights into the pathology and molecular biology of human germ cell tumors. *World journal of urology* **22**, 15-24 (2004).
28. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **15**, 594-603 (1997).

chapter **TWO**

MICRORNA REGULATION BY RNA-BINDING PROTEINS AND ITS IMPLICATIONS FOR CANCER

Marieke van Kouwenhove[1], Martijn Kedde[1] and Reuven Agami[1]‡

[1]Division of Gene Regulation, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, Amsterdam,The Netherlands.
‡The Centre for Biomedical Genetics, UMCU, Stratenum 3.217, Universiteitsweg 100, 3584CG Utrecht, The Netherlands.

ABSTRACT

Non-protein-coding transcripts have been conserved throughout evolution, indicating that crucial functions exist for these RNAs. For example, microRNAs (miRNAs) have been found to modulate most cellular processes. The protein classes of RNA-binding proteins include essential regulators of miRNA biogenesis, turnover and activity. RNA-RNA and protein-RNA interactions are essential for post-transcriptional regulation in normal development and may be deregulated in disease. In reviewing emerging concepts of the interplay between miRNAs and RNA-binding proteins, we highlight the implications of these complex layers of regulation in cancer initiation and progression.

MicroRNAs (miRNAs) are small, non-coding RNAs that repress gene expression through interaction with 3′ untranslated regions (3′ UTRs) of mRNAs (reviewed in REF. 1). MiRNAs are predicted to target over 50% of all human protein-coding genes, enabling them to have numerous regulatory roles in many physiological and developmental processes[2]. Global downregulation of miRNA expression is an emerging feature in cancer, and the specific deregulation of certain miRNAs is seen in specific tumour types[3,4] (BOX 1). RNA-binding proteins (RBPs) are key components in the determination of miRNA function, as they control different stages of miRNA biogenesis and their localization, degradation and activity. Indeed, alteration of RBP function can lead to impairment in any of the crucial steps of the miRNA pathway. Deregulation of RBP expression or activity has been reported in several malignancies. Recently, several groups obtained evidence for more specific miRNA-RBP interplay under various physiological conditions or in response to external stimuli. Such regulatory mechanisms rely on miRNA and RBP binding activity to common target RNAs and are probably under tight spatiotemporal control. These insights uncover a wide variety of new mechanisms in RNA regulation, which could have relevance for cancer development and progression.

RBPS MODULATE POST-TRANSCRIPTIONAL REGULATION

RBPs are essential players in RNA metabolism, regulating RNA splicing, transport, localization, stability, translation and degradation. Some RBPs recognize common mRNA features such as the 5′ cap or the 3′ poly(A) tail, but most RBPs contain RNA-binding domains for recognition of specific sequence motifs or secondary structures in mRNA (reviewed in REF. 5) (TABLE 1). In the past decade, numerous roles of RBPs in miRNA processing and function have emerged.

Table 1 | Classes of proteins with RNA-binding capacity and their motifs

Classes of RBPs	Structure elements	RNA recognition	Examples
RNA recognition motif (RRM)	β-α-β-β-α-β (canonical)	~4-bp ssRNA nucleotide	DAZ, DAZL, DND1, ELAV, GW182, hnRNPs, IGF2BP1, PABP1 and SRSF1
dsRNA-binding domain (dsRBD)	α-β-β-β-α (canonical)	dsRNA	DGCR8, Dicer, Drosha and TARBP2
HNRNPK homology (KH) domain	Variable loop of α/β and Gly-X-X-Gly	~4-bp ssRNA nucleotide	FMRP, FXR1, FXR2, IGF2BP1, KHSRP and STAR
PIWI, AGO and Zwille (PAZ) domain	OB-like β-barrel fold	3′ single-stranded overhangs	AGO and Dicer
PIWI domain	RNase H-like fold	5' phosphate group*	AGO
MID domain	Rossman-like fold	5' phosphate group*	AGO
GW domain	Gly-Trp repeats	-	GW182
Zinc-binding motif	Cys^n/His^n in a β-β-α structure	-	LIN28, TRIM proteins with NHL domain and TTP
PUF repeat	Eight base-specific repeats of three α-helices	UGUAHAUA	PUM1 and PUM2
DEAD-box motif	Asp-Glu-Ala-Asp	-	DDX5, DDX6, DDX17, DDX20 and DDX42
DExD/H-box motif	Asp-Glu-X-Asp/His	-	MOV10

α, alpha helix in secondary protein structure; β, beta sheet in secondary protein structure; AGO, Argonaute; DAZ, deleted in azoospermia; DAZL, DAZ-like; DGCR8, DiGeorge syndrome critical region 8; DDX, DEAD-box; DND1, dead end 1; dsRNA: double-stranded RNA; ELAV, embryonic lethal abnormal vision; FMRP, fragile X mental retardation protein; FXR, fragile X mental retardation syndrome-related protein; hnRNP, heterogenous nuclear ribonucleoprotein; IGF2BP1, insulin-like growth factor 2 binding protein 1; KHSRP, KH-type splicing regulatory protein; MOV10, Moloney leukaemia virus 10; PABP1, poly(A) binding protein 1; PUM1, pumilio 1; SRSF1, serine/arginine-rich splicing factor 1; ssRNA, single-stranded RNA; STAR, steroidogenic acute regulatory protein; TARBP2, transactivation responsive RNA-binding protein 2; TRIM, tripartite motif; TTP, tristetraprolin.
*The 5′-phosphate binding pocket lies at the interface between the MID and PIWI domains; the MID domain alone binds nucleotides with low affinity.

Box 1 | MicroRNAs as regulators of gene expression

At present, almost 1,500 unique microRNAs (miRNAs) are produced from mostly evolutionarily conserved regions in the human genome (miRBase, release April 2011)[161–163]. miRNAs regulate gene expression by seed base-pairing to one or more partially complementary sites in target mRNAs (reviewed in REF. 1). Gene expression is mainly and most efficiently controlled via target sites located in the mRNA 3′ untranslated region (UTR), although a small number of target sites in open reading frames (ORFs) or 5′ UTRs have also been reported[164]. During animal development, expression of miRNAs is pivotal for the timing and regulation of many processes[165]. Subsequent differentiation is associated with a global increase in expression of miRNAs that define tissue-specific gene expression patterns[3,166]. Moreover, miRNA expression profiles can discriminate tumour tissue from normal tissue, associating low miRNA expression levels with a loss of cellular differentiation in tumours (reviewed in REF. 167). In almost all cancer types, alterations in miRNA-mediated regulation are implicated in key processes of tumorigenesis, such as apoptosis[168], proliferation[169], angiogenesis[170], migration[171] and invasion[172,173]. Whether miRNAs act as oncogenes or as tumour suppressor genes is dependent on the presence of their targets and the cellular context[174] (reviewed in REFS 175,176). In addition, expression profiling of human tumours has identified miRNA signatures associated with diagnosis, progression, prognosis and treatment response (reviewed in REF. 177). In cancer, miRNA-mediated repression can be altered owing to genetic variation and small- or large-scale mutations in miRNA genes and mRNA target genes (reviewed in REFS 178,179). Also, the production of miRNAs can be induced or disturbed by changes in miRNA transcription or further processing (reviewed in REFS 178,180,181).

More than 500 human RBPs are known, but only a few have been assigned an oncogenic or tumour-suppressive function (reviewed in REF. 6). Examples of RBPs implicated in tumorigenesis are the TET family of RBPs[7], the STAR family of RBPs (such as Src-associated in mitosis 68 kDa protein (SAM68; also known as KHDRBS1)[8], β-catenin[9] and multiple RBPs involved in alternative splicing (see REF. 10 for a review of alternative splicing in cancer). In addition, many RBPs are involved in RNA stability, in which simultaneous assembly of these RBPs on target RNA has either a synergistic or antagonistic effect. RNA-binding activity can be rapidly modulated in response to external stimuli, for example through RBP expression levels, nucleocytoplasmic translocation, post-translational modifications or changes in secondary structure. The specifics of RNA binding remain largely undiscovered, but the function of various classes of RBPs in miRNA-mediated post-transcriptional regulation has started to receive more attention.

Seed

Six to eight nucleotides at the 5′ end of the mature microRNA that are involved in the recognition of target mRNAs.

THE BIOGENESIS OF MIRNAS

The biogenesis of miRNAs starts with RNA polymerase II-dependent transcription of a miRNA gene locus, generating a long primary transcript (pri-miRNA) that folds into a hairpin structure. These pri-miRNAs are first 5′ 7-methyl-guanosine (m7G) capped and 3′ polyadenylated before further processing occurs (FIG. 1). In the nucleus,

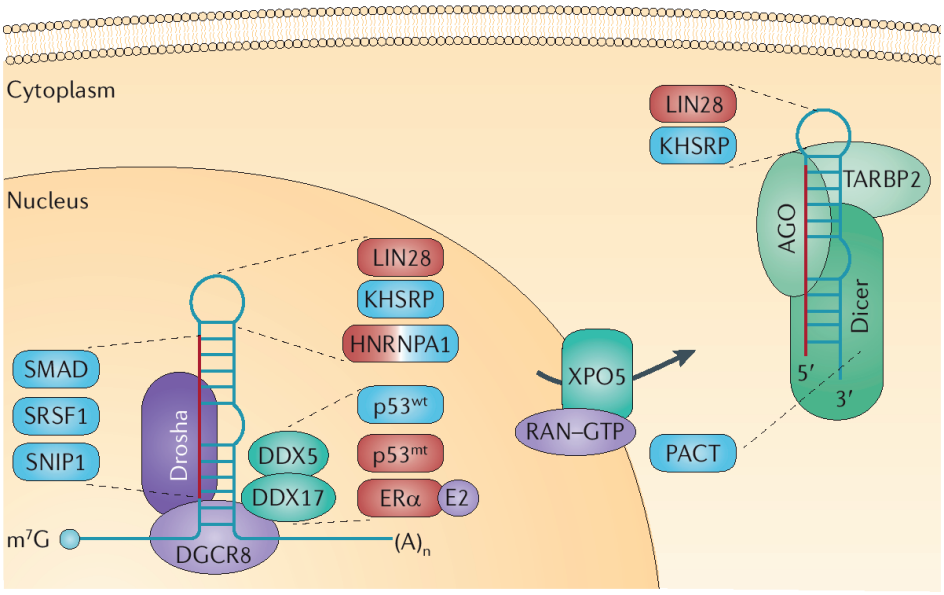


Figure 1 | Regulation of the microRNA biogenesis pathway by processing factors. Several RNA-binding proteins (RBPs) influence the processing of (a subset of) microRNAs (miRNAs) by recognizing specific RNA features or by associating with key components of the miRNA biogenesis pathway. Some of these factors are overexpressed or lost in tumours. RBPs can influence miRNA processes both in the nucleus and in the cytoplasm, depending on which steps in the biogenesis of miRNAs they affect. Blue and red boxes indicate a stimulative or inhibitive effect, respectively, of RBPs on miRNA processing. AGO, Argonaute; DGCR8, syndrome critical region 8; E2, 17β-oestradiol; ERα, oestrogen receptor-α; HNRNPA1, heterogeneous nuclear ribonucleoprotein A1; KHSRP, KH-type splicing regulatory protein; m^7G, 7-methyl-guanosine; $p53^{mt}$, mutant p53; $p53^{wt}$, wild-type p53; SNIP1, SMAD nuclear interacting protein 1; SRSF1, serine/arginine-rich splicing factor 1; TARBP2, transactivation responsive RNA-binding protein 2; XPO5, exportin 5.

recognition by the microprocessor complex results in cleavage of the pri-miRNA, which is attained through catalytic cleavage of the double-stranded stem by the RNase III endonuclease Drosha, while the hairpin is correctly positioned by DiGeorge syndrome critical region 8 (DGCR8; also known as Pasha)[11,12]. After cleavage, the secondary structure of the resulting ~70 nucleotide (nt) precursor miRNA (pre-miRNA) is recognized by a complex of exportin 5 (XPO5) and RAN-GTP[13,14]. The stabilized pre-miRNAs are then shuttled to the cytoplasm and released on GTP hydrolysis.

In the cytoplasm, the pre-miRNA terminal loop is cleaved by another double-stranded RNA (dsRNA)-specific RNase III, Dicer, in collaboration with the human immunodeficiency virus transactivation responsive RNA-binding protein 2 (TARBP2)[15–18]. A PAZ domain within Dicer binds the pre-miRNA 2-nt 3′ overhang, while the RNA-binding domain of Dicer binds the double-stranded stem and defines the site of cleavage by measuring 22 nt from the 3′ overhang[19,20]. The cofactor TARBP2 uses two dsRNA-binding domains (dsRBDs) to interact with the pre-miRNA, stimulating Dicer-mediated cleavage, and a third dsRBD to increase the stability

of the Dicer–RNA complex[21,22]. Another cofactor, the protein activator PACT (also known as PRKRA), has a similar function to TARBP2 in that it recognizes the same binding domain of Dicer[23]. TARBP2 and PACT therefore selectively promote miRNA processing but are not essential for Dicer-mediated cleavage. After cleavage of the ~22-nt RNA duplex, now consisting of two 5′ phosphorylated sequence strands with 3′ overhangs, the functional strand, referred to as the guide strand, is loaded into an Argonaute (AGO) protein[24]. TARBP2 secures the recruitment of the AGO protein and the formation of a ternary complex together with Dicer[18]. All AGO proteins are characterized by evolutionarily conserved MID and PAZ domains involved in RNA binding and an RNase H-like PIWI domain for endonuclease activity[25,26]. The 5′ phosphate group of the miRNA guide strand is stably bound by the MID domain while the PAZ domain recognizes the dinucleotide 3′ overhang that is characteristic of Dicer-mediated cleavage[27]. Occasionally, a processing intermediate is generated by AGO2-mediated cleavage of a pre-miRNA that probably facilitates removal of the passenger strand[28]. Rarely, a miRNA with a short stem region that cannot be recognized by Dicer, as occurs for pre-miR-451 in mice and zebrafish, is trimmed by the endonuclease activity of AGO2 (REFS 29,30). Although this provides evidence for an miRNA biogenesis pathway that does not require Dicer cleavage, the extent to which such phenomena occur is unclear. In general, pre-miRNA characteristics determine which strand is retained, whereby the suffixes 5p and 3p (or *) designate the 5′ and 3′ duplex arms, respectively[31,32].

Passenger strand

The strand of the microRNA duplex that is complementary to the guide strand and is destined for degradation upon loading of the guide strand into the microRNA-induced silencing complex (miRISC).

GLOBAL CONTROL OF MIRNA BIOGENESIS IN CANCER

In human tumours, global downregulation of miRNA expression is an apparent feature[3,4]. Another striking observation in primary tumours is the accumulation of pri-miRNAs compared with normal tissue[33]. Hence, impairment of crucial steps in miRNA production, either in the nucleus or in the cytoplasm, could be the underlying cause. In recent years, an increasing amount of evidence has been obtained for cancer-related alterations in RBPs that are intimately involved in miRNA biogenesis.

At the genetic level, copy number abnormalities of *DICER1*, *AGO2*, *XPO5* and other genes that are essential for miRNA biogenesis occur often in breast and ovarian cancer, as well as in melanoma[34,35]. In various human cancer cell lines, mature miRNA expression levels are inconsistent with pre-miRNA expression levels owing to nuclear retention of pre-miRNAs[36]. Mutations that inactivate *XPO5* in human tumours lead to precursor accumulation in the nucleus and lower levels of mature miRNAs[37]. Moreover, *XPO5* knockdown enhances the tumorigenicity of cells injected into mice, and the reverse effect is seen on overexpression of wild-type *XPO5* in colorectal cancer cells expressing mutant *XPO5* (REF. 37).

Intriguingly, disruption of miRNA production by depletion of any of the miRNA processing factors Drosha, DCGR8 or Dicer has been shown to promote oncogenesis[38]. In a mouse model of retinoblastoma, the upregulation of an miRNA

subset in response to *Rb1* inactivation is abolished on monoallelic loss of *Dicer1*, resulting in accelerated tumour formation[39]. Indeed, the frequent occurrence of heterozygous, but not homozygous, genetic deletions in various human tumours implicate *DICER1* as a haploinsufficient tumour suppressor[40,41]. Piccolo and colleagues[42] have also recently shown that the miRNA family miR-103/107 abrogates miRNA maturation by targeting the 3′ UTR of *DICER1*. Low Dicer protein levels result in stimulation of migration and metastasis that is at least partially due to blocked processing of miR-200.

Several reports suggest that TARBP2 function is also impaired in cancer. The occurrence of frame-shift mutations in *TARBP2* in colon tumours with micro- satellite instability correlates with lower levels of Dicer and mature miRNAs[43,44]. Indeed, the downregulation of TARBP2 expression by RNA interference (RNAi) destabilizes the Dicer protein, resulting in concurrent impairment of miRNA biogenesis[18]. As expected, restoration of wild-type TARBP2 levels in cell lines expressing truncated TARBP2 reconstitutes normal levels of Dicer and mature miRNAs[43]. Importantly, rescue of miRNA production by TARBP2 restoration is accompanied by a decline in tumour growth *in vivo*, showing that the normal production of miRNAs is tumour suppressive in this setting. Interestingly, under normal growth conditions TARBP2 is phosphorylated, which increases the stability of TARBP2 and Dicer[45]. On growth factor stimulation, the MAPK–ERK pathway increases TARBP2 phosphorylation, resulting in increased levels of a subset of miRNAs but lower levels of the let-7 miRNA family[45]. This illustrates how mitogenic signalling can be translated into changes in cell viability and proliferation through the miRNA biogenesis pathway. Thus, the role of TARBP2 in miRNA processing is important for preservation of a normal, untransformed cell state.

SPECIFIC CONTROL OF MIRNA BIOGENESIS IN CANCER

In addition to global changes in miRNA levels, differential expression of specific miRNAs is also apparent in tumours, and this could result from changes in various RBPs.

DDX5 and DDX17. The stability of the microprocessor complex is controlled by post-transcriptional cross-regulation between Drosha and DGCR8 (REF. 46). Furthermore, Drosha-mediated cleavage of specific pri-miRNAs is modulated by accessory RBPs such as DEAD-box 5 (DDX5; also known as p68) and DDX17 (also known as p72), a number of heterogeneous nuclear ribonucleoproteins (hnRNPs) and other factors[47]. Interestingly, both DDX5 and DDX17 are often highly expressed in human breast, prostate and colon tumours (reviewed in REF. 48). In mice, both DDX5 and DDX17 are required for efficient processing of a subset of pri-miRNAs, as depletion of either protein results in lower levels of mature forms[49]. This has consequences for cell proliferation and survival, as *Ddx5*-deficient or *Ddx17*-deficient mouse embryonic fibroblasts (MEFs) are characterized by slower cell growth and increased apoptosis[49]. Moreover, joint knockdown of *DDX5* and *DDX17* in human cervical carcinoma cells suppressed cell proliferation, whereas overexpression of wild-type DDX5 stimulated keratinocyte proliferation, and overexpression of a

constitutively phosphorylated DDX5 mediated cell proliferation and epithelial-to-mesenchymal transition (EMT) on growth factor stimulation[48]. These functions provide a causal explanation for the over-expression of DDX5 and DDX17 in human cancer, but evidence for a specific role of the miRNAs regulated by DDX5 and DDX17 in tumorigenesis is still lacking.

Intriguingly, DDX5 and DDX17 can act as a bridge between Drosha activity and other regulators through protein–protein interactions. The tumour suppressor p53, for example, interacts with DDX5 to enhance processing of miRNAs that function in growth suppression[50]. Cancer-related mutations in p53 result in loss of DDX5 interaction with Drosha and inefficient miRNA maturation[50]. As a transcription factor, p53 is known to activate transcription of a subset of miRNA genes that partially overlaps with the group of miRNAs that is targeted by p53 during processing (reviewed in REF. 51).

Another case is illustrated by oestrogen receptor-α (ERα), a transcription factor that is overexpressed in the largest subgroup of breast tumours. Yamagata *et al.*[52] reported that the processing of a set of DDX5- and DDX17-dependent pre-miRNAs is blocked by the binding of activated ERα to DDX5 and DDX17. Thus, steroid hormones can affect miRNA maturation through their cognate nuclear receptors, resulting in the stabilization and efficient expression of ERα-target mRNAs.

SMAD signal transducers also facilitate processing of a subset of miRNAs following transforming growth factor-β (TGFβ) and bone morphogenetic protein (BMP) growth factor activation[53]. The SMADs (in a complex with the microprocessor component DDX5) promote miRNA maturation by binding a consensus sequence in the pri-miRNA stem region, which is similar to the SMAD-binding sequences in gene promoters[54]. SMAD nuclear interacting protein 1 (SNIP1) induces processing of some miRNAs as well, either by pri-miRNA binding or by direct interaction with Drosha[55]. Whether enhanced processing of these specific pri-miRNAs contributes to TGFβ-induced tumour progression remains to be established (reviewed in REF. 56).

The serine/arginine-rich splicing factor 1 (*SRSF1*) gene encodes the mRNA splicing factor SRSF1 (also known as SF2 and ASF), which recognizes the stem region of specific pri-miRNAs, resulting in enhanced cleavage by Drosha[57]. Amplification of the *SRSF1* protooncogene is found in various tumours, along with increased levels of miR-221 and miR-222, which are targets of SRSF1. However, the contribution of miRNAs to the tumorigenic capacity of SRSF1 is uncertain[58,59]. Some common ALL1 (also known as MLL) leukaemogenic fusion proteins interact with Drosha and have a role in the biogenesis of specific miRNAs, which could contribute to the abundant levels of some miRNAs observed in acute myeloid leukaemias[60].

The let-7 miRNA family. Another regulatory circuit is provided by the let-7 miRNA family and LIN28, an RBP exclusively expressed in undifferentiated cells[61]. The role of LIN28 in development is well studied and involves suppression of let-7 miRNA levels by binding to the terminal loop of pri-let-7, thereby blocking Drosha cleavage (reviewed in REF. 62). In the cytoplasm, LIN28 recruits terminal uridyltransferase (TUT4; also known as ZCCHC11) to a 4-nt sequence motif in the terminal loop of pre-let-7, and of other pre-miRNAs, resulting in oligo-uridylation. This causes resistance to Dicer cleavage and the subsequent precursor degradation decreases let-7 abundance (see REF. 62). In response to inflammation, upregulation

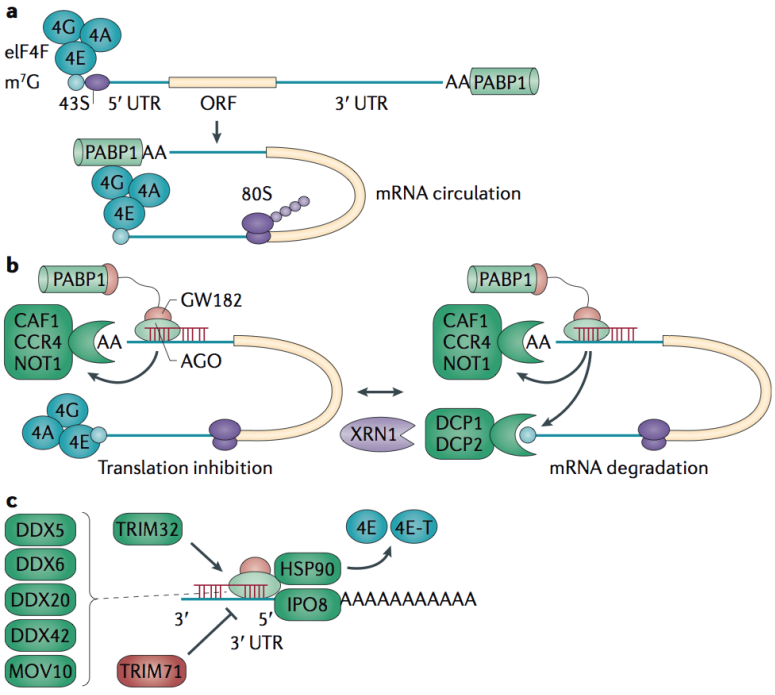


Figure 2 | RISC-associated factors regulate efficient microRNA-mediate drepression. (a) The common process of cap-dependent mRNA translation begins with modifications at the mRNA 5′ methylated guanosine cap structure and 3′ poly(A) tail. The cap-initiation complex eukaryotic translation initiation factor 4F (eIF4F, which consists of the RNA helicase eIF4A, the cap-binding protein eIF4E and the scaffolding protein eIF4G) associates with the 5′ cap and recruits the 43S pre-initiation complex (which contains the 40S ribosomal subunit) to the mRNA 5′ cap. The cytoplasmic poly(A) binding protein 1 (PABP1) is bound to the 3′ poly(A) tail and interacts with the cap-initiation complex through eIF4G. This interaction effectively establishes mRNA circularization, which stimulates translation. The 43S pre-initiation complex starts scanning in a 3′ direction and then, on recognition of an initiation codon, the 60S large ribosomal subunit is recruited to form a ribosome (80S) that enables translation elongation. **(b)** When human miRNA-induced silencing complex (miRISC) recognizes the 3′ untranslated region (UTR) of a target mRNA, interference with the process of translation can occur either at the initiation step or during mRNA translation processes following initiation (reviewed in REF. 189). According to the conventional model, the mechanism of gene repression by miRISC is through translational inhibition at initiation and removal of the target poly(A) tail, which is facilitated by the cytoplasmic deadenylase complex made up of CC chemokine receptor type 4 (CCR4), CCR4-associated factor 1 (CAF1; also known as CNOT7) and negative regulator of transcription subunit 1 (NOT1; also known as CNOT1). It is plausible that this deadenylation is enabled by direct interaction between the carboxy-terminal silencing domain of trinucleotide repeat-containing gene 6 (TNRC6) and PABP1, which interferes with the function of eIF4F and PABP1 and impedes mRNA circularization (reviewed in REFS 78,190). Actual target degradation can follow recruitment of the DCP1–DCP2 decapping complex and several decapping activators to the 5′ cap, allowing mRNA degradation by the major cytoplasmic 5′-to-3′ exonuclease XRN1. **(c)** Some miRISC-associated RNA helicases, such as specific DEAD-box (DDX) proteins and Moloney leukaemia virus 10 (MOV10), may facilitate miRISC loading or target binding by RNA unwinding. The RNA-binding tripartite motif (TRIM) proteins, heat shock proteins (HSPs) and importin 8 (IPO8) co-regulate mRNA translation at various points, and this function can be deregulated in cancer. Green boxes indicate a stimulatory effect and red boxes an inhibitory effect on miRNA function. 4E-T, eIF4E transporter; AGO, Argonaute; m7G, 7-methyl-guanosine.

of LIN28 blocks let-7-mediated repression of interleukin-6 (*IL6*). This activates STAT3 and reinforces LIN28 expression, both of which can lead to a transformed state in the absence of external signals[63].

Interestingly, in differentiated cells lacking LIN28 expression, mature let-7 levels differ substantially, pointing towards a more complex regulation of miRNA processing by distinct RBPs. KH-type splicing regultory protein (KHSRP; also known as KSRP) recognizes a conserved sequence in the terminal loop of let-7 and promotes its maturation and that of other miRNAs by facilitating an association with Drosha in the nucleus or with Dicer in the cytoplasm[64]. By contrast, the binding of the splicing regulator HNRNPA1 to the same sequence in the pri-let-7 terminal loop represses Drosha-mediated processing[65]. In differentiated cells that lack LIN28 expression, KHSRP and HNRNPA1 compete for pri-let-7 binding to regulate the extent of Drosha cleavage. In addition, the pre-miR-18a hairpin is recognized by HNRNPA1, and this promotes microprocessor interaction and maturation of pre-miR-18a[66]. Other members of the *mir-17-92* polycistron are not subject to regulation by HNRNPA1.

Altogether, the findings described above illustrate that RBP binding can have a dual effect on miRNA processing, and thus on development and cancer. Whether the eventual effect is determined by the relative abundance of each RBP, by differences in RNA binding affinity or by both generally requires additional clarification. Interestingly, both the pri-miRNA stem region and the terminal loop harbour specific RNA-binding motifs serving as control regions[67,68]. It remains a challenge to unravel sequence and structural characteristics that confer RBP binding specificity and dictate further processing.

RBPS, MIRNA-MEDIATED REPRESSION AND CANCER

The RNA silencing process involves assembly of the miRNA-induced silencing complex (miRISC), also called the miRNA ribonucleoprotein complex (miRNP) (reviewed in REF. 69). This complex minimally consists of three components: an active miRNA strand associated with an AGO protein and a GW182 protein[70]. The incorporated miRNA strand guides the RISC machinery to a matching mRNA, resulting in specific target repression[71,72]. The activity of miRISC is regulated by interacting proteins, which either directly bind AGO or GW182 or are integrated through binding to RNA molecules[74].

The AGO protein family is active in embryogenesis, germ cell maintenance and cell differentiation[74]. The four ubiquitously expressed human AGO proteins (AGO1–4), with AGO2 as the best-characterized member[24], seem to have overlapping functions in miRNA-mediated repression[75,76]. The second essential RISC component, GW182, is probably responsible for translational repression, as GW182-deficient cells have impaired miRNA function (reviewed in REF. 77). In humans, the PIWI domain of AGO proteins interacts with trinucleotide repeat-containing gene 6A (TNRC6A), TNRC6B and TNRC6C (homologues of *Drosophila melanogaster* GW182). All three homologues contain a bipartite carboxy-terminal silencing domain that is required for translational repression (see REF. 77).

Modulation of miRISC function by RBPs in cancer. According to the current view, miRNAs accomplish gene repression by inhibiting translation or by reducing mRNA stability, which generally results in target degradation (reviewed in REFS 78,79) (FIG. 2a,b). It is unknown what determines the decision between mRNA storage in a translationally repressed state or destabilization. However, although inhibition of translation results in decreased protein levels, the changes in target protein levels following miRNA binding are found to mainly coincide with actual destabilization of the target mRNA[80]. Remarkably, in mammalian cells, observed changes in ribosome occupancy on miRNA targeting largely reflect concurrent mRNA destabilization, rather than persistence of repressed mRNAs[81].

Human miRISC is found associated with several proteins characterized by RNA helicase domains (FIG. 2c). By RNA unwinding, RNA helicase activity could facilitate either the incorporation of the active miRNA in AGO or miRISC target binding. The DEAD-box RNA helicase DDX6 (also known as RCK and p54) interacts with AGO2 and has been assigned a function both in translational repression and decapping[82,83]. Overexpression of DDX6 is found in colon cancer and RNAi-mediated downregulation results in antitumour effects, suggesting an oncogenic function[84]. Another DEAD-box member, DDX5, has also been shown to have helicase activity when in the cytoplasm and facilitates miRISC loading by unwinding the let-7 precursor duplex[85]. The proteins described next are also reported to co-purify with miRISC, but their exact role in miRNA-mediated repression has not been validated yet. The DExD-box RNA helicase Moloney leukaemia virus 10 (MOV10) interacts with either of the core miRISC components in the cytoplasm, whereas when in the nucleus it is thought to bind chromatin[86,87]. Furthermore, the co-purification of the DEAD-box helicases DDX20 (also known as gemin 3) and DDX42 (also known as gemin 4) from HeLa cells suggests that they reside in a complex with AGO2 and miRNAs[71,88]. Whereas knockdown of MOV10, DDX20 or DDX42 does not disrupt miRISC localization in P-bodies, an intact helicase domain of DDX6 is required for P-body assembly and for miRNA-mediated repression[82]. Interestingly, HeLa cells in which P-bodies have been disrupted by depletion of the P-body component LSM1 retain functional miRNA-mediated repression[89]. Thus, the function of DDX6 in miRNA-mediated targeting is probably not dependent on P-body localization. Moreover, these data indicate that the aggregation of stalled translation complexes in cytoplasmic foci is merely a consequence of translational repression, but leave the exact role of RNA helicase activity in facilitating miRNA function unexplained.

P-bodies

Cytoplasmic foci containing proteins involved in diverse post-transcriptional processes, such as mRNA degradation.

A number of heat shock proteins, a class of proteins that is frequently overexpressed in cancer, were found to reside in complex with AGO proteins[73,90] (FIG. 2c). For example, heat shock protein 90 (HSP90), an essential member of an ATP-dependent chaperone complex, stabilizes unloaded AGO2 but also influences miRISC, eukaryotic translation initiation factor 4E (eIF4E) and eIF4E transporter (eIF4E-T) localization[91–93]. eIF4E-T is thought to compete with eIF4G for binding to eIF4E, thereby preventing mRNA circularization[94]. Inhibition of HSP90 prevents P-body formation and, through AGO protein destabilization, affects miRNA function indirectly[91,95].

Several tripartite motif (TRIM) domain proteins, which are known for ubiquitin ligase activity, influence the function of specific miRNAs. *TRIM71* is a mammalian homologue of *Caenorhabditis elegans lin-41* and a target of let-7. TRIM71 drives AGO degradation through ubiquitylation, thereby interfering with miRNA function[96] (FIG. 2c). Together with LIN28, TRIM71 forms a feedback circuit for let-7 regulation. As a positive modulator, TRIM32 stimulates the function of a specific set of miRNAs in mice, including let-7a, independently of its E3 ubiquitin ligase domain[97] (FIG. 2c). However, the mechanism by which TRIM32 operates, or whether TRIM32 binds RNA, is unknown. Whereas TRIM32 and TRIM71 do not change miRNA levels, interaction of a TRIM protein that contains an NCL1, HT2A and LIN41 (NHL) domain and AGO1 interferes with miRNA biogenesis in *D. melanogaster*[98]. In head and neck carcinoma, TRIM32 is overexpressed and promotes tumour growth in part through its ubiquitin activity[99].

Importin 8 (IPO8) binds to AGO proteins and influences AGO nuclear import. It also functions as a chaperone and targets AGO2-associated mRNAs, possibly through RNA-independent interactions with miRISC that may alter the structure of the complex[100] (FIG. 2c). In addition, 5′–3′ exoribonuclease 1 (XRN1), a cellular exoribonuclease that is essential for efficient decay of uncapped mRNA[101], is often lost in osteogenic sarcomas and may act as a tumour suppressor protein in these types of tumours[102].

MIRNA–RBP INTERPLAY ON THE TARGET MRNA

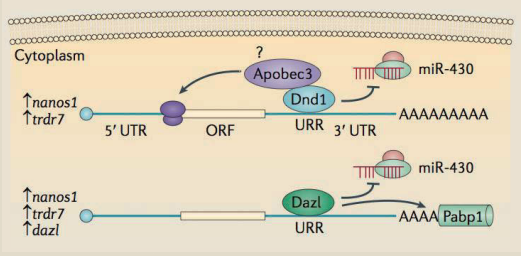
Although several proteins can modulate the efficacy of miRNA biogenesis, two observations support the existence of an alternative regulatory mechanism that influences miRNA activity. First, certain genes are exclusively subject to miRNA regulation in particular conditions, but no significant change is observed in the level of the targeting miRNA. Second, of all potential high affinity targets of a specific miRNA, only a subset is subjected to miRNA regulation. An expanding number of reports now provide mechanistic explanations, often demonstrating interplay between miRNAs and RBPs on target 3′ UTRs under specific conditions, some of which are linked to differentiation (BOX 2) or oncogenesis (FIG. 3).

The ELAV RBP family. Positive modulators of RNA stability include the Hu proteins. They share homology with the *D. melanogaster* embryonic lethal abnormal vision (ELAV) protein and have a broad function in stabilizing AU-rich element (ARE)-containing mRNAs in the cytoplasm (reviewed in REF. 103). Whereas HuB, HuC, and HuD are neuronal or gonadal proteins, HuR is ubiquitously expressed and mediates cellular responses to different types of stress (FIG. 3a). Bhattacharyya and colleagues[104] found that HuR relieves cationic amino acid transporter 1 (*CAT1*) mRNA from miR-122-mediated repression under stress conditions in human liver cells. On stress induction, nuclear HuR is translocated to the cytoplasm and specifically recruits target mRNA to polysomes to secure translation[104,105]. In rare cases, HuR and/or HuD inhibit target expression[106–108]. HuR can bind to untranslated mRNA regions

AU-rich element (ARE)

An element that is present in certain 3′ untranslated regions. It often contains a repeat of the AUUUA motif, which has a destabilizing effect on the mRNA in which it resides.

Box 2 | RNA–RNA-binding protein interplay during differentiation



Specifically expressed in germ cells, dead end 1 (DND1) is a negative modulator of microRNA-induced silencing complex (miRISC) activity. In zebrafish and human germ cells, DND1 binds uridine-rich regions (URRs) in the 3′ untranslated regions (UTRs) of germline-specific genes, which either sequesters mRNAs or physically displaces miRISC to alleviate microRNA (miRNA)-mediated suppression[182]. In zebrafish, the miR-430 family represses translation of the germ cell factors nanos1 (also known as nanos3), tdrd7 and deleted in azoospermia-like (dazl). The human orthologues of the miR-430 family, which comprises miR-371–373 and miR-518–520, are oncogenic in germ cells and associated with increased proliferation, migration, invasion and metastasis[183,184]. DND1 alleviates miR-372-mediated repression of large tumour suppressor 2 (LATS2) and also miR-221- and miR-222-mediated repression of cyclin-dependent kinase inhibitor 1B (CDKN1B)[182]. In mouse germ cells, DND1 directly interacts with apolipoprotein B mRNA-editing complex 3 (APOBEC3), but whether these RNA-binding proteins (RBPs) cooperatively de-repress germ-cell-specific genes is not known[185]. DAZL also antagonizes miRNA function in human germ cells. During embryogenesis, DAZL binds URRs and drives polyadenylation through interaction with poly(A)-binding protein 1 (PABP1), thereby antagonizing the miRNA effect[186]. DAZL in humans is found associated with pumilio 2 (PUM2), either as an RNA-binding complex or bound to separate motifs[187,188]. Although PUM proteins generally function in gene silencing, the fate of mRNAs jointly targeted by DAZL and PUM2 in the context of human germ cells is unclear. The RBPs DND1, DAZL and PUM2 are essential for germline development and form part of an extended network of post-transcriptional regulation for the maintenance of stemness. Although these RBPs are mainly expressed in germ cells, their regulatory mechanisms could be relevant in tumours with acquired multipotency or pluripotency. This hypothesis needs to be examined in the future. ORF, open reading frame.

or can modulate miRNA binding to a nearby site, as illustrated by the requirement of HuR binding for let-7-mediated repression of *MYC*[109]. Interestingly, the activity of HuRis controlled by phosphorylation and other post-translational modifications that affect HuR subcellular localization and RNA binding activity[110]. Furthermore, miR-125a and miR-519 levels in cells generally inversely correlate with HuR target levels in various tumours. In fact, re-expression of these miRNAs decreased HuR protein levels and tumorigenicity *in vitro* and in a nude mouse xenograft model[111–113]. The relevance of HuR to cancer is further shown by the high levels of HuR protein in various tumours and direct or miRNA-mediated regulation of many mRNA targets involved in cell proliferation, survival, evasion of immune recognition, metastasis, invasion and local angiogenesis (reviewed in REF. 110).

Occasionally, more complex interactions between HuR and other RBPs in post-transcriptional regulation are observed[114]. For example, HuR and the cap-binding protein eIF4E cooperatively stimulate translation of proteins involved in growth, survival and malignancy, whereas *eIF4E* mRNA itself is concurrently stabilized by

HuR binding[115]. As eIF4E is overexpressed in many cancers and closely correlates with HuR expression levels and poor prognosis, these mechanisms are likely to influence tumour progression[115–117].

hnRNPs. Another important RBP family is the hnRNPs. These are sequence-specific repressors of mRNA splicing but are being increasingly associated with a broader range of functions. Whereas some hnRNPs strictly localize to the nucleus, others constantly shuttle between the nucleus and the cytoplasm, indicating a putative role in translational control[118]. Nuclear AU-rich element RNA-binding protein 1 (AUF1; also known as HNRNPD) and HuR simultaneously bind separate regions in the 3′ UTR of common targets but competitively bind shared targets in the cytoplasm in a manner dependent on the abundance of either RBP[115,119]. HuR-bound mRNAs are localized to polysomes for protein synthesis, whereas mRNAs bound by AUF1 are destined for degradation[120]. Several environmental cues can cause cytoplasmic enrichment of these, and other, RBPs[105]. However, the contrasting fate of the common targets cyclin-dependent kinase inhibitor 1A (*CDKN1A*) and cyclin D1 (*CCND1*) after ultraviolet C (UVC) irradiation argues for target-specific regulatory mechanisms in addition to RBP enrichment[119]. Indeed, access of AUF1, but not of HuR, to AREs in target 3′ UTRs can be inhibited by local changes in secondary ARE structure[121]. However, RBPs tend to induce a local change in secondary RNA structure on binding, possibly modulating access of other *trans*-acting factors[122,123].

An interesting interplay has been observed between HNRNPE2 (also known as PCBP2) and miR-328 (REF. 128) (FIG. 3c). The binding of HNRNPE2 to C-rich regions in the 5′ UTR of CCAAT/enhancer binding protein-α (*CEPBA*) mRNA, a key regulator of differentiation, represses protein translation and myeloid differentiation[129]. The C-rich mature form of miR-328 competes with *CEPBA* for binding of HNRNPE2 in a RISC-independent manner[128]. Intriguingly, this competition is induced by CEPBA-stimulated transcription, mainly of miR-328 and other miRNA genes[128,130]. It is conceivable that miR-328 sequesters other C-rich region-binding proteins, such as PCBP4 and HNRNPK, in a similar way or that other miRNAs containing an RBP recognition motif function likewise. The importance of this interaction in the context of cancer is stressed by the loss of miR-328 in chronic myelogenous leukaemia. Here, HNRNPE2 activity is upregulated by oncogenic signalling and results in lower miR-328 levels. Subsequent deregulation of direct HNRNPE2 targets, as well as miR-328 targets (such as *MYC* and *PIM1*), contributes to cancer progression[128,129,131].

Tristetraprolin. Another ARE-targeting RBP is tristetraprolin (TTP; also known as ZFP36), an mRNA decay factor that recruits components of the mRNA degradation machinery to the bound target and directs ARE-mediated decay[120,132]. Tumour necrosis factor (*TNF*), *VEGFA*, hypoxia inducible factor 1α (*HIF1A*), cyclooxygenase 2 (*COX2*; also known as *PTGS2*), large tumour suppressor 2 (*LATS2*) and *MYC* are among numerous TTP targets, establishing links with apoptosis and proliferation[133–138] (FIG. 3d). Interestingly, human tumours frequently lack TTP expression, and low *TTP* mRNA levels correlate with a poor prognosis[139]. Several studies have also reported interactions between TTP and components of the RISC machinery, as well as overlap between the associated mRNA degradation machinery (reviewed in REF. 140). For example, TTP can induce mRNA decay by decapping in a manner that requires both binding to an ARE in the 3′ UTR and interaction of the RBP with the decapping

complex. However, it has been reported that TTP interaction with AGO2 instead of binding to AREs enables miR-16-mediated repression of *TNF* and *COX2* in human cells[141]. These data suggest that there may be multiple mechanisms whereby TTP and miRNAs can synergistically promote mRNA degradation.

Cell cycle control by RBP–miRNA interplay. On activation of β-catenin signalling, the insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1; also known as CRDBP or IMP1) increases levels of MYC and an F-box protein, βTrCP1 (also known as FBW1A), by targeting a recognition motif in the protein coding region[142,143] (FIG. 3e). IGF2BP1 captures the mRNAs in cytoplasmic particles, thereby providing protection from decay, possibly with assistance from IGF2BP1-associated RBPs[144]. Moreover, IGF2BP1 was found to control degradation of *BTRC* (which encodes βTrCP1) by disrupting the association of miR-183 and AGO2 with a target site in the *BTRC* coding region[145]. Additionally, IGF2BP1 targets the 3′ UTR of a microphthalmia-induced transcription factor (*MITF*) isoform that is predominantly expressed in melanoma and prevents miR-340-mediated repression of *MITF*[146]. Whereas IGF2BP1 proteins are barely detectable in normal cells, overexpression is seen in human tumour cells and correlates with high *MYC* mRNA levels and poor prognosis[145,147]. Moreover, knockdown of IGF2BP1 in colorectal cancer cells reduces colony formation and stimulates apoptosis[142]. Thus, part of IGF2BP1's multifunctional activity is to bind mRNAs, prevent miRNA-mediated repression and regulate tumour progression.

The mRNA 3′ UTR of *CDKN1B* (also known as p27), a crucial inhibitor of cell cycle progression, harbours an 8-nt recognition motif for pumilio 1 (PUM1) near a miR-221 and miR-222 target region[59,148] (FIG. 3f). Under starved conditions, these reverse complementary regions form a stem-loop RNA structure, prohibiting miRNA binding[149]. However, on growth factor stimulation, both phosphorylation and upregulation of PUM1 promote RNA binding activity, exposing the miR-221 and miR-222 site and allowing repression of *CDKN1B*. In fact, several other cell cycle regulators have been identified as human PUM1 targets and are repressed through their mRNA 3′ UTR[148,150]. Furthermore, enrichment of PUM motifs in low-accessibility target 3′ UTRs of miR-410, together with consistent target expression data in various cancer cell lines, support the idea that PUM1 cooperates with miRNAs by inducing conformational changes[151].

As reported by Steitz and colleagues[152], the repression of ARE-containing mRNAs switches to translational activation on growth factor deprivation and subsequent cell cycle exit. Whereas miRNAs act as repressors under proliferative conditions, the binding of miRNAs to AREs now stabilizes target mRNAs by an unspecified mechanism[153]. Strikingly, besides complementary miRNA binding, this process requires recruitment of the RBPs AGO2 and fragile X mental retardation syndrome-related protein 1 (FXR1) to the ARE. FXR1 may function as a translational repressor by binding to the ARE in proliferating cells[154]. However, despite these and other data, the mechanism for activation has not yet been elucidated.

Alternative splicing and polyadenylation. Genome-wide analysis of mature mRNAs reveals alternative transcripts of individual genes with diverse composition and length. About half of the mammalian genes express mRNA isoforms varying in 3′ UTR length or sequence as a result of alternative polyadenylation (APA)[155]. APA can occur in two modes: 3′ exon switching, which requires splicing-dependent

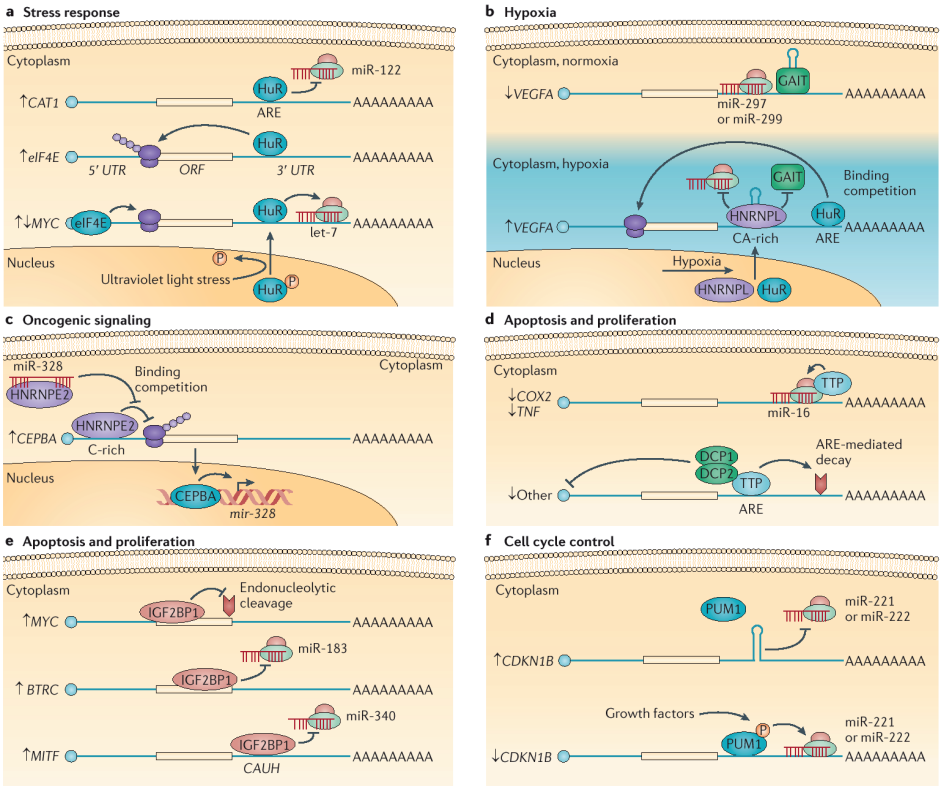


Figure 3 | Mechanisms of microRNA–RNA-binding protein interplay in various cellular processes. Schematics of the versatile roles of RNA-binding proteins (RBPs) in regulation of microRNA (miRNA)-mediated repression under various conditions. Deregulation of miRNAs or RBPs in cancer can be due to altered expression, localization, activity or stability of these regulators. Also, both RBPs and miRNAs are dependent on the secondary structure of their target RNA for accessibility. The mRNA is presented linearly for simplicity. **(a)** In response to cellular stress, HuR is dephosphorylated and translocates, which relieves cationic amino acid transporter 1 (CAT1) mRNA from miR-122-mediated repression. HuR is also required for let-7-mediated repression of MYC mRNA. Binding of HuR to AU-rich elements (AREs) in the 3′ untranslated region (UTR) of eukaryotic translation initiation factor 4E (eIF4E) mRNA promotes production of eIF4E protein, which is a positive modulator of MYC mRNA stability. **(b)** Under hypoxic conditions, heterogeneous nuclear ribonucleoprotein L (HNRNPL) and HuR translocate and target the 3′ UTR of vascular endothelial growth factor (VEGF) mRNA. Through competitive binding of CA-rich sites, HNRNPL relieves miR-297- or miR-299-mediated repression of VEGF mRNA. Also, HNRNPL binding causes a change in secondary structure of VEGF mRNA that impedes translational repression by interferon-γ-activated inhibitor of translation (GAIT). VEGF mRNA can be stabilized by HuR binding to AREs downstream of the 3′ UTR. **(c)** The binding of HNRNPE2 to C-rich regions in the 5′UTR of CCAAT/enhancer binding protein-α (CEPBA) mRNA represses protein translation. The C-rich mature form of miR-328 competes with CEPBA for binding of HNRNPE2, which is induced by CEPBA-stimulated transcription of miR-328. **(d)** Tristetraprolin (TTP) interaction with an Argonaute (AGO) protein enables miR-16-mediated repression of tumour necrosis factor (TNF) and cyclooxygenase 2 (COX2). TTP is also known to bind AREs and promote ARE-mediated mRNA decay by recruiting components of the degradation machinery, as is demonstrated for VEGFA, hypoxia inducible factor 1α (HIF1A), large tumour suppressor 2 (LATS2) and MYC. **(e)** On binding of insulin-like growth

factor 2 mRNA-binding protein 1 (IGF2BP1) to the coding region of MYC mRNA, the mRNA is sequestered and protected from cleavage and subsequent decay by an endoribonuclease within a region termed the coding region determinant (CRD). IGF2BP1 also relieves miR-183-mediated targeting of the coding region of BTRC mRNA and miR-340-mediated targeting of the 3′ UTR of microphthalmia-induced transcription factor (MITF) mRNA, where IGF2BP1 is thought to bind on recognition of a CAUH motif. **(f)** Phosphorylation of pumilio 1 (PUM1) activates binding of this RBP to the UGUANAUA pumilio recognition element (PRE). The change in secondary structure of cyclin-dependent kinase inhibitor 1B (CDKN1B) mRNA enables miR-221 or miR-222 to access its binding site in the 3′ UTR.

terminal exon selection, and tandem UTRs, where different polyadenylation sites (PASs) occur in the same terminal exon. Recognition of a PAS by the polyadenylation machinery is directed by RNA cleavage factors, which select the site of cleavage and determine 3′ UTR length. Interestingly, switching to shorter 3′ UTR forms may circumvent 3′ UTR-mediated regulation by miRNAs and RBPs, conveying changes in mRNA and protein abundance[156,157].

Whereas in the majority of 3′ UTRs with APA the canonical poly(A) signal is distally located, substantial expression of mRNAs with shorter 3′ UTRs is observed in proliferating and less differentiated cells[146,156,158]. For several genes, mRNA isoforms with shorter 3′ UTRs are expressed at higher levels in transformed cells than in non-transformed cells, possibly owing to changes in RBP activity[158]. In human colorectal cancer cell lines, binding of TTP to the *COX2* 3′ UTR promotes the use of the proximal PAS[134]. Overexpression of a shorter, but not full length mRNA isoform of *IGF2BP1* leads to oncogenic transformation[158]. However, the exact changes in APA that correlate with tumour development, as well as the mechanisms that operate to control them, are largely unknown.

CONCLUSIONS AND PERSPECTIVES

MiRNAs are often deregulated in cancer. However, miRNA deregulation is rarely caused by gene amplification or disruption, most probably owing to redundant functions in different genomic loci. Frequently, changes in transcription rate and processing, as well as miRNA activity, are observed. Although transcription factors and chromatin modulators account for alterations in miRNA production, which explain several cases of miRNA overexpression in cancer, RBPs and their interacting partners mostly cause the changes observed in miRNA processing and activity. In cancer these range from disruption of miRNA biogenesis core components to changes in the secondary structure of miRNA target sites, and from a global effect on miRNA processing to specific regulation of select 3′ UTRs. Interestingly, in response to external or internal stimuli, RBPs can dynamically shape the extent of miRNA-mediated repression to maintain robust gene expression. In this way, the 3′ UTR can be considered a multi-faceted docking platform for post-transcriptional regulators that either synergistically or antagonistically fine-tune gene expression in time and space. The fact that 3′ UTRs frequently contain multiple evolutionary conserved binding sites for both miRNAs and RBPs suggests that the interplay between RBPs and miRNAs is a crucial component of gene regulation. Novel high-throughput

Table 2 | Current methods for isolating RNPs to identify interactions between RNA and RBPs or miRISC

Method	Procedure	Ref
RIP-chip	Microarray profiling of endogenous mRNAs associated with immunoprecipitated RBPs and identification of sequence motifs among the bound targets	191
SELEX	Immunoprecipitation of RBPs bound to artificial 52-nt RNAs in vitro, followed by cDNA sequencing to identify sequence motifs	192
Genomic SELEX	Analogous to SELEX, but using a genome-based RNA pool, generated by random priming and in vitro transcription to reduce complexity and increase sensitivity	193
HITS-CLIP	RBP immunoprecipitation with prior in vivo ultraviolet-light-mediated crosslinking of RNA–protein interactions, followed by deep sequencing of linked RNA fragments	194
Ribotrap	Expression of a reporter mRNA containing a 3' untranslated region recognition site for a known RBP, followed by RBP immunoprecipitation and analysis of associated RNP components by mass spectrometry	195
RNA competition	Definition of RBP binding preferences by microarray analysis of RBP-bound sequences in vitro from an abundant pool of 29–38-nt small RNAs. This approach yields binding preferences for either structured or unstructured RNA	196
PAR-CLIP	Analogous to HITS-CLIP, but using incorporation of the nucleotide analogue 4SU into the RNA, which allows efficient crosslinking and direct identification of the RBP binding site by deep sequencing	143
RaPID	Identification of RNP components associated to RNA-aptamer tagged mRNA in vivo by mass spectrometry, which allows detection of different RNA species captured in the same RNP by quantitative real-time PCR	197

4SU, 4-thiouridine; HITS-CLIP, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation; miRISC, microRNA-induced silencing complex; nt, nucleotide; PAR-CLIP, photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation; RaPID, RBP purification and identification; RBP, RNA-binding protein; RIP-chip, RBP immunoprecipitation on cDNA array chip; RNP, ribonucleo protein; SELEX, systemic evolution of ligands by exponential enrichment.

techniques to measure RNA–RNA and RNA–protein interactions (reviewed in REF. 159) (TABLE 2), as well as to monitor mRNA secondary structure[160], should enable us to connect networks of post-transcriptional regulation and decipher their relevance for cancer initiation and progression. This exciting playground of RBPs and miRNAs still holds secrets that, when uncovered, hopefully will reveal networks with potential therapeutic benefits.

ACKNOWLEDGEMENTS

We thank A. Morris for critical reading of the manuscript and the members of the R.A. group for their suggestions. This work was supported by the EURYI (European Research Young Investigator Award), ERC (European Research Council), KWF (Koningin Wilhelmina Fonds; Dutch Cancer Foundation) and Horizon-NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; R.A.). We apologize to colleagues whose work was not cited because of space limitations.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

FURTHER INFORMATION

CLIPZ (open access database and analysis environment): http://www.clipz.unibas.ch
Homo sapiens in the UCSC Genome Browser: http://genome.ucsc.edu/cgi-bin/hgGateway
miRBase: http://www.mirbase.org
RBPDB (RNA Binding Protein Database): http://rbpdb.ccbr.utoronto.ca
RNAfoldWebServer: http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi
starBase: http://starbase.sysu.edu.cn

REFERENCES

1. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233 (2009).
2. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105 (2009).
3. Lu, J. et al. MicroRNA expression profiles classify human cancers. Nature 435, 834–838 (2005). **The results presented in this paper had a big impact. It was found that miRNA expression profiles could be used to discriminate between normal and tumour tissue and, in addition, they reflected the developmental origin and differentiation state of the tumour. This led to discussion of the possible application of this finding in diagnosis and raised queries about the cause of global miRNA deregulation in tumours.**
4. Volinia, S. et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc. Natl Acad. Sci. USA 103, 2257–2261 (2006). **Based on a panel of 540 solid tumour samples, an miRNA expression signature was presented that contains miRNAs that are differentially expressed in tumours. Many of their predicted targets were tumour suppressors or oncogenes, turning the miRNA genes into novel cancer genes.**
5. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. Nature Rev. Mol. Cell Biol. 8, 479–490 (2007).
6. Lukong, K. E., Chang, K. W., Khandjian, E. W. & Richard, S. RNA-binding proteins in human genetic disease. Trends Genet. 24, 416–425 (2008).
7. Riggi, N., Cironi, L., Suvà, M. L. & Stamenkovic, I. Sarcomas: genetics, signalling, and cellular origins. Part 1: the fellowship of TET. J. Pathol. 213, 4–20 (2007).
8. Rajan, P. et al. Regulation of gene expression by the RNA-binding protein

Sam68 in cancer. Biochem. Soc. Trans. 36, 505–507 (2008).

9. Lee, H. K. et al. β-catenin regulates multiple steps of RNA metabolism as revealed by the RNA aptamer in colon cancer cells. Cancer Res. 67, 9315–9321 (2007).
10. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev. 24, 2343–2364 (2010).
11. Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. Nature 432, 231–235 (2004).
12. Gregory, R. I. The Microprocessor complex mediates the genesis of microRNAs. Nature 432, 235–240 (2004).
13. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 17, 3011–3016 (2003).
14. Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. Science 303, 95–98 (2004).
15. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentateribonuclease in the initiation step of RNA interference. Nature 409, 363–366 (2001).
16. Grishok, A. et al. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell 106, 23–34 (2001).
17. Hutvágner, G. et al. A cellular function for the RNA- interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293, 834–838 (2001).
18. Chendrimada, T. P. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 436, 740–744 (2005).
19. Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. Cell 118, 57–68 (2004).
20. MacRae, I. J., Zhou, K. & Doudna, J. A. Structural determinants of RNA recognition and cleavage by Dicer. Nature Struct. Mol. Biol. 14, 934–940 (2007).
21. Haase, A. D. et al. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. EMBO Rep. 6, 961–967 (2005).
22. Chakravarthy, S., Sternberg, S. H., Kellenberger, C. A. & Doudna, J. A. Substrate-specific kinetics of dicer-catalyzed RNA processing. J. Mol. Biol. 404, 392–402 (2010).
23. Lee, Y. et al. The role of PACT in the RNA silencing pathway. EMBO J. 25, 522–532 (2006).
24. Meister, G. et al. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol. Cell 15, 185–197 (2004).
25. Cerutti, L., Mian, N. & Bateman, A. Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. Trends Biochem. Sci. 25, 481–482 (2000).
26. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305, 1434–1437 (2004).
27. Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature 429, 318–322 (2004).
28. Diederichs, S. & Haber, D. A. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell 131, 1097–1108 (2007).
29. Cheloufi, S., Dos Santos, C. O., Chong, M. M. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature 465, 584–589 (2010).
30. Cifuentes, D. et al. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science 328, 1694–1698 (2010).
31. Ambros, V. et al. A uniform system for microRNA annotation. RNA 9, 277–279 (2003).
32. Hutvágner, G. Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. FEBS Lett. 579, 5850–5857 (2005).
33. Thomson, J. M. et al. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. Genes Dev. 20, 2202–2207 (2006).
34. Zhang, L. et al. microRNAs exhibit high frequency genomic alterations in human cancer. Proc. Natl Acad. Sci. USA 103, 9136–9141 (2006).

35. Kim, M. S. et al. Somatic mutations and losses of expression of microRNA regulation-related genes AGO2 and TNRC6A in gastric and colorectal cancers. J. Pathol. 221, 139–146 (2010).
36. Lee, E. J. et al. Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. RNA 14, 35–42 (2008).
37. Melo, S. A. et al. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. Cancer Cell 18, 303–315 (2010).
38. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nature Genet. 39, 673–677 (2007). **This report provided evidence that impairment of miRNA biogenesis contributes to oncogenic transformation and is not merely a result of it. Independent knockdown of miRNA pathway components caused a global loss of miRNA (as observed in reference 3) and promoted tumour formation.**
39. Lambertz, I. et al. Monoallelic but not biallelic loss of Dicer1 promotes tumorigenesis in vivo. Cell Death Differ. 17, 633–641 (2010).
40. Hill, D. A. et al. DICER1 mutations in familial pleuropulmonaryblastoma. Science 325, 965 (2009).
41. Kumar, M. S. et al. Dicer1 functions as a haploinsufficienttumor suppressor. Genes Dev. 23, 2700–2704 (2009).
42. Martello, G. et al. A microRNA targeting Dicer for metastasis control. Cell 141, 1195–1207 (2010).
43. Melo, S. A. et al. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. Nature Genet. 41, 365–370 (2009).
44. Garre, P., Pérez-Segura, P., Díaz-Rubio, E., Caldés, T. & de la Hoya, M. Reassessing the TARBP2 mutation rate in hereditary nonpolyposis colorectal cancer. Nature Genet. 42, 817–818 (2010).
45. Paroo, Z., Ye, X., Chen, S. & Liu, Q. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erksignaling. Cell 139, 112–122 (2009).
46. Han, J. Posttranscriptional crossregulation between Drosha and DGCR8. Cell 136, 75–84 (2009).
47. Obernosterer, G., Leuschner, P. J., Alenius, M. & Martinez, J. Post-transcriptional regulation of microRNA expression. RNA 12, 1161–1167 (2006).
48. Fuller-Pace, F. V. & Moore, H. C. RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development. Future Oncol. 7, 239–251 (2011).
49. Fukuda, T. et al. DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. Nature Cell Biol. 9, 604–611 (2007).
50. Suzuki, H. I. et al. Modulation of microRNA processing by p53. Nature 460, 529–533 (2009).
51. Feng, Z., Zhang, C., Wu, R. & Hu, W. Tumor suppressor p53 meets microRNAs. J. Mol. Cell Biol. 3, 44–50 (2011).
52. Yamagata, K. Maturation of microRNA is hormonally regulated by a nuclear receptor. Mol. Cell 36, 340–347 (2009).
53. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. Nature 454, 56–61 (2008).
54. Davis, B. N., Hilyard, A. C., Nguyen, P. H., Lagna, G. & Hata, A. Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. Mol. Cell 39, 373–384 (2010).
55. Yu, B. The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc. Natl Acad. Sci. USA 105, 10073–10078 (2008).
56. Wendt, M. K., Allington, T. M. & Schiemann, W. P. Mechanisms of the epithelial-mesenchymal transition by TGF-β. Future Oncol. 5, 1145–1168 (2009).
57. Wu, H. A splicing-independent function of SF2/ASF in microRNA processing. Mol. Cell 38, 67–77 (2010).
58. Karni, R. et al. The gene encoding the splicing factorSF2/ASF is a proto-oncogene. Nature Struct. Mol. Biol.14, 185–193 (2007).
59. le Sage, C. et al. Regulation of the p27(Kip1) tumorsuppressor by miR-221 and miR-222 promotes cancer cell proliferation. EMBO J. 26, 3699–3708 (2007).
60. Nakamura, T., Canaani, E. & Croce, C. M. Oncogenic All1 fusion proteins target Drosha-mediated microRNA processing. Proc. Natl Acad. Sci. USA 104, 10980–10985 (2007).

61. Rybak, A. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nature Cell Biol. 10, 987–993 (2008).
62. Viswanathan, S. R. & Daley, G. Q. Lin28: a microRNA regulator with a macro role. Cell 140, 445–449 (2010).
63. Iliopoulos, D., Jaeger, S. A., Hirsch, H. A., Bulyk, M. L. & Struhl, K. STAT3 activation of miR-21 and miR-181b-1 via PTEN and CYLD are part of the epigenetic switch linking inflammation to cancer. Mol. Cell 39, 493–506 (2010).
64. Trabucchi, M. The RNA-binding protein KSRPpromotes the biogenesis of a subset of microRNAs. Nature 459, 1010–1014 (2009).
65. Michlewski, G. & Cáceres, J. F. Antagonistic role ofhnRNP A1 and KSRP in the regulation of let-7a biogenesis. Nature Struct. Mol. Biol. 17, 1011–1018 (2010).
66. Guil, S. & Cáceres, J. F. The multifunctional RNA- binding protein hnRNP A1 is required for processing of miR-18a. Nature Struct. Mol. Biol. 14, 591–596 (2007).
67. Michlewski, G., Guil, S., Semple, C. A. & Caceres, J. F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. Mol. Cell 32, 383–393 (2008).
68. Zhang, X. & Zeng, Y. The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res. 38, 7689–7697 (2010).
69. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Rev. Genet. 9, 102–114 (2008).
70. Eulalio, A., Huntzinger, E. & Izaurralde, E. Getting to the root of miRNA-mediated gene silencing. Cell 132, 9–14 (2008).
71. Hutvágner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. Science 297, 2056–2060 (2002).
72. Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110, 563–574 (2002).
73. Höck, J. et al. Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. EMBO Rep. 8, 1052–1060 (2007).
74. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. 16, 2733–2742 (2002).
75. Azuma-Mukai, A. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc. Natl Acad. Sci. USA 105, 7964–7969 (2008).
76. Su, H., Trombly, M. I., Chen, J. & Wang, X. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. Genes Dev. 23, 304–317 (2009).
77. Eulalio, A., Tritschler, F. & Izaurralde, E. The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. RNA 15, 1433–1442 (2009).
78. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nature Rev. Genet. 12, 99–110 (2011).
79. Djuranovic, S., Nahvi, A. & Green, R. A parsimonious model for gene regulation by miRNAs. Science 331, 550–553 (2011).
80. Baek, D. et al. The impact of microRNAs on protein output. Nature 455, 64–71 (2008).
81. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466, 835–840 (2010).
82. Chu, C. Y. & Rana, T. M. Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol. 4, e210 (2006).
83. Eulalio, A. et al. Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. Genes Dev. 21, 2558–2570 (2007).
84. Lin, F. et al. Knockdown of RCK/p54 expression by RNAi inhibits proliferation of human colorectal cancer cells in vitro and in vivo. Cancer Biol. Ther. 7, 1669–1676 (2008).
85. Salzman, D. W., Shubert-Coleman, J. & Furneaux, H. P68 RNA helicase unwinds the human let-7 microRNA precursor duplex and is required for let-7-directed silencing of gene expression. J. Biol. Chem. 282, 32773–32779 (2007).

86. Meister, G. et al. Identification of novel argonaute- associated proteins. Curr. Biol. 15, 2149–2155 (2005).
87. Chendrimada, T. P. et al. MicroRNA silencing through RISC recruitment of eIF6. Nature 447, 823–828 (2007).
88. Mourelatos, Z. et al. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. 16, 720–728 (2002).
89. Eulalio, A., Behm-Ansmant, I., Schweizer, D. & Izaurralde, E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. Mol. Cell. Biol. 27, 3970–3981 (2007).
90. Landthaler, M. et al. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA 14, 2580–2596 (2008).
91. Pare, J. M. Hsp90 regulates the function of argonaute 2 and its recruitment to stress granules and P-bodies. Mol. Biol. Cell 20, 3273–3284 (2009).
92. Suzuki, Y. et al. The Hsp90 inhibitor geldanamycin abrogates colocalization of eIF4E and eIF4E- transporter into stress granules and association of eIF4E with eIF4G. J. Biol. Chem. 284, 35597–35604 (2009).
93. Iwasaki, S. et al. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol. Cell 39, 292–299 (2010).
94. Andrei, M. A. et al. A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. RNA 11, 717–727 (2005).
95. Johnston, M., Geoffroy, M. C., Sobala, A., Hay, R. & Hutvágner, G. HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. Mol. Biol. Cell 21, 1462–1469 (2010).
96. Rybak, A. The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. Nature Cell Biol. 11, 1411–1420 (2009).
97. Schwamborn, J. C., Berezikov, E. & Knoblich, J. A. The TRIM-NHL protein TRIM32 activates microRNAs and prevents self-renewal in mouse neural progenitors. Cell 136, 913–925 (2009).
98. Neumuller, R. A. Mei-P26 regulates microRNAs and cell growth in the Drosophila ovarian stem cell lineage. Nature 454, 241–245 (2008).
99. Kano, S., Miyajima, N., Fukuda, S. & Hatakeyama, S. Tripartite motif protein 32 facilitates cell growth and migration via degradation of Abl-interactor 2. Cancer Res. 68, 5572–5580 (2008).
100. Weinmann, L. Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. Cell 136, 496–507 (2009).
101. Parker, R. & Song, H. The enzymes and control of eukaryotic mRNA turnover. Nature Struct. Mol. Biol. 11, 121–127 (2004).
102. Zhang, K. et al. The human homolog of yeast SEP1 is a novel candidate tumor suppressor gene in osteogenic sarcoma. Gene 298, 121–127 (2002).
103. Hinman, M. N. & Lou, H. Diverse molecular functions of Hu proteins. Cell. Mol. Life Sci. 65, 3168–3181 (2008).
104. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA- mediated translational repression in human cells subjected to stress. Cell 125, 1111–1124 (2006). **In this paper, the first example of stress-induced relief of miRNA-mediated repression through binding of an RBP to the 3′ UTR was described.**
105. Wang, W. et al. HuR regulates p21 mRNA stabilization by UV light. Mol. Cell. Biol. 20, 760–769 (2000).
106. Kullmann, M., Göpfert, U., Siewe, B. & Hengst, L. ELAV/Hu proteins inhibit p27 translation via an IRES element in the p27 5′UTR. Genes Dev. 16, 3087–3099 (2002).
107. Leandersson, K., Riesbeck, K. & Andersson, T. Wnt-5a mRNA translation is suppressed by the Elav-like protein HuR in human breast epithelial cells. Nucleic Acids Res. 34, 3988–3999 (2006).
108. Saunus, J. M. et al. Posttranscriptional regulation of the breast cancer susceptibility gene BRCA1 by the RNA binding protein HuR. Cancer Res. 68, 9469–9478 (2008).
109. Kim, H. H. et al. HuR recruits let-7/RISC to repress c-Myc expression. Genes Dev. 23, 1743–1748 (2009). **In contrast to the stabilizing function of HuR described in reference 104, this RBP was found to be required for let-7-mediated target repression.**

110. Abdelmohsen, K. & Gorospe, M. Posttranscriptional regulation of cancer traits by HuR. RNA 1, 214–229 (2010).

111. Abdelmohsen, K., Srikantan, S., Kuwano, Y. & Gorospe, M. miR-519 reduces cell proliferation by lowering RNA-binding protein HuR levels. Proc. Natl Acad. Sci. USA 105, 20297–20302 (2008).

112. Guo, X., Wu, Y. & Hartley, R. S. MicroRNA-125a represses cell growth by targeting HuR in breast cancer. RNA Biol. 6, 575–583 (2009).

113. Abdelmohsen, K. et al. miR-519 suppresses tumor growth by reducing HuR levels. Cell Cycle 9, 1354–1359 (2010).

114. Pullmann, R. et al. Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs. Mol. Cell. Biol. 27, 6265–6278 (2007).

115. Topisirovic, I. et al. Stability of eukaryotic translation initiation factor 4E mRNA is regulated by HuR, and this activity is dysregulated in cancer. Mol. Cell. Biol. 29, 1152–1162 (2009). **This paper reports that HuR and eIF4E cooperatively stabilize mRNA targets, while eIF4E mRNA is stabilized by HuR. Such regulative loops add a layer of complexity onto the interplay of RBPs**.

116. Wendel, H. G. et al. Dissecting eIF4E action in tumorigenesis. Genes Dev. 21, 3232–3237 (2007).

117. Furic, L. et al. eIF4E phosphorylation promotes tumorigenesis and is associated with prostate cancer progression. Proc. Natl Acad. Sci. USA 107, 14134–14139 (2010).

118. Mili, S., Shu, H. J., Zhao, Y. & Piñol-Roma, S. Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA. Mol. Cell. Biol. 21, 7307–7319 (2001).

119. Lal, A. et al. Concurrent versus individual binding of HuR and AUF1 to common labile target mRNAs. EMBO J. 23, 3092–3102 (2004).

120. Chen, C. Y. et al. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell 107, 451–464 (2001).

121. Fialcowitz, E. J., Brewer, B. Y., Keenan, B. P. & Wilson, G. M. A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics. J. Biol. Chem. 280, 22406–22417 (2005).

122. Wilson, G. M., Sutphen, K., Chuang, K. & Brewer, G. Folding of A+U-rich RNA elements modulates AUF1 binding. Potential roles in regulation of mRNA turnover. J. Biol. Chem. 276, 8695–8704 (2001).

123. Wilson, G. M., Sutphen, K., Moutafis, M., Sinha, S. & Brewer, G. Structural remodelling of an A + U-rich RNA element by cation or AUF1 binding. J. Biol. Chem. 276, 38400–38409 (2001).

124. Shih, S. C. & Claffey, K. P. Regulation of human vascular endothelial growth factor mRNA stability in hypoxia by heterogeneous nuclear ribonucleoprotein L. J. Biol. Chem. 274, 1359–1365 (1999).

125. Jafarifar, F., Yao, P., Eswarappa, S. M. & Fox, P. L. Repression of VEGFA by CA-rich element-binding microRNAs is modulated by hnRNP L. EMBO J. 30, 1324–1334 (2011). **The binding of HNRNPL to the VEGFA 3′ UTR competes with miRNA binding and changes the RNA conformation.**

126. Ray, P. S. et al. A stress-responsive RNA switch regulates VEGFA expression. Nature 457, 915–919 (2009). **Under hypoxic conditions, the binding of HNRNPL to the VEGFA 3′ UTR changes the RNA secondary structure and relieves IFNγ-induced repression of VEFGA by GAIT, showing how the response to combined stimuli is orchestrated. Together with reference 125, these data demonstrate how HNRNPL can relieve two repressive mechanisms simultaneously.**

127. Goldberg-Cohen, I., Furneauxb, H. & Levy, A. P. A 40-bp RNA element that mediates stabilization of vascular endothelial growth factor mRNA by HuR. J. Biol. Chem. 277, 13635–13640 (2002).

128. Eiring, A. M. miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. Cell 140, 652–665 (2010). **This paper uniquely describes a decoy function of an miRNA. miR-328 competes with CEPBA mRNA for binding of HNRNPE2. CEPBA is a transcriptional activator of the miR-328 gene, thus forming a positive feedback loop.**

129. Chang, J. S. et al. High levels of the BCR/ABL oncoprotein are required for the MAPK-hnRNP-E2 dependent suppression of C/EBPα-driven myeloid differentiation. Blood 110, 994–1003 (2007).

130. Fazi, F. et al. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPα regulates human granulopoiesis. Cell 123, 819–831 (2005).

131. Evans, J. R. et al. Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo. Oncogene 22, 8012–8020 (2003).

132. Fenger-Grøn, M., Fillman, C., Norrild, B. & Lykke-Andersen, J. Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. Mol. Cell 20, 905–915 (2005).

133. Carballo, E., Lai, W. S. & Blackshear, P. J. Feedback inhibition of macrophage tumor necrosis factor-α production by tristetraprolin. Science 281, 1001–1005 (1998).

134. Sawaoka, H., Dixon, D. A., Oates, J. A. & Boutaud, O. Tristetraprolin binds to the 3′-untranslated region of cyclooxygenase-2 mRNA. A polyadenylation variant in a cancer cell line lacks the binding site. J. Biol. Chem. 278, 13928–13935 (2003).

135. Lai, W. S., Parker, J. S., Grissom, S. F., Stumpo, D. J. & Blackshear, P. J. Novel mRNA targets for tristetraprolin (TTP) identified by global analysis of stabilized transcripts in TTP-deficient fibroblasts. Mol. Cell. Biol. 26, 9196–9208 (2006).

136. Marderosian, M. et al. Tristetraprolin regulates Cyclin D1 and c-Myc mRNA stability in response to rapamycin in an Akt-dependent manner via p38 MAPK signaling. Oncogene 25, 6277–6290 (2006).

137. Lee, H. H. et al. Stability of the LATS2 tumor suppressor gene is regulated by tristetraprolin. J. Biol. Chem. 285, 17329–17337 (2010).

138. Stoecklin, G. et al. Genome-wide analysis identifies interleukin-10 mRNA as target of tristetraprolin. J. Biol. Chem. 283, 11689–11699 (2008).

139. Brennan, S. E. et al. The mRNA-destabilizing protein tristetraprolin is suppressed in many cancers, altering tumorigenic phenotypes and patient prognosis. Cancer Res. 69, 5168–5176 (2009).

140. vonRoretz, C. & Gallouzi, I. E. Decoding ARE- mediated decay: is microRNA part of the equation? J. Cell Biol. 181, 189–194 (2008).

141. Jing, Q. et al. Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell 120, 623–634 (2005). **These data indicate that the RBP TTP, an mRNA decay factor that binds AREs and generally operates by uncapping, assists miR-16 in targeting AREs by interacting with AGO2 or other RISC factors, showing the versatility of RBPs.**

142. Noubissi, F. K. et al. CRD-BP mediates stabilization of βTrCP1 and c-myc mRNA in response to β-catenin signalling. Nature 441, 898–901 (2006).

143. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–141 (2010).

144. Weidensdorfer, D. et al. Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. RNA 15, 104–115 (2009).

145. Elcheva, I., Goswami, S., Noubissi, F. K. & Spiegelman, V. S. CRD-BP protects the coding region of βTrCP1 mRNA from miR-183-mediated degradation. Mol. Cell 35, 240–246 (2009). **This paper shows that the RBP IFG2BP1 can bind a coding region in a specific target and can alleviate the interaction between miR-183 in complex with RISC and its target site, which is also located in the coding region.**

146. Goswami, S. et al. MicroRNA-340-mediated degradation of microphthalmia-associated transcription factor mRNA is inhibited by the coding region determinant-binding protein. J. Biol. Chem 285, 20532–20540 (2010). **The RBP IGF2BP1 can also antagonize miRNA-mediated repression through the 3′ UTR, illustrating the multifunctional character of this protein.**

147. Köbel, M. et al. Expression of the RNA-binding protein IMP1 correlates with poor prognosis in ovarian carcinoma. Oncogene 26, 7584–7589 (2007).

148. Galgano, A. et al. Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. PLoS ONE 3, e3164 (2008).

149. Kedde, M. et al. A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. Nature Cell Biol. 12, 1014–1020 (2010). **This is the first report of a reverse complementary RBP recognition site and miRNA site that form a stem–loop structure in the absence of the RBP PUM1. This emphasizes the**

importance of RNA secondary structure for miRNA accessibility.

150. Morris, A. R., Mukherjee, N. & Keene, J. D. Ribonomic analysis of human Pum1 reveals cis-trans conservation across species despite evolution of diverse mRNA target sets. Mol. Cell. Biol. 28, 4093–4103 (2008).
151. Leibovich, L., Mandel-Gutfreund, Y. & Yakhini, Z. A structural-based statistical approach suggests a cooperative activity of PUM1 and miR-410 in human 3′-untranslated regions. Silence 1, 17 (2010).
152. Vasudevan, S. & Steitz, J. A. AU-rich-element- mediated upregulation of translation by FXR1 and Argonaute 2. Cell 128, 1105–1118 (2007).
153. Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. Science 318, 1931–1934 (2007).
154. Garnon, J. et al. Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level. J. Biol. Chem. 280, 5750–5763 (2005).
155. Zhang, H., Lee, J. Y. & Tian, B. Biased alternative polyadenylation in human tissues. Genome Biol. 6, R100 (2005).
156. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science 320, 1643–1647 (2008). **This paper shows that shorter 3′ UTRs, as expressed in proliferating and undifferentiated cells, confer post-transcriptional mRNA stability and higher protein levels. This implies that the function of the 3′ UTR is mainly repressive and that regulation by repressors acting on the 3′ UTR seems to be essential for maintaining correct gene expression.**
157. Legendre, M., Ritchie, W., Lopez, F. & Gautheret, D. Differential repression of alternative transcripts: a screen for miRNA targets. PLoSComput. Biol. 2, e43 (2006).
158. Mayr, C. & Bartel, D. P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138, 673–684 (2009). **In this paper, the consequences of alternative PAS usage for oncogenic transformation are shown by isoforms of IGF2BP1, in agreement with the data presented in reference 156.**
159. Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: global insights into biological networks. Nature Rev. Genet. 11, 75–87 (2010).
160. Kertesz, M. et al. Genome-wide measurement of RNA secondary structure in yeast. Nature 467, 103–107 (2010).
161. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34, D140–D144 (2006).
162. Berezikov, E., Cuppen, E. & Plasterk, R. H. Approaches to microRNA discovery. Nature Genet. 38, S2–S7 (2006).
163. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401–1414 (2007).
164. Forman, J. J. & Coller, H. A. The code within the code: microRNAs target coding regions. Cell Cycle 9, 1533–1541 (2010).
165. Stefani, G. & Slack, F. J. Small non-coding RNAs in animal development. Nature Rev. Mol. Cell Biol. 9, 219–230 (2008).
166. Farh, K. K. et al. The widespread impact of mammalian microRNAs on mRNA repression and evolution. Science 310, 1817–1821 (2005).
167. Kloosterman, W. P. & Plasterk, R. H. The diverse functions of microRNAs in animal development and disease. Dev. Cell 11, 441–450 (2006).
168. Lima, R. T. et al. MicroRNA regulation of core apoptosis pathways in cancer. Eur. J. Cancer 47, 163–174 (2011).
169. Negrini, M., Nicoloso, M. S. & Calin, G. A. MicroRNAsand cancer — new paradigms in molecular oncology.Curr. Opin. Cell Biol. 21, 470–479 (2009).
170. Urbich, C., Kuehbacher, A. & Dimmeler, S. Role ofmicroRNAs in vascular diseases, inflammation, andangiogenesis. Cardiovasc. Res. 79, 581–588 (2008).
171. Gregory, P. A., Bracken, C. P., Bert, A. G. & Goodall, G. J. MicroRNAs as regulators of epithelial-mesenchymaltransition. Cell Cycle 7, 3112–3118 (2008).
172. Nicoloso, M. S., Spizzo, R., Shimizu, M., Rossi, S. & Calin, G. A. MicroRNAs — the micro steering wheel of tumour metastases. Nature Rev. Cancer 9, 293–302 (2009).
173. Olson, P. et al. MicroRNA dynamics in the stages of tumorigenesis correlate with hallmark capabilities of cancer. Genes Dev. 23, 2152–2165 (2009).

174. Volinia, S. et al. Reprogramming of miRNA networks in cancer and leukemia. Genome Res. 20, 589–599 (2010).

175. Voorhoeve, P. M. MicroRNAs: oncogenes, tumor suppressors or master regulators of cancer heterogeneity? Biochim. Biophys. Acta 1805, 72–86 (2010).

176. Farazi, T. A., Spitzer, J. I., Morozov, P. & Tuschl, T. miRNAs in human cancer. J. Pathol. 223, 102–115 (2011).

177. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. Nature Rev. Cancer 6, 857–866 (2006).

178. Ventura, A. & Jacks, T. MicroRNAs and cancer: short RNAs go a long way. Cell 136, 586–591 (2009).

179. Ryan, B. M., Robles, A. I. & Harris, C. C. Genetic variation in microRNA networks: the implications for cancer research. Nature Rev. Cancer 10, 389–402 (2010).

180. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. Nature Rev. Mol. Cell Biol. 10, 126–139 (2009).

181. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. Nature Rev. Genet. 11, 597–610 (2010).

182. Kedde, M. et al. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell 131, 1273–1286 (2007). **This paper describes an RBP, which interacts with the 3′ UTR of some germ cell genes and allows their expression by preventing miRNA-mediated repression during germ cell development in zebrafish.**

183. Voorhoeve, P. M. et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell 124, 1169–1181 (2006).

184. Huang, Q. et al. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. Nature Cell Biol. 10, 202–210 (2008).

185. Bhattacharya, C., Aggarwal, S., Kumar, M., Ali, A. & Matin, A. Mouse apolipoprotein B editing complex 3 (APOBEC3) is expressed in germ cells and interacts with dead-end (DND1). PLoS ONE 3, e2315 (2008).

186. Takeda, Y., Mishima, Y., Fujiwara, T., Sakamoto, H. & Inoue, K. DAZL relieves miRNA-mediated repression of germline mRNAs by controlling poly(A) tail length in zebrafish. PLoS ONE 4, e7513 (2009). **This paper states that miRNA-mediated repression can be counteracted through poly(A)-tail elongation controlled by an RBP during germ cell development in zebrafish.**

187. Moore, F. L. et al. Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. Proc. Natl Acad. Sci. USA 100, 538–543 (2003).

188. Fox, M., Urano, J. & ReijoPera, R. A. Identification and characterization of RNA sequences to which human PUMILIO-2 (PUM2) and deleted in Azoospermia-like (DAZL) bind. Genomics 85, 92–105 (2005).

189. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. Annu. Rev. Biochem. 79, 351–379 (2010).

190. Tritschler, F., Huntzinger, E. & Izaurralde, E. Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of déjà vu. Nature Rev. Mol. Cell Biol. 11, 379–384 (2010).

191. Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc. Natl Acad. Sci. USA 97, 14085–14090 (2000).

192. Galarneau, A. & Richard, S. Target RNA motif and target mRNAs of the Quaking STAR protein. Nature Struct. Mol. Biol. 12, 691–698 (2005).

193. Lorenz, C., von Pelchrzim, F. & Schroeder, R. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. Nature Protoc. 1, 2204–2212 (2006).

194. Licatalosi, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456, 464–469 (2008).

195. Beach, D. L. & Keene, J. D. Ribotrap: targeted purification of RNA-specific RNPs from cell lysates through immunoaffinity precipitation to identify regulatory proteins and RNAs. Methods Mol. Biol. 419, 69–91 (2008).

196. Ray, D. et al. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nature Biotech. 27, 667–670 (2009).

197. Slobodin, B. & Gerst, J. E. A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. RNA 16, 2277–2290 (2010).

chapter **THREE**

REGULATION OF MICRORNA-371-373 EXPRESSION AT THE LEVEL OF TRANSCRIPTION

Marieke van Kouwenhove[1], Joachim A.F. Oude Vrielink[1], Amaia Lujambio[2], Manel Esteller[2], Ran Elkon[1], Leendert H.J. Looijenga[3], Reuven Agami[1]

[1]Division of Gene Regulation, The Netherlands Cancer Institute, Amsterdam, The Netherlands.
[2]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute, Barcelona, Catalonia, Spain.
[3]Department of Pathology, Josephine Nefkens Institute, Erasmus MC, Rotterdam, The Netherlands.

ABSTRACT

The expression of microRNAs (miRNAs) is dynamically regulated during development, but sometimes appears deregulated in cancer. The gene cluster of miR-371-373 is specifically expressed during embryonic development, and its expression is also found in testicular germ cell tumours (GCTs). MiRNAs 372 and 373 can overcome a tumour-suppressive response to oncogenic stress in the presence of wildtype p53 and contribute to oncogenic transformation of human cells[1]. A putative promoter region immediately upstream of the miRNA gene showed transcriptional activation in GCT cell lines. This region comprises functional binding motifs for pluripotency-related transcription factors (TFs) Sox2 and Klf4. In somatic cell lines, combined expression of pluripotency-related TFs establishes transcriptional activation mediated by this region. Also, methylation of CpGs in vicinity of the miRNA gene decreased in somatic cells upon overexpression of both Sox2 and Klf4. Correlation analysis between expression of miRNA-372/373 and the subset of predicted mRNA targets indicate that these miRNAs are functional antagonists of target expression in GCT cell lines. We conclude that the expression of miR-371-373 can be driven by binding of Sox2 and Klf4 to the promoter region, possibly in combination with Oct3/4.

INTRODUCTION

MiRNAs are small non-coding RNAs of ~21 nucleotides in length that negatively regulate gene expression of specific target genes. In eukaryotic organisms, the mechanism of miRNA-mediated gene repression is based on partial complementarity of the miRNA sequence to sites in the target mRNA 3' untranslated region (3' UTR) and subsequent inhibition of translation and mRNA decay[2]. Over 50% of the human genome is estimated to be targeted by miRNAs, consequently these regulators have a role in most cellular processes. The expression levels of miRNAs are tightly controlled in time and space, whereby post-transcriptional regulation of multiple miRNA target genes can be coordinated in parallel. Disruption of miRNA-mediated gene regulation is implicated in many diseases, and also in cancer. In various tumour types, differential expression of miRNAs has been reported resulting in changes in target expression levels (reviewed in REF. 3). Furthermore, miRNAs can serve as markers for diagnosis, prognosis and treatment response in several tumours (reviewed in REF. 4).

Moreover, miRNA expression is essential during human embryonic development. Knockdown of principal factors in miRNA biogenesis results in embryonic lethality[5,6]. Whereas the expression of most miRNAs is gradually upregulated during embryonic development alongside a loss of pluripotency, one particular miRNA family is highly expressed during early development covering 80% of the total miRNA expression[7]. This miRNA family comprises four genes or gene clusters in total (miR-93, 302-367, 371-373 and 518-520) and its expression is restricted to embryonic stem cells (ESCs) and primordial germ cells (PGCs). Furthermore, the expression of miR-371-373 as well as miR-302 and miR-518-520 is deregulated in some testicular GCTs, a malignancy that has its origin around the time of PGC migration during embryonic development[8]. In a screen for miRNAs as potential oncogenes, the miRNAs 372 and 373 showed resistance to a RAS^{V12}-induced oncogenic stress response and stimulated transformation of primary human cells in the presence of wildtype p53[1]. Moreover, miR-373 and family member miR-520c are implicated in tumour cell migration and invasion[9].

In this project, we investigated whether inappropriate transcriptional activation of miR-371-373 can explain their expression in GCTs**.** The location and characterization of the core promoter region of miRNA genes is essential to investigate mechanisms and conditions of their activation. MiRNA genes have different locations in the genome: intergenic miRNA genes contain their own transcriptional units where as intronic miRNAs are co-transcribed with their host genes[10]. For many miRNA genes, one or more promoter regions or transcription start sites (TSS) have been inferred from *in silico* predictions, but few miRNA promoters have been empirically characterized[11-17]. For human miRNA cluster 371-373 and for the mouse homolog miR-290-295, predictions of the promoter region have so far been based on nucleosome positioning, histone marks and RNA polymerase II association (overview in Table 1).

The main question is when and how miRNA levels are maintained or re-established during germ cell tumorigenesis (reviewed in REF. 22). During human embryonic development, ESCs start to differentiate into PGCs; both these cell types are characterized by the expression of the pluripotency markers octamer transcription factor 3/4 (Oct3/4), SRY-box containing gene 2 (Sox2) and Nanog, and also miRNAs 371-373. After PGC migration, the expression of miR-371-373 decreases and the

Table 1 | Summary of promoter prediction approaches and results for miR-371-373 in human or miR-290-295 in mouse.

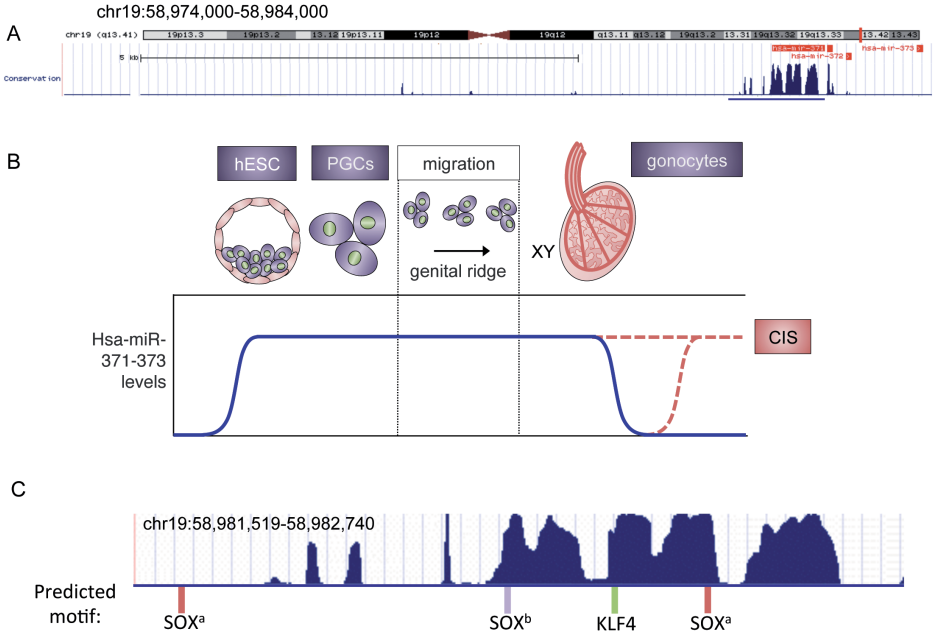
system	cell type	predicted promoter	predicted TSS	parameters	ref
Hsa-miR-371-373 - chr19:58,982,741-58,983,839 Putative promoter region - chr19:58,981,519-58,982,740					
Hsa (miR-371-373)	n/a	chr19:58,981,858-58,982,024	chr19:58,982,010	Comparative genomics	13
Hsa (miR-371-373)	MALME, MCF-7, UACC62	-	chr19:58,970,865	H3K4me3, H3K19/14Ac, nucleosome occupancy, Pol II	18
Hsa (miR-371-373)	ESC	-	chr19:58,980,348-58,981,893 chr19:58,981,893-58,984,412	H3K4me3	7
Hsa (miR-371-373)	CD4+ T cells	-	chr19:58,983,105 (CpG) chr19:58,980,685 (no-CpG)	Histone marks, DNA sequence	19
Hsa (miR-371-373)	n/a	chr19:58,981,942-58,982,291	-	Comparative genomics	20
Hsa (miR-371-373)	HuH-7, Hep3b, HEK293T	-	chr19:54,120,041	H3K4me3, CAGE tags	21
Mmu-miR-290-295 - chr7:3,218,627-3,220,842					
Mmu (miR-290-295)	ESC	chr7:3,217,656-3,217,817	chr7:3,217,803	Comparative genomics, Northern	13
Mmu (miR-290-295)	ESC	-	chr7:3,218,001-3,219,675	H3K4me3	7

TSS = transcription start site, CAGE tag = cap analysis of gene expression tags, ESC = embryonic stem cells, Pol II = RNA polymerase II.

germ cells are referred to as gonocytes. Late PGCs or early gonocytes can arrest in differentiation, thereby forming a carcinoma *in situ* (CIS) that is the typical precursor of type II GCT presenting later in life. These type II GCTs are subdivided in two types: seminomas, which are proliferative CIS cells, and non-seminomas, which are preceded by dedifferentiation of CIS cells into more pluripotent EC cells.

The basis of pluripotency resides in regulatory networks that maintain ESCs in a pluripotent and undifferentiated state. In a context-dependent manner, key regulators in these networks can direct embryonic development by switching between pluripotency and differentiation. Alterations in the expression of these factors or stoichiometry between them can promote or restrict cellular differentiation. Oct3/4 and Sox2 have been shown to perform a core function in maintaining pluripotency[23]. These TFs can bind a gene promoter in complex and activate transcription; about two-third and one-third of their target genes are bound by both factors,

respectively[23]. Oct3/4 and Sox2 are not only involved in maintenance of a pluripotent state, but have also a mayor role in inducing dedifferentiation of somatic cells (a process called reprogramming). Retroviral expression of Oct3/4 and Sox2 together with Krüppel-like factor 4 (Klf4) has been shown to be sufficient for induction of pluripotent stem (iPS) cells from mouse and human fibroblast cells (overview in Supplementary table S1). After retroviral infection of somatic cells, dedifferentiated cells can be selected for the expression of a pluripotency marker (e.g. Nanog) to increase iPS yield (Supplementary table S1). It should be noted that the efficiency of reprogramming is generally low and can be considerably increased by co-introduction of the transcription factor c-Myc[24,25]. Reactivation of the c-Myc retrovirus,



motif	dsDNA sequence
SOX^a	ATTTTTGTATTT TAAAAACATAAA
SOX^b	TTTCTTT AAAGAAA
KLF4	CTTTTTTTTT GAAAAAAAAA
SOX^a	ACTTTTGTTCTA TGAAAACAAGAT

Figure 1 | Characterization of the miR-371-373 gene and expression pattern. (a) View of human genomic region (10kb) where the level of mammalian conservation (17 species) is represented by blue peaks. The data are derived from UCSC genome browser hg18 assembly. A 1222 bp region which is specified in **(c)** is indicated by a blue line. **(b)** Timing of miR-371-373 expression during human embryonic development (blue line) and development of carcinoma *in situ* (CIS) lesion. hESC = human embryonic stem cells, PGCs = primordial germ cells, XY = male. **(c)** Predicted binding motifs for Sox proteins (SOX type a and b) and Klf4 (KLF4) in the 1222 bp region upstream of the miR-371-373 gene based on Transfac, JASPAR, and EPD databases. The specific dsDNA sequence motifs are presented in a table.

however, increases teratocarcinoma incidence in iPS chimeras and progeny mice, hindering clinical applications[26]. Dysregulated expression or function of c-Myc is one of the most common abnormalities in human malignancy. Thus, c-Myc promotes both iPS cell generation and tumorigenicity, but these functions can be separated as demonstrated by introduction of transformation-deficient c-Myc mutants[27].

We aimed at a better understanding of the genetic and epigenetic mechanisms that control miRNA expression in development and in disease. In our search for cis-acting and trans-acting factors responsible for miR-371-373 expression in PGCs and GCTs, we found a putative promoter region that activates transcription and contains binding motifs for the transcription factors Sox2 and Klf4. Also, Sox2 and Klf4 were identified to induce expression of miR-371-373.

RESULTS

Characterization of the miR-371-373 gene and expression pattern

MiRNA genes exhibit overall a high degree of evolutionary conservation. The gene cluster containing miR-371-373 is conserved in eutherian species[13]. Inside the cluster, the seed sequences are conserved, but considerable conservation is found up to ~1 kb upstream of the gene cluster as well (Fig. 1a). The homolog of this miRNA cluster in mouse is miR-290-295, and in zebrafish miR-430. Deficiency of miR-290-295 results in embryonic lethality in some, but not all mouse embryos, demonstrating a relevant role for these miRNAs in early embryonic development. These miRNAs are expressed from the start of embryogenesis in zebrafish, mice and human, and remain expressed in PGCs until these cells reach the genital ridge and differentiate into gonocytes (Fig. 1b, Table 2). In human GCTs, miR-371-373 is found expressed in type II GCT, both in precursor CIS and in invasive seminomas or non-seminomas (Table 3). In non-seminomas, miR-371-373 is found expressed in the undifferentiated EC component and in YS components, but not in TC and CH components that also lack expression of some pluripotency markers. In human EC cell lines, miR-371-373 expression is found in 2102Ep, 833KE, and Tera-1 that do not have differentiation potential, whereas Tera-2 and NCCIT that are capable of significant somatic or extra-embryonic differentiation lack miR-371-373 expression[28].

We aimed to determine which genes transcriptionally regulate miR-371-373 expression during embryonic development, and possibly, during GCT development. TFs with an evident role in pluripotency are the first candidates to be involved in transcriptional activation. The 1.2 kb upstream region was scanned for potential binding sites of the pluripotency-associated genes Oct3/4, Sox2, Klf4 and c-Myc using various databases for transcription factor binding sites. Analysis of the region upstream of miR-371-373 for predicted TF binding motifs applying multiple TF databases yielded two different SRY-box containing gene (SOX) motifs and one gut-enriched Krüppel-like factor (GKLF, identical to Klf4) motif (Fig. 1c). One SOX motif occurs twice in the putative promoter region (SOX^{a}), the second SOX motif is present in one site (SOX^{b}). Three of these motifs are located in conserved parts of the promoter region, yet other conserved parts exist that lack predicted motifs for the mentioned TFs.

Table 2 | Expression of pluripotency-related genes during various stages of gametogenesis.

	hESC	PGC	Gonocyte	Spermatogonia	Spermatocyte	ref
	blastocyst	*epiblast*	*genital ridge*	*male gonads*		
miR-371-373	+	+	+	+/-	-	28, 29
Oct3/4	+	+	+	-	-	28, 30
Sox2	+	-	-	-	-	28, 30, 31
Klf4	+	+	-	-	-	32
c-Myc	+	+	-	-	-	
LIN28	+	+	+	+	+	30
NANOG	+	+	+	-	-	30

The embryonal sites of expression are in italic.

Table 3 | Expression of pluripotency-related genes in various GCT types.

GCT	Type 1	CIS	Type II	Type II	Type III	ref
		precursor type II	seminoma	non-seminoma (EC, TC, YS, CH)		
miR-371-373	-	+	+	+EC +/- YS -TC, CH	-	1, 29
Oct3/4	-	+	+	+EC - TC, YS, CH	-	28, 30, 33, 34
Sox2	+	-	-	+ EC +/- TC, YS - CH	-	28, 30, 35, 36
Klf4		+	+	+	-	32, 37
c-Myc		+	+	-	-	37
Lin28	-	+	+	+ EC, YS, CH - TC	-	30
Nanog	-	+	+	+ EC, CH +/- YS - TC	-	30, 38

In non-seminomas, EC cells are found but various components of differentiated tissue as well, which are characterized by differential expression of the pluripotency genes. By histology, TC, YS or CH components from GCT type I or type II tumours cannot be distinguished, but they may be characterized by different expression of the pluripotency genes as well. TC = teratocarcinoma, YS = yolk-sac tumour, CH = choriocarcinoma, CIS = carcinoma *in situ*.

Transcription activation capacity of the putative promoter region

We selected a 1.2 kb region upstream of the miR-371-373 gene cluster, cloned it into a luciferase reporter vector without promoter or enhancer region and tested its ability to activate transcription. This DNA region showed significant promoting activity when transfected in 833KE cells, where miR-371-373 are endogenously expressed, whereas transcription activation in HeLa, MCF-7 or other somatic cell types was comparable to the negative control vector (Fig. 2a). The level of transcriptional promoting activity by the putative promoter region is at least comparable to that induced by the SV40 promoter as positive control. Furthermore, activation is restricted to GCT cell lines and can be induced in a concentration-dependent manner (data not shown).

Next we cloned fragments of the 1.2 kb region, based on evolutionary conserved elements, to pinpoint the promoter activity to more defined loci. Transcription promoting activity is observed in the 833KE and Tera-1 cell lines; in Tera-1 cells the most upstream half of the region does not confer luciferase activity as compared to the downstream half. The other EC cell lines Tera-2, 2102Ep, NCCIT and the seminoma cell line TCam-2 did not show significant luciferase activity (Fig. S1). Also, regions of decreasing length were tested in this setup and gave a reproducible pattern of transcription activation in 833KE cells (Fig. 2b).

In a concurrent study, miRNA TSSs were predicted based on genome-wide chromatin signatures and nucleosome positioning[18]. We tested a 300 bp region including the bioinformatically inferred TSS for miR-371-373 (0.3 kb TSS) in our luciferase reporter assay, but no transcription activation was detected in 833KE cells compared to the SV40 promoter (Fig. 2a). Together, these data implicate a conserved region directly upstream of the miR-371-373 gene in transcriptional activation in EC cell lines.

Binding of Sox2 and Klf4 transcription factors to predicted motifs

We performed electromobility shift assays (EMSA) to examine direct binding of TFs Sox2 and Klf4 to the predicted binding sites *in vitro*. Incubation of a ^{32}P-labeled DNA probe with nuclearlysate of 833KE cells (with endogenous miR-371-373 expression) yielded a single DNA:protein complex using the wild-type, but not the mutated,Klf4 binding site (Fig. 3a). In a next experiment, increasing amounts of unlabeled probe added to the reaction resulted in binding competition with the labeled probe, but unlabeled probe harboring a mutation showed less binding competition (Fig. 3b). These results were obtained with probes containing the KLF4 binding site and the most downstream SOX[a] binding site. Using 833KE lysate, two ^{32}P-DNA: protein complexes were observed, but only the upper one was considered specific to the binding site (Fig. 3b). To validate specific binding of Klf4 and Sox2 to the used probes, binding reactions were repeated using lysate of MCF-7 cells overexpressing Klf4 or Sox2. Separation by gel electrophoresis revealed that amounts of ^{32}P-DNA: protein complex increased upon overexpression of either Klf4 or Sox2 (Fig. 3c).

To assess binding of TFs to the putative promoter region *in vivo*, chromatin immunoprecipitation (ChIP) assays were performed for Sox2 and Klf4 in miR-371-373-expressing EC cell lines and enrichment of genomic regions containing the

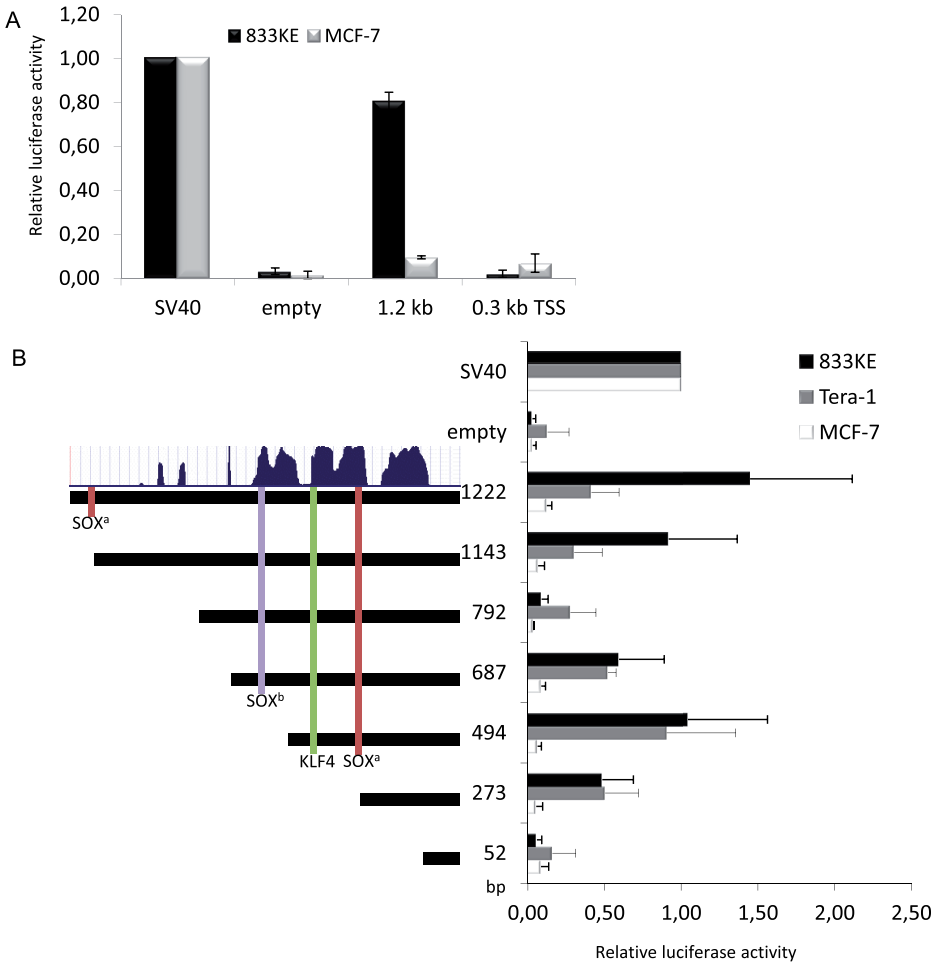


Figure 2 | Transcription activation capacity of the putative promoter region. (a) 833KE and MCF-7 cells were transfected with the indicated luciferase reporter constructs and compared to a control promoter (SV40) construct. Relative luciferase activity was measured after 72 hours and is the ratio between firefly luciferase and *Renilla* luciferase, the positive control is set at 1. Error bars represent SD for at least three independent experiments. **(b)** 833KE, Tera-1 and MCF-7 cells were transfected with luciferase reporter constructs of indicated length, as in **a**.

SOX^a, SOX^b or KLF4 motifs was measured by qPCR (Fig. 3d). Validated Sox2 and Klf4 binding regions on the Sox2 (Sox-responsive region 2; SRR2) and p53 (21-bp element; PE21) gene promoters were used as positive control. The KLF4 region and PE21 were detected in anti-Klf4 precipitates in 2102Ep cells, when compared to control IgG precipitates (Fig. 3d, upper panel). In addition, a control region in the downstream part of the putative promoter (-124:0) was much less associated with Klf4. Association of Sox2 to the SOX^a motif-containing region, and to SRR2, was detectable in 833KE and Tera-1 cells. The downstream control region was not

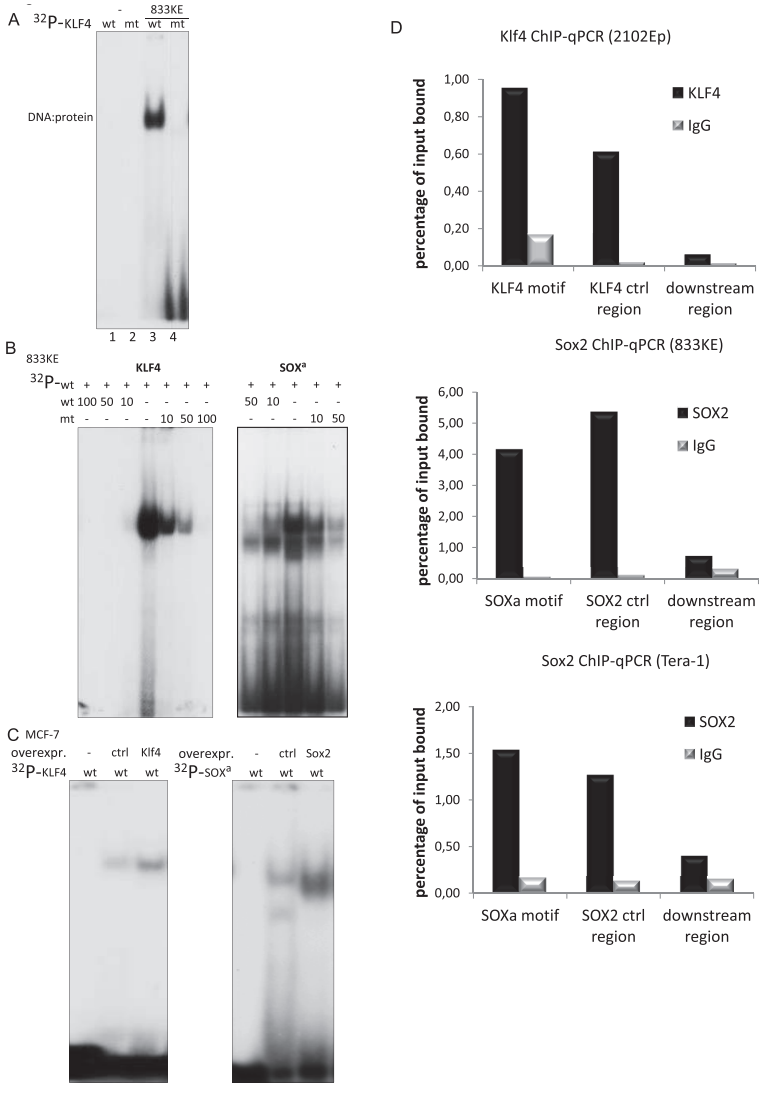


Figure 3 | Binding of Sox2 and Klf4 transcription factors to predicted binding motifs. **(a)** Nuclear protein fractions of 833KE cells were incubated with ^{32}P-labeled wild-type and mutant DNA probe versions of the KLF4 motif and DNA:protein complexes were visualized on gel. **(b)** Pre-cleared protein lysate of 833KE cells was incubated as in **a**, together with different amounts of unlabeled wild-type and mutant probe. **(c)** Equally concentrated pre-cleared protein lysates of MCF-7 cells transfected with constructs coding for Klf4, Sox2, or control were incubated with ^{32}P-labeled wild-type DNA probes and DNA:protein complexes were visualized by gel electrophoresis. **(d)** Chromatin immunoprecipitation was performed in pre-cleared lysate from 2102Ep, 833KE and Tera-1 cells using antibodies against Sox2 and Klf4 and using rabbit IgG as negative control. Association of indicated genomic regions were measured by qPCR, association of validated Sox2 and Klf4 bound regions on the Sox2 (Sox-responsive region 2; SRR2) and p53 (21-bp element; PE21) gene promoters was measured as positive control, respectively, and a downstream promoter region as negative control. Amount of associated DNA is plotted as percentage; levels in input were set at 100%.

detected in the Sox2 ChIP, neither did the IgG control contain the SOX[a] motif or SRR2 (Fig. 3d, lower panels). In immunoprecipitates of Sox2 in 2102Ep cells and of Klf4 in 833KE and Tera-1 cells, no specifically enriched regions could be detected. Together, these results suggest *in vitro* and *in vivo* association of Sox2 and Klf4 to regions in the putative promoter region of miR-371-373, with varying binding activity in the specific EC cell lines.

Impact of Sox2 and Klf4 on transcription activation mediated by the promoter region

Next, we characterized the importance of the predicted SOX and KLF4 sequence motifs with luciferase reporter vectors carrying point mutations in these binding sites. In 833KE cells, luciferase activity was significantly reduced (~50%) upon mutating the most upstream SOX[a] motif, indicating that transcription activation of the promoter region is dependent on at least one intact SOX[a] motif. A double point mutation in the KLF4 motif does not affect activity mediated by the 1.2kb promoter region (Fig. 4a). A reporter construct containing both mutated SOX[a] and KLF4 motifs showed ~50% reduction in activity compared to the wild-type reporter, which is similar to the reporter containing one mutated SOX[a] motif (Fig. 4a). The SOX[b] motif was excluded from analysis based on negative results by ChIP-qPCR. Next, we tested whether siRNA-mediated or shRNA-mediated silencing of Sox2 or Klf4 affects luciferase reporter activity in 833KE cells. Upon knockdown of Sox2, activation of transcriptionby the 1.2 kb region (-1222:0) is ~50% reduced being consistent with the effect of mutating one of two SOX[a] motifs (Fig. S2). Interestingly, Sox2 knockdown does not affect luciferase activity mediated by the reporter construct lacking one of two predicted SOX[a] motifs (-1143:0). We did not observe an effect upon shRNA-mediated knockdown of Klf4.

We then tested the effect of Sox2 and Klf4 overexpression by co-transfection of expression vectors and luciferase reporter vectors in MCF-7 cells. When compared to the negative control reporter, transcription was not significantly activated upon single or combined Sox2 and Klf4 overexpression (Fig. 4b). To check the possible contribution of Oct3/4 and c-Myc, expression vectors encoding these pluripotency TFs were co-transfected. Expression of Oct3/4 in addition to Sox2 and Klf4 resulted in an increase in luciferase activity to ~50% compared to the SV40 promoter vector, whereas Oct3/4 expression solely does not significantly increase transcription (Fig. 4b). Combined expression of Sox2 and Oct3/4 increases luciferase activity as much as the sum of their individual effects. Co-transfection of c-Myc with Sox2, Klf4 and Oct3/4 does not further induce transcription.

These results indicate that binding of Sox2 to the miR-371-373 promoter activates transcription, which minimally involves the most upstream SOX[a] motif. Also, Klf4 is important for transcription activation via the promoter, but not necessarily via binding to the predicted KLF4 motif or via direct DNA binding at all. Oct3/4 enhances transcriptional activation of miR-371-373, but how Oct3/4 is involved remains unclear. An unnoticed consensus sequence for Oct3/4 may be present, or Oct3/4 may act via interaction with Sox2.

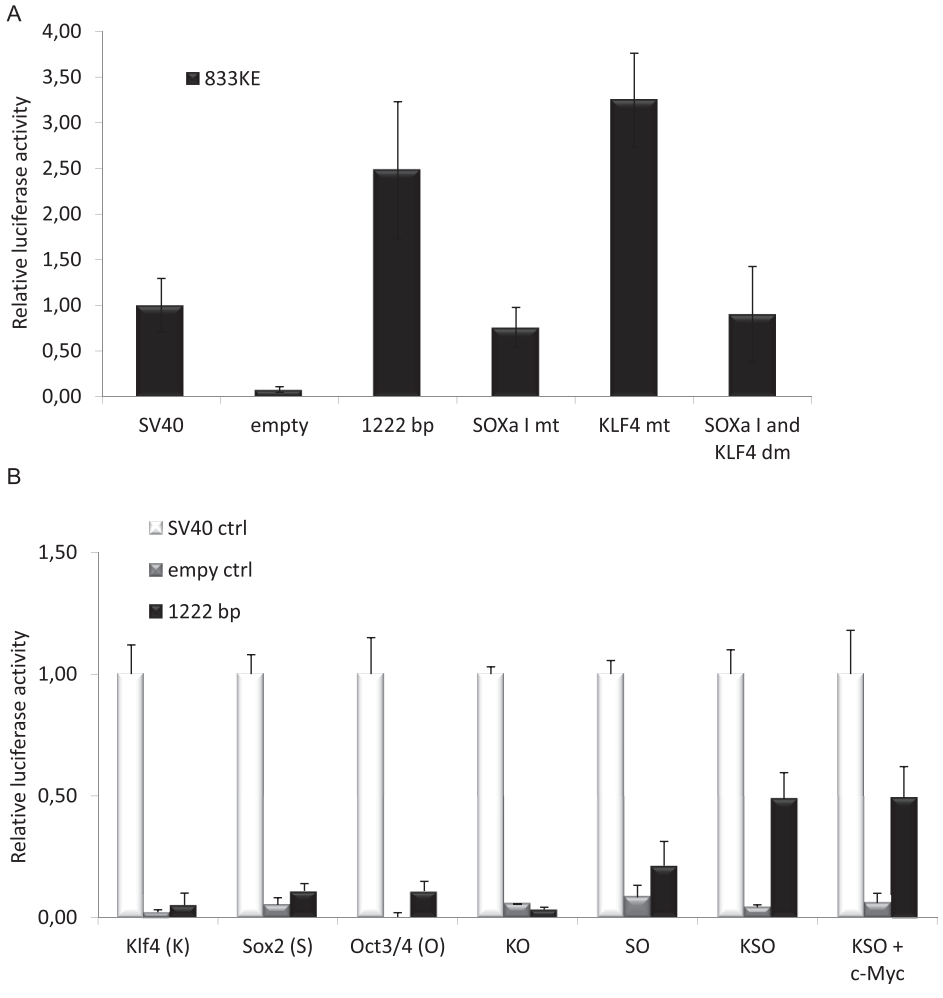


Figure 4 | Impact of Sox2 and Klf4 on transcription activation mediated by the promoter region. (a) 833KE cells were transfected with the indicated luciferase reporter constructs, the double mutant (dm) construct contains both mutated (mt) binding motifs. Relative luciferase activity was measured after 72 hours and is the ratio between firefly luciferase and *Renilla* luciferase, the SV40 control promoter is set at 1. Error bars represent SD for triplicate experiments. **(b)** MCF-7 cells were co-transfected with a luciferase reporter construct and expression constructs coding for Klf4 (K), Sox2 (S), Oct3/4 (O), and c-Myc as indicated. As in **a**.

Epigenetic changes upon overexpression of Sox2 and Klf4 in somatic cells

Access of transcription factors to gene promoters is dynamically controlled by a variety of epigenetic factors, such as modifications of the DNA. DNA CpG dinucleotide methylation is an epigenetic modification that is associated with two-third of miRNA promoters[18] resulting in transcriptional repression. Methyl residues on cytosines in

CpG-rich sites (CpG islands) potentially interfere with the binding of the transcriptional machinery or transcription factors. Since the global methylation pattern varies across GCT types[39], we assessed the methylation status of two known CpG is lands in the region around the miR-371-373 gene by bisulfite sequencing (BiS). In 833KE and Tera-1 cells, no methylated CpGs were detected in either of two islands whereas in U2OS and MCF-7 cells at least 60% of CpGs was methylated, implying a more inactive transcriptional state (Fig. 5a). However, upon prolonged overexpression of both Sox2 and Klf4 in U2OS and MCF-7 cells, CpG methylation decreased with 8% and 15%, respectively (Fig. 5b). Similar results were obtained for the methylation status of the intergenic CpG island (not shown), which is consistent with the methylation profile of this region that was previously demonstrated in somatic HCT-116 cells[40]. Expectedly, partial demethylation facilitates TF binding, but it is unknown how or whether TF expression actively contributes to demethylation. By qRT-PCR, we could detect endogenous levels of miR-372 after prolonged Sox2 and Klf4 expression in somatic cells, whereas we did not detect expression of miRNA family members (not shown) or changes in expression of unrelated miRNAs (Fig. 5c). Next, we tested whether direct inhibition of methylation gives a comparable effect on miR-371-373 expression. Therefore, we treated cells with 5-aza-2'-deoxycytidine reagent (5-ACR) and subsequently measured changes in miRNA expression by qRT-PCR analysis. Upon 5-ACR treatment, miR-372 was not found with detectable expression by qRT-PCR at different time points (Fig. S3), neither could expression of miR-93/302/518-520 family members be observed (not shown). By microarray analysis of miRNA expression in HEK293T and HeLa cells, we observed an increase in expression of normally methylated miRNAs upon 8 hours of 5-ACR treatment[41], but expression signals for the miR-93/302/371-373/518-520 family members did not change (data not shown). This indicates that demethylation of the CpG islands nearby the miR-371-373 gene is not sufficient to unleash transcription, but requires other activating factors.

Analysis of genome-wide mRNA and miRNA expression data of GCT cell lines

Intriguingly, human ESC lines exhibit GCT-like features after prolonged culturing time, which is readily noticeable in elevated expression of the pluripotency factors and miR-371-373[42]. Therefore, factors involved in this *in vitro* process may be found in genes that are differentially expressed in GCT cell lines with high endogenous miR-371-373 expression (833KE, 2102Ep, Tera-1) compared to GCT cell lines with relatively low (Tera-2, NCCIT) or no endogenous expression (HEK293T, MCF-7) of this miRNA cluster. SAM analysis on mRNA expression data sets of these cell lines yielded a list of 17 annotated genes that are significantly upregulated in cell lines with endogenous miR-371-373 expression (Fig. 6a, Table S2). Of these genes, many are regulated directly upstream or downstream of the pluripotency transcription factors (Fig. S4). For example, the Sox15 and VSNL1 genes are targeted by Sox2, DPPA3 is targeted by Oct3/4, and ZSCAN1 and the Oct3/4-Sox2 ---complex are reciprocal targets[43,44]. In summary, genes that are coordinately expressed with miR-371-373 in GCT cell lines may contribute to regulatory complexity with regard to miR-371-373 expression.

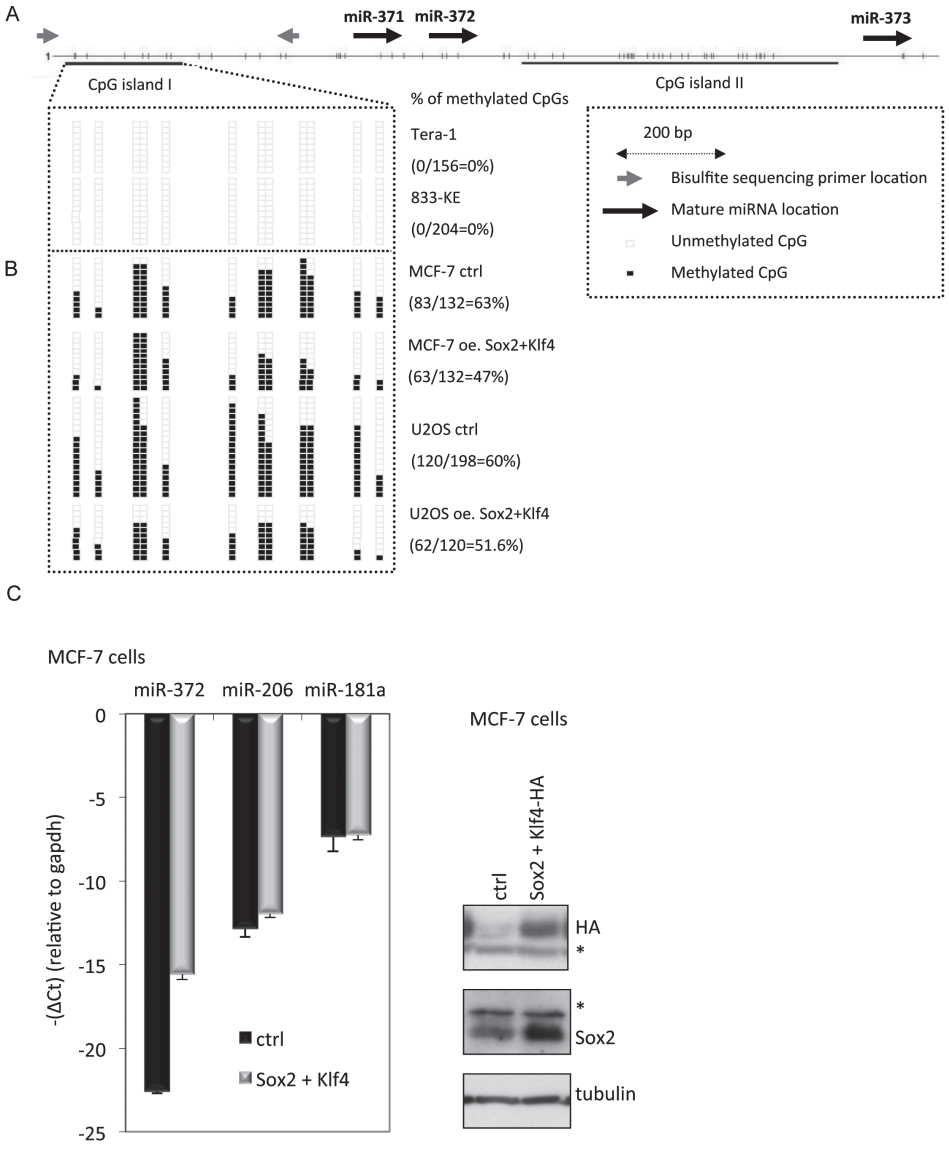


Figure 5 | Epigenetic changes upon overexpression of Sox2 and Klf4 in somatic cells. **(a)** In 833KE and Tera-1 cells, the fraction of methylated CpG residues was examined in two CpG islands upstream and in between the miR-371-373 genes, respectively. Sodium bisulfite-mediated conversion of DNA was detected by methylation-specific PCR and the methylation profile of the CpG regions was quantified as the percentage of methylated (black boxes) CpG residues. **(b)** As in **a**. MCF-7 and U2OS cells were transfected twice with constructs coding for Sox2 and Klf4-HA, or control. **(c)** MCF-7 cells were transfected as in **b**. Endogenous levels of miR-372 and of unrelated miR-206 and miR-181a were analyzed by qRT-PCR, and the results are displayed as –(**Δ**Ct). Immunoblots were performed with antibodies against Sox2, HA (12ca5) and tubulin.

In addition, we used genome-wide miRNA and mRNA expression data sets to predict active miRNAs in GCT cell lines by a computational approach. Given mRNA and miRNA expression data sets were recorded under the very same conditions, our computational method identifies active miRNAs by seeking those miRNAs whose expression pattern is significantly anti-correlated with the expression level of their (predicted) targets. Therefore, expression data of GCT cell lines were combined with databases of predicted miRNA targets (TargetScan version 4.0). For each miRNA in the dataset, we calculated the correlation of its expression with the expression of all mRNAs, and then statistically compared the distribution of those correlations between two sets of genes comprising the predicted miRNA targets and non-target genes. This resulted in a top-ten of miRNAs (including miR-372, miR-373, miR-302a-3p) for which the correlation between miRNA levels and predicted mRNA targets levels significantly differs from the correlation between miRNA levels and non-target mRNA levels (Kolmogorov-Smirnov test, Table 4). These miRNAs represent five distinct seed sequences; mRNAs that carry a complementary sequence have a significant propensity to be downregulated when compared to other mRNAs (Fig. 6b, Table 4). The expression of predicted targets of miR-371 (dissimilar seed sequence) was also found anti-correlated, however, with a lower significance (data not shown). These results indicate that expression of miR-372 and miR-373, amongst others, is relevant for post-transcriptional regulation of gene expression in GCT cell lines, and possibly in human GCTs as well.

Identification of genetic and epigenetic regulators of miRNA transcription adds to the understanding of differential miRNA expression in development and in disease. In this study, we identified a promoter element upstream of miR-371-373 capable of transcription activation. TFs Sox2 and Klf4 showed binding capacity to motifs in this region, and induce transcriptional activation, possibly in cooperation with Oct3/4. In somatic cells, partial demethylation of CpG islands nearby miR-371-373

Table 4 | Prediction of miRNAs with activity in GCT cell lines.

hsa-miR-	p-value	seed sequence
17-5p	4.87E-08	CAAAGUGCUUACAGUGCAGGUAG
181b-5p	1.62E-07	AACAUUCAUUGCUGUCGGUGGGU
23b-3p	1.47E-06	AUCACAUUGCCAGGGAUUACC
181a-5p	1.84E-06	AACAUUCAACGCUGUCGGUGAGU
181d	2.79E-06	AACAUUCAUUGUUGUCGGUGGGU
302a-3p	3.05E-06	UAAGUGCUUCCAUGUUUUGGUGA
125a-5p *)	5.36E-06	UCCCUGAGACCCUUUAACCUGUGA
33-5p *)	6.55E-06	GUGCAUUGUAGUUGCAUUGCA
373-3p	3.74E-05	GAAGUGCUUCGAUUUUGGGGUGU
372	6.72E-05	AAAGUGCUGCGACAUUUGAGCGU

MiRNAs whose expression in GCT cell lines is significantly anti-correlated with the expression of their predicted target mRNAs ($p<0.0001$) are listed, based on the comparison between the correlation distributions for their predicted targets and non-target mRNAs. MiRNAs marked by an * show borderline expression in the GCT cell lines.

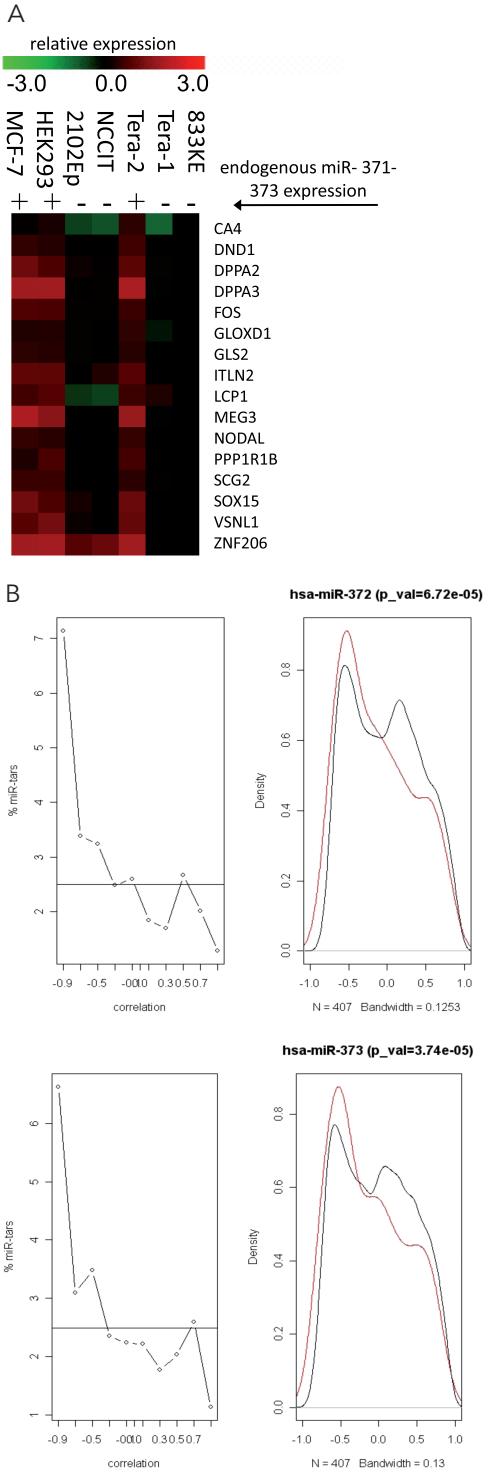


Figure 6 | Genome-wide mRNA and miRNA expression in GCT cell lines. (a) Supervised classification was performed on mRNA expression data of 7 human cancer cell lines using SAM analysis. Genes that are significantly overexpressed in EC cell lines highly expressing miR-371-373 (2102Ep, 833KE, Tera-1) compared to other cell lines were selected and their relative expression is represented in a heat map, in alphabetical order. **(b)** Significant anti-correlation between the expression of miR-372/373 and its predicted target genes in the analyzed data set is demonstrated using two types of plots: 1) in the plot to the left, all mRNAs were binned into 10 bins in the [-1.0, 1.0] interval according to the correlation between their expression and the miRNA expression, and the fraction of predicted miR-372/373 targets in each bin was calculated. For example, the proportion of targets in the -0.9 bin is ~ 7.0% whereas this proportion in the +0.9 bin is lower than 1.0%. 2) In the plot to the right, the distribution of correlation coefficients between the expression of miR-372 or miR-373 and the expression of their predicted target mRNAs (miR-tars, red) and non-targets (black) is shown. P-values for the comparison between the two distributions were calculated using the Kolmogorov-Smirnov test.

may facilitate miRNA expression. In part, Sox2, Klf4 and Oct may be responsible for miR-371-373 expression in PGCs and GCTs, possibly in collaboration with factors specifically expressed in these cell types. Besides transcriptional activation, miRNA function can be controlled at other levels of regulation, e.g. miRNA processing, miRNA stability or miRNA target access.

DISCUSSION

For human miRNA cluster 371-373 and for the homologous miR-290-295 cluster in mouse, predictions of the promoter region and TSSs are based on comparative genomics or on experimentally demonstrated nucleosome positioning, histone marks and PolII binding (Table 1). The predicted promoter positions in the genomes do not merely overlap, also it is worth mentioning that some of these predictions are based on data derived from cell lines without endogenous miR-371-373 expression. Transcriptional activation demonstrated by our luciferase reporter studies is mostly consistent with miR-371-373 expression per cell type. Except for 2102Ep and TCam-2 cells with endogenous miR-371-373 expression, lack of transcription activation in luciferase may be due to extra regulatory complexity or to technical issues.

In the luciferase reporter assays, the variance in transcription activation mediated by the genomic regions of increasing length may be due to presence or absence of TF binding sites. The differences between cell lines may be explained by differential expression of (combinations of) functional transcription activators or repressors. Since introduction of the transcription factors Sox2 and Klf4 in somatic cell lines does not fully mimic endogenous levels of miR-371-373 being detected in EC cell lines, a more complex regulation may apply. The results of ChIP-sequencing in human and mouse ES cells revealed that a subset of transcriptionally active miRNAs is co-occupied by Sox2 and Oct3/4[45], implying that binding of these factors is dependent on each other. Sox2 and Oct3/4 bind their promoters reciprocally, and likewise, Sox2 is suggested to stabilize Oct3/4 expression[45,46]. Therefore, the described effects upon overexpression of Sox2 exclusively, could either be independent of Oct3/4 or, in case of Oct3/4 dependency, be attributed to stabilization of Oct3/4.

In mouse, miR-290-295 family is strongly upregulated in early mouse embryogenesis from two-cell stage onwards[47], this is the stage when Oct3/4 expression increases. Also, in certain stages of human germ cell development, Oct3/4 is detected when Sox2 is not expressed anymore (Table 2). However, gain or loss of expression of Sox2 and Oct3/4 occurs in diverse GCT types, illustrating the heterogeneity of these tumours. Here, another known member of the SRY transcription factor family might substitute the function of Sox2, such as Sox15[48] or Sox17[35]. It is equally well possible that Oct3/4 targets a different subset of promoters in developmental stages where it is solely expressed. Concerning miR-371-373, it remains to be established whether transcription is activated by collaboration of Sox2 and Oct3/4, or is dependent on often-associated factors such as Nanog or p300[45,46,49].

In vitro transcriptional activation of miR-371-373 can be accomplished by a combination of the TFs Sox2, Klf4 and Oct3/4. This implies that inappropriate transcriptional activation of miR-371-373 may be due to expression or activation

of the implied factors. Since each TF targets hundreds of genes, this would have a widespread effect supplemented by the effect on the functional targets of the activated miRNAs. More than 500 genes are found differentially expressed in GCTs compared to other tumours, and expression of just as many genes significantly differs between subtypes[50,51]. By genome-wide mRNA expression analysis in GCT cell lines and clinical samples, a hexamer motif complementary to the shared seed of the miR-302 family and miR-372/373 was found overrepresented in the 3' UTRs of downregulated genes[52]. This is consistent with the universal overexpression of these miRNAs in GCTs and also with our data. Noteworthy, the seed sequence of miR-371 differs from those of miR-372/373 and the most abundant mature miR-302 forms (Table S3), suggesting that an additional group of mRNAs is targeted upon expression of the complete miR-371-373 gene cluster. However, miR-372/373, and not the seed mutants, can alleviate senescence and transform primary cells. Moreover, miR-373, and also miR-520c with a matching seed sequence, has been found to drive proliferation and metastasis. Hence in the context of transcriptional activation of the miR-371-373 cluster in GCTs, supported by miRNA-mRNA expression correlation analysis, the role of miR-371 seems less important.

MATERIAL AND METHODS

Constructs and antibodies

Promoter fragments were cloned by using a forward primer corresponding to the positions (-1222:-1193), (-1143:-1116), (-792:-765), (-687:-656), (-494:-464), (-273:-244), and (-52:-22) relative to the start of the miR-371 gene, and a universal reverse primer (+26:+55). A 0.3 kb region around a predicted TSS was cloned using the following primers: GCGGTACCAATGTCAACATCGGGAGCCTCATCTGC (forward) and GCAAGCTTTACCTGTTAACCCAGGCTGGACTGGAA (reverse). Each of the fragments was ligated into the KpnI-HindIII digested pGL3-Basic vector to allow transcription of the firefly luciferase gene under the control of these fragments (Promega). The common site in the SOX[a] motifs was mutated to TT**G**TG**C**,the KLF4 motif was mutated to CTTT**G**TTTT**G** (mutations in bold). Mutants were made with the Stratagene multisite-directed mutagenesis kit, all constructs were sequence-verified. A Sox2 construct was made by cloning the cDNA plus a double carboxy-terminal HA tag into the Invitrogen pcDNA3.1+ vector, the pPthC-Oct3/4 construct is described elsewhere[53], the peGFP(N3)-Klf4-HAHA and pRc/CMV-c-Myc constructs were given by I. Kolfschoten and K. Berns, respectively.

Antibodies used were Sox2 (R&D Systems, AF-2018), Klf4 H-180 X (Santa Cruz, sc-20691), HA Y-11 (Santa Cruz, sc-805), and tubulin (YL1/2 ECACC). On-TargetPlus SmartPool siRNAs for Sox2 were purchased from Dharmacon and a control siRNA from Ambion.

Cell culture, transfections, and dual luciferase activity analysis

The 833KE, 2102Ep, NCCIT, Tera-1, Tera-2 cell lines are derived from nonseminomatous tumours, TCam-2 is a seminoma cell line. HEK293 embryonic kidney, MCF7 breast adenocarcinoma, U2OS osteosarcoma, and 2102EP cell lines

were cultured in DMEM, Tera-1 cell line was cultured in McCoy's and 2102Ep, NCCIT, Tera-2, and TCam-2 cell lines were cultured in RPMI-1640. All media were supplemented with 10% heat-inactivated fetal calf serum (FCS) and all cells were cultured in 5% CO^2 at 37 °C. U2OS and MCF-7 cells were transiently transfected by using Fugene6 (Roche). For luciferase analysis, cells were transfected with 5 ng of firefly luciferase reporter, 5 ng of *Renilla* control plasmid, 150 ng of either expression construct or control construct up to 600 ng per reaction using Fugene6 (Roche). Dual luciferase activity assays were performed 72 h after transfection in accordance with the manufacturer's instructions (Promega). siRNAs were transfected in a final concentration of 50 nM (Dharmacon) with the use of Fugene6 (Roche) reagent, in accordance with the manufacturer's instructions. Cell lines were treated with the demethylating agent 5-aza-2′ deoxycytidine (Sigma) at a concentration of 10 μM for the indicated amount of time.

mRNA expression array and quantitative RT-PCR analysis

Total RNA was extracted with Trizol (Invitrogen), DNase-treated and hybridized to RNA expression arrays. For microarray expression profiling of mRNA we used Illumina Sentrix BeadChip human-6 v2 arrays, for expression profiling of miRNA we used Exiqon miRCURY LNA v2 arrays in accordance with the manufacturer's instructions (http://microarrays.nki.nl). For mRNA qRT–PCR, cDNA (from 2 μg RNA) was synthesized with SuperScript III and primed with oligo(dT) in accordance with the manufacturer's instructions (Invitrogen). QPCR primers for miR-372, miR-302a, miR-206, miR-181a, GAPDH and 18S were from Applied Biosystems. Analysis was performed with SYBR Green PCR master mix or TaqMan UNG master mix (Applied Biosystems) and Chromo 4 system (Bio-Rad Laboratories).

Electromobility shift assay (EMSA)

Probes were constructed using the following oligos for SOXa (5'-GGTAATTTTTGT ATTTTTTT-3' and 3'-CCATTAAAAACATAAAAAAA-5') and for KLF4 (5'-TCTGCTTT TTTTTTTTTT-3' and 3'-AGACGAAAAAAAAAAAAA-5'). Radiolabeled DNA probes were generated by ^{32}P-labeling of the annealed probe at the 5′ end using T4 polynucleotide kinase. ^{32}P-dsDNA probe (20,000 cpm) was incubated with 10 μg protein extract in EMSA binding buffer (f.c. 100 mM Tris, 50% glycerol, 10 mM DTT, 500 mM NaCl, 20 mM $MgCl_2$, 10 mM EDTA) in the presence of 100 μg/mg poly(dI•dC) (Amersham) for 15 min at room temperature. Cellular fractionation was performed on MCF7 cells with NE-PER kits from Pierce, according to manufacturer's instructions. The reactions were loaded on a pre-run native gel (5% polyacrylamide [29:1] and 0.5X TBE) for 2-3 hours at 200 V. The gel was exposed to a Phosphor Imaging Plate and developed on the BAS-2500 system (both Fujifilm).

Immunoblotting, chromatin immunoprecipitation

For immunoblot analysis, extracts were separated on 10% SDS–PAGE gels, and transferred to Immobilon-P membranes (Millipore). Western blots were developed using enhanced chemiluminescence (ECL; Amersham Biosciences) and exposed to film (Kodak).

833KE, Tera-1 and 2102Ep cells were cross-linked for 10 min with 1% formaldehyde, neutralized with 125 mM glycine and washed in PBS. The cells were lysed (50 mMTris pH 8.0, 10 mM EDTA, 1% SDS), nuclei were collected by centrifugation and sonicated for 15 min to yield chromatin with 500–1000 base pairs (bp) average fragment size. A 5% aliquot was taken as input material. Cell extracts were incubated in 1:3 RIPA buffer (10 mM Tris pH 7.5, 140 mM NaCl, 1mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, and protease inhibitor mixture (Roche Applied Science)) with antibodies against Sox2 or Klf4 (1 µg per immunoprecipitation) in a tumbler placed at 4 °C for 4 hours, as a negative control rabbitIgG antibody (1 µg per immunoprecipitation, Santa Cruz, sc-2027) was used. Sox2 and Klf4 were immunoprecipitated with GammaBind G Sepharose (GE Healthcare). Beads were preblocked with yeast tRNA (Invitrogen) and BSA (Ambion), incubated for 2 hours and then washed in RIPA buffer and in TE (10 mM Tris pH 8.0, 10 mM EDTA). Chromatin was eluted from the beads (20 mM Tris pH 7.5, 5 mM EDTA, 50 mM NaCl, 20 mM Na butyrate, 1% SDS, 50 ug/ml proteinase K (Roche)) at 65 °C overnight, then proteinase K activity was inhibited by 5 mM PMSF (Sigma). DNA was extracted and resuspended in water, a 5% aliquot was subjected to qPCR analysis. Real-time PCR was performed on a StepOne Plus cycler (Applied Biosystems), using SYBR Green Universal PCR Master Mix (Applied Biosystems). Promoter occupancy was calculated as the percentage of input chromatin. Primers for qPCR were KLF4 motif (5'-GCGCACCACCATTCCCATGTGCGTTTT-3' (forward) and 5э-TACGGTGGTAGGATTGCTCTTG-3' (reverse)) and KLF4 ctrl (5'-GCCCTTACTTGT CATGGCGA-3' (forward) and 5'-CAATCCCATCAACCCCTGC-3' (reverse)), SOX^{a} motif (5'-GCGGTACCGCAATTCTCTTGCCTCAGCCTCCCAAGCAG-3' (forward) and 5'-GCAACATGGCGAAACCCCGTCTCTACCAA-3' (reverse)) and SOX^{a} ctrl (5'-GGA TAACATTGTACTGGGAAGGGACA-3' (forward) and 5'-CAAAGTTTCTTTTATTCGTAT GTGTGAGCA-3' (reverse)), GAPDH (5′-TGCACCACCAACTGCTTAGC-3′ (forward) and (5′-GGCATGGACTGTGGTCATGAG-3′ (reverse)) and β-actin (5′-CCTGGCA CCCAGCACAAT-3′ (forward) and 5′ GGGCCGGACTCGTCATACT-3′ (reverse)).

Bisulfite sequencing

The CpG Island Searcher (http://cpgislands.usc.edu/) was used to define CpG islands in vicinity to the miR-371-373 genes. DNA methylation status was established by PCR analysis of bisulfite-modified genomic DNA, which induces chemical conversion of unmethylated, but not methylated, cytosine to uracil as described elsewhere[40]. Primers used for genomic bisulfite sequencing (BS) and methylation-specific PCR analyses were: BS-371/2-s (TAGGATGAAGTGTATAGGTAGGATG), BS-371/2-as (AACCTCCAAATACAAACCATAA), BS-373-s (TTGGGGAAGGGAAGGGGGTTTT), BS-373-as (CCTACCTCAACCTCCCAAATAAC).

Bioinformatical approaches

Prediction of binding motifs for transcription factors was based on position-weight matrices listed in Transfac[54], JASPAR[55], and EPD[56] databases. The MotifScanner module within TOUCAN 2 software[57] was used to scan and visualize motifs. Supervised classification was performed with Cluster 3.0 software using the C clustering library

version 1.38 and results were visualized with Treeview[58]. SAM software[59] was applied to calculate a SAM score for gene expression in EC cell lines compared to other cell lines. Calculation of correlation coefficients was based on mRNA and miRNA expression data sets generated from 833KE, TCam-2 and MCF-7 cells. Distribution plots and Kolmogorov-Smirnov tests were generated using the R statistical package.

ACKNOWLEDGEMENTS

We thank all members of the Agami laboratory for technical help and discussions. We also thank A. Sparmann, I. Kolfschoten and K. Berns for providing antibodies and constructs. Further we are grateful to the NKI Central genomics facility for mRNA and miRNA microarray analysis on our samples. This work was supported by ERC (European Research Council), KWF (Koningin Wilhelmina Fonds; Dutch cancer foundation) and Horizon-NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; R.A.).

AUTHOR CONTRIBUTIONS

M.v.K. performed most of the experimental work. R.A. supervised the project. J.O.V. provided technical assistance. A.L. conducted bisulfite sequencing under supervision of M.E. R.E. performed Kolmogorov-Smirnov analyses. L.L. gave advice on the GCT cell lines. M.v.K. and R.A. wrote the manuscript.

REFERENCES

1. Voorhoeve, P.M. et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124**, 1169-81 (2006).
2. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
3. Farazi, T.A., Spitzer, J.I., Morozov, P. & Tuschl, T. miRNAs in human cancer. *The Journal of pathology* **223**, 102-15 (2011).
4. Calin, G.A. & Croce, C.M. MicroRNA signatures in human cancers. *Nature reviews. Cancer* **6**, 857-66 (2006).
5. Bernstein, E. et al. Dicer is essential for mouse development. *Nature genetics* **35**, 215-7 (2003).
6. Liu, J. et al. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-41 (2004).
7. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521-33 (2008).
8. Gillis, A. et al. High-throughput microRNAome analysis in human germ cell tumours. *The Journal of pathology* (2007).
9. Huang, Q. et al. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat Cell Biol* **10**, 202-10 (2008).
10. Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. *Genome research* **14**, 1902-10 (2004).
11. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA (New York, N.Y.)* **10**, 1957-66 (2004).
12. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-60 (2004).
13. Houbaviy, H.B., Dennis, L., Jaenisch, R. & Sharp, P.A. Characterization of a highly variable eutherian microRNA gene. *RNA (New York, N.Y.)* **11**, 1245-57 (2005).

14. Woods, K., Thomson, J.M. & Hammond, S.M. Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. *The Journal of biological chemistry* **282**, 2130-4 (2007).
15. Barroso-delJesus, A. et al. Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter. *Molecular and cellular biology* **28**, 6609-19 (2008).
16. Card, D.A. et al. Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Molecular and cellular biology* **28**, 6426-38 (2008).
17. Neves, R. et al. Role of DNA methylation in miR-200c/141 cluster silencing in invasive breast cancer cells. *BMC research notes* **3**, 219 (2010).
18. Ozsolak, F. Chromatin structure analyses identify miRNA promoters. *Genes Dev.* **22**, 3172-3183 (2008).
19. Wang, X., Xuan, Z., Zhao, X., Li, Y. & Zhang, M.Q. High-resolution human core-promoter prediction with CoreBoost_HM. *Genome research* **19**, 266-75 (2009).
20. Zhou, X., Ruan, J., Wang, G. & Zhang, W. Characterization and identification of microRNA core promoters in four model species. *PLoS computational biology* **3**, e37 (2007).
21. Chien, C.H. et al. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic acids research* (2011).
22. Looijenga, L.H., Gillis, A.J., Stoop, H., Hersmus, R. & Oosterhuis, J.W. Relevance of microRNAs in normal and malignant development, including human testicular germ cell tumours. *International journal of andrology* **30**, 304-5 (2007).
23. Boyer, L.A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56 (2005).
24. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature biotechnology* **26**, 101-6 (2008).
25. Wernig, M., Meissner, A., Cassady, J.P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell stem cell* **2**, 10-2 (2008).
26. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-7 (2007).
27. Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T. & Yamanaka, S. Promotion of direct reprogramming by transformation-deficient Myc. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14152-7 (2010).
28. Sperger, J.M. et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13350-5 (2003).
29. Novotny, G.W. et al. Analysis of gene expression in normal and neoplastic human testis: new roles of RNA. *International journal of andrology* **30**, 316-7 (2007).
30. Gillis, A.J. et al. Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis. *International journal of andrology* (2011).
31. Perrett, R.M. et al. The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture. *Biology of reproduction* **78**, 852-8 (2008).
32. Godmann, M. et al. The pluripotency transcription factor Krüppel-like factor 4 is strongly expressed in intratubular germ cell neoplasia unclassified and seminoma. *Molecular human reproduction* **15**, 479-88 (2009).
33. Cools, M., Drop, S.L., Wolffenbuttel, K.P., Oosterhuis, J.W. & Looijenga, L.H. Germ cell tumors in the intersex gonad: old paths, new directions, moving frontiers. *Endocrine reviews* **27**, 468-84 (2006).
34. Rajpert-De Meyts, E. et al. Developmental expression of POU5F1 (OCT-3/4) in normal and dysgenetic human gonads. *Human reproduction (Oxford, England)* **19**, 1338-44 (2004).
35. de Jong, J. et al. Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications. *The Journal of pathology* **215**, 21-30 (2008).
36. Sonne, S.B. et al. Analysis of SOX2 expression in developing human testis and germ cell neoplasia. *The International journal of developmental biology* **54**, 755-60 (2010).
37. Laguna, M.P., Albers, P. & Richie, J.P. Cancer of the testis. *Springer*.
38. Hoei-Hansen, C.E. et al. Stem cell pluripotency factor NANOG is expressed

in human fetal gonocytes, testicular carcinoma in situ and germ cell tumours. *Histopathology* **47**, 48-56 (2005).

39. Wermann, H. et al. Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance. *The Journal of pathology* **221**, 433-42 (2010).
40. Lujambio, A. et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer research* **67**, 1424-9 (2007).
41. Weber, B., Stresemann, C., Brueckner, B. & Lyko, F. Methylation of human microRNA genes in normal and neoplastic cells. *Cell cycle (Georgetown, Tex)***6**, 1001-5 (2007).
42. Harrison, N.J., Baker, D. & Andrews, P.W. Culture adaptation of embryonic stem cells echoes germ cell malignancy. *International journal of andrology* **30**, 275-81; discussion 281 (2007).
43. Yu, H.B., Kunarso, G., Hong, F.H. & Stanton, L.W. Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. *The Journal of biological chemistry* **284**, 31327-35 (2009).
44. Sharov, A.A. et al. Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC genomics* **9**, 269 (2008).
45. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-17 (2008).
46. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-635 (2007).
47. Tang, F. et al. Maternal microRNAs are essential for mouse zygotic development. *Genes & development* **21**, 644-8 (2007).
48. Maruyama, M., Ichisaka, T., Nakagawa, M. & Yamanaka, S. Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *The Journal of biological chemistry* **280**, 24371-9 (2005).
49. Almstrup, K. et al. Embryonic stem cell-like features of testicular carcinoma in situ revealed by genome-wide gene expression profiling. *Cancer research* **64**, 4736-43 (2004).
50. Juric, D. et al. Gene expression profiling differentiates germ cell tumors from other cancers and defines subtype-specific signatures. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17763-8 (2005).
51. Palmer, R.D. et al. Pediatric malignant germ cell tumors show characteristic transcriptome profiles. *Cancer research* **68**, 4239-47 (2008).
52. Palmer, R.D. et al. Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets. *Cancer research* **70**, 2911-23 (2010).
53. Masui, S. et al. An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic acids research* **33**, e43 (2005).
54. Wingender, E., Dietze, P., Karas, H. & Knüppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic acids research* **24**, 238-41 (1996).
55. Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic acids research* **32**, D91-4 (2004).
56. Périer, R.C., Praz, V., Junier, T., Bonnard, C. & Bucher, P. The eukaryotic promoter database (EPD). *Nucleic acids research* **28**, 302-3 (2000).
57. Aerts, S. et al. TOUCAN 2: the all-inclusive open source workbench for regulatory sequence analysis. *Nucleic acids research* **33**, W393-6 (2005).
58. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14863-8 (1998).
59. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5116-21 (2001).

SUPPLEMENTARY TABLES AND FIGURES

Supplementary table 1 | Overview of currently available strategies to generate induced pluripotent stem (iPS) cells.

system	expression of reprogramming factors	selection	ref
MEFs	Oct3/4, Sox2, Klf4, c-Myc	Fbx15	1
MEFs	Oct3/4, Sox2, Klf4, c-Myc	Nanog	2
MEFs	Oct3/4, Sox2, Klf4, c-Myc	Oct3/4, Nanog	3
MEFs	Oct3/4, Sox2, Klf4	Oct3/4, Nanog	4
MEFs	Oct3/4, Sox2, Klf4, miR-290-295	-	5
MEFs	Oct3/4, Sox2, Klf4, miR-93, miR-106b		6
Hsa fibroblasts	Oct3/4, Sox2, Klf4, c-Myc	Fbx15	7
Hsa fibroblasts	Oct3/4, Sox2, Lin28, Nanog	Oct3/4	8
Hsa hDF fibroblasts	Oct3/4, Sox2, Klf4	-	9
Hsa BJ fibroblasts	Oct3/4, Sox2, Klf4, miR-372 or miR-302b	-	10
Hsa MRC-5 fibroblasts	Oct3/4, Sox2, Klf4, c-Myc, miR-372 or miR-302b	-	10

Selection for gene expression is applied to increase efficiency. MEF = mouse embryonic fibroblasts, hDF = adult human dermal fibroblasts, BJ = primary human foreskin fibroblasts, MRC-5 = primary human lung fibroblasts.

Supplementary table 2 | Supervised classification was performed on mRNA expression data of seven human cancer cell lines using SAM-software.

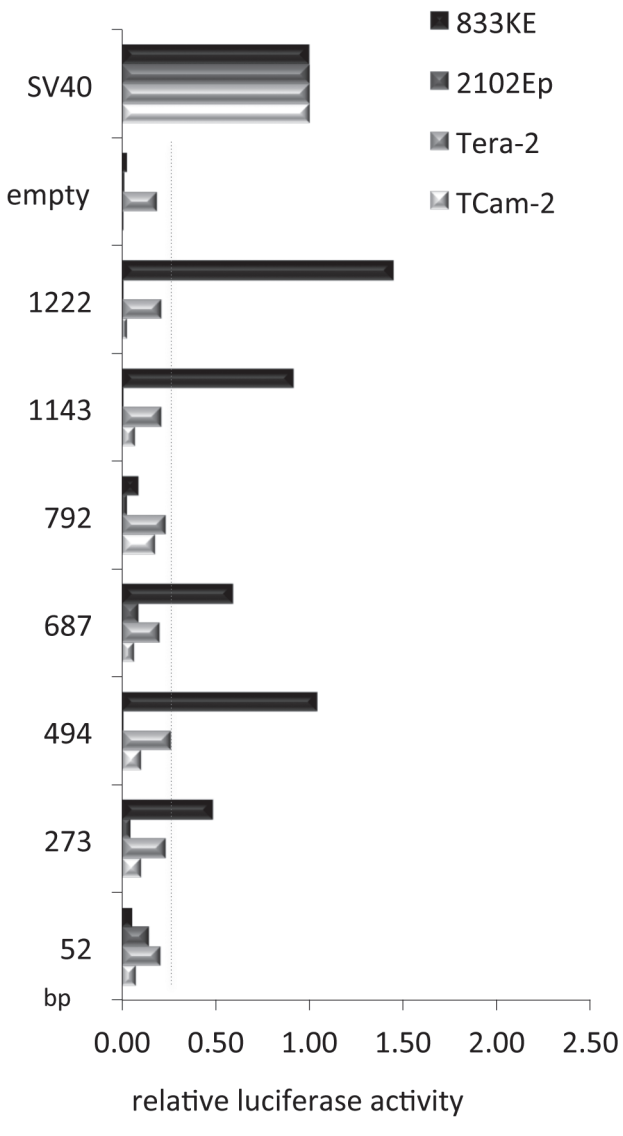
	symbol	gene name	A	B
1	DPPA3	developmental pluripotency associated 3 STELLA	3.7	3.4
2	MEG3	maternally expressed 3 on chr 14	3.6	3.5
3	LOC644273	PREDICTED: similar to STELLA	2.7	2.6
4	VSNL1	visinin-like 1	2.2	-
5	SOX15	SRY (sex determining region Y)-box 15	2.1	-
6	DPPA2	developmental pluripotency associated 2	2.1	-
7	ITLN2	intelectin 2	2.0	-
8	PPP1R1B	protein phosphatase 1, regulatory subunit 1B	1.8	-
9	GLOXD1	4-hydroxyphenylpyruvate dioxygenase-like	1.7	1.7
10	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	1.7	-
11	DND1	dead end homolog 1 (zebrafish)	1.7	-
12	GLS2	glutaminase 2 (mitochondrial) nuclear gene	1.6	1.7
13	NODAL	nodal homolog (mouse)	1.6	1.6
14	LCP1	lymphocyte cytosolic protein 1 (L-plastin)	-	3.1
15	CA4	carbonic anhydrase IV	-	2.6
16	SCG2	secretogranin II (chromogranin C)	-	1.6
17	ZNF206	zinc finger and SCAN domain containing 10	-	1.6

Genes that are significantly overexpressed in EC cell lines highly expressing miR-371-373 (833KE, 2102Ep, Tera-1) compared to other cell lines (A= versus Tera-2/NCCIT, B= versus Tera-2/NCCIT/HEK293T/MCF7) were selected for a SAM score > 1.5.

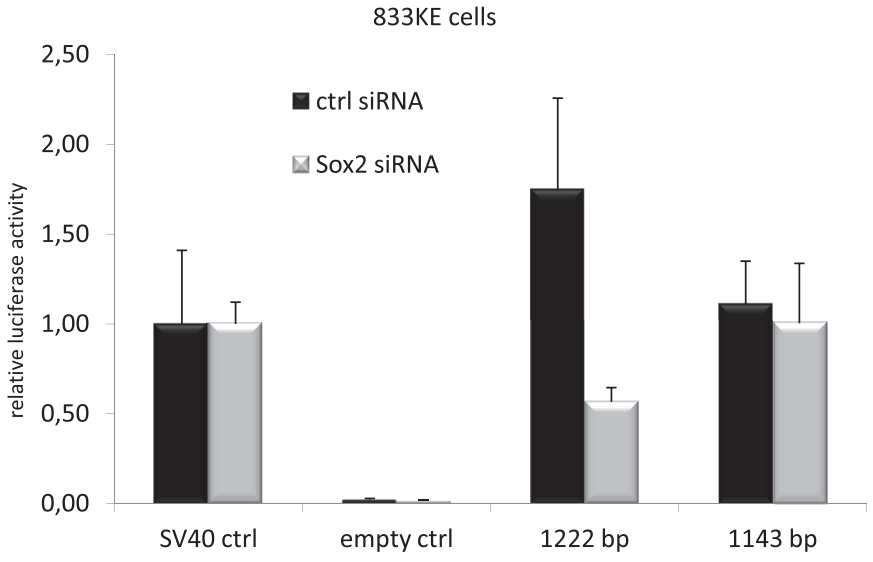
Supplementary table 3 | Seed sequences of members of the miR-302 and miR-371-373 clusters.

hsa-miR	seed sequence
302a-5p	ACUUAAACGUGGAUGUACUUGCU
302a-3p	UAAGUGCUUCCAUGUUUUGGUGA
302b-5p	ACUUUAACAUGGAAGUGCUUUC
302b-3p	UAAGUGCUUCCAUGUUUUAGUAG
302c-5p	UUUAACAUGGGGGUACCUGCUG
302c-3p	UAAGUGCUUCCAUGUUUCAGUGG
302d-5p	ACUUUAACAUGGAGGCACUUGC
302d-3p	UAAGUGCUUCCAUGUUUGAGUGU
371a-5p	ACUCAAACUGUGGGGGCACU
371a-3p	AAGUGCCGCCAUCUUUUGAGUGU
371b-5p	ACUCAAAAGAUGGCGGCACUUU
371b-3p	AAGUGCCCCCACAGUUUGAGUGC
372	AAAGUGCUGCGACAUUUGAGCGU
373-5p	ACUCAAAAUGGGGGCGCUUUCC
373-3p	GAAGUGCUUCGAUUUUGGGGUGU

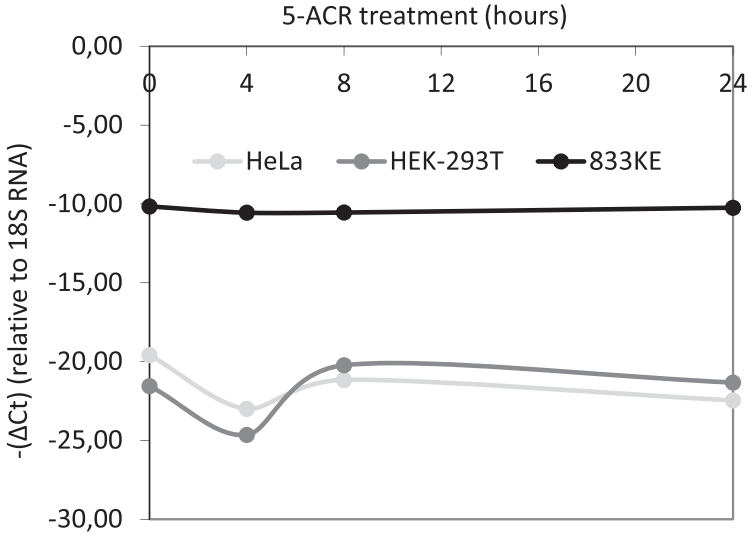
Underlined sequence indicates the seed, which are taken from miRbase release 18. MiRNA cluster members in bold indicate the most abundantly expressed form.



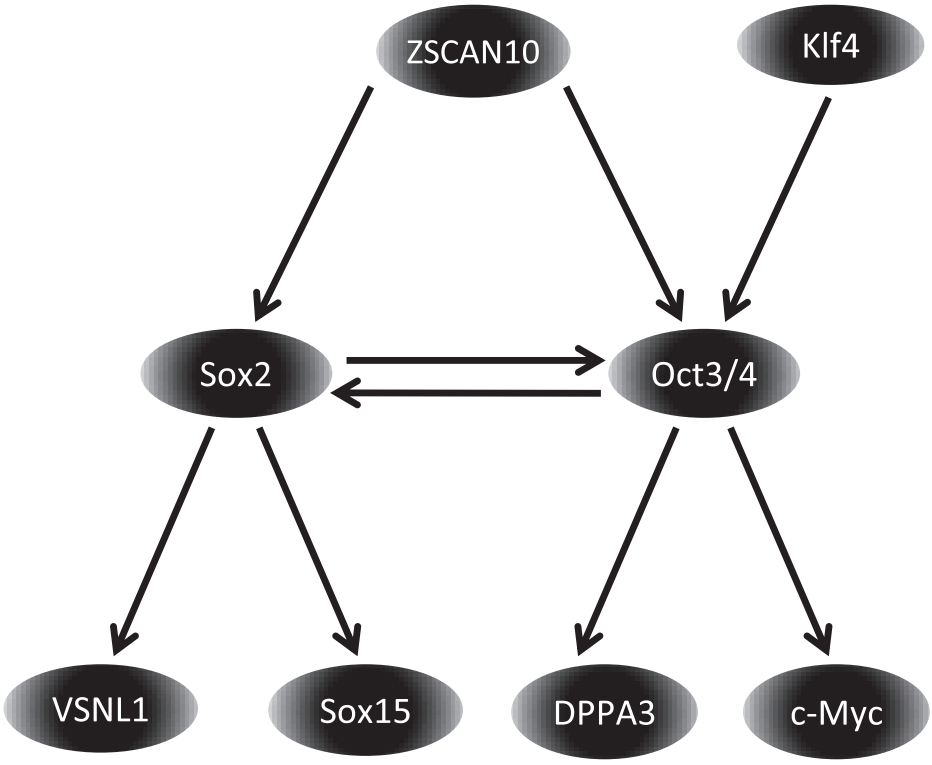
Supplementary figure 1 | Transcription activation capacity of the putative promoter region. 833KE, 2102Ep, Tera-2 and TCam-2 cells were transfected with the luciferase reporter constructs containing a genomic region of indicated length, a control promoter (SV40) construct, or a empty vector. Relative luciferase activity was measured after 72 hours and is the ratio between firefly luciferase and *Renilla* luciferase as control, the SV40 control is set at 1. Error bars represent SD for triplicate experiments. The red dotted line the maximum luciferase activity that is produced by any region in all cell lines except 833KE.



Supplementary figure 2 | Impact of Sox2 knockdown on transcription activation mediated by the promoter region. 833KE cells were co-transfected with the indicated luciferase reporter construct and with Sox2 or control siRNAs. Relative luciferase activity was measured after 72 hours and is the ratio between firefly luciferase and *Renilla* luciferase, the SV40 positive control is set at 1. Error bars represent SD for triplicate experiments.



Supplementary figure 3 | Effect of DNA methylation inhibition on miRNA levels. 5-ACR saturates the cell's methylation machinery thereby interfering placement of methyl marks in dividing cells. After several hours of 5-ACR treatment, endogenous levels of miR-372 in HeLa, HEK293T and 833KE cells were measured by qRT-PCR, and the results are displayed as **–(Δ**Ct).



Supplementary figure 4 | Connections between SAM genes. Of the genes that are significantly overexpressed in EC cell lines highly expressing miR-371-373 compared to other cell lines, eight have been previously reported to reside in a common regulatory network. An arrow between genes indicates transcriptional activation of the gene (arrowhead) by the indicated TF (line).

SUPPLEMENTARY REFERENCES

1 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p2235 (2006).

2 Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-317, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p16522 (2007).

3 Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-324, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p18030 (2007).

4 Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell stem cell* **2**, 10-12, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p16412 (2008).

5 Judson, R. L., Babiarz, J. E., Venere, M. & Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature biotechnology* **27**, 459-461, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p14790 (2009).

6 Li, Z., Yang, C. S., Nakashima, K. & Rana, T. M. Small RNA-mediated regulation of iPS cell generation. *The EMBO journal* **30**, 823-834, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p18043 (2011).

7 Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p2561 (2007).

8 Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science*, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p2530 (2007).

9 Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature biotechnology* **26**, 101-106, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p18104 (2008).

10 Subramanyam, D. *et al.* Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature biotechnology* **29**, 443-448, doi:papers://92636505-AFAD-4AF7-B1F9-D10C211B61AC/Paper/p15807 (2011).

chapter **FOUR**

A PUMILIO-INDUCED RNA STRUCTURE SWITCH IN P27-3' UTR CONTROLS MIR-221 AND MIR-222 ACCESSIBILITY

Martijn Kedde[1,*], Marieke van Kouwenhove[1,*], Wilbert Zwart[2], Joachim A. F. Oude Vrielink[1], Ran Elkon[1] and Reuven Agami[1,3]

[1]Division of Gene Regulation, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. [2]Division of Cell Biology II, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. [3]Centre for Biomedical Genetics UMCU, Stratenum 3.223, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands.

*These authors contributed equally to this work

ABSTRACT

Key regulators of 3' untranslated regions (3' UTRs) are microRNAs and RNA-binding proteins (RBPs)[1,2]. The p27 tumour suppressor is highly expressed in quiescent cells, and its downregulation is required for cell cycle entry after growth factor stimulation[3,4]. Intriguingly, p27 accumulates in quiescent cells despite high levels of its inhibitors miR-221 and miR-222 (REFS 5, 6). Here we show that miR-221 and miR-222 are underactive towards p27-3' UTR in quiescent cells, as a result of target site hindrance. Pumilio-1 (PUM1) is a ubiquitously expressed RBP that was shown to interact with p27-3' UTR[7,8]. In response to growth factor stimulation, PUM1 is upregulated and phosphorylated for optimal induction of its RNA-binding activity towards the p27-3' UTR. PUM1 binding induces a local change in RNA structure that favours association with miR-221 and miR-222, efficient suppression of p27 expression, and rapid entry to the cell cycle. We have therefore uncovered a novel RBP-induced structural switch modulating microRNA-mediated gene expression regulation.

RESULTS

MiR-221/222 is underactive towards p27 in quiescent cells

MicroRNAs (miRNAs) are genes involved in normal development and in cancer, mainly by associating with 3' untranslated regions (3' UTRs) of messenger RNAs, regulating their expression[9,10]. In a similar manner to miRNAs, RBPs can interact with 3' UTRs in a sequence-specific manner and can both stimulate and inhibit gene expression[1,2]. In particular, a member of the *Caenorhabditis elegans* Pumilio family

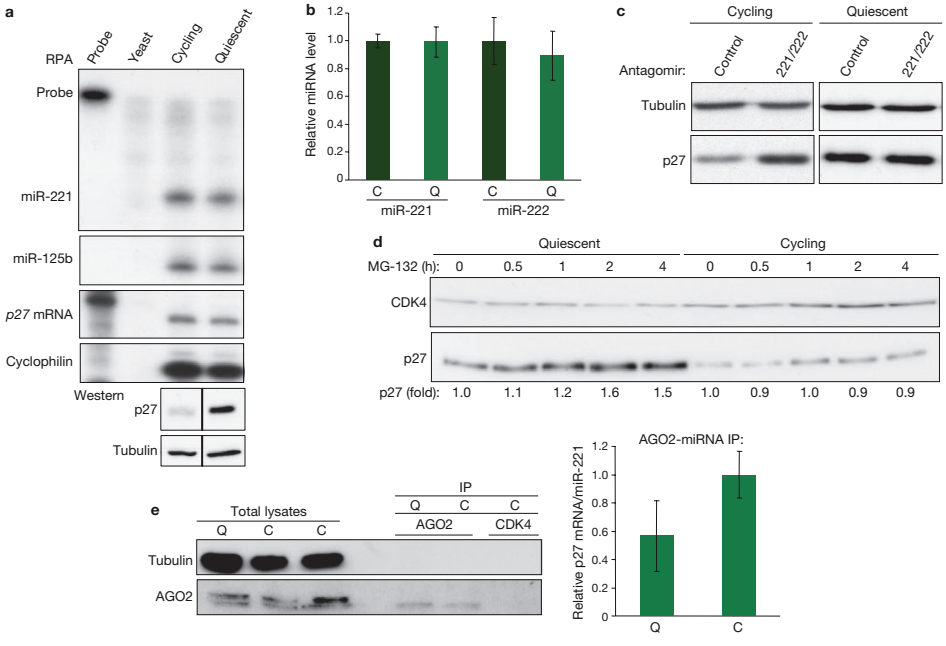


Figure 1 | miR-221 and miR-222 are underactive towards p27 in quiescent cells. (a) RNA was extracted from quiescent and cycling BJ primary fibroblasts and was subjected to RPA analysis for p27, miR-221 and miR-125b, with cyclophilin as control. Immunoblots were performed with antibody against p27, with anti-tubulin as control. Bands were spliced together from different parts of the same blot as indicated by the line. **(b)** The amounts of miR-221 and miR-222 were measured by qRT–PCR in cycling (C) and quiescent (Q) BJ cells. Error bars represent SD for triplicate reactions. **(c)** Quiescent and cycling BJs were treated with cholesterol-conjugated control or miR-221/miR-222 antagomirs. Immunoblot analysis was performed as in **a**. **(d)** Quiescent and cycling BJs were treated with the proteasome inhibitor MG-132. Time points are indicated; a densitometric analysis is shown below. Immunoblot analysis was performed with antibodies against p27 and CDK4. **(e)** BJ cell extracts were used for immunoprecipitation analysis with antibodies against AGO2 and CDK4. Immunoblots were performed with antibody against AGO2, with anti-tubulin as control. The amounts of miR-221 and *p27* mRNA were measured by qRT–PCR in the immunoprecipitates. Results are presented as relative p27/miR-221 ratio. The ratio in the immunoprecipitates (IP) from cycling BJs was set to 1. Enrichment factors of miR-221 and *p27* mRNA in AGO2 immunoprecipitates over CDK4 immunoprecipitates are shown in Supplementary Information, Fig. S2b. Error bars represent SEM for triplicate reactions. Uncropped images of blots are shown in Supplementary Information, Fig. S10.

(Puf-9) is required for 3' UTR-mediated regulation of the *let-7* target *hbl-1* (REF. 11). By association with hundreds of mRNAs, many coding for cell cycle regulators, Pumilio RBPs potentially influence expression by an as yet unknown mechanism[7,8]. High levels of miR-221 and miR-222 are required in many different cancer types to inhibit the expression of *p27* (*CDKN1B*; cyclin-dependent kinase inhibitor 1b) and stimulate proliferation[5,6]. p27 is a cyclin-dependent kinase inhibitor that negatively regulates cell cycle progression by association with cyclin-dependent kinase 2 (CDK2) and cyclin E complexes, resulting in the inhibition of the transition from G1 to S phase[4]. Accumulation of p27 protein is required for entry into quiescence (G0), and, on stimulation with growth factor, p27 levels must decrease to allow proper S-phase entry[3,4].

We asked whether the miR-221/miR-222 cluster is involved in p27 regulation in quiescence, because it is a negative regulator of p27 translation in many cancer cell types. We therefore examined p27 and miR-221/miR-222 levels in both quiescent and cycling BJ primary fibroblasts by RNase protection assays (RPAs), quantitative RT–PCR (qRT–PCR), and northern blot and expression array analyses. Although p27 protein level was clearly elevated in quiescent cells, *p27* mRNA and miR-221/miR-221 levels remained constant (Fig. 1a, b; Supplementary Information, Fig. S1a–d). We next inhibited miR-221 and miR-222 function by using miR-221 and miR-222 antagomirs (validated in REF. 5). Addition of miR-221 and miR-222 antagomirs, but not a control antagomir, to cycling BJ cells resulted in an increase in p27 levels (Fig. 1c). In contrast, addition of miR-221 and miR-222 antagomirs to quiescent BJ cells did not affect the level of p27 protein, suggesting that in quiescent cells miR-221 and miR-222 is less functional in suppressing its target, p27. Effective uptake of antagomirs in quiescent cells was demonstrated by a control directed against p53 short hairpin RNAs (shRNAs), in BJ-p53kd cells (Supplementary Information, Fig. S1e)[12]. Indeed, despite similar *p27* mRNA levels (Supplementary Information, Fig. S1b–d), p27 translation is increased in quiescent cells (Fig. 1d), indicating that the production of p27 protein is not inhibited in quiescent cells despite the presence of its miRNA inhibitor.

The activity of miRNAs can be dependent on accessibility to their target mRNAs[13]. To examine the association of miR-221 and miR-222 with p27-3' UTR in quiescent and cycling cells, we immunoprecipitated endogenous Argonaute 2 (AGO2, the main component of the RNA-induced silencing complex (RISC), directing miRNA target inhibition[14]) and measured the relative amounts of associated *p27* mRNA and miR-221/miR-222. As controls we used anti-CDK4 antibody for immunoprecipitation, and both glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and 18S ribosomal RNA for qRT–PCR. Although similar amounts of AGO2 were expressed and immunoprecipitated (Fig. 1e), less *p27* mRNA was associated with AGO2–miR-221/miR-222 in quiescent cells than in cycling cells (Fig. 1e; Supplementary Information, Fig. S2a, b). Previous formaldehyde crosslinking yielded similar results (Supplementary Information, Fig. S2c). Analysis of a control miRNA (miR-29a) and its target mRNA (collagen 3A1)[15] revealed opposite association ratios, indicating the specificity of this assay (Supplementary Information, Fig. S2d). These data suggest that *p27* mRNA in cycling cells is more accessible for interaction with miR-221 and miR-222.

Pumilio is required for miR-221/222 function

Recently, screens for mRNA targets of the RBP Pumilio revealed, among many genes, *p27* (refs 7, 8). The p27-3' UTR harbours two evolutionarily conserved Pumilio recognition elements (PREs); one is located close to the miR-221 and miR-222 target sites (Fig. 2a). The human Pumilio family contains two members, PUM1 and PUM2. We knocked down PUM1, the most abundant Pumilio family member, in miR-221/miR-222-expressing HEK293 cells[13]. Immunoblot analysis revealed elevated p27 protein levels in cells transfected with either of two functional PUM1 knockdown constructs (Fig. 2b). These data suggest that PUM1 inhibits p27 expression in HEK293 cells.

Because both PUM1 and miR-221/miR-222 inhibit p27 expression, we measured the effect of PUM1 on miR-221-induced repression of a luciferase reporter gene coupled to the 3' UTR of p27 in MCF7 cells, which endogenously express PUM1 but not miR-221 and miR-222 (ref. 13). As expected, co-transfection of miR-221 resulted in decreased luciferase activity (Fig. 2c). On knockdown of PUM1 (Supplementary Information, Fig. S3a), miR-221 function was compromised. No effect of PUM1 knockdown on the reporter was seen in the absence of miR-221 (Fig. 2c). In addition, inactivating mutations in the miR-221 and miR-222 target sites also resulted in a loss of the PUM1 knockdown effect (Supplementary Information, Fig. S3b). Moreover, mutating the PREs in p27-3' UTR also compromised the PUM1 knockdown effect, whereas miR-221 function remained intact (Fig. 2d). The fact that PUM1 knockdown abolished miR-221 function, but loss of its binding sites on the p27-3' UTR did not, suggests that PUM1-induced changes in mRNA structure are involved in regulating miR-221 function (see below). Taken together, these results indicate, first, that both PUM1 and miR-221 inhibit p27 expression post-transcriptionally through sites in p27-3' UTR, and second, that efficient suppression of p27 expression by miR-221 requires Pumilio.

Stimulation of PUM1 RNA-binding activity during cell cycle entry from quiescence

We next examined whether PUM1 RNA-binding activity is altered between quiescent and cycling cells. To test this we developed an assay to measure the RNA-binding capacity of PUM1 *in vivo* with a Cy3-tagged RNA oligonucleotide corresponding to the 5' PRE in the p27-3' UTR (Cy3-RNA) and a green fluorescent protein (GFP)-tagged PUM1 (Fig. 3a; Supplementary Information, Fig. S4a, b). These were microinjected into both quiescent and cycling BJ fibroblasts and revealed by fluorescent confocal microscopy. GFP–PUM1 showed a granular localization pattern, as reported previously[8]. Similar localization was observed with endogenous PUM1 (Supplementary Information, Fig. S4c). These granules are juxtaposed to P-bodies, which contain miRNAs and repressed mRNAs and are thought to be sites of translational repression[8,16,17]. We found PUM1 and AGO2 to co-localize to granules in both quiescent and cycling cells, although no direct interaction between the two could be shown (Supplementary Information, Fig. S4c, and data not shown). As a result of non-specific adhesion of Cy3-RNA oligonucleotides to chromatin, a partial nuclear localization was observed. GFP–PUM1 and Cy3-tagged wild-type RNA (Cy3-wt-RNA) showed strong co-localization in the cytosol of cycling cells

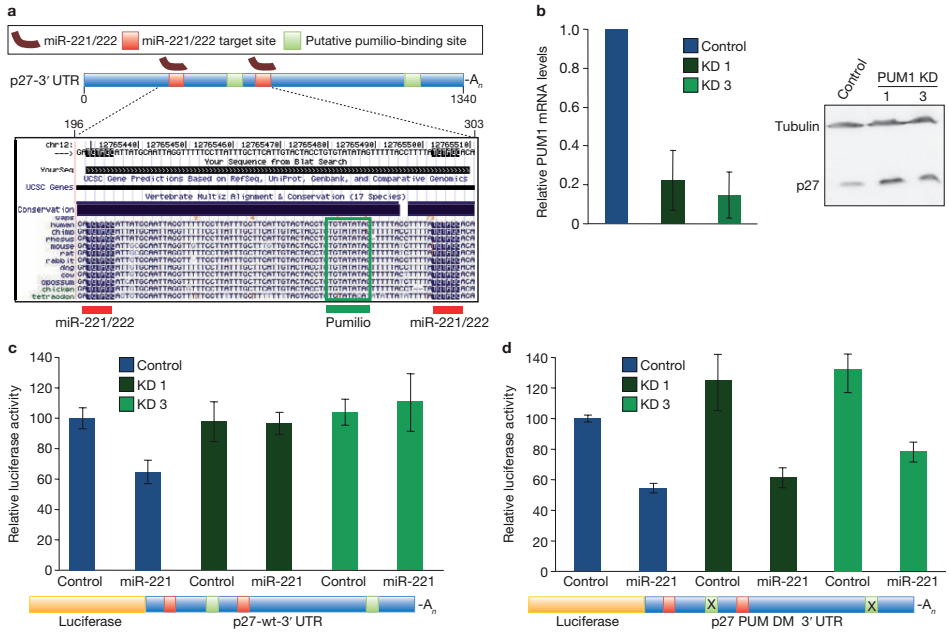


Figure 2 | Pumilio is required for miR-221 and miR-222 function. (a) Conservation analysis of p27-3' UTR from human to fish[25]. The positions of the binding sites for miR-221 and miR-222 and the PREs (consensus sequence 5'-UGUANAUA-3') are marked. **(b)** HEK293 cells were transfected with shRNA constructs targeting PUM1 and control. Cells were subjected to quantitative RT–PCR for PUM1 and actin control. Error bars represent SD for triplicate reactions. Right: immunoblot analysis as in Fig. 1a. KD, knock-down. **(c)** MCF7 cells were co-transfected with expression vectors coding for luciferase coupled to the wild-type p27-3' UTR, miR-221 and hTR control, and shRNA vectors against PUM1 or control. Relative luciferase activity is the ratio between firefly luciferase and *Renilla* control, adjusted to 100%. A schematic representation of the p27-3' UTR is shown below. Error bars represent SD for triplicate experiments. **(d)** Luciferase assay performed as in **c**, with a luciferase construct coupled to the p27-3' UTR mutated for the PREs. Error bars represent SD for triplicate experiments. Uncropped images of blots are shown in Supplementary Information, Fig. S10.

(Fig. 3a, middle panel; Supplementary Information, Fig. S4b). When GFP–PUM1 and Cy3-wt-RNA were injected together into quiescent BJ cells (Fig. 3a, top panel), or when an RNA oligonucleotide with two nucleotide alterations in the PRE was injected with GFP–PUM1 into cycling cells, no co-localization was observed (Fig. 3a, bottom panel). PUM1 co-localization with Cy3-wt-RNA was specific, because it was not observed with several other RBPs (Supplementary Information, Fig. S4d, e). Direct and specific binding of PUM1 to wild-type, but not mutant, RNA is shown in immunoprecipitation-binding assays with radioactively labelled probes and PUM1–TAP (tandem affinity purification; Fig. 3b). These results indicate that the RNA binding of PUM1 is specific and its RNA-binding capacity, at least towards the p27-3' UTR, is low in quiescent cells and high in cycling cells.

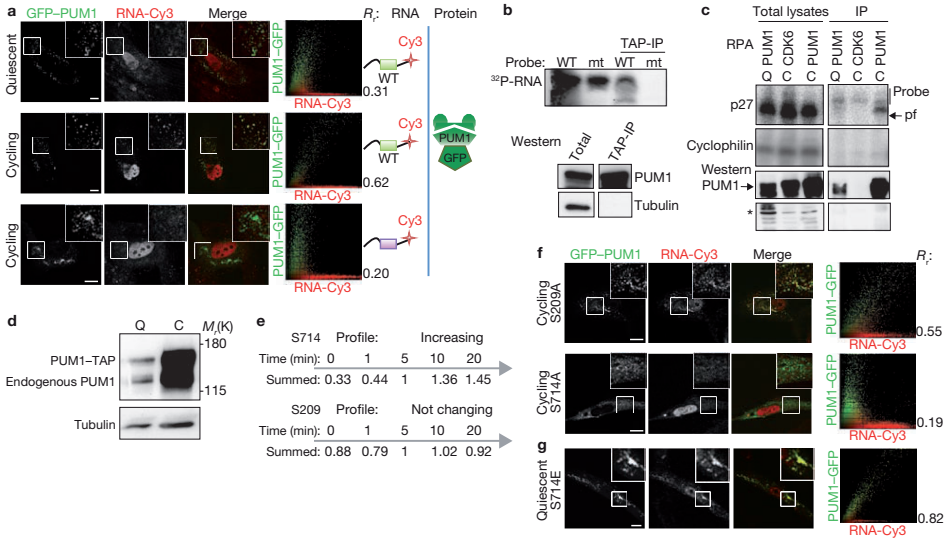


Figure 3 | PUM1 RNA-binding activity is enhanced in cycling versus quiescent cells. (a) Quiescent or cycling BJ primary fibroblasts were microinjected with GFP–PUM1 constructs and a Cy3-labelled RNA. After incubation overnight, cells were fixed and revealed by confocal laser scanning microscopy (CLSM). Co-localization is shown in the merge panel, and the correlations between GFP and Cy3 signals within the same cell were ascertained with a scatter plot. Representative pictures are shown. Insets: enlargements of the highlighted areas. Pearson's correlation coefficients (R_r) are shown in the scatter plots. green = wild-type; purple = mutant. Scale bar, 10 mm. **(b)** Binding assay of immunoprecipitated PUM1–TAP from HEK293 cells and ^{32}P-labelled wild-type (wt) and mutant (mt) p27 RNA. Gel image displays both unbound and bound probes. Total lysate and immunoprecipitate (TAP-IP) were analyzed by immunoblotting with antibody against PUM1, with anti-tubulin as control. **(c)** RPA for *p27* mRNA and cyclophilin negative control was performed on immunoprecipitates of PUM1 complexes and of CDK6 control from cycling (C) and quiescent (Q) BJ cells. pf, protected fragment. Quantification is shown in Supplementary Information, Fig. S4f. Total lysates and immunoprecipitates (IP) were analyzed by immunoblotting with antibody against PUM1. Asterisk, loading control. **(d)** Immunoblot analysis of endogenous and overexpressed PUM1 on growth factor stimulation with antibody against PUM1, with anti-tubulin as control. **(e)** Data adapted from the PHOSIDA phosphorylation site database showing the phosphorylation of PUM1 at Ser 209 and Ser 714 along the course of stimulation with epidermal growth factor (in minutes). **(f)** Cycling BJ cells were microinjected with inactivated phospho-mutants of GFP–PUM1 (S209A, S714A) in combination with Cy3-tagged RNA. The experiment was performed as in **a**. **(g)** Quiescent BJ cells were microinjected with a phospho-mimic mutant (S714E) of GFP–PUM1, in combination with Cy3-tagged *p27* RNA. The experiment was performed as in **a**. Uncropped images of blots are shown in Supplementary Information, Fig. S10.

We also confirmed this conclusion by immunoprecipitations coupled to RPA of endogenous PUM1 in quiescent and cycling BJ cells. Immunoprecipitation with anti-PUM1 antibody, but not an anti-CDK6 control, from cycling BJ cells confirmed binding of PUM1 to *p27* mRNA (Fig. 3c; quantification is shown in Supplementary Information, Fig. S4f). The cyclophilin RNA-negative control was not enriched in the immunoprecipitations, indicating the specificity of PUM1 binding. In contrast,

in quiescent BJ cells no *p27* mRNA was detected in PUM1 immunoprecipitation. Immunoblot analysis revealed higher levels of endogenous PUM1 in cycling BJ cells than in quiescent cells (Fig. 3c). This effect was observed with both endogenous and stable exogenous tagged PUM1 (Fig. 3d), indicating post-translational modifications. Taken together, our observations show that, on cell cycle entry from quiescence, PUM1 levels increase and its RNA-binding activity is turned on.

A study of phosphorylated proteins in HeLa cells reported unchanged phosphorylation of PUM1 Ser 209 on stimulation with epidermal growth factor, whereas Ser 714 phosphorylation was rapidly increased up to about fivefold (Fig. 3e)[18,19]. To examine whether these phosphorylation sites affect the RNA-binding activity of PUM1 in cycling cells, GFP–PUM1 phospho-mutants (Supplementary Information, Fig. S4a) were microinjected together with the Cy3-RNA oligonucleotides into cycling BJ cells. Mutation of Ser 714 to alanine (S714A) decreased the RNA-binding activity of PUM1 for Cy3-wt-RNA in cycling cells (Fig. 3f, lower panel), whereas the S209A mutant was as active as wild-type PUM1 (Fig. 3f, upper panel). Furthermore, a phospho-mimic mutation of Ser 714 to glutamic acid (S714E) showed persistent RNA-binding activity in quiescent cells (Fig. 3g and Supplementary Information, Fig. S4g). In contrast, the RNA-binding domain of PUM1 (GFP–PUM1(HD); HD: homology domain) is mostly nuclear, and the cytoplasmic fraction does not co-localize to Cy3-wt-RNA (Supplementary Information, Fig. S4h). This suggests that residues outside the HD domain are essential for cytoplasmic localization and binding specificity. Although the results above do not exclude the involvement of other modification events in the activation process of PUM1, they suggest that both PUM1 upregulation and phosphorylation of Ser 714 in response to stimulation with growth factor are necessary and sufficient to increase the RNA-binding activity of PUM1 in BJ fibroblasts.

Pumilio is required for proper cell cycle re-entry in a p27-dependent manner

Next we examined the effect of Pumilio on endogenous p27 expression and cell cycle re-entry from quiescence. PUM1 knockdown resulted in a delayed re-entry into the cell cycle from quiescence (Supplementary Information, Fig. S5a) despite modest differences in p27 levels (see below); this can be explained by haploinsufficiency of p27 (ref. 20). We noticed that the levels of PUM2, a homologue of PUM1 that is also expressed in BJ and HEK293 cells, increased when PUM1 was suppressed by RNA-mediated interference (Fig. 4a and data not shown). The presence of two PREs in the 3' UTR of PUM2 could explain this[7]. Suppression of PUM2, like that of PUM1, led to an increase in p27 levels and a comparable delay in cell cycle re-entry, suggesting a redundant activity with PUM1 (Supplementary Information, Fig. S5b). Knockdown of both PUM1 and PUM2 significantly increased p27 protein levels by elevating translation (Fig. 4b) and halted proliferation (Fig. 4c). Knockdown of PUM1 and PUM2 in BJ cells caused a delayed entry into S-phase on stimulation with growth factor, whereas BJ cells containing a stable p27 knockdown were insensitive to the loss of PUM1 and PUM2 (Fig. 4d; Supplementary Information, Fig. S5c). These results indicate that Pumilio proteins control cell cycle re-entry in response to growth factors, and that this function is in part mediated by controlling p27 expression.

Pumilio binding alters local p27-3'UTR structure and miR-221/222 accessibility

Using the secondary structure prediction RNAfold software (Vienna RNA package version 1.8.3)[21], we noticed that the PRE and the miR-221 and miR-222 target site could form a stem-loop structure with considerable base-pair probability (Fig. 5a; Supplementary Information, Fig. S6). We therefore speculated that PUM1-binding to the PRE favours opening of the stem-loop structure, allowing miR-221 and miR-222 to gain access to the p27-3' UTR in cycling cells. To study changes in RNA secondary structure *in vivo*, we tagged RNA oligonucleotides containing the p27-3' UTR PRE and the proximal miR-221/miR-222-binding site, with both 3' (fluorescein) and 5' (Cy3) fluorophores. On microinjection of this RNA, the fluorescein lifetime in cycling BJ cells was significantly longer than in quiescent cells, as a result of decreased energy transfer (FRET; fluorescence resonance energy transfer) to the Cy3 fluorophore (Fig. 5b). This suggests an increased distance between the two fluorophores and thus a more open conformation of the stem-loop structure in cycling cells. To examine the potential differences in donor-fluorophore lifetime and to test the specificity of this assay, we microinjected mutant RNAs with strong and weak predicted secondary structures (Fig. 5b; Supplementary Information, Fig. S7). A PRE mutant RNA that was energetically more stable than the wild-type RNA (strong mutant) showed a short fluorescein lifetime in both quiescent and cycling cells, suggesting an unchanged, closed, RNA conformation. In contrast, an energetically weak structured RNA mutated in the miRNA site (weak mutant) maintained a longer fluorescein lifetime in both quiescent and cycling cells, suggesting an open conformation in both conditions. Because the changes in FRET observed with the wild-type RNA were within the range indicated by the mutant RNAs, our results imply that the measured changes in FRET represent actual structural differences. We also tested changes in luciferase–p27-3' UTR reporter activity on mutation of either PUM binding site with strong or weak complementarity to the miRNA sites. Whereas the weak mutant of both PUM sites permitted miRNA-mediated repression, altering either or both of the PUM sites to strong mutants abrogated miRNA activity (Supplementary Information, Fig. S8). These results suggest a functional interaction between Pumilio and miR-221/miR-222 through their binding sites on the p27-3' UTR and indicate that both Pumilio sites in p27-3' UTR may contribute to the miRNA inhibitory structure *in vivo*.

On PUM1 knockdown in quiescent BJ cells, donor lifetime was not affected when compared with transfection of control siRNAs, which is consistent with an inactive state of PUM1 (Fig. 5c). In contrast, knockdown of PUM1 in cycling cells abolished the increase in donor lifetime, suggesting that the changes in conformation observed with the wild-type oligonucleotide are dependent on PUM1 protein. These results are supported by *in vivo* crosslinking of BJ cells and RT–PCR with a primer designed to detect the structured RNA loop specifically. A PCR product indicating a closed p27-3' UTR conformation was observed in quiescent cells and in PUM1 and PUM2 knockdown cells but not in cycling cells (Fig. 5d; Supplementary Information, Fig. S9).

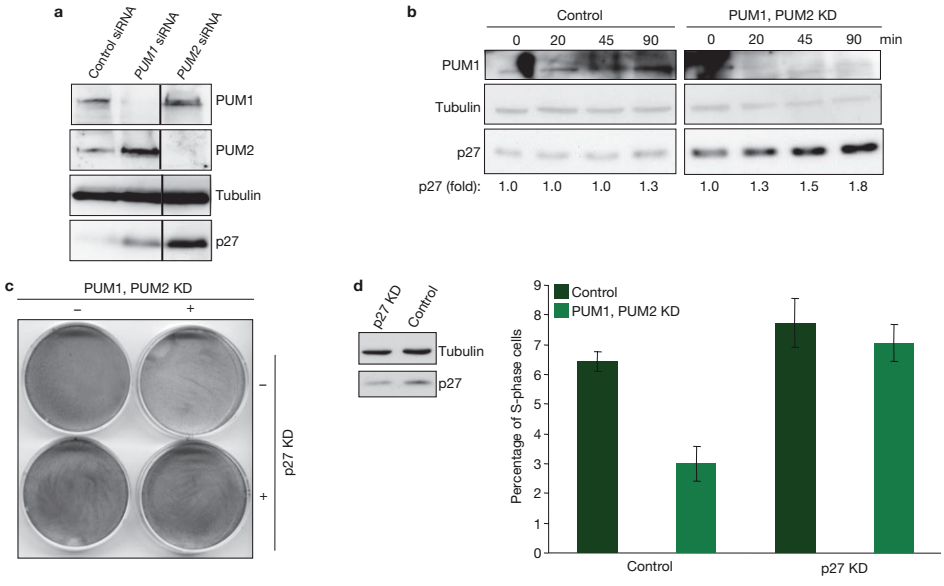


Figure 4 | Pumilio regulates p27-dependent cell cycle re-entry from quiescence. (a) HEK293 cells were transiently transfected with *PUM1* siRNA, *PUM2* siRNA and scrambled control siRNA, and immunoblot analysis was performed with antibodies against PUM1, PUM2 and p27, with anti-tubulin as control. **(b)** HEK293 cells transfected with *PUM1*- and *PUM2*-siRNA, or control, were treated with the proteasome inhibitor MG-132. A densitometric analysis at the indicated time points is shown below. Immunoblot analysis was performed as in **a**. **(c)** HEK293 cells were transfected with shRNA vectors against p27 or control, and with either *PUM1*- and *PUM2*-siRNA or control siRNA. After 3 days the cell densities were revealed by staining with Coomassie blue. **(d)** Wild-type BJ cells and BJ cells containing a stable p27 knockdown were transfected with either *PUM1*- and *PUM2*-siRNA or control siRNA, and deprived of growth factors for 72 h. After 16 h of subsequent stimulation with growth factor, the percentage of cells in S-phase was determined by flow cytometric analysis of bromodeoxyuridine incorporation. The percentage of cells in G1 and G2/M are shown in Supplementary Information, Fig. S5c. Error bars represent SD for three independent experiments. Immunoblot analysis of BJ cells stably expressing p27 KD or control was performed with antibody against p27, with anti-tubulin as control. Uncropped images of blots are shown in Supplementary Information, Fig. S10.

CONCLUSION

Taken together, our results provide evidence in support of a model in which, on stimulation by growth factors, Pumilio levels are increased and RNA-binding activity is further enhanced by phosphorylation inducing a conformational change in the p27-3' UTR (Fig. 5e). These changes permit a more efficient binding of miR-221 and miR-222 specifically to their target sites on the p27-3' UTR and tuning of cell cycle progression by repressing p27 expression. In addition, miRNA upregulation in response to growth factors has been reported in cancer cells, resulting in global target downregulation, implying distinct modes of regulation to achieve target specificity[22]. Our results reveal a highly conserved, specific case of complementarity

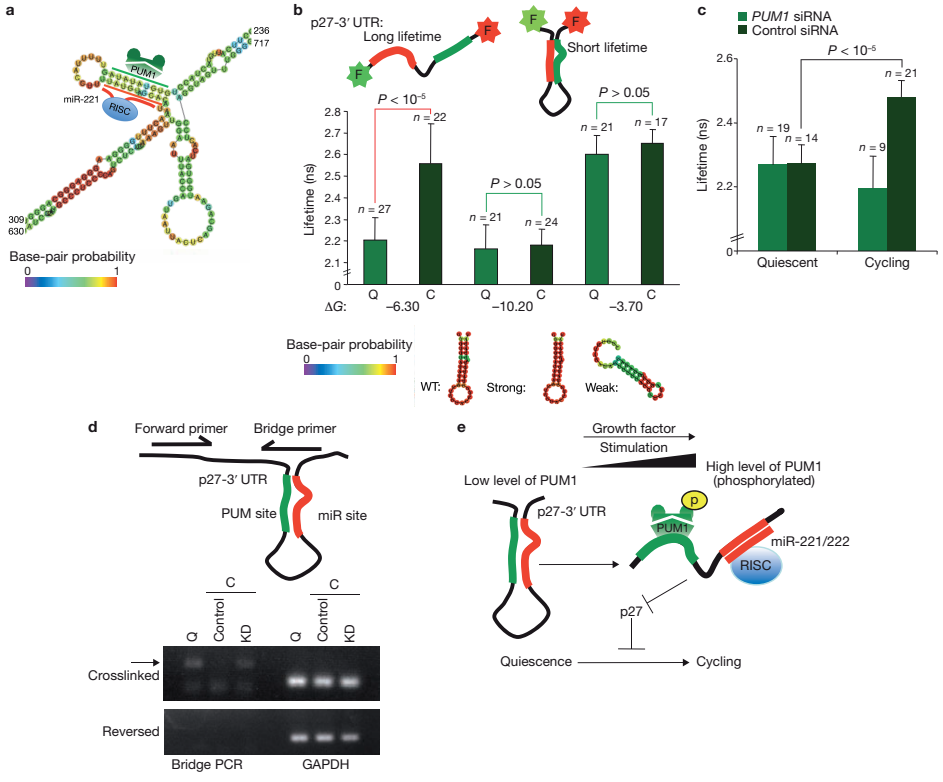


Figure 5 | Pumilio binding alters local p27-3' UTR structure and miR-221 and miR-222 accessibility. (a) Schematic representation of the conformation of a region of the p27-3' UTR containing a PRE and a miR-221/miR-222 site as predicted by RNAfold software. Base-pair probability is indicated in the key. **(b)** Quiescent (Q) and cycling (C)BJ cells were microinjected with short RNAs containing the p27-3' UTR-PRE and the proximal miR-221/miR-222-binding site, and tagged with both 3' (fluorescein) and 5' (Cy3) fluorophores. The amount of conformational free energy (DG in kcal mol-1) is listed for the wild-type (WT) and the two mutant short RNAs (named accordingly 'strong' and 'weak'). Differences in Cy3 fluorophore lifetime (in ns) due to FRET are shown, and *P* values are calculated for the differences in lifetime. Error bars represent SD, **(c)** BJ cells were transfected with *PUM1* siRNA and control siRNA and microinjected with the wild-type short RNA as in **b**. Error bars represent SD. **(d)** Model representing part of the p27-3' UTR, indicating the sequences recognized by the bridge primer used for RT and PCR. Ethidium-bromide-stained gels of PCRs performed on bridge or GAPDH reverse primer-primed cDNA from crosslinked or crosslink-reversed RNA isolated from quiescent (Q), cycling (C) and cycling PUM1 and PUM2 knockdown (KD) BJ cells. **(e)** Model proposing a role for Pumilio RBPs in mammalian somatic cells. See the text for details. Uncropped images of blots are shown in Supplementary Information, Fig. S10.

of an RBP target motif to a miRNA-binding site. To our knowledge, this is the only demonstration of an RBP that modulates miRNA activity by inducing a local structural switch in mRNA. Considering the generally high conservation of some 3' UTR regions, we expect that other RBPs may be found to modulate miRNA regulation of other genes in a similar manner.

METHODS

Constructs and antibodies

MiR-Vec constructs and the pGL3-p27-3' UTR and miR mutants were described previously[5,23]. The PREs in the p27-3' UTR were mutated (weak) to the following sequences using the Stratagene multisite-directed mutagenesis kit: PRE1, 5'-tgtatata-3' to 5'-ggtatgta-3'; PRE2, 5'-tgtacata-3' to 5'-ggtacgta-3' (strong mutants are shown below). Constructs for RPA detection of hTR, cyclophilin, p27 and miR-221 were described previously[13]; the RPA probe sequence for miR-125b was 5'-CUCAGUCCCUGAGACCCUAACUUGUGAUGUUU-3'. Probes were prepared in accordance with the manufacturer's instructions (Ambion mirVana probe construction kit). shRNA for p27 was described previously[5]; the shPUM1.1 sequence was 5'-AATCCAACATGTACTGGAGCA-3', the shPUM1.3 sequence was 5'-AACAGACCACCCCACAGGCTC-3', the shPUM1.4 sequence was 5'-AATT CAGCTAATCAACAGACC-3'; these were cloned in pRETRO-SUPER. siRNAs ordered from Ambion were against PUM1 (no. 138317), PUM2 (no. 138319) and a scrambled control (5'-CUGUAGCCGUAUCAAGUCGUUCCUGTT-3') from Invitrogen. The PUM1–TAP construct was a gift from A. Gerber. The GFP–PUM1(HD) and GFP–RBP constructs were made by cloning the cDNA into the Clontech eGFP vector; mutants were made with the Stratagene multisite-directed mutagenesis kit. All constructs were sequence-verified. The Cy3-wt-RNA oligonucleotide (5'-ACUACCU GUGUAUAUAGUUUUU-3') and the Cy3-mt-RNA oligonucleotide (5'-acuaccucucca uauaguuuuu-3') were labelled 3' (Dharmacon). Labelled RNA oligonucleotides (3' (fluorescein) and 5' (Cy3)) used for FLIM were wild-type (5'-CUGUGUAUAU AGUUUUUACCUUUUAUGUAGCACAU-3'), strong mutant (5'-CUGUGCACAUAG UUUUUACCUUUUAUGUAGCACAU-3') and weak mutant (5'-CUGUGUAUAUAGUUU UUACCUUUUAGGUCGGAGAU-3') (Dharmacon). Antibodies used were AGO2 (Transduction Labs and Abcam), actin (Abcam), p27 (Transduction Labs), CDK4 (C22), p53 (DO1) and CDK6 (Santa Cruz), PUM1 and PUM2 (Bethyl Labs), tubulin (YL1/2 ECACC), rabbit GFP and bromodeoxyuridine (Dako).

Cell culture, transfections, dual luciferase activity analysis and cell cycle profile analysis

HEK293, MCF7 and BJ primary fibroblast cells were cultured in DMEM supplemented with 10% heat-inactivated fetal calf serum (FCS) in 5% CO_2 at 37°C. HEK293 cells were transiently transfected by using calcium phosphate precipitation. MCF7 cells were transfected with Fugene (Roche) for luciferase analysis with 10 ng of reporter, 5 ng of *Renilla* control plasmid, 250 ng of either miR-Vec or miR-Vec control, and 250 ng of knockdown construct for PUM1 or control. Dual luciferase activity assays were performed 72 h after transfection in accordance with the manufacturer's instructions (Promega). BJ cells were transfected with siRNAs in a final concentration of 50 nM with the use of Dharmafect reagent (Dharmacon), in accordance with the manufacturer's instructions. To obtain quiescent BJ cells, cells were cultured for 72 h in DMEM containing 0.25% FCS. Antagomir sequences were described previously[5] and applied to the cells overnight at a final concentration of 15 mM.

The proteasome inhibitor MG-132 was from Sigma, used at a final concentration of 10 mM. For cell cycle profile analysis, quiescent BJs were stimulated with growth factors. Cell cycle analysis was performed as described previously[24].

Immunoprecipitation, immunoblotting, RNAse protection assays and qRT–PCR analysis

PUM1 and AGO2 were immunoprecipitated from BJ cell extracts with GammaBind G Sepharose (GE Healthcare). Beads were preblocked with yeast tRNA (Invitrogen) and RNase-free BSA (Ambion) and then washed; extracts were sonicated and cleared in lysis buffer (100 mM KCl, 10 mM Tris-HCl pH 7.5, 0.1% Nonidet P40, 0.5% Tween 20, 5 mM $MgCl_2$, 2 mM β-glycerophosphate, 0.5 mM dithiothreitol, protease inhibitor mixture (Roche Applied Science) and RNase-OUT (Invitrogen). Extracts were incubated for 4 h with antibodies against AGO2, CDK4, CDK6 or PUM1 (1 mg per immunoprecipitation) in a tumbler placed at 4 °C. Thereafter, beads were washed and a 10% aliquot was used for immunoblot analysis; from the remainder, RNA was extracted (Trizol, Invitrogen) to be subjected to RPA or qRT–PCR analysis. PUM1–TAP was precipitated from transiently transfected HEK293 cells by using rabbit IgG Sepharose (Sigma) in lysis buffer with 125 mM NaCl instead of KCl as described above. Beads were washed and incubated for 20 min with ^{32}P-labelled oligonucleotides (wild-type and mutant, described above) at 30 °C. Beads were washed, and bound RNA and proteins were revealed on gel.

For immunoblot analysis, extracts were separated on 10% SDS–PAGE gels, and transferred to Immobilon-P membranes (Millipore). Western blots were developed with Supersignal (Pierce) or by enhanced chemiluminescence (ECL; Amersham Biosciences) and exposed to film (Kodak). Densitometric analysis was performed with AIDA software (Raytest).

RPAs for p27 and cyclophilin were performed with the HybSpeed RPA and MAXIscript kits (Ambion) as described[13]. For miRNAs, we used mirVana kits (Ambion) in accordance with the manufacturer's instructions[13]. Northern analysis was performed with standard protocols and RPA probe for p27.

For mRNA qRT–PCR, cDNA (from 3 μg RNA) was synthesized with SuperScript III and primed with oligo(dT) in accordance with the manufacturer's instructions (Invitrogen). For combined miRNA and mRNA qRT–PCR, about 100 ng of input RNA and 20% of immunoprecipitated RNA was used for cDNA synthesis with random primers from a Taqman High Capacity cDNA kit (Applied Biosystems), in accordance with the manufacturer's instructions. Primers for qPCR were PUM1 (5'-AAAAACCTGAGAAGTTTGAATTGT-3' (forward) and 5'-GCAAGACCAAAAGCAGAGTTG-3' (reverse)) and COL3A1 (5'-AACA CGCAAGGCTGTGAGACT-3' (forward) and 5'-GCCAACGTCCACACCAAATT-3' (reverse)); p27, GAPDH and b-actin primers were as described[13]. QPCR primers for miR-221, miR-222, miR-29a, GAPDH and 18S were from Applied Biosystems. Analysis was performed with SYBR Green PCR master mix or TaqMan UNG master mix (Applied Biosystems) and Chromo 4 system (Bio-Rad Laboratories).

Crosslink bridge RT–PCRs were performed with bridge reverse (5'-CTTCCCCAAAGTTTATAGGTAG-3') and GAPDH reverse primer; PCR forward

primer was 5'-TATAAGCACTGAAAAACAACAACAC-3'. BJ cells were crosslinked for 15 min with 1% formaldehyde (Sigma), inactivated with 330 mM glycine (Sigma), sonicated and cleared. Cleared lysate was treated for 1 h with proteinase K (Invitrogen) at 37 °C and inactivated with phenylmethylsulphonyl fluoride (Sigma). RNA was extracted, and reverse crosslinking was performed for 1 h at 70 °C.

Fluorescence lifetime imaging microscopy (FLIM)

Before FLIM experiments, cells were grown on coverslips and microinjected with RNA labelled 3' with fluorescein and 5' with Cy3. Subsequently, cells were mounted in bicarbonate-buffered saline (140 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 23 mM $NaHCO_3$, 10 mM glucose, 10 mM HEPES at pH 7.3) in a heated tissue-culture chamber at 37 °C under 5% CO_2. FLIM experiments were performed on a Leica inverted DM-IRE2 microscope equipped with a Lambert Instruments frequency domain lifetime attachment (Leutingewolde), controlled by the vendor's LI FLIM software. Fluorescein was excited with about 4 mW of 488-nm light from a light-emitting diode modulated at 40 MHz; emitted light was collected at 490–550 nm with an intensified charge-coupled-device camera (CoolSNAP HQ; Roper Scientific). To calculate the fluorescein lifetime, the intensities from 12 phase-shifted images (modulation depth about 70%) were fitted with a sinus function, and lifetimes were derived from the phase shift between excitation and emission. Differences in lifetimes were assigned *P* values with Student's *t*-test.

CLSM analysis

For CLSM analysis, BJs were microinjected with GFP–PUM1 or its mutants, in combination with Cy3-labelled RNA. After expression of the GFP–PUM1 overnight, cells were fixed with 3.7% formaldehyde in PBS, and coverslips were mounted in Vectashield mounting medium (Vector Laboratories). The specimens were imaged with a Leica TCS SP2 System equipped with a 63x oil-immersion objective. Endogenous stainings for PUM1 and AGO2 were performed in accordance with the manufacturer's instructions. Scatter plots for co-localization analysis were generated with ImageJ WCIF software (http://www.uhnresearch.ca/wcif).

ACKNOWLEDGEMENTS

We thank all members of the Agami laboratory for technical help and discussions. We also thank André Gerber for constructs, Kees Jalink for advice on fluorescence lifetime imaging microscopy measurements, and R. B. Israel for assistance with statistical analysis. This work was supported by the EURYI (European research young investigator award), ERC (European Research Council), KWF (Koningin Wilhelmina Fonds; Dutch cancer foundation) and Horizon-NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; R.A.) and an EMBO long-term fellowship (R.E.).

AUTHOR CONTRIBUTIONS

M.K. and M.v.K. performed most of the experimental work. R.A. supervised the project. W.Z. performed fluorescence lifetime imaging microscopy and confocal laser scanning microscopy analyses. J.O.V. provided technical assistance. R.E. performed bioinformatical analyses. M.K., M.v.K. and R.A. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* **9,** 102–114 (2008).
2. Kedde, M. & Agami, R. Interplay between microRNAs and RNA-binding proteins determines developmental processes. *Cell Cycle* **7,** 899–903 (2008).
3. Hengst, L. & Reed, S. I. Translational control of p27Kip1 accumulation during the cell cycle. *Science* **271,** 1861–1864 (1996).
4. Chu, I. M., Hengst, L. & Slingerland, J. M. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat. Rev. Cancer* **8,** 253–267 (2008).
5. le Sage, C. *et al.* Regulation of the $p27_{Kip1}$ tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation. *EMBO J.* **26,** 3699–3708 (2007).
6. Lotterman, C. D., Kent, O. A. & Mendell, J. T. Functional integration of microRNAs into oncogenic and tumor suppressor pathways. *Cell Cycle* **7,** 2493–2499 (2008).
7. Galgano, A. *et al.* Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS ONE* **3,** e3164 (2008).
8. Morris, A. R., Mukherjee, N. & Keene, J. D. Ribonomic analysis of human Pum1 reveals *cis-trans* conservation across species despite evolution of diverse mRNA target sets. *Mol. Cell. Biol.* **28,** 4093–4103 (2008).
9. Kloosterman, W. P. & Plasterk, R. H. The diverse functions of microRNAs in animal development and disease. *Dev. Cell* **11,** 441–450 (2006).
10. Voorhoeve, P. M. & Agami, R. Classifying microRNAs in cancer: the good, the bad and the ugly. *Biochim. Biophys. Acta* **1775,** 274–282 (2007).
11. Nolde, M. J., Saka, N., Reinert, K. L. & Slack, F. J. The Caenorhabditis elegans *pumilio* homolog, *puf-9*, is required for the 3'UTR-mediated repression of the *let-7* microRNA target gene, *hbl-1*. *Dev. Biol.* **305,** 551–563 (2007).
12. Brummelkamp, T. R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2,** 243–247 (2002).
13. Kedde, M. *et al.* RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131,** 1273–1286 (2007).
14. Pillai, R. S. *et al.* Inhibition of translational initiation by Let-7 microRNA in human cells. *Science* **309,** 1573–1576 (2005).
15. van Rooij, E. *et al.* Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc. Natl Acad. Sci. USA* **105,** 13027–13032 (2008).
16. Leung, A. K., Calabrese, J. M. & Sharp, P. A. Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. *Proc. Natl Acad. Sci. USA* **103,** 18125–18130 (2006).
17. Kedersha, N. *et al.* Stress granules and processing bodies are dynamically

linked sites of mRNP remodeling. *J. Cell Biol.* **169,** 871–884 (2005).

18. Gnad, F. *et al.* PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8,** R250 (2007).
19. Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127,** 635–648 (2006).
20. Fero, M. L., Randel, E., Gurley, K. E., Roberts, J. M. & Kemp, C. J. The murine gene $p27_{Kip1}$ is haplo-insufficient for tumour suppression. *Nature* **396,** 177–180 (1998).
21. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31,** 3429–3431 (2003).
22. Medina, R. *et al.* MicroRNAs 221 and 222 bypass quiescence and compromise cell survival. *Cancer Res.* **68,** 2773–2780 (2008).
23. Voorhoeve, P. M. *et al.* A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124,** 1169–1181 (2006).
24. Duursma, A. & Agami, R. p53-dependent regulation of Cdc6 protein stability controls cellular proliferation. *Mol. Cell. Biol.* **25,** 6937–6947 (2005).
25. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).

SUPPLEMENTARY FIGURES

Figure S1a

M	A	P	primaryAccession	description
0.061836	9.61413	0.23458	11023	hsa-miR-222|mo-miR-222|mmu-miR-222

M	A	P	primaryAccession	description
0.191585	7.81673	0.24234	11022	hsa-miR-221|mo-miR-221|mmu-miR-221

Figure S1b

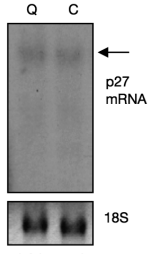


Figure S1c

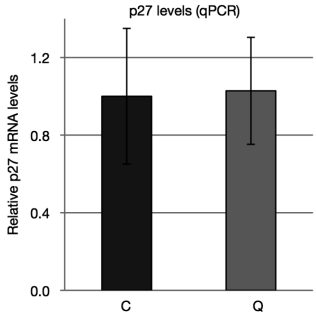


Figure S1d

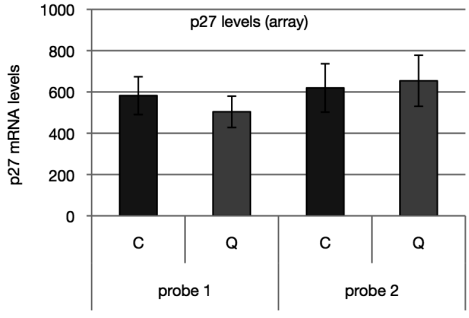


Figure S1e

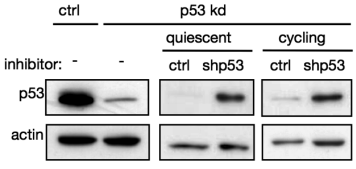


Figure S1 | MiR-221/222 and p27 levels in quiescent (Q) versus cycling (C) cells. (a) MiR-221 and miR-222 expression analysis on the Exiqon v2 microRNA microarray platform. M represents fold change (log2) as detected in quiescent versus cycling BJ cells. Hybridization was performed using a standard protocol (http://microarrays.nki.nl). **(b)** Northern blot for p27 mRNA and ethidium bromide staining for 18S ribosomal RNA in quiescent versus cycling BJ cells. Densitometric analysis resulted in the normalized amounts displayed below. (**c**) The amount of p27 mRNA was measured by qRT-PCR in cycling and quiescent BJ cells. Error bars represent SD from triplicate reactions. **(d)** Expression analysis of p27 on the Illumina Sentrix BeadChip v3 microarray platform. Absolute expression values were obtained with two probes in quiescent versus cycling BJ cells. The data is a representative of a duplo experiment. Hybridization was performed using a standard protocol (http://microarrays.nki.nl). **(e)** An antago-p53kd was administered to quiescent and cycling BJ-p53kd cells and immunoblot analysis on BJ-p53kd and control cells was performed with p53 and actin control antibodies.

Figure S2a

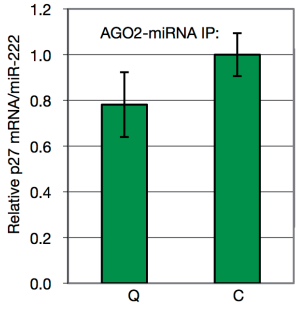


Figure S2b

Enrichment in AGO2 IP/CDK4 IP

	C	+/-		C	+/-
p27	**4.9**	2.1	p27	**3.2**	0.5
miR-221	**8.3**	0.9	miR-222	**3.7**	0.5
gapdh	**1.4**	0.1	gapdh	**1.2**	0.4

Figure S2c

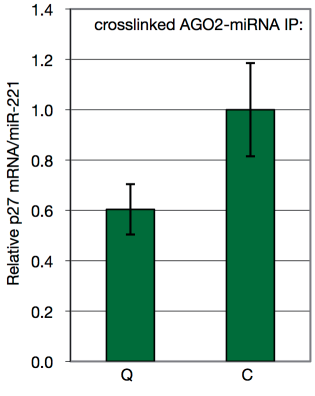
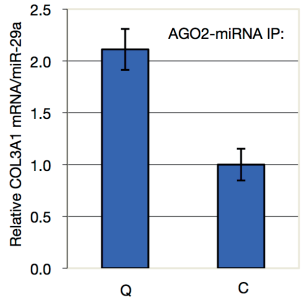


Figure S2d



	C	+/-
p27	**4.7**	1.3
miR-221	**5.0**	0.7
gapdh	**1.0**	0.2

Figure S2 | Relative amounts of miRNAs and associated mRNAs in AGO2 IPs. (a) The amounts of miR-222 and p27 mRNA were measured by qRT-PCR in the IPs shown in Fig. 1e. Results are presented as relative p27/miR-222 ratio. The ratio in the IPs from cycling BJ cells was set to 1. Error bars represent SEM from triplicate reactions. **(b)** Enrichment factors and SEM of miR-221, miR-222, p27 and gapdh control mRNA in AGO2 IPs over CDK4 IPs. **(c)** qRT-PCR performed as in **a**, for miR-221 and p27 mRNA in AGO2 IPs from formaldehyde crosslinked BJ cells. Enrichment factors and SEM of miR-221, p27 and gapdh control mRNA in AGO2 IPs over CDK4 IPs are shown in the table. **(d)** qRT-PCR performed as in **a** for miR-29a and COL3A1 mRNA.

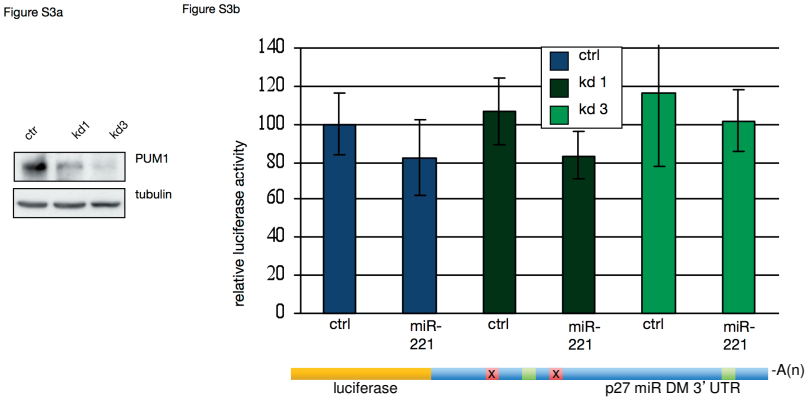


Figure S3 | Effect of PUM1 knockdown on miR-221 mediated regulation towards the p27-3'UTR. (a) MCF7 cells were transfected with shRNA vectors against PUM1 or control and selected, immunoblot analysis was performed with PUM1 and tubulin control antibodies. **(b)** Luciferase assay performed as in Fig. 2c, with a luciferase construct coupled to the p27-3'UTR mutated for the miR-221/222 sites. Error bars represent SD from three independent experiments.

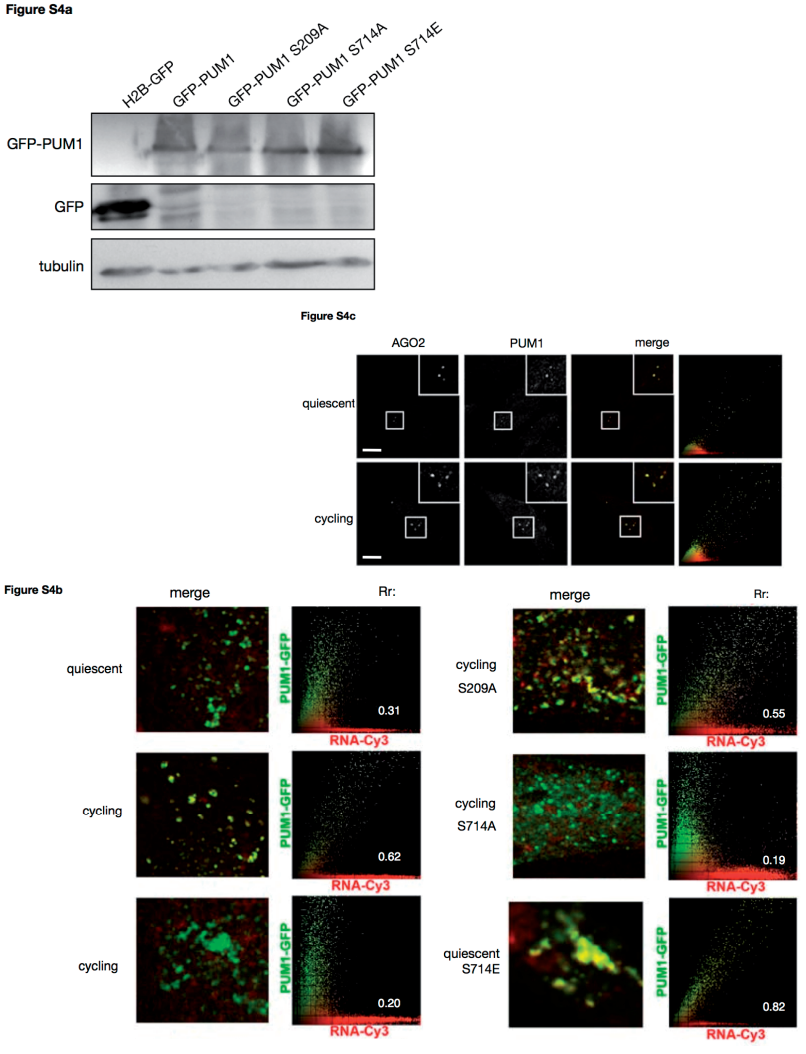


Figure S4 | Localization and expression level analysis of PUM1 and control RBPs. (a) Confirmation of expression levels of GFP-PUM1 constructs by immunoblot for GFP and tubulin. **(b)** Zoom-in of the inset in merge panel and scatterplots in Fig. 3a,f,g. **(c)** Immunostaining for endogenous PUM1 and AGO2 in quiescent and cycling BJ cells. Colocalisation analyses within the same cell were performed through a scatter plot. Inset shows zoom-in on the highlighted area. Scalebar represents 10 mm. Representative pictures are shown. **(d)** Cycling BJ cells were microinjected with GFP-RBP control constructs and a Cy3-labeled RNA as in Fig. 3a. Representative pictures are shown. **(e)** Confirmation of expression levels of GFP-PUM1HD and GFP-RBP constructs by immunoblot for GFP and tubulin. **(f)** Quantification of RPA signals as shown in Fig. 3c, quantifications were performed with Phosphoimager software. **(g)** Binding assay of immunoprecipitated PUM1–TAP from quiescent BJ cells and ^{32}P-labelled wild-type (wt) p27 RNA. Gel image displays both unbound and bound probe. Total lysate and immunoprecipitate (TAP-IP) were analyzed by immunoblotting with antibody against PUM1, with anti-tubulin as control. **(h)** Quiescent or cycling BJ cells were microinjected with the GFP-PUM1HD construct and a Cy3-labeled RNA as in Fig. 3a. Representative pictures are shown.

Figure S4d

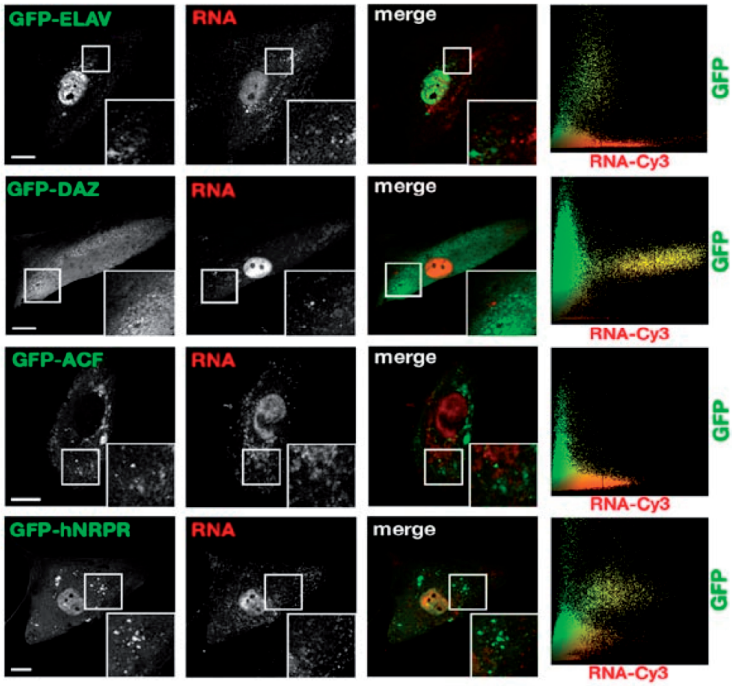


Figure S4e

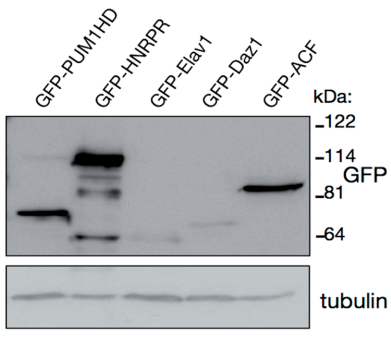


Figure S4f

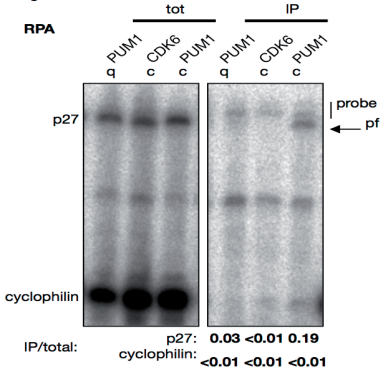


Figure S4g

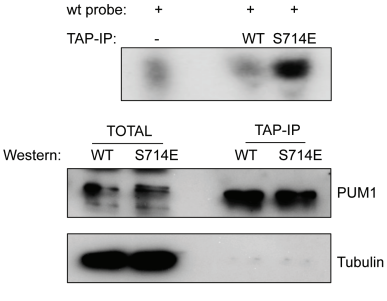


Figure S4h

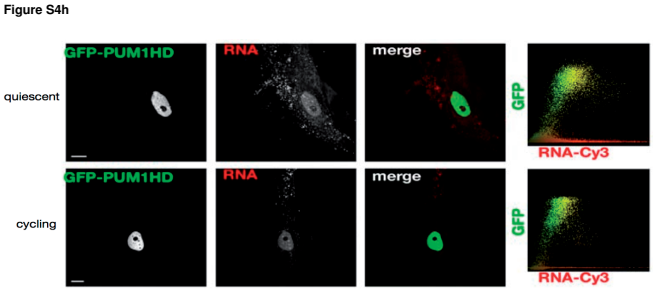
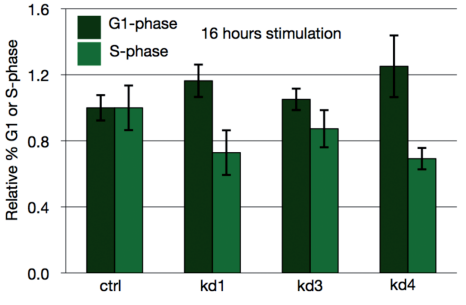
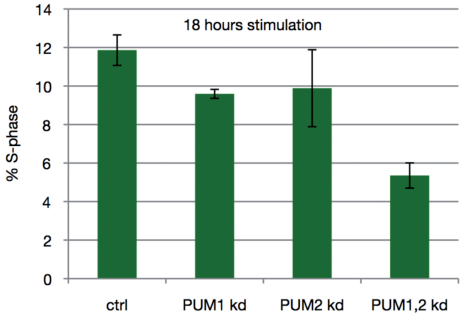


Figure S5a



%	ctrl	PUM1 kd1	PUM1 kd3	PUM1 kd4
G1	62.0	71.9	65.0	77.2
S	33.9	24.4	29.4	23.4
G2/M	2.9	3.0	3.6	2.7

Figure S5b



%	ctrl	PUM1 kd	PUM2 kd	PUM1,2 kd
G1	73.1	74.2	76.2	79.5
S	11.9	9.6	9.9	5.4
G2/M	14.8	16.0	13.4	14.4

Figure S5c

	%	ctrl	PUM1,2 kd
wt	G1	84.5	85.1
	S	6.5	3.0
	G2/M	8.4	11.3
p27 kd	%	ctrl	PUM1,2 kd
	G1	86.1	85.3
	S	7.7	7.1
	G2/M	5.8	7.2

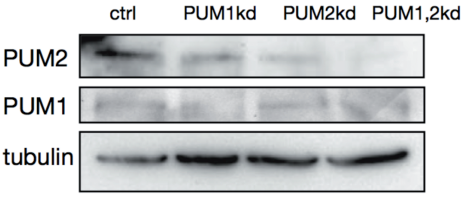


Figure S5 | Loss of Pumilio affects cell cycle progression. (a) BJ cells containing stable PUM1 knockdowns or control were growth factor deprived and then stimulated for 16 hours with growth factors. The percentage of cells in S-phase was determined by flow cytometric analysis of BrdU incorporation. G1- and G2/M-phase percentages as measured by propidium iodine are shown below. Error bars represent the SD of triplicate experiments. **(b)** BJ cells were transfected with either siPUM1, siPUM2, both, or control siRNA, and growth factor deprived for 72 hours. After 18 hours of subsequent growth factor stimulation, the percentage of cells in S-phase was determined by flow cytometric analysis of BrdU incorporation. G1- and G2/M-phase percentages as measured by propidium iodine are shown below. Error bars represent the SD of triplicate experiments. Immunoblot analysis of quiescent BJ cells stimulated with growth factors for PUM1&2 and tubulin control. **(c)** G1- and G2/M- phase percentages as measured by propidium iodine from Fig. 4d.

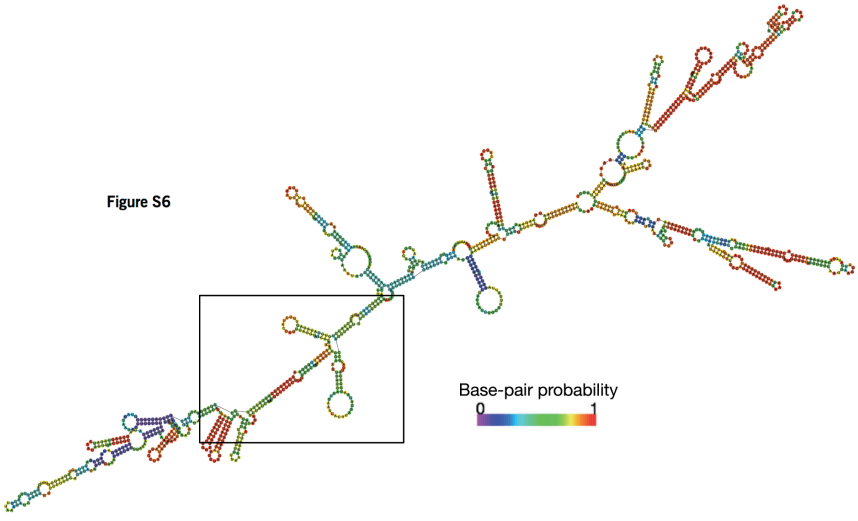


Figure S6 | Predicted conformation of the complete p27-3'UTR. A schematic representation of the conformation of the complete p27-3'UTR as predicted by RNAfold software. Base pair probability is indicated in the legend.

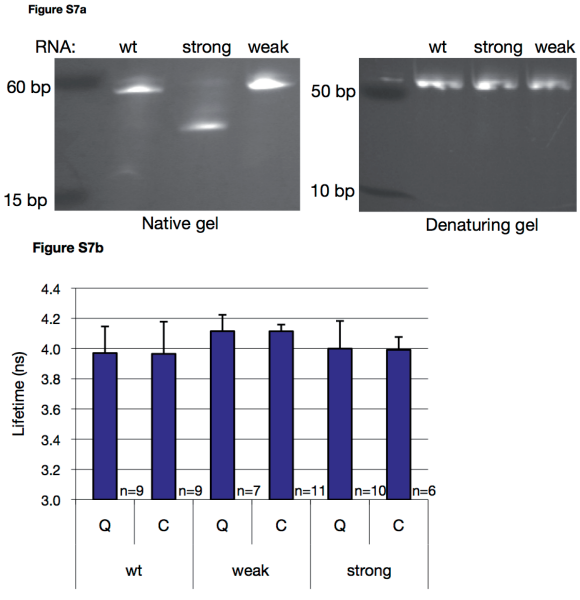


Figure S7 | Conformation of tagged RNA oligos and single labelled FLIM control. (a) Gel migration analysis of RNA oligos containing wild-type, strong and weak mutant p27-3'UTR-PRE and the proximal miR-221/222 binding site. **(b)** Quiescent (Q) and cycling (C) BJ cells were microinjected with short RNAs containing the p27-3'UTR-PRE and the proximal miR-221/222 binding site, and tagged with 3' (fluorescein) fluorophore. The wild-type, weak and strong mutants described in the text were used. Cy3 fluorophore lifetimes (in ns) due to FRET are displayed. Error bars represent SD.

Figure S8

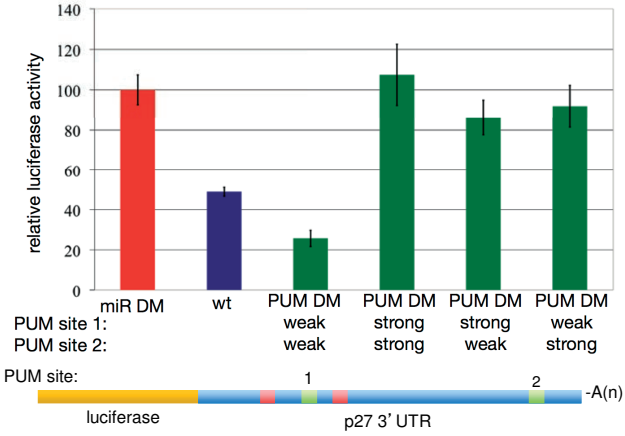


Figure S8 | Functional interaction between Pumilio and miRNA sites in the p27-3'UTR. Luciferase assay performed as in Fig. 2c, in HEK293 cells (endogenously expressing miR-221/222), with luciferase constructs coupled to the p27-3'UTR mutated for the miR-221/222 sites (miR DM), and several constructs mutated for both Pumilio sites (see schematic representation below). Pumilio sites 1 and 2 are shown in green, miRNA-221/222 sites are shown in red, weak and strong PUM site mutants are as described in the text. Error bars represent SD from three independent experiments.

Figure S9a

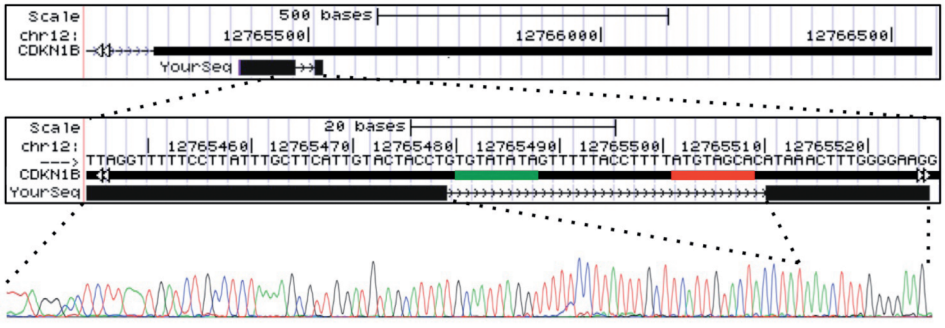


Figure S9b

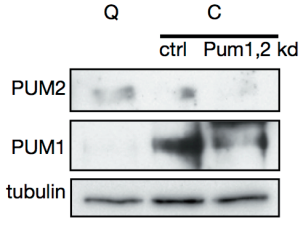


Figure S9 | *In vivo* crosslinking reveals predicted secondary p27-3'UTR structure. (a) Schematic representation of the bridge PCR product (yourseq) adapted from BLAT search function[23] with nucleotide numbers shown. Sequence is shown below, as expected the miR-221/222 site (red) and the Pumilio site (green) are missing from the PCR product. **(b)** Immunoblots showing levels of PUM1&2 and control tubulin from crosslinked BJ cells from Fig. 5d.

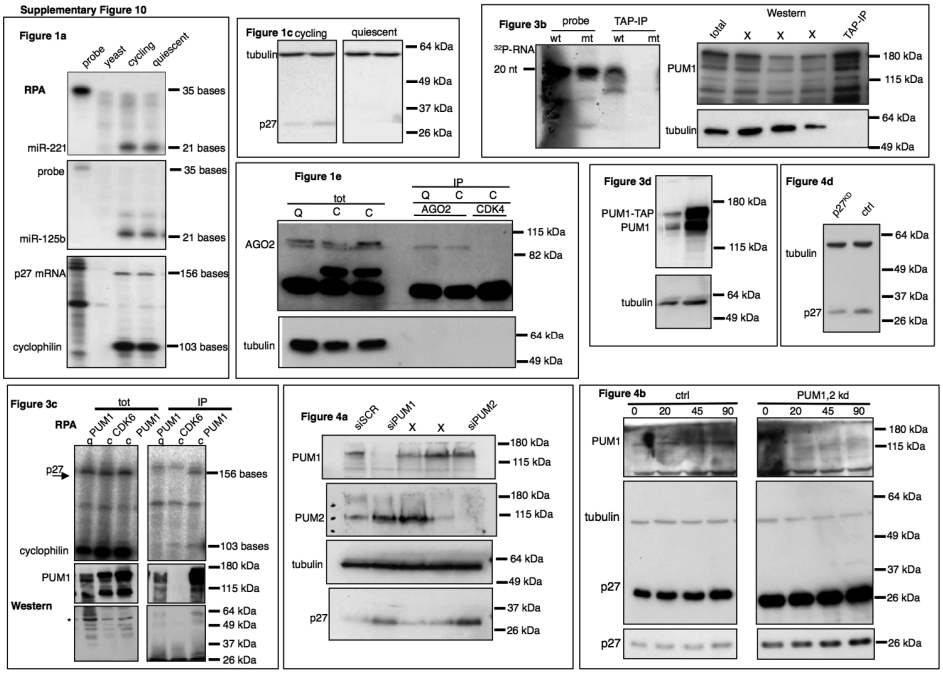


Figure S10 | Full scans.

chapter **FIVE**

INTRONIC POLYADENYLATION RESULTS IN A FUNCTIONALLY DEFICIENT DND1 PROTEIN VARIANT

Marieke van Kouwenhove[1], Joachim A.F Oude Vrielink[1], Mathias Jenal[1], Ran Elkon[1], Gijs van Haaften[1], Koos Rooijers[1], Martijn Kedde[1], Henning Urlaub[2], L. Ayse Erozenci[1], Ad J.M. Gillis[3], Leendert H.J. Looijenga[3], Reuven Agami[1]

[1]Division of Gene Regulation, The Netherlands Cancer Institute, Amsterdam, The Netherlands,
[2]Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany,
[3]Department of Pathology, Josephine Nefkens Institute, Erasmus MC, Rotterdam, The Netherlands.

ABSTRACT

The RNA-binding protein Dead end 1 (DND1) is highly expressed in germ cells and a negative modulator of microRNA (miRNA) activity. DND1 binds uridine-rich regions (URRs) in the 3′ untranslated region (3′ UTR) of germline-specific mRNAs and relieves miRNA-mediated suppression. Here we identified using global sequencing of 3′ ends of mRNAs a transcript variant of DND1 that terminates intronically (DND1^{IPA}). This DND1^{IPA} variant generates a protein that contains a full RNA recognition motif (RRM), but lacks the carboxyl terminal domain. Consequently, DND1^{IPA} binds RNA as effective as the full-length DND1 (DND1^{FL}), but is unable to relieve miRNA-mediated repression. As the lack of the DND1 carboxyl (C) terminus region results in reduced self-oligomerization capability, our results suggest that oligomerization is required for relieving miRNA-mediated repression.

In the mouse 129/SvJ strain, truncation of DND1 due to the *Ter* germline mutation increases susceptibility to spontaneous germ cell tumours. To examine a role for DND1^{IPA} in cancer, we measured the expression in germ cell tumour cell lines and clinical samples. We found that DND1^{IPA} mRNA is expressed in both in normal testis and in germ cell tumours, but the relative abundance of DND1^{IPA} is higher in tumours. Our results suggest that intronic polyadenylation of DND1 may be relevant with respect to germ cell tumorigenesis.

INTRODUCTION

Germ cell development is essential for reproduction and various aspects of RNA biology have central roles in the development of the germ cell lineage. The RNA-binding protein DND1 is important for survival and migration of primordial germ cells (PGCs) during embryonic development in vertebrates from zebrafish to mouse[1,2]. DND1 is a negative modulator of miRNA-induced silencing complex (miRISC) activity and binds URRs in 3' UTRs to support expression of specific mRNAs[3]. In zebrafish, DND1 relieves miR-430-mediated translational repression of germ cell factors nanos1, trdr7 and dazl. Furthermore, DND1 relieves the effect of members of the human miRNA family ortholog of miR-430, comprising miR-93, miR-302, miR-371-373 and miR-518-520, some of which were identified as oncogenic and associated with enhanced proliferation, migration, invasion and metastasis[4,5]. The stabilizing effect on protected targets such as LATS2 and CDKN1B/p27 is evident, but DND1 supposedly functions on a larger scale[6]. Nevertheless, it is not entirely clear which structural components of the DND1 protein fulfill its function. DND1 contains two conserved RRMs, one of which has been shown to be required for cytoplasmic localization of DND1, PGC survival, and miRNA-mediated repression[3,7]. Introduction of DND1 proteins lacking the less conserved C-terminus into zebrafish embryos in which endogenous DND1 was blocked using morpholino antisense oligonucleotides revealed a dramatic reduction in PGC survival[7]. Deletion mapping revealed an ATPase domain in the final 91 amino acids of the DND1 C-terminus in zebrafish and by rescue experiments in zebrafish embryos, DND1 ATPase activity was shown to be required for normal PGC development and survival[8].

As discovered in mouse, the Ter mutation introduces a premature stop codon in the DND1 gene causing inactivation of DND1 expression and driving PGCs to exit the germ cell lineage[9,10]. Moreover in the 129 mouse strain background, inactive DND1 results in high testicular germ cell tumour (GCT) incidence[10,11]. Most DND1-deficient germ cells die, with sterility as a consequence, and a small residue migrates to the genital ridge (future gonads)[1,2]. Germ cell carcinoma *in situ* (CIS) is initiated at this point in embryonic development by oncogenic transformation of the residual PGCs due to unknown factors[12]. Whereas the introduction of a stop codon in the DND1 gene results in loss of expression of DND1 in homozygous Ter mice, post-transcriptional silencing of DND1 expression from the wildtype allele is observed in affected testes, but not in normal testes of heterozygous Ter mice[10]. The human DND1 gene has been extensively scanned in familiar and sporadic testicular GCTs, but no common germ line mutation was found[13,14].

In addition to genetic defects, misregulation at the level of transcription or translation can also contribute to pathogenesis. Here, we describe mRNA polyadenylation variants of DND1 discovered by a 3' RNA sequencing approach. Polyadenylation is an essential part of nuclear mRNA processing and facilitates efficient mRNA export, translation, and stability[15]. About half of the mammalian genes express mRNA isoforms varying in length or sequence due to usage of distinct polyadenylation signals (PAS) which initiate mRNA cleavage and addition of a poly(A)tail by recruitment of polyadenylation machinery[16,17]. Only a few cis-acting factors or proteins regulating PAS selection have been identified thus far[18,19].

Besides more common alternative polyadenylation (APA) within mRNA 3′ UTRs, instances of regulated intronic polyadenylation (IPA) have been reported[20-23]. IPA occurs concurrently and competitively with pre-mRNA splicing; factors favoring IPA are a weak 5′ splice site signal and long intron length[24]. The usage of intronic PASs leads to shortened coding sequence, alternate 3′ UTRs and smaller protein isoforms. While expression of truncated protein due to a premature stop codon or by out-of-frame splicing is prevented by nonsense-mediated mRNA decay (NMD), mRNA variants by IPA are naturally immune to this[25]. As complete exons are excluded due to IPA, the function of truncated proteins lacking essential C-terminal domains can be dramatically different from their full-length counterparts.

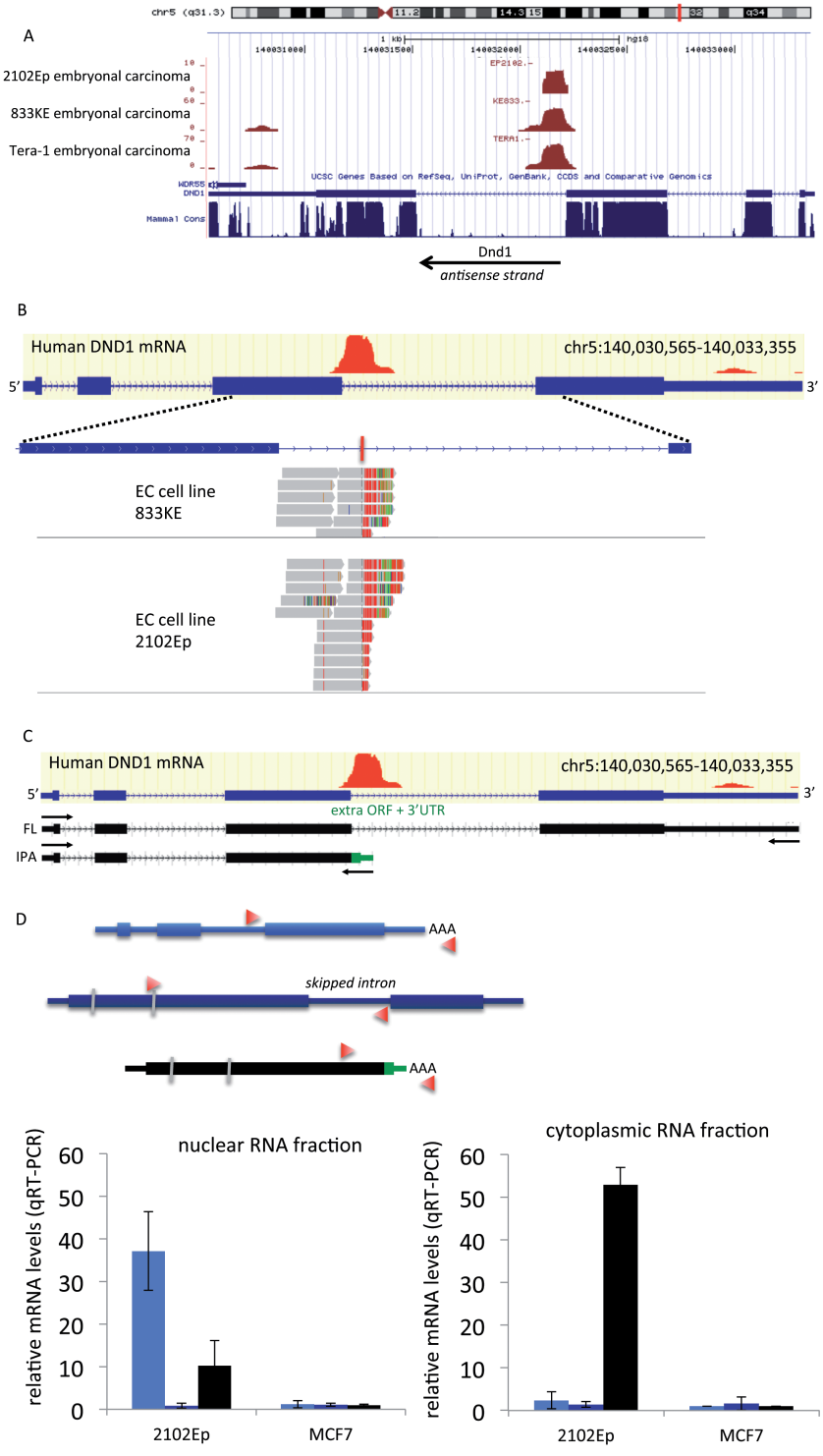
5

RESULTS

Detection of DND1 mRNA variants generated by intronic polyadenylation

A panel of human EC cell lines was subjected to a 3′ RNA sequencing approach to study variation in mRNA polyadenylation (Fig. S1a, REF. 19). The results revealed expression of a shortened mRNA variant of DND1 in the EC cell lines 2102Ep, 833KE and Tera-1 whereas this variant was not detected in other cell lines (Fig. 1a, Fig. S1b). In addition to the full-length mRNA (DND1^{FL}), a shorter intronically polyadenylated mRNA isoform (DND1^{IPA}) is expressed (Fig. 1b, fig. S1c). DND1 expression is typical for germ cells, but is also found in other tissues such as heart and brain, and in some human carcinoma cell lines. We analyzed several of these cell lines for expression of either mRNA variant and concluded that whereas DND1^{FL} is more generally expressed, DND1^{IPA} expression is restricted to the EC cell lines 2102Ep and 833KE (Fig. S1d). For Tera-1 cells, the result of RT-PCR is inconsistent with 3′ RNA sequencing, which we believe is a technical issue. The process generating intronically polyadenylated mRNA variants can be either intronic polyadenylation

Figure 1 | Expression of an intronically polyadenylated DND1 mRNA. (a) Results of 3′ RNA sequencing of polyadenylated mRNAs in human EC cell lines for region q31.3 on chromosome 5. Red-coloured sequence reads represent transcripts encoded on the reverse DNA strand, so the transcripts are cleaved left of the peak. The existence of two DND1 pseudogenes in the human genome disallowed mapping of sequence reads to the distal end of the DND1 gene at chromosome 5, because only uniquely aligned reads were taken into account[19]. Blue peaks show evolutionary conservation across mammals (17 species, UCSC genome browser hg18 assembly). **(b)** Sequence reads of 3′ RNA sequencing data that uniquely aligned to the genome were visualized in the Integrative Genomics Viewer (IGV) version 1.5. Grey colour indicates aligned nucleotides, red colour indicates untemplated (T), which is the reverse strand equivalent of untemplated poly(A). **(c)** The complete 5′-3′ full-length (DND1^{FL}) and intronically polyadenylated (DND1^{IPA}) transcripts in EC cell lines 833KE and 2102Ep were amplified by PCR with transcript-specific primers and sequenced. The sequences are schematically represented: green regions indicates the extended ORF and 3′ UTR exclusive to the IPA transcript. **(d)** QRT-PCR was performed to discriminate between DND1 mRNA processing intermediates in the nuclear and cytoplasmic RNA fractions of 2102Ep and MCF7 cells as a control. The signal detected for DND1 in MCF7 cytoplasmic fraction was set to 1. Red arrows indicate specific PCR primers for each mRNA intermediate, the second forward primer spans an intron.


A
2102Ep embryonal carcinoma
833KE embryonal carcinoma
Tera-1 embryonal carcinoma
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
Mammal Cons
Dnd1
antisense strand
B
Human DND1 mRNA
chr5:140,030,565-140,033,355
5′
3′
EC cell line
833KE
EC cell line
2102Ep
C
Human DND1 mRNA
chr5:140,030,565-140,033,355
extra ORF + 3′UTR
FL
IPA
D
AAA
skipped intron
AAA
nuclear RNA fraction
cytoplasmic RNA fraction
relative mRNA levels (qRT-PCR)
2102Ep
MCF7

followed by regular splicing, or a splicing defect with intronic polyadenylation as a consequence. By qRT-PCR with primers targeting a specific intermediate of either of these processes, we could detect abundant levels of an intronically polyadenylated, but not yet spliced, mRNA intermediate. This transcript was found in the nuclear RNA fraction, whereas it was absent in the cytoplasmic RNA fraction or in control cells (Fig. 1c). Therefore, DND1^{IPA} mRNA is likely due to active involvement of the cleavage and polyadenylation machinery. This mRNA variant contains a stretch of intronic sequence in its ORF, which would result in a DND1 protein with extra amino acids but lacking the complete fourth exon (Fig. 1d).

Truncated DND1 protein binds RNA but does not alleviate miRNA-mediated repression

To examine the protein stability of both DND1 variants, U2OS cells overexpressing either protein form were treated with translation inhibitor cycloheximide (CHX) and cell lysates were subjected to analysis by western blot (Fig. S2a). No difference was observed between levels of either DND1 protein up to six hours of CHX treatment, indicating that there is no difference in the rate of protein degradation. Furthermore, subcellular localization can be dramatically changed if the truncated protein lacks a localization signal. We therefore assessed DND1 mRNA and protein levels in nuclear and cytoplasmic RNA and protein fractions (Fig. S2b), and conclude that DND1 localization is not different between 2102Ep cells endogenously expressing DND1 and U2OS overexpressing either form. This is also in agreement with the presence of a cytoplasmic localization signal in the N-terminus of DND1 as was found in zebrafish[7].

Since both DND1 protein forms are robustly expressed, we examined their functionality using a luciferase reporter assay with the p27 3′ UTR and LATS2 3′ UTR coupled to a luciferase reporter gene. While overexpression of DND1^{FL} alleviated miRNA-mediated repression, DND1^{IPA} overexpression had no effect (Fig. 2a). However, gradually increasing amounts of co-transfected DND1^{IPA} did not compete with the function of DND1^{FL} in miRNA derepression (Fig. 2b). These findings suggest that DND^{IPA} is inactive in blocking miRNA activity and does not have a dominant negative effect on DND1^{FL}.

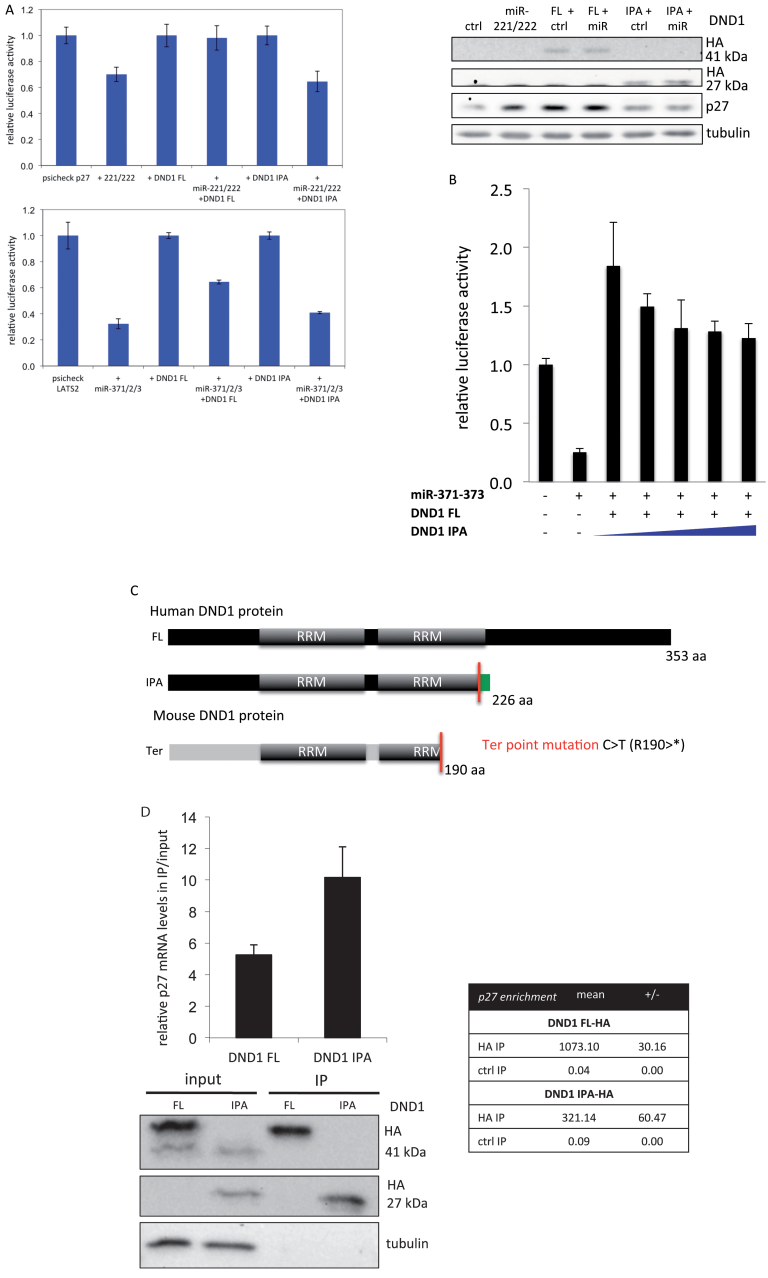
The truncated form of DND1 would still contain at least one intact RRM (Fig. 2c), therefore it is likely that its RNA-binding capacity is intact. To ascertain this, U2OS cells overexpressing either DND1 protein form were crosslinked with formaldehyde and immunoprecipitated DND1 was subjected to qRT-PCR to analyze crosslinked RNA species (Fig. S2c). In both IPs, p27 mRNA levels were enriched compared to their levels in input RNA or to IPs with control antibodies (Fig. 2d). The level of p27 mRNA seemed to be repeatedly more enriched in IPs of DND1^{IPA}, thus we conclude that the binding capacity of DND1^{IPA} to RNA is at least equal to that of DND1^{FL}. In order to examine the accessibility of the p27 3′ UTR to miRISC in the presence of either DND1 protein form, endogenous AGO2 as main miRISC component was immunoprecipitated from formaldehyde-crosslinked U2OS cells and subjected to mRNA and miRNA expression analysis. Enrichment of miR-221/222 is apparent in IPs from cells with miR-221/222 expression, implying proper loading of miRNAs into the RISC complex (Fig. 2e, upper chart). However, the enrichment of p27 in cells expressing DND1^{FL}, but not in cells expressing DND1^{IPA}, is decreased compared to

control (Fig. 2e). We plotted the ratio of p27 mRNA/miR-221 enrichment measured in the IPs (Fig. 2e, lower chart) and the changes in this ratio are consistent with the effect of DND1 in relieving miRNA-mediated repression. As a control we measured enrichment of mRNA/miRNA pairs that are expressed in U2OS cells but are not regulated by DND1, there we did not observe a difference (Fig. S2d). We conclude that DND1^{IPA} lacks a functional domain that is required for counteracting miRNA-mediated expression in addition to binding of the 3' UTR.

The truncated DND1 protein lacks oligomerization properties

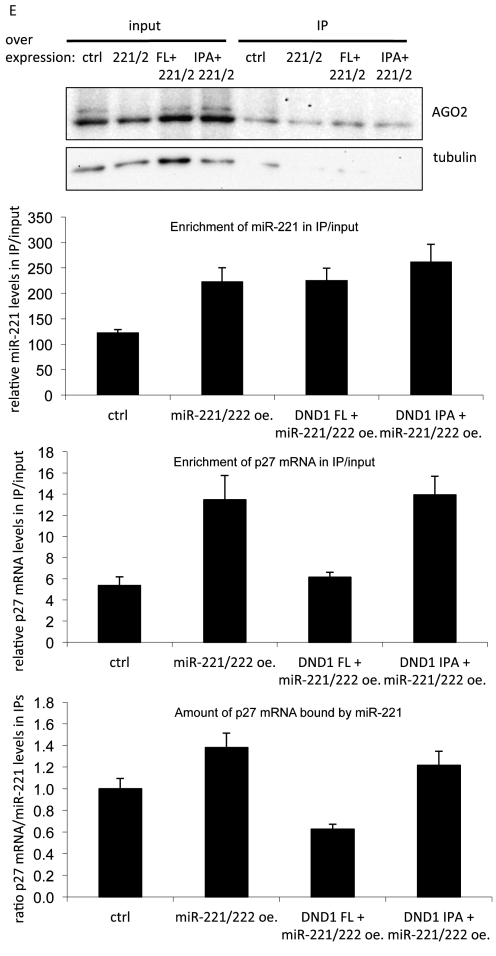
Our results thus far indicate that the DND1 protein that is processed following intronic polyadenylation lacks the C-terminus and functionality. Apart from a functional ATPase domain in zebrafish DND1, no other domains in the C-terminus have been annotated. Alignment of structural predictions for the human, mouse and zebrafish DND1 C-terminal protein structure using the HHpred module of the HH-suite software package[26], showed homology to various proteins (Fig. 3a). The homologous secondary structure that these proteins have in common is an $\alpha\beta\beta\beta\alpha$ fold, which is designated a double stranded RNA-binding domain (dsRBD) in tertiary structure (Fig. S3a, table S1). A dsRBD is known for its versatility and can function in RNA-binding, but also in DNA-binding and protein interactions. Eight of the proteins with structure homology are reported to interact with each other, but whether their dsRBD is employed for this interaction is not known (Fig. S3b). Since the RRMs have an established function in RNA-binding and the truncated protein retains its RNA-binding capacity, a possible difference in protein interaction was investigated. We compared oligomerization capacity of either protein form by visualization of DND1 monomers and larger complexes on denaturing and native protein gels, respectively. This revealed larger complexes in case of DND1^{FL} overexpression compared to DND1^{IPA} (Fig. 3b). The bands on the native gel may represent DND1 monomers, dimers and oligomers, but whether these bands include other proteins cannot be excluded. Upon co-transfection of both DND1 forms, a distinct extra band is visible, implying the formation of a DND1^{FL}:DND1^{IPA} dimer (Fig. 3b, lane 2). The possibility of interaction between these forms is confirmed by co-immunoprecipitation of both tagged DND1 proteins (Fig. S3c). Since DND1^{IPA} lacks the C-terminal, these results suggest a role for the C-terminus in protein complex formation. However, DND1^{FL} was not clearly co-immunoprecipitated with the DND1 C-terminus (DND1^{CT}; Fig. S3c). This raises the question whether complex formation via the C-terminal part of DND1 depends on interaction with RNA. When DND1^{FL} or DND^{CT} native protein lysates were treated with RNase for 30 minutes, a small decrease in larger protein complexes was observed for DND1^{FL} whereas for DND1^{CT}, most of the larger complexes had disappeared (Fig. 3c). This suggests that oligomerization via the C-teminal part of DND1 may be dependent on the presence of RNA. If both DND1^{IPA} and DND1^{CT} possess the ability to dimerize, then we hypothesize that two incompatible domains are involved in oligomerization.

Another function of the putative ~80 aa spanning dsRBD may be interaction with other proteins. We overexpressed HA-tagged human DND1 in HEK293 cells, performed IPs and subjected these to mass spectrometric analysis. This yielded a list with certain interesting interaction candidates with respect to DND1 function



p27 enrichment	mean	+/-
DND1 FL-HA		
HA IP	1073.10	30.16
ctrl IP	0.04	0.00
DND1 IPA-HA		
HA IP	321.14	60.47
ctrl IP	0.09	0.00

Figure 2 | The truncated DND1 protein lacks function in antagonizing miRNA-mediated repression. (a) U2OS cells were co-transfected with expression vectors coding for luciferase coupled to the wild-type p27-3′ UTR, wild-type LATS2-3′ UTR, miR-221/222, miR-371-373, HA-tagged DND1^{FL} or DND1^{IPA}, and hTR control. Relative luciferase activity was measured after 72 hours and is the ratio between *Renilla* luciferase and firefly luciferase as control, adjusted to 1. Error bars represent SD for triplicate experiments. Immunoblot was performed with antibodies against HA (12ca5) to check for DND1$^{FL/IPA}$ overexpression, p27 and tubulin. **(b)**



U2OS cells were co-transfected as in **a** with increasing amounts of vector coding for DND1^{IPA}. **(c)** The consequence of intronic polyadenylation for protein sequence is schematically presented; the protein lacks 150 amino acids and has 23 new amino acids. The first of two RRMs, both spanning ~80 amino acids, is intact, the second RRM lacks a small part. The Ter mutation detected in mice involves a C-to-T point mutation which changes arginine at position 190 into a stop codon, resulting in loss of 162 aa (mouse DND1 is 352 aa). RRM = RNA recognition motif. **(d)** Extract of crosslinked U2OS cells transfected with constructs coding for HA-tagged DND1^{FL} or DND1^{IPA} was used for immunoprecipitation with anti-HA (12ca5) and anti-CDK4 antibody. QRT-PCR for p27 and gapdh levels was performed to calculate enrichment of p27 mRNA in crosslinked RNA. Results are presented as ratio of p27 levels in IP/input where a value >1 indicates enrichment. Error bars represent the SEM for triplicate reactions. The enrichment of p27 as measured in HA versus control IPs is displayed in the table +/- SEM Immunoblots were performed with antibodies against HA (12ca5) and tubulin. **(e)** U2OS were transfected with constructs coding for miR-221/222, DND1^{FL}, DND1^{IPA} and hTR control. Extracts of crosslinked cells was used for immunoprecipitation with anti-AGO2 antibody. The amounts of miR-221, p27 and gapdh in the input and immunoprecipitates were measured by qRT-PCR. Results are presented as enrichment in IP/input, where a value >1 indicates enrichment. In the lower graph, p27/miR-221 ratios as calculated from the upper graphs are displayed. Error bars represent the SEM for triplicate reactions. Immunoblots were performed with antibodies against AGO2 and tubulin.

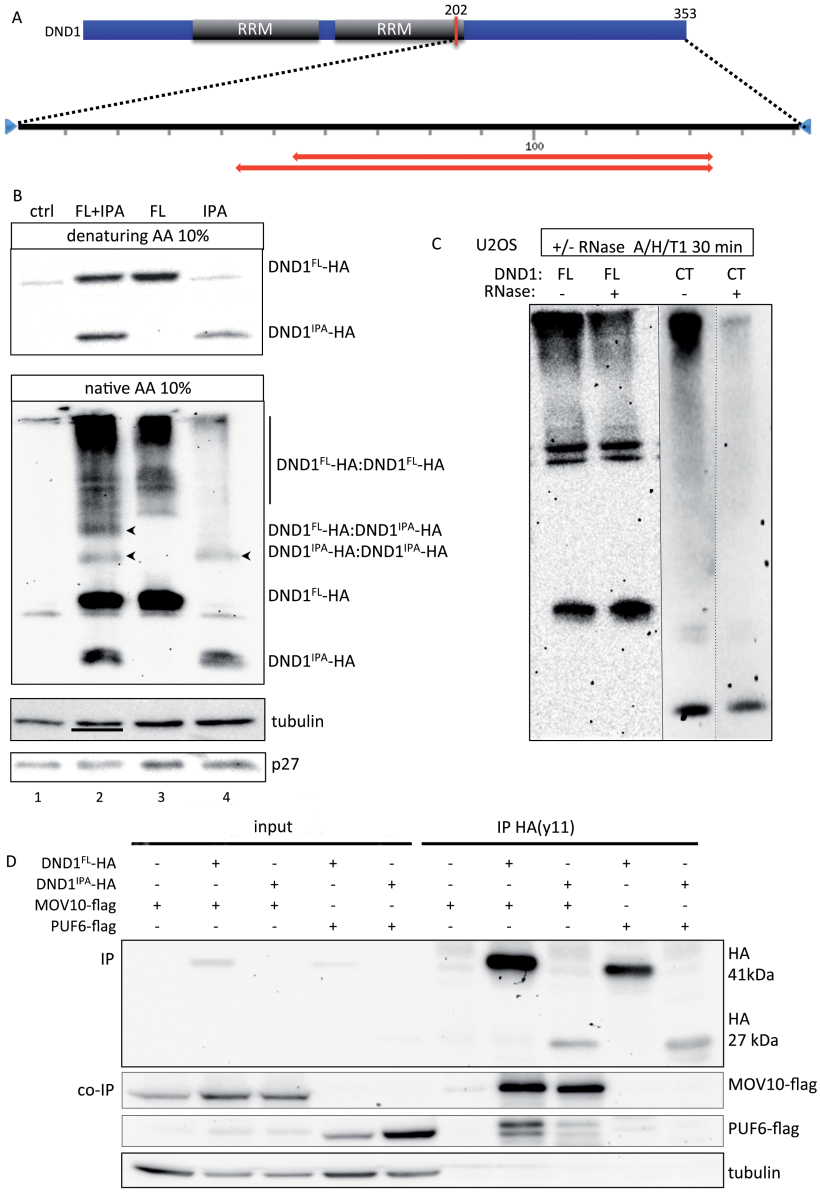


Figure 3 | The truncated DND1 protein lacks oligomerization properties. (a) Predicted structural homology between the human DND1 C terminus (amino acids 202-353) and other proteins or protein domains (in red) by HHpred algorithm. **(b)** DND1^{FL} and DND1^{IPA} were overexpressed in U2OS cells, and DND1 monomers and larger protein complexes were visualized on native and denaturing protein gels using antibodies against HA (12ca5), p27 and tubulin. **(c)** Total lysate from U2OS cells transfected with DND1^{FL} and DND1^{CT} encoding constructs was treated with RNase A/H/T1 for 30 min and analyzed by immunoblotting with antibodies against flag. **(d)** U2OS cells overexpressing flag-tagged MOV10 or Puf6 and HA-tagged DND1^{FL}, DND1^{IPA} or control vector were used for immunoprecipitation using anti-flag antibody. Immunoblot analysis was performed with antibodies against flag and HA(12ca5).

(Table S2). Of these candidates, we validated DExD-box containing RNA helicase MOV10 to interact with $DND1^{FL}$ and with $DND1^{IPA}$ in a co-immunoprecipitation assay (Fig. 3d). These results imply that the function of DND1 may involve direct interaction with MOV10, but that the compromised function of the truncated protein is not due to lack of interaction with MOV10. Together, these results suggest that the C-terminal domain of DND1 structurally resembles a dsRBD which is required for oligomerization by enabling protein:RNA or protein:protein interactions.

Expression of DND1 intronic polyadenylation variant in germ cell tumours

As $DND1^{IPA}$ is not capable of blocking miRNA accessibility to target mRNAs, and a truncated form ($DND1^{Ter}$) generates germ cell tumours in mice, we asked whether the expression of $DND1^{IPA}$ correlates with cancer development. Therefore we analyzed expression of $DND1^{IPA}$ in available GCT cell lines: 2102Ep, 833KE, Tera 1, Tera-2, NCCIT and TCam-2. Also, the total of both $DND1^{FL}$ and $DND1^{IPA}$ mRNA was estimated using primers recognizing a common 5′ mRNA region, and the levels of exclusively $DND1^{FL}$ using primers recognizing a region downstream of the IPA site. The levels of $DND1^{IPA}$ mRNA positively correlate with levels of total or full-length DND1 mRNA (R^2=0.63 and R^2=0.93, data not shown). Predictably, the level of $DND1^{IPA}$ mRNA compared to the total mRNA is higher in cell lines with generally high DND1 expression (2102Ep, 833KE, Tera-1) (Fig. 4a). In other words, in the 2102Ep and 833KE cell lines, DND1 is expressed but $DND1^{IPA}$ is also relatively high. Since expression of $DND1^{IPA}$ was evident in 833KE and 2102Ep cell lines, we examined total RNA from a panel of clinical GCTs and normal testes as well. Consistently with the findings in cell lines, the expression of DND^{IPA} mRNA correlated with total DND1 expression. In normal testes, $DND1^{FL}$ expression is higher than in seminoma or non-seminoma. However, the relative abundance of $DND1^{IPA}$ mRNA is higher in most tumour samples than in normal testes (Fig. 4b). Also, the relative abundance of $DND1^{IPA}$, but not of $DND1^{FL}$, correlates with miR-372 levels in GCTs (not shown). MiR-372 is not detected in normal testes,and anti-correlated withhigh $DND1^{FL}$ levels (data not shown). Individual GCTs can be separated from normal testes (Fig. 4c) based on total DND1 mRNA expression, but also on $DND1^{IPA}$ mRNA levels although less effectively.

DISCUSSION

We described an intronically polyadenylated DND1 transcript, which is relatively highly expressed in GCTs, and results in a truncated protein with intact RNA-binding capacity but not capable of relieving miRNA-mediated repression (Fig. 5a). Intronic polyadenylation occurs co-transcriptionally: binding of the cleavage and polyadenylation machinery causes transcription termination, followed by actual cleavage and poly-A tailing (Fig. 5b). The polyadenylation machinery competes with the spliceosome especially when the 5′ splice site is weak or when the intron is relatively long. With regard to DND1, the 5′ splice sites are identical, but the cleaved intron is relatively longer. For introns bearing a proximal PAS, binding of the

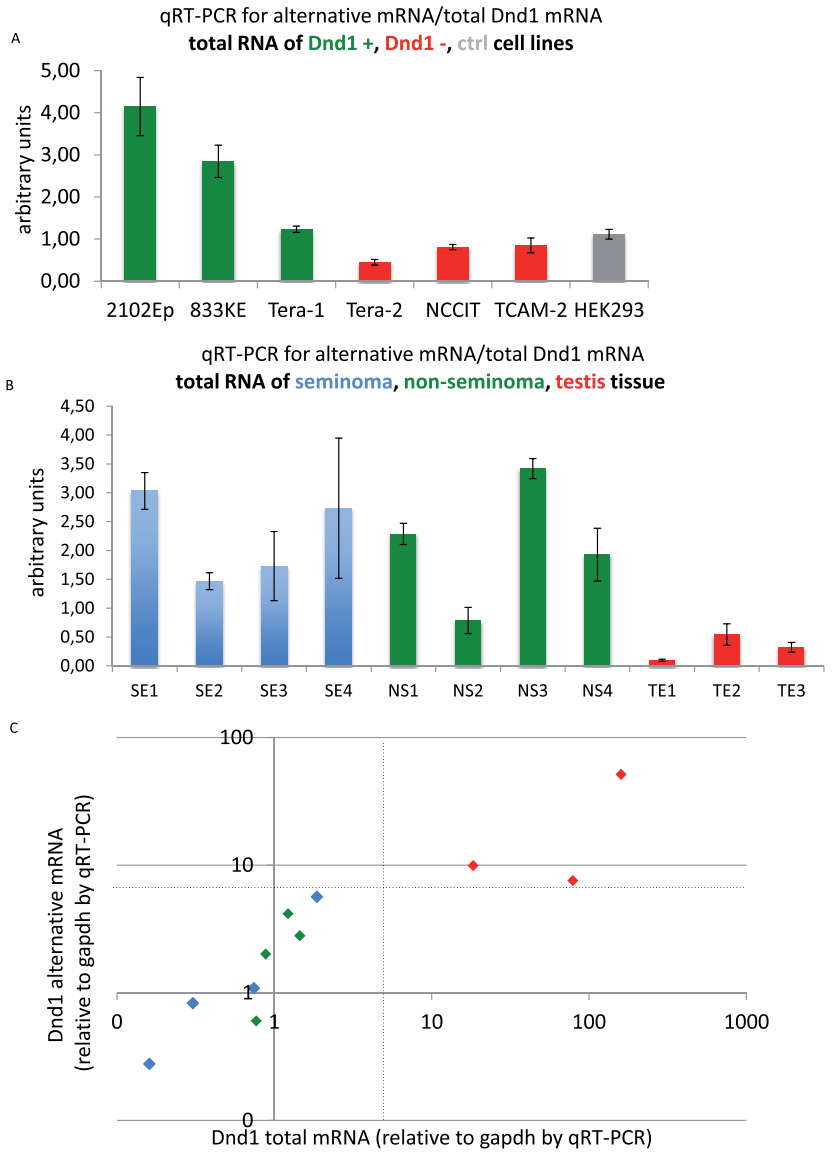


Figure 4 | Relative expression of DND1 polyadenylated mRNA in human GCTs. (a) Total RNA of GCT cell lines was used for cDNA synthesis using oligo(d)T primers specific for two variants of polyadenylated DND1 mRNA. Amounts of DND1^{IPA} mRNA and of total DND1 mRNA (by using a primer pair recognizing both forms) were calculated using qRT-PCR. The plotted value represents the amount DND1^{IPA} relative to the amount of total DND1 mRNA as detected by qRT-PCR, errors bars represent the SEM for triplicate experiments. Green, red and grey bars represent cell lines with high, moderate and absent expression of DND1, respectively. **(b)** As in **a**, using total RNA of human GCTs that was fragmented prior to analysis. Blue, green and red bars represent seminomas (SE), non-seminomas (NS) and normal testes (TE), respectively. **(c)** The qRT-PCR data obtained in **b** is presented in a XY scatterplot. On the x-axis, total DND1 mRNA measurements are displayed, on the y-axis, DND1^{IPA} mRNA measurements.

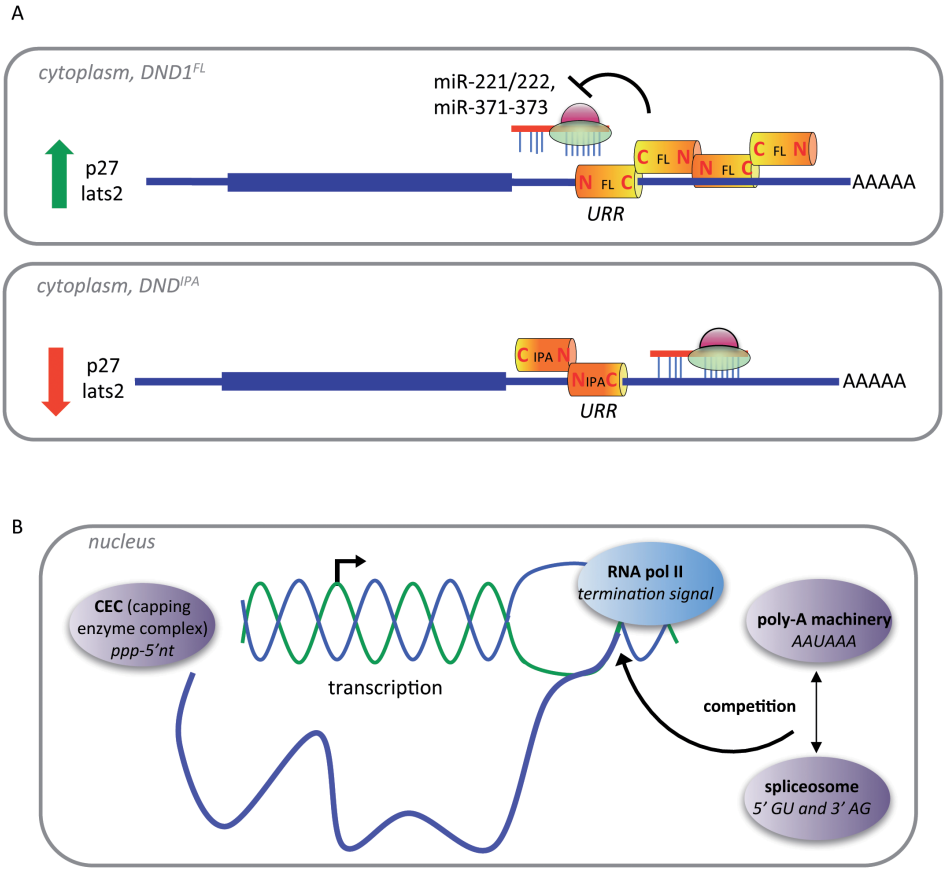


Figure 5 | A model for the effects of intronic polyadenylation of DND1. (a) Schematic representation of co-transcriptional mRNA processing and competition between the spliceosome and polyadenylation machinery. After addition of a 5′ cap to nascent transcript by the capping enzyme complex, binding of the polyadenylation machinery causes interaction of cleavage factors with RNA Pol II. This triggers transcription termination, followed by actual cleavage and poly-A tailing. The polyadenylation machinery is also physically linked with the spliceosome, which recognizes poly-A tails of transcript to-be-spliced. **(b)** Model of DND1 function and implications of DND1^{IPA} expression for post-transcriptional regulation of target genes. URR = U-rich region. Green arrows indicate upregulation, red arrows indicate downregulation. See text for further details.

polyadenylation machinery may depend on whether this PAS is protected or not. This can involve protection by RBPs or, putatively, by secondary RNA structure prohibiting access to PASs or surrounding sequences. Upstream of the intronic cleavage site in DND1, no consensus PAS can be pointed out. Although polyadenylation is nearly universal, male germ cell mRNAs have been found to exhibit increased alternative polyadenylation[27], decreased use of the canonical polyadenylation signal[28], and reduced dependence on downstream sequence elements[29]. These differences suggest a different mechanism for polyadenylation in male vertebrate germ cells, but

regulating factors have not been identified. Except for the RNA-binding subunit of cleavage stimulation factor (CstF), X-chromosomally located CstF-64, which interacts with sequences downstream the cleavage sites of many pre-mRNAs. Remarkably, during meiosis in males testis-specific τCstF-64 takes over this role and τCstF-64 deficient mice have impaired spermatogenesis and are infertile[30].

Although intronic sequence is less conserved, the extension of ORFs into introns may result in additional protein domains and alternative 3′ UTRs exclusive to the IPA form. In the case of DND1, the IPA mRNA contains an 8-mer putative target site for miR-372/373 (cggcacuu) in the extended intronic ORF. To test whether this miRNA can repress this DND1 variant, miR-372 was co-transfected with a luciferase construct containing the wildtype or mutated target site, but no effect was observed that is mediated by this site (data not shown). Also, DND1 contains URRs in the full-length mRNA 3′ UTR which are not present in the DND1^{IPA} 3′ UTR. This is particularly interesting in the light of the post-transcriptional silencing of wild type DND1 reported in heterozygous Ter mice. However, overexpression of either DND1 protein form did not affect luciferase activity as a measure for DND1^{FL} stability mediated by the DND1^{FL} 3′ UTR (data not shown). In RNA fractions associated to DND1, p27 mRNA was found more enriched in DND1^{IPA} IPs; this was reproducible but not investigated into more detail. It can be explained by a higher binding affinity of truncated DND1 towards URRs that is not conferred through the binding domain but through surrounding protein structure. Or, it could imply that by one DND1^{IPA} molecule, more p27 transcript is bound than by one DND1^{FL} molecule in a situation where the amount of p27 transcript is restrictive. Whether this involves oligomerization capacity, or stoichiometry in general, remains to be investigated.

When analyzing sequence homology between the DND1 C-terminus and other proteins, only ACF shows a considerable protein sequence identity, but when analyzing structure homology more proteins are identified. Moreover, searching for predicted structure homology gives similar results for human, mouse and zebrafish C-terminal DND1, despite the lack of sequence homology. This works so well because structures diverge much more slowly than sequences and homologous proteins may have very similar structures even when their sequences have diverged beyond recognition within or between species. Especially the RRMs of DND1 show high sequence homology to ACF, which is the essential RNA-binding component that constitutes the RNA editing enzyme complex (editosome) with APOBEC1 protein[31]. Interestingly, interaction was found between DND1 and another family member, APOBEC3, in mouse and between DND1 and APOBEC3G in human. This interaction was independent of RNA, but the function has not been elucidated[31]. The APOBEC3 family is expressed in testes and adult germ cells, but whether DND1 is involved RNA-editing is a matter of speculation.

Among the interaction candidates of DND1 as based on mass spectrometry data, the RNA helicase MOV10 was validated to interact with DND1. MOV10 interacts with either of the main RISC components and promotes miRNA-mediated translational repression. A helicase can serve as an RNA chaperone and assist in RNA duplex strand separation and/or protein displacement from RNA molecules. The presence of both an RRM or dsRBD and ATPase domain raises the possibility that DND1 itself

might have RNA helicase activity as well, supported by the predicted structural homology between DND1 and a number of RNA helicases.

Our current hypothesis is that DND1 binds URRs in the 3' UTR via one or both RRMs, and then physically blocks miRISC access to a miRNA target site by self-oligomerization, with or without interaction with MOV10 as a miRISC-associated factor (Fig. 5a). DND1-MOV10 interaction could either function as a signal; once docked to an URR the presence of MOV10 ignites DND1 oligomerization, or as decoy; fixation of miRISC to the 3' UTR prevents it from binding elsewhere. Unfortunately, experiments to analyze the role of MOV10 are limited, since manipulation of MOV10 compromises the effect of DND1 on miRNA-mediated repression.

The supposed DND1 oligomers deserve follow-up to analyze 3D and 4D conformation, preferably by using synthetic peptides representing either DND1 form. Based on the current data, we postulate that DND1 form oligomers via two incompatible domains, one common to both protein forms and the other one exclusively present in DND1^{FL}. In that case, DND1^{IPA} is restricted to dimerization, whereas DND1^{FL} can oligomerize to larger complexes, but probably dependent on the presence of RNA. The dimensional structure is unknown; DND1 dimerization could also interconnect several DND1-bound transcripts into dense complexes.

MATERIAL AND METHODS

Constructs and antibodies

The p27 and LATS2 3' UTR luciferase reporter constructs were made by subcloning the 3' UTRs from pGL3[3] into the psiCHECK *Renilla* luciferase reporter with firefly luciferase as internal control (Promega). The pCS2+-DND1^{FL} construct containing a double carboxy-terminal HA tag, the variant of this vector containing one mutated RRM, the miRvec construct and the flag-tagged MOV10 are described elsewhere[3]. The pCS2+-DND1^{IPA} construct was made by cloning a duplex oligo containing 120 intronic nucleotides, replacing the C-terminal part of the protein, into the DND1^{FL}. The DND1^{CT} construct was made by cloning the last exon (152 aa) of DND1 and a double C-terminal flag tag into the peGFP-C1 vector (Clontech), also DND1^{FL} and DND1^{IPA} with double C-terminal flag tags were subcloned into peGFP. The Antibodies used were against AGO2 (Transduction Labs and Abcam), actin (Abcam), p27 (Transduction Labs), cdc6 (Santa Cruz, sc-9964), HA Y-11 (Santa Cruz, sc-805), HA 12ca5 (Santa Cruz, sc-57592), flag M2 (Sigma, F3165), and tubulin (YL1/2 ECACC).

Cell culture, transfections, and dual luciferase activity analysis

The 833KE, 2102Ep, NCCIT, Tera-1, Tera-2 cell lines are derived from nonseminomatous tumours. HEK293 embryonic kidney, MCF7 breast adenocarcinoma, U2OS osteosarcoma, HCT116 colon carcinoma, BJ primary fibroblasts and 2102EP cell lines were cultured in DMEM, Tera-1 cell line was cultured in McCoy's and 2102Ep, NCCIT, Tera-2, and RPE1 retinal pigment epithelial cells were cultured in RPMI-1640. All media were supplemented with 10% heat-inactivated fetal calf serum (FCS, Gibco) and antibiotics, all cells were cultured in 5% CO_2 at 37 °C. Non-transformed mammary epithelial MCF10A cells were cultured in DMEM:F12

Ham's medium (Sigma) 1:1 supplemented with 10% FCS, insulin (10µg/ml, Sigma), hydrocortisone (0.5 µg/ml) and epidermal growth factor (20 ng/ml, Peprotech). U2OS and MCF-7 cells were transiently transfected using Fugene6 (Roche). HEK293 cells were transiently transfected using calcium phosphate precipitation, siRNAs were transfected in a final concentration of 50 nM. For luciferase analysis, cells were transfected with 5 ng of psiCHECK construct, 250 ng of either expression construct or control construct up to 500 ng per reaction using Fugene6 (Roche). Dual luciferase activity assays were performed 72 h after transfection in accordance with the manufacturer's instructions (Promega). The translation inhibitor cycloheximide (Sigma) was used at a final concentration of 25 µg/ml.

3′ RNA sequencing and quantitative RT-PCR analysis

Schematic representation of the 3′ RNA sequencing method is provided in Fig. S1A and is described elsewhere[19]. In summary, total RNA was extracted, enriched for polyadenylated RNA, heat-fragmented and reverse transcribed using the P7-t25-vn oligo-dT primer (5′-CAAGCAGAAGACGGCATACGAGATTTTTTTTTTTTTTTT TTTTTTTTTTVN-3′), and eventually subjected to RNA deep sequencing. For mRNA qRT–PCR, cDNA (from 2 µg RNA) was synthesized with SuperScript III and primed with oligo(dT) or P7-t25-vn in accordance with the manufacturer's instructions (Invitrogen). For qPCR of DND1 in cell lines, primers for FL (5′-CACATGGCATTCTGGTTTTG-3′ (forward) and 5′-TTTTTTTTTTTTTTTTTTTTTTTTTTTTATATACTAACCAGGGG-3′ (reverse) and IPA (5′-AGGCCCTGGTGGAAGGTAG-3′ (forward) and 5′-TTTTTTTTTTTTTTTTTTTTT TTTTAAAAGGTTGCAA-3′ (reverse)) were used. For qPCR of mRNA intermediates, forward primers (5′-TGCATACCCTCTTCCCACAG-3′ and 5′-GCCACCCCCAGGCT GGGT-3′) and reverse primers (5′-GCTCGGGAAGTGTCACAAAT-3′ and 5′-TTTTTTTT TTTTTTTTTTTTTTTTTAAAAGGTTGCAA-3′) were used. Q-PCR primers for miR-221, miR-29a, GAPDH and 18S were from Applied Biosystems, primers for GAPDH (5′-TGCACCACCAACTGCTTAGC-3′ (forward) and (5′-GGCATGGACTGTGGTCATG AG-3′ (reverse)), β-actin (5′-CCTGGCACCCAGCACAAT-3′ (forward) and 5′ GGGCC GGACTCGTCATACT-3′ (reverse)) and COL4A2 (5′-CCGGGCTAAAAGGAGATAG-3′ (forward) and 5′-GACCGGGTATTCCGAAAAAT-3′ (reverse)) were ordered at Invitrogen. Analysis was performed with SYBR Green PCR master mix or TaqMan UNG master mix (Applied Biosystems) and Chromo 4 system (Bio-Rad Laboratories). For sequencing of PCR products, the following primers (5′-GGAGGAAGTTATAAAGAGGGTACGA-3′ (forward) and 5′-AAAAGGTTGCAAATTCTGGC-3′ (reverse)) were used.

Immunoblotting, immunoprecipitation, crosslinked protein:RNA immunoprecipitation

HA-tagged DND1 was precipitated from transiently transfected U2OS cells with antibodies against HA in ELB lysis buffer (125 mM NaCl, 0.1% Nonidet P-40, 50 mM Hepes pH 7.3) containing 1 mM NaF, 1 mM NaOV, 0.5 mM EDTA, 4 mM $MgCl_2$, 2 mM β-glycerophophate, protease inhibitor mixture (Roche Applied Science)). We used GammaBind G Sepharose (GE Healthcare) and protein A-Sepharose (Amersham) beads, these were washed and incubated for 2 hours at 4 °C, and bound proteins were revealed on gel.

For protein: RNA immunoprecipitation, U2OS and HEK293 cells were transfected and after 72 hours cross-linked for 10 min with 1% formaldehyde, neutralized with 125 mM glycine and washed in PBS. Cell lyates were sonicated and cleared in lysis buffer (140 mM NaCl, 50 mM HEPES pH 7.5, 1% Triton X-100, 0.1% sodiumdeoxycholate, 1mM dithiothreitol, protease inhibitor mixture (Roche Applied Science) and RNase-OUT (Invitrogen)). Extracts were incubated overnight with antibodies against AGO2 (1 μg per immunoprecipitation) in a tumbler placed at 4 °C. AGO2 was immunoprecipitated with GammaBind G Sepharose (GE Healthcare) and protein A-Sepharose (Amersham). Beads were preblocked with yeast tRNA (Invitrogen) and RNase-free BSA (Ambion) and then washed. Thereafter, beads were washed in lysis buffer and in wash buffer (10 mM Tris pH 8.0, 250 mM LiCl, 0.5% Nonidet P40, 0.5% sodium deoxycholate, 1 mM EDTA) and a 10% aliquot was used for immunoblot analysis. The remainder was crosslink-reversed for 1 hour at 70 °C and RNA was extracted (Trizol, Invitrogen) in the presence of glycogen. Real-time PCR was performed on a StepOne Plus cycler (Applied Biosystems), using SYBR Green Universal PCR Master Mix (Applied Biosystems). RNA enrichment was calculated as a ratio of IP/input RNA.

For immunoblot analysis, extracts were separated on 10% SDS–PAGE gels, and transferred to Immobilon-P membranes (Millipore). Western blots were developed using enhanced chemiluminescence (ECL; Amersham Biosciences) and exposed to film (Kodak). Cellular fractionation was performed on U2OS cells with NE-PER kits from Pierce, according to manufacturer's instructions.

Bioinformatics

We assessed the predicted structural homology between the human DND1 and other proteins or protein domains by HHpred module[26] as part of the HH-suite package. For creating interaction networks, we searched the protein interaction database BioGRID version 3.1 (http://thebiogrid.org/).

ACKNOWLEDGEMENTS

We thank all members of the Agami laboratory for technical help and discussions. We also thank Danny Sahtoe for assistance with structural homology prediction. Further we are grateful to the NKI Central genomics facility for deep sequencing our samples. This work was supported by the EURYI (European research young investigator award), ERC (European Research Council), KWF (Koningin Wilhelmina Fonds; Dutch cancer foundation) and Horizon-NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; R.A.) and an EMBO long-term fellowship (R.E.).

AUTHOR CONTRIBUTIONS

M.v.K. performed most of the experimental work. R.A. supervised the project. J.O.V. generated most constructs. M.J., R.E., G.v.H. and K.R. generated the 3' RNA sequencing data set, G.v.H. provided technical advice. M.K. and H.E. set up and performed the mass spec analysis. A.E. helped with luciferase assays and immunoblots. A.G. and L. L. selected and provided clinical GCT samples. M.v.K. and R.A. wrote the manuscript.

REFERENCES

1. Sakurai, T., Katoh, H., Moriwaki, K., Noguchi, T. & Noguchi, M. The ter primordial germ cell deficiency mutation maps near Grl-1 on mouse chromosome 18. *Mammalian genome : official journal of the International Mammalian Genome Society* **5**, 333-6 (1994).
2. Weidinger, G. et al. dead end, a novel vertebrate germ plasm component, is required for zebrafish primordial germ cell migration and survival. *Current biology : CB* **13**, 1429-34 (2003).
3. Kedde, M. et al. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**, 1273-86 (2007).
4. Voorhoeve, P.M. & Agami, R. The tumor-suppressive functions of the human INK4A locus. *Cancer cell* **4**, 311-9 (2003).
5. Huang, Q. et al. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat Cell Biol* **10**, 202-10 (2008).
6. Zhu, R., Iacovino, M., Mahen, E., Kyba, M. & Matin, A. Transcripts that associate with the RNA binding protein, DEAD-END (DND1), in embryonic stem (ES) cells. *BMC molecular biology* **12**, 37 (2011).
7. Slanchev, K. et al. Control of Dead end localization and activity - Implications for the function of the protein in antagonizing miRNA function. *Mechanisms of development* (2008).
8. Liu, W. & Collodi, P. Zebrafish dead end possesses ATPase activity that is required for primordial germ cell development. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **24**, 2641-50 (2010).
9. Asada, Y., Varnum, D.S., Frankel, W.N. & Nadeau, J.H. A mutation in the Ter gene causing increased susceptibility to testicular teratomas maps to mouse chromosome 18. *Nature genetics***6**, 363-8 (1994).
10. Youngren, K.K. et al. The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. *Nature* **435**, 360-4 (2005).
11. Bhattacharya, C. et al. The mouse dead-end gene isoform alpha is necessary for germ cell and embryonic viability. *Biochemical and biophysical research communications* **355**, 194-9 (2007).
12. Looijenga, L.H., Gillis, A.J., Stoop, H., Biermann, K. & Oosterhuis, J.W. Dissecting the molecular pathways of (testicular) germ cell tumour pathogenesis; from initiation to treatment-resistance. *International journal of andrology* (2011).
13. Linger, R. et al. Analysis of the DND1 gene in men with sporadic and familial testicular germ cell tumors. *Genes, chromosomes & cancer* **47**, 247-52 (2008).
14. Sijmons, R.H. et al. Screening for germline DND1 mutations in testicular cancer patients. *Familial cancer* **9**, 439-42 (2010).
15. Licatalosi, D.D. & Darnell, R.B. RNA processing and its regulation: global insights into biological networks. *Nature reviews Genetics* **11**, 75-87 (2010).
16. Tian, B., Hu, J., Zhang, H. & Lutz, C.S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic acids research* **33**, 201-12 (2005).
17. Yan, J. & Marr, T.G. Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome research* **15**, 369-75 (2005).
18. Chien, C.H. et al. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic acids research* (2011).

19. Jenal, M. et al. The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. *Cell*, 1-16 (2012).
20. Lou, H., Neugebauer, K.M., Gagel, R.F. & Berget, S.M. Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Molecular and cellular biology* **18**, 4977-85 (1998).
21. Takagaki, Y., Seipelt, R.L., Peterson, M.L. & Manley, J.L. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**, 941-52 (1996).
22. Vorlová, S. et al. Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Molecular Cell* **43**, 927-39 (2011).
23. Godmann, M., Kromberg, I., Mayer, J. & Behr, R. The mouse Krüppel-like Factor 4 (Klf4) gene: four functional polyadenylation sites which are used in a cell-specific manner as revealed by testicular transcript analysis and multiple processed pseudogenes. *Gene* **361**, 149-56 (2005).
24. Tian, B., Pan, Z. & Lee, J.Y. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome research* **17**, 156-65 (2007).
25. Lejeune, F. & Maquat, L.E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Current opinion in cell biology* **17**, 309-15 (2005).
26. Söding, J., Biegert, A. & Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* **33**, W244-8 (2005).
27. Zhang, H., Lee, J.Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome biology* **6**, R100 (2005).
28. Wallace, A.M. et al. Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6763-8 (1999).
29. Liu, D. et al. Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis. *Nucleic acids research* **35**, 234-46 (2007).
30. Dass, B. et al. Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20374-9 (2007).
31. Bhattacharya, C., Aggarwal, S., Kumar, M., Ali, A. & Matin, A. Mouse apolipoprotein B editing complex 3 (APOBEC3) is expressed in germ cells and interacts with dead-end (DND1). *PLoS ONE* **3**, e2315 (2008).

SUPPLEMENTARY TABLES AND FIGURES

Table 1 | Prediction of DND1 C-terminus structural homologs.

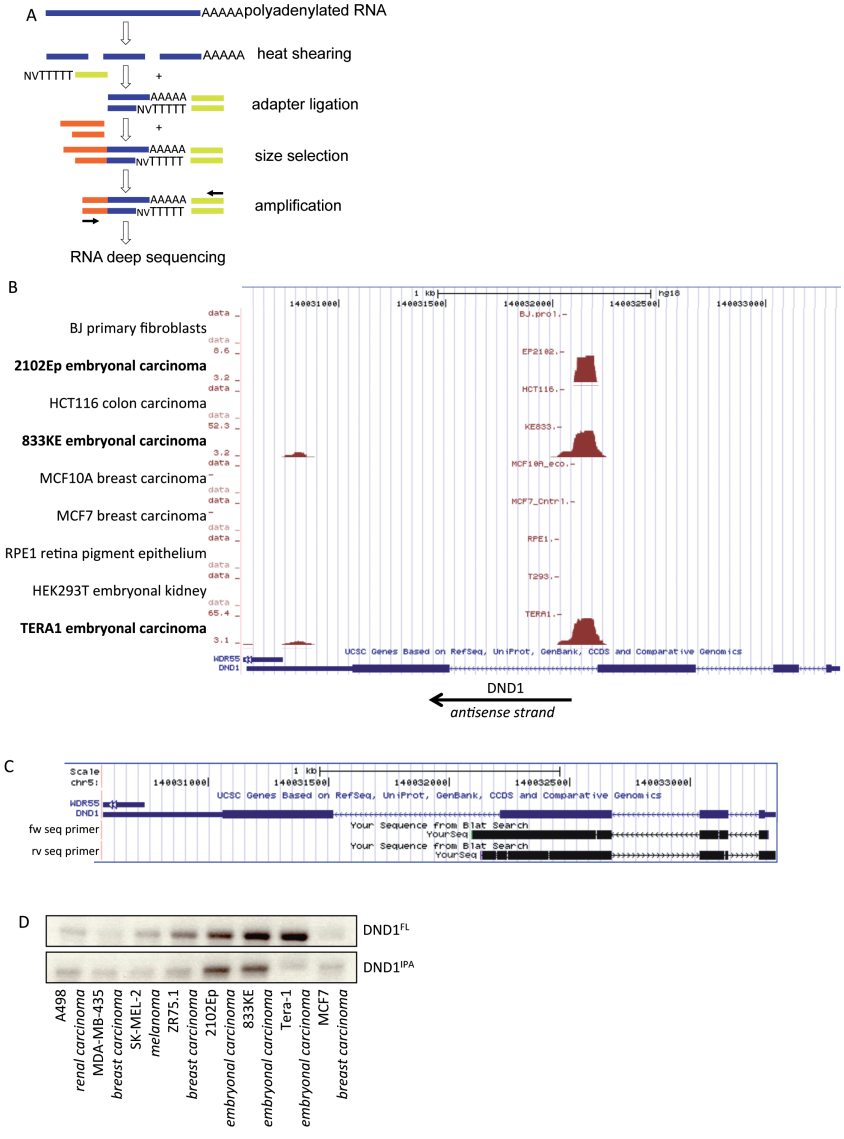
Gene symbol	Gene name
ACF	Apobec-1 complementation factor
RBM46	RNA-binding protein 46
Staufen	Staufen (RNA-binding protein)
PACT (PRKRA)	interferon-inducible dsRNA-dependent protein kinase activator A
TRBP2	HIV transactivation responsive RNA-binding protein 2
DGCR8	DiGeorge syndrome critical region gene 8
HYL1	dsRNA-binding hyponastic leave 1 protein
SON	DNA-binding protein isoform F
DDX9 (RHA)	DEAH (Asp-Glu-Ala-His) box polypeptide 9 (RNA helicase A)
ADAR1	dsRNA-specific editase 1
ADAR2	dsRNA-specific editase 2
ILF3	interleukin enhancer binding factor 3
STRBP	spermatid perinuclear RNA-binding protein
EIF2AK2 (PRKR)	eukaryotic translation initiation factor 2-alpha kinase 2
DUS2L	dihydrouridine synthase 2-like
ADAD1	testis nuclear RNA-binding protein
Dicer	endoribonuclease
DDX29	DEAH (Asp-Glu-Ala-His) box polypeptide 29

List of closest homologs by comparing predicted secondary structure homology between the human DND1 C-terminus (amino acids 202-353) and other proteins using HHpred algorithm. Comparison of mouse and zebrafish DND1 C-termini give similar results (data not shown).

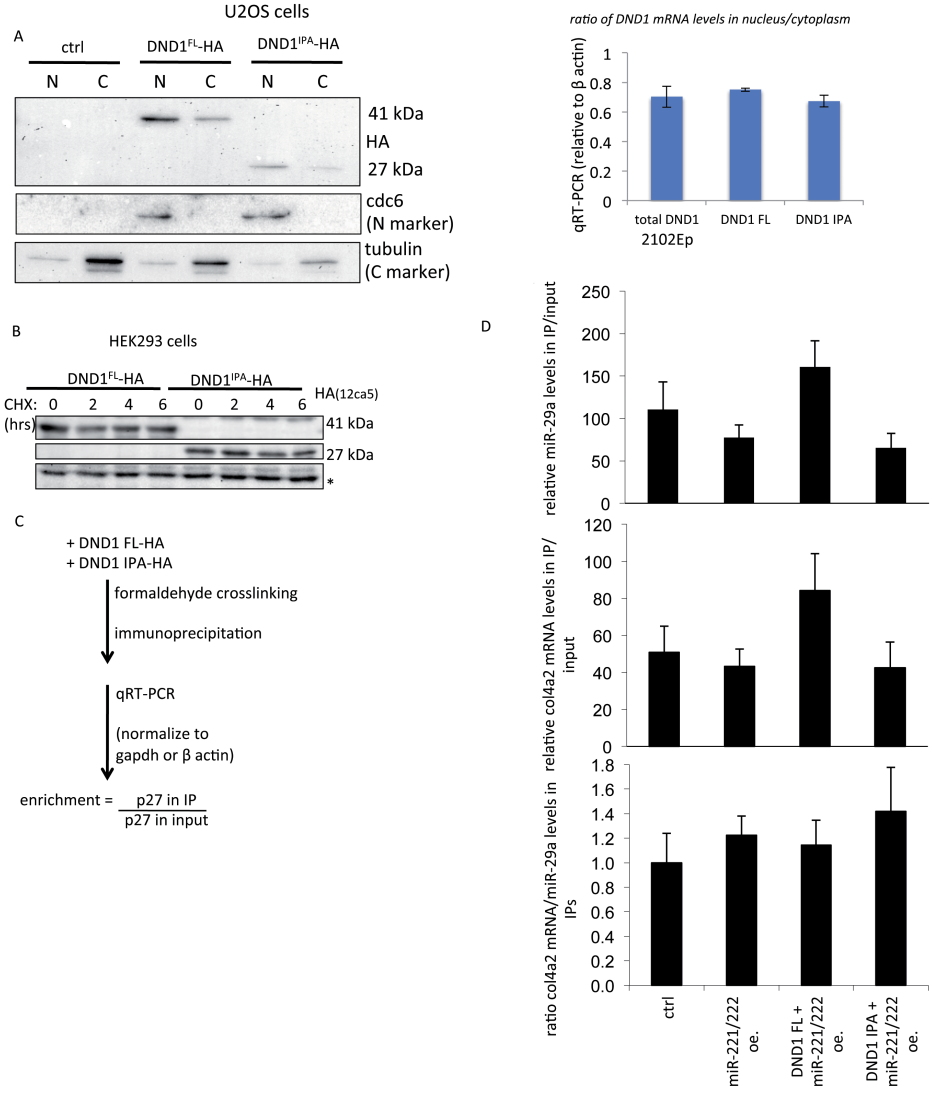
Table 2 | Mass spectrometry analysis of DND1-associated proteins.

Gene symbol	Cellular localization
DDX5	nucleus
PABP1	cytoplasm
EIF2C2	nucleus, cytoplasm
DHX36	cytoplasm
FUS	nucleus
DDX17	nucleus
DHX15	nucleus
XRN2	cytoplasm
DHX30	cytoplasm
MOV10	cytoplasm (SG)
HNRNPA1	nucleus
TIA1	cytoplasm (SG)
EWS	nucleus, cytoplasm

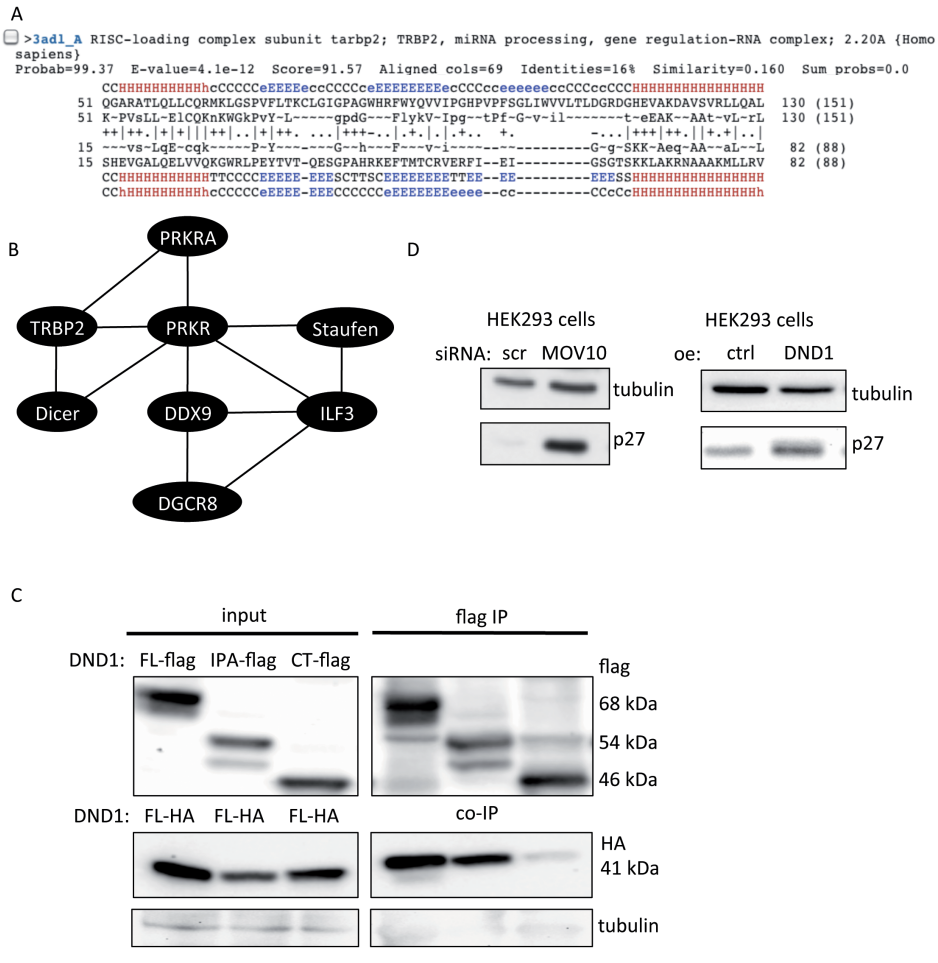
Selection of potential relevant proteins found associated to HA-tagged $DND1^{FL}$ by mass spectrometry analysis. SG = stress granules



Supplementary figure 1 | Expression of an intronically polyadenylated DND1 mRNA. **(a)** 3′ RNA sequencing approach to detect polyadenylated mRNA transcripts. Sequencing from the 3′ end allows precise mapping of poly(A) tails by alignment of sequenced reads that represent both 3′ ends of transcripts and the beginning of untemplated poly(A) tails. **(b)** Results of 3′ RNA sequencing of polyadenylated mRNAs in a panel of human primary or cancer cell lines for region q31.3 on chromosome 5. Red-colored reads represent transcripts encoded on the reverse DNA strand, so the transcripts are cleaved left of the peak. **(c)** The complete 5′-3′ full-length ($DND1^{FL}$) and intronically polyadenylated ($DND1^{IPA}$) transcripts in EC cell lines 833KE and 2102Ep were amplified by PCR with transcript-specific primers and sequenced. **(d)** GelRed-stained gels of PCRs performed on cDNA generated from total RNA of cell lines with primer pairs targeting either of the two differently polyadenylated DND1 transcripts. The amplicons were sequence-verified.



Supplementary figure 2 | The truncated DND1 protein lacks function in antagonizing miRNA-mediated repression. (a) U2OS cells were transfected with constructs coding for HA-tagged DND1$^{FL/IPA}$ and harvested after 72 hrs. Immunoblot was performed to confirm separation of the nuclear (N) and cytoplasmic (C) RNA fraction with antibodies against cdc6 and tubulin, respectively. U2OS cells were transfected as indicated, and also from 2102Ep cells, nuclear and cytoplasmic fractions were separated and the levels of specific DND1 transcripts were measured by qRT-PCR. The amounts are displayed as a ratio in nuclear/cytoplasmic extract, error bars represent the SEM for triplicate reactions. **(b)** U2OS cells overexpressing HA-tagged DND1^{FL} or DND1^{IPA} were treated with translation inhibitor cycloheximide for the time indicated. Immunoblot analysis was performed with antibodies against HA (12ca5). **(c)** Schematic representation of the protein:RNA crosslink and IP procedure. **(d)** Similar as performed in **fig. 2c** instead of p27 and miR-221, col4a2 and miR-29a levels and ratio was measured.



Supplementary figure 3 | Truncated DND1 protein lacks oligomerization properties. **(a)** Example of predicted secondary structure homology between human DND1 C terminus and the dsRBD of TARBP2 over a region of ~80 amino acids. Red H/h indicates a predicted α-helix structure and blue E/e indicates a predicted β-sheet structure. A dsRBD is annotated as an αββββα fold. **(b)** Of the closest structural homologs of DND1 C terminus, eight proteins have been reported to interact which each other. **(c)** U2OS cells co-transfected with constructs encoding flag-tagged or HA-tagged DND1 were lysed and used for immunoprecipitation using anti-flag(M2) antibody. Immunoblot analysis was performed using antibodies against flag(M2), HA(12ca5) and tubulin. **(d)** HEK293T cells were transfected with constructs encoding HA-tagged DND1^{FL} or siRNAs against MOV10. Immunoblots were performed using antibodies against MOV10, HA (12ca5), p27 and tubulin.

chapter SIX

GENERAL DISCUSSION

In the previous chapters I presented three different studies that give new insights in mechanisms of gene expression regulation. While these studies were performed, many discoveries in the field of transcription, post-transcriptional regulation and translation were made. In this chapter, I will put the new findings in the perspective of what is known at present and point out a direction for future research in the particular fields. Furthermore, I present a model merging specific parts of the studies.

TRANSCRIPTION ACTIVATION OF MIRNA GENES

Post-transcriptional regulation of gene expression by miRNAs can affect numerous target genes at the same time. Therefore, we aim to understand how the expression of miRNA genes is regulated, which may also reveal connections between different biological systems. Transcriptional regulation involves combinations of transcription factors (TFs) and their interactions. *In vitro* transcriptional activation of miR-371-373 can be accomplished by a combination of the TFs Sox2, Klf4 and Oct3/4. TFs and miRNAs both regulate many genes, and in some instances they reciprocally regulate one another[1]. In summary, the combination of transcriptional and post-transcriptional processes control many downstream targets simultaneously.

The miRNA clusters 371-373, 302-367, 520 and 93 are special in that they are expressed in ESCs and downregulated during embryonic development. However, transcription of miR-371-373 is not necessarily promoted under the same conditions as described for miR-302-367[2,3]. All the family members exhibit a similar expression pattern when comparing GCT types, with overall higher levels in EC components of GCTs than in more differentiated parts[4,5]. Especially the family members sharing the same seed sequence are similar when comparing cell type specificity, but not when comparing p53 status. MiR-371-373 is predominantly expressed in GCTs and cell lines having wild-type p53, whereas miR-302 expression is not linked to the p53 status[6]. This implies that genetic factors can drive differences in miRNA expression. Furthermore, the miR-302-367 and miR-371-373 clusters were reported to have a similar beneficial effect on reprogramming via similar mRNA targets[7]. Yet, the effects on either reprogramming or proliferation may be mediated via different target subsets of these miRNAs, of which some members have been validated[3,6,7]. Also, preferences in target binding may also be specified by regions flanking the miRNA seed and mRNA recognition site, which are usually not taken into account in prediction algorithms.

Co-introduction of c-Myc increases reprogramming efficiency, however, also increases teratocarcinoma incidence in iPS mice[8-10]. Although c-Myc is a prominent candidate for regulation of miR-371-373 expression, we found no effect in our luciferase reporter assays. C-Myc has a proposed function in chromatin accessibility and may allow Oct3/4 and Sox2 to bind to their specific target loci *in vivo*. Moreover, c-Myc overexpression stabilizes p53, which stimulates apoptosis, but also represses p21 thereby preventing a G1 cell cycle arrest. Klf4 is actively involved in cell cycle regulation and differentiation. Dependent on the cellular environment[11], Klf4 acts either as a proliferation-inhibiting tumour suppressor by stabilizing p21 levels or as a proliferation-promoting oncogene by repressing the tumour suppressor p53 directly via the p53 gene promoter in a p21-deficient context[12]. Upon p21 repression by c-Myc, the dual functionality of Klf4 might be

shifted to the pro-proliferation modus whereby p53 is repressed[13]. Consistent with this hypothesis, instead of introduction of c-Myc, suppression of p53 significantly enhances iPS cell generation[14]. Yet, whether the effect of c-Myc on reprogramming is mediated by Klf4-promoted proliferation or by other effects of c-Myc is unclear. In conclusion, the balance between c-Myc and Klf4 function is important for reprogramming, and putatively for teratocarcinoma formation.

In eukaryotes, multiple variables affect the rate of transcription of genes, and also transcription of miRNA genes. Clearly, miRNA synthesis, as orchestrated by transcription factors and RNA-binding proteins, is important for levels of mature miRNAs. However, the extent of miRNA-mediated repression is dependent on presence and accessibility of the target sites in the 3′ UTR and the presence and activity of RBPs that either stimulate or inhibit miRNA-mediated translational inhibition.

6

REGULATION OF RNA STRUCTURE AND MIRNA-MEDIATED REPRESSION BY RNA-BINDING PROTEINS

Over the past decade, we gained a much greater insight in post-transcriptional gene regulation by miRNAs and RBPs. It has become increasingly clear that interplay between these factors fine-tunes gene expression by modest, though effective adjustments in the levels of gene expression of a target (subset). We discovered a mechanism by which RBP Pumilio regulates miRNA-mediated repression of targets comprising a miRNA site and a reverse complementary Pumilio recognition element (PRE). Due to the complexity of RNA conformation in the context of a complete 3′ UTR, it remains a challenge to establish the contribution of the upstream PRE and downstream miR-221/222 site. In the p27 3′ UTR, the stem-loop sequence bound by Pumilio and miR-221/222, as well as the surrounding sequence, are fully conserved among vertebrates. For example, the U-rich region in the loop in between the PRE and miR-221/222 site can be bound by DND1 in cells that express this RBP. Also, PREs in other Pumilio targets are characterized by high conservation of surrounding sequences[15], suggesting co-regulation by other factors or by sequence elements that form mRNA secondary structures.

More generally, additional modes of structural regulation by RBPs on binding of other RBPs or miRNAs are likely to exist. Direct complementarity between the two binding sites may not be a requisite, but surrounding sequence elements may influence binding of miRNAs or RBPs by changing the local structure. For example, hnRNPL binds the VEGFA 3′ UTR in competition with certain miRNAs thereby disrupting the adjacent GAIT element, which makes binding of hnRNPL and repressive GAIT complex mutually exclusive (see chapter 2, Fig. 3b). Also, RBP binding can have opposing effects on miRNA-mediated repression (e.g. HuR can allow or block miRNA-mediated regulation towards distinct targets, chapter 2, Fig. 3a), which could be due to different structural changes upon binding. By such mechanisms, RBPs can act as sensors of cellular homeostasis, whilst miRNAs provide the regulatory factors that are directed towards or blocked from specific target subsets. In conditions where mature miRNA levels are not changed, structural switches could provide an explanation for a change in miRNA-mediated effects, or for the exclusive repression of only a

subset of targets. Both 5′ and 3′ UTRs of mRNAs contain sequence or structural elements that affect translation efficiency in response to environmental stimuli or RBP binding[16]. We expect that 2D and 3D functional structures will explain the presence of many blocks of highly conserved RNA sequence that have no ascribed function yet. Moreover, RNA secondary structure is more evolutionary conserved than RNA sequence, rendering comparative RNA structure prediction a relevant research tool. Very recent approaches to examine transcriptome-wide mRNA secondary structure *in vivo* have revealed structural organization in 5′ UTRs, 3′ UTRs and coding regions[17]. Whether differences in structural properties of distinct mRNA components are related to length, nucleotide composition, or function remains unravelled. In any case, these technical advances will give us insight into a relatively unknown layer of gene regulation applicable to many, if not all, UTRs.

CHANGING DND1 PROTEIN FUNCTION BY INTRONIC POLYADENYLATION

Alternative polyadenylation, by usage of a more proximal PAS than the canonical one, results in the processing of a shorter mRNA transcript. Usage of an alternative PAS in the 3′ UTR usually shortens the 3′ UTR, thereby removing sites mainly used for negative post-transcriptional regulation thus increasing protein expression. On the contrary, intronic polyadenylation excludes coding sequence, which is likely to result in an incomplete protein and loss of function. RBPs or mRNA structure can serve to protect proximal PASs from co-transcriptional polyadenylation, but not many such mechanisms are currently known. Specifically for intronic PASs, rapid splicing of introns by the splicing machinery may also protect from cleavage and polyadenylation.

A truncated protein form of DND1 increases susceptibility to spontaneous testicular teratoma in the mouse 129/SvJ strain, but is neither necessary nor sufficient for oncogenic transformation[18]. On other mouse strain backgrounds, the Ter mutation causes PGC death, but the remaining PGCs do not transform into EC cells. This implies that a second hit is required for tumorigenesis. Also, spontaneous GCT development in other mouse strains with intact DND1 points towards additional factors[19]. Zhu et al. implicate mRNA splicing factor 1 (SF1), which shows a stronger oncogenic effect in the absence of DND1[20]. Another question is why the 129/SvJ strain develops teratomas or testicular type I germ cell tumours, but not the common type II germ cell tumours found in young human males. Although the answer is yet unknown, the speculation is that this may be related to the time frame of germ cell development in mouse compared to that in man. PGC development in the embryo progresses rapidly in the mouse (within days) compared to that in humans (in months). Also, the time to fertility is extremely short in mice compared to that in humans, giving germ cell CIS more time to acquire additional genetic or epigenetic changes that cause dedifferentiation and transformation.

A rapid transition from G1- to S-phase is required to maintain ES cells in a pluripotent state, and is dependent on the suppression of p27, p21, Lats2 and other cell cycle inhibitor proteins[21]. In the 129/SvJ mouse strain, many DND1 mutant germ cells fail to enter mitotic arrest and the pluripotency markers Sox2,

Oct3/4 and Nanog are not downregulated. These and other data suggest that cell-cycle progression and pluripotency maintenance are mechanistically linked. In the C57BL/6J mouse strain, which is not susceptible to GCT formation, germ cells express negative regulators of the cell cycle (e.g. p21 and p27) in high levels relative to 129/SvJ strain[22]. In addition, a few hundred genes related to other processes than cell cycle are differentially expressed between the two cell lines[23]. Also, how levels of miR-290-295 compare in the different mouse strains is not clear.

COMPARISON OF MIR-290-295 AND DND1 EXPRESSION DURING MOUSE EMBRYONIC DEVELOPMENT

Low expression of the total miRNA pool is a characteristic of embryonic development, indicating that the expression of most target genes is desirable to maintain a pluripotent cell state or a high differentiation potential. This does not apply to the miR-371-373 cluster and its homologs in other species, which contribute to the maintenance of pluripotency during embryonic development. We compared the available data on timing of gene expression during embryonic development for the murine homologs. In mouse, miR-290-295 deficiency can result in embryonic lethality, or alternatively affects PGC migration and survival and female fertility (Fig. 1a)[24]. However, at some point in development, these miRNAs are co-ordinately downregulated to enable cellular differentiation. Intriguingly, formation of a CIS lesion due to an arrest in PGC differentiation is usually observed in a developmental time frame that parallels changes in both miR-290-295 and DND1 levels. CIS formation can occur until 15.5 dpc (days post coitus), which parallels the ultimate phase of miR-290-295 expression in male gonocytes. Preceding sexual differentiation, cell cycle inhibitors are upregulated even though repressive miRNAs are present and the germ cells arrest in mitosis around 12.5 dpc[25]. An increase in expression or activity of DND1, as observed in males at 12.5 dpc, may prevent target repression by miR-290-295. Expression of miR-290-295 and DND1 is differently regulated in male and female germ cells. The key events of sexual differentiation are entry into meiosis for XX gonocytes and entry into quiescence for XY gonocytes. We postulate that, if a desired increase in expression of functional DND1 in males after 12 dpc is lacking, the mRNA targets remain suppressed and mitosis continues. In that case, the PGCs arrest in development, which may also involve continued expression of miR-371-373 (Fig. 1b). Thus, downregulation of miR-371-373 expression is important, but counteracting its effect also seems important. However, it is not clear whether the predicted or validated mRNA targets of these miRNAs are repressed in$DND1^{Ter}$ mice, or which other functional miRNA targets are affected.

The miRNA cluster 17-92 is also highly expressed throughout PGC development and its overexpression is linked to cell cycle progression and proliferation. Most family members are downregulated in XX gonocytes at mitotic arrest and subsequent entry into meiosis, but remain expressed in XY gonocytes[25]. By contrast, expression of the let-7 miRNA family is upregulated between 11.5 dpc and 13.5 dpc and only in

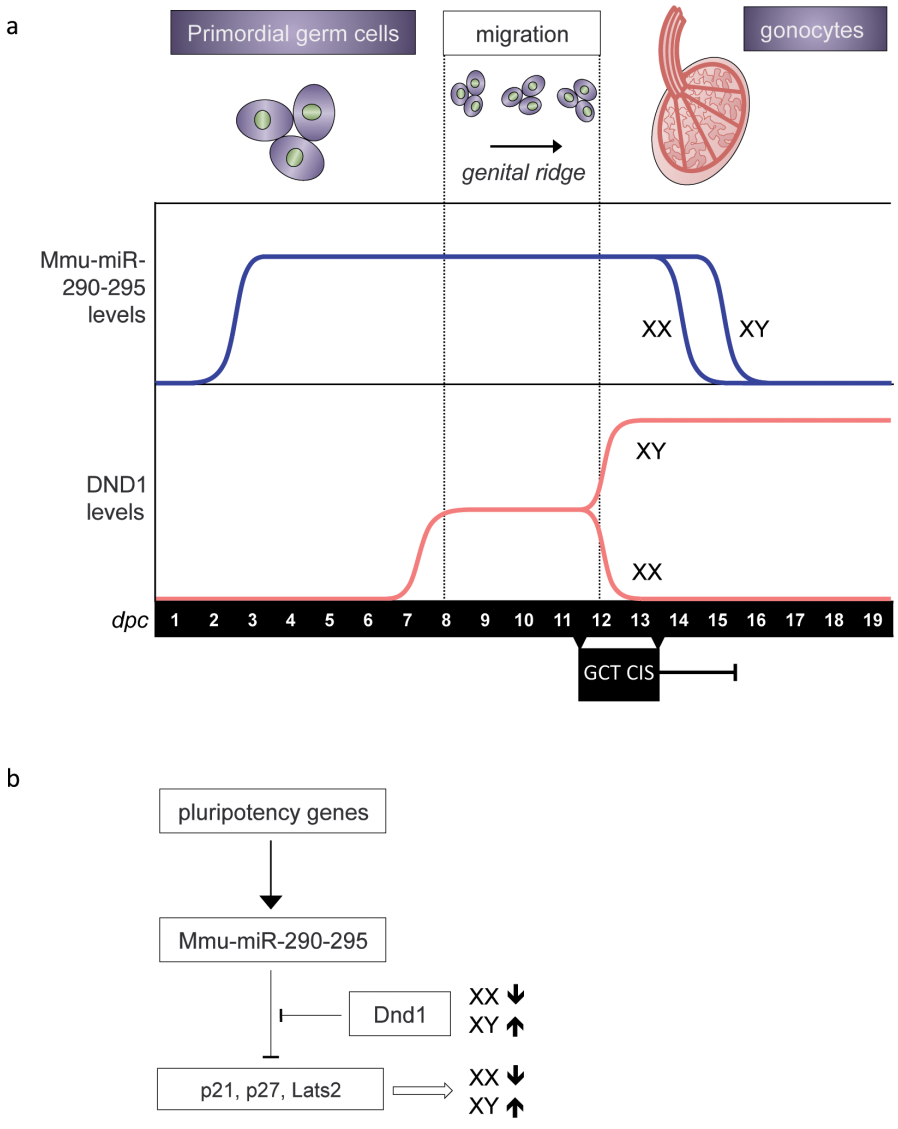


Figure 1 | Comparison of miR-290-295 and DND1 expression during mouse embryonic development. (a) The timing of expression of the miR-290-295 cluster and of DND1 during mouse embryonic development is plotted for 1-19 days post coitus. Over this period, primordial germ cells (PGCs) are derived from epiblast cells and migrate towards the developing gonads. They undergo mitosis, which is followed by meiosis at a certain time point to generate mature gametes. An arrest in development of late PGCs or early gonocytes may result in the formation of carcinoma *in situ* (CIS), which is the precursor of germ cell tumours (GCTs). CIS formation is typically observed between 11.5-13.5 dpc, but is still possible until 15.5 dpc as demonstrated by teratoma formation from PGCs grafted on older gonocytes. **(b)** Possible regulatory link between differentiation and cell cycle regulation. DND1 levels in gonocytes start to differ between XX and XY mice around 11.5 dpc, this could result in a difference in miR-290-295-mediated repression of cell cycle repressor genes. References can be found throughout the text in chapters 3, 5 and 6. XX = female, XY = male.

XY PGCs/gonocytes[25]. It is known that the let-7 family stimulates differentiation and inhibits cell proliferation, and its expression is often lost in cancer cells. Changes in expression of important genes for sex determination (e.g. SRY and Sox9)[26] or the distribution of differentiation-inducing retinoic acid during this exact time frame have not been mentioned here, but are also important.

Despite high cure rates of GCTs, diagnosis at the CIS stage would enable intervention before an invasive tumour develops and intensive therapy is required. Surgical testicular biopsy seems the only reliable diagnostic method to detect CIS. The treatment of unilateral CIS is orchiectomy, or localised irradiation in bilateral cases. Patients with cryptorchidism or fertility problems are more at risk of developing a GCT. Further research is warranted to obtain a method of non-invasive CIS detection and to identify causal factors explaining the increasing incidence of GCTs.

RNA & CLINICAL APPLICATIONS

The activity of mature miRNAs can be modulated at different levels, starting from sequence identity (RNA editing) to interaction with mRNA. The range of intrinsic regulatory mechanisms opens a window of therapeutic opportunities at the RNA level. A current strategy to inhibit endogenous miRNA expression is through administration of antisense synthetic oligonucleotides complementary to the miRNA. Oligonucleotide-modified anti-miRNAs, also known as antagomiRs, are miRNA inhibition tools that are currently being improved for clinical applications[27]. The modifications are essential for antagomiR stability in blood and in the cellular environment. Further, attachment of a cholesterol molecule to the 3' end improves uptake by cells. Inhibition of oncomiRs by using antagomiRs has been demonstrated in cell culture models, and systemic treatment with antagomiRs reduced metastasis in a mouse model for mammary tumours[28]. Still, two major difficulties for antagomiR application in clinical settings are local release in the target tissue and undesirable off-target effects. Another approach to neutralize overexpressed miRNAs in cancer is by transient expression of "miRNA sponges"[29]. These are transcripts that contain multiple miRNA target sequences to serve as decoy. For tumour suppressive miRNAs with reduced expression in tumours, local restoration of miRNA levels should provide a therapeutic benefit. The effect of introduction of double-stranded miRNAs mimics, equivalent to Dicer products, or vector-based RNA expression needs to be evaluated.

REFERENCES

1. Shalgi, R., Lieber, D., Oren, M. & Pilpel, Y. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS computational biology* **3**, e131 (2007).
2. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521-33 (2008).
3. Card, D.A. et al. Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Molecular and cellular biology* **28**, 6426-38 (2008).
4. Gillis, A. et al. High-throughput micro RNAome analysis in human germ cell tumours. *The Journal of pathology* (2007).
5. Barroso-delJesus, A. et al. Embryonic stem cell-specific miR302-367 cluster:

human gene structure and functional characterization of its core promoter. *Molecular and cellular biology* **28**, 6609-19 (2008).

6. Voorhoeve, P.M. et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124**, 1169-81 (2006).
7. Subramanyam, D. et al. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature biotechnology***29**, 443-8 (2011).
8. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-7 (2007).
9. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature biotechnology* **26**, 101-6 (2008).
10. Wernig, M., Meissner, A., Cassady, J.P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell stem cell* **2**, 10-2 (2008).
11. Evans, P.M. & Liu, C. Roles of Krüpel-like factor 4 in normal homeostasis, cancer and stem cells. *Acta biochimica et biophysica Sinica* **40**, 554-64 (2008).
12. Rowland, B.D., Bernards, R. & Peeper, D.S. The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. *Nat Cell Biol* **7**, 1074-82 (2005).
13. Seoane, J., Le, H.V. & Massagué, J. Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* **419**, 729-34 (2002).
14. Zhao, Y. et al. Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell stem cell* **3**, 475-9 (2008).
15. Galgano, A. et al. Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS ONE* **3**, e3164 (2008).
16. Keene, J.D. & Tenenbaum, S.A. Eukaryotic mRNPs may represent posttranscriptional operons. *Molecular cell* **9**, 1161-7 (2002).
17. Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. & Chang, H.Y. Understanding the transcriptome through RNA structure. *Nature reviews Genetics* **12**, 641-55 (2011).
18. Asada, Y., Varnum, D.S., Frankel, W.N. & Nadeau, J.H. A mutation in the Ter gene causing increased susceptibility to testicular teratomas maps to mouse chromosome 18. *Nature genetics***6**, 363-8 (1994).
19. Zhu, R., Ji, Y., Xiao, L. & Matin, A. Testicular germ cell tumor susceptibility genes from the consomic 129. MOLF-Chr19 mouse strain. *Mammalian genome : official journal of the International Mammalian Genome Society* **18**, 584-95 (2007).
20. Zhu, R., Heaney, J., Nadeau, J.H., Ali, S. & Matin, A. Deficiency of splicing factor 1 suppresses the occurrence of testicular germ cell tumors. *Cancer research* **70**, 7264-72 (2010).
21. Wang, Y. et al. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nature genetics* **40**, 1478-83 (2008).
22. Cook, M.S., Munger, S.C., Nadeau, J.H. & Capel, B. Regulation of male germ cell cycle arrest and differentiation by DND1 is modulated by genetic background. *Development (Cambridge, England)* (2010).
23. Sharova, L.V. et al. Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains. *Developmental biology* **307**, 446-59 (2007).
24. Medeiros, L.A. et al. Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14163-8 (2011).
25. Hayashi, K. et al. MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis. *PLoS ONE***3**, e1738 (2008).
26. Munger, S.C. et al. Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. *Genes & development* **23**, 2521-36 (2009).
27. Elmén, J. et al. LNA-mediated microRNA silencing in non-human primates. *Nature* **452**, 896-9 (2008).
28. Ma, L. et al. Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. *Nature biotechnology* **28**, 341-7 (2010).
29. Ebert, M.S. & Sharp, P.A. MicroRNA sponges: progress and possibilities. *RNA (New York, NY)* **16**, 2043-50 (2010).

&

ADDENDUM

SUMMARY

A human body is made up of many cells, which all arose from one zygote with the potential to differentiate into any kind of cell. During development of the embryo, the cells' pluripotency is gradually lost when cells differentiate to fulfil a specific function. Together, all cell types and organs cooperate to sustain the germ cells and secure reproduction.

Whereas pluripotent cells are rapidly dividing, most differentiated cells are inactive in replication (replicative senescence) except the cell types in regenerative organs with a high turnover. Still, all cells can experience genetic damage every day caused by internal or external stress factors. The tumour suppressor p53 is responsible for interception of damaged cells, but cells with dysfunctional p53 or cells that can otherwise escape this checkpoint will accumulate damage and continue to divide. Also, growth factors stimulate progression through cell cycle; in the absence of growth factors cells enter a reversible growth arrest (quiescence). Tumour cells are often self-sufficient in growth signals and insensitive to anti-growth signals.

DNA is the crucial collection of hereditary material that is passed on from generation to generation and contains the complete information for the synthesis of all types of proteins in an organism. Small parts of DNA are copied into a transient and mobile type of information storage, messenger RNA (mRNA). An mRNAs is transported to the cytoplasm where translation into a protein takes place. Over the last decade, many classes of RNAs have been identified that differ from mRNAs in that they do not code for proteins (non-coding RNAs). The evolutionary conservation of DNA sequences that are transcribed into non-coding RNAs implies important functions for these RNAs.

In this thesis, one class of small, non-coding RNAs (microRNAs or miRNAs) is studied. MiRNAs inhibit the expression of a gene into a protein at the post-transcriptional level. Besides the nucleotide sequence that actually codes for a protein, an mRNA transcript contains extra regions such as the 5' and the 3' untranslated regions (UTRs). MiRNAs regulate gene expression through binding 3' UTRs, which may be viewed,much like promoters for DNA, as regulatory domains controlling expression. The function of a miRNA results from the effect of translational repression of its mRNA targets, which do not accomplish their role in information transport. MiRNAs can function as oncogene or tumour-suppressor, and aberrant expression of miRNAs has been found in many tumours. We aim to understand how miRNA expression and function is regulated in human disease.

In **chapter 1**, we introduce the theories and concepts that are further discussed throughout the thesis. Also, nine general hallmarks of cancer cells and malignant growth are introduced. In **chapter 2**, we provide an overview of connections between miRNAs and RNA-binding proteins (RBPs) in relation to cancer that have been demonstrated thus far. Global downregulation of miRNA expression is an apparent feature of many tumours. Alterations in key players of miRNA biogenesis may affect mature miRNA levels in a global manner, whereas other RBPs regulate the production of specific (subsets of) miRNAs. Also, interplay between miRNAs and RBPs on target 3' UTRs rapidly modulates target expression under specific conditions or in certain cellular processes, some of which are linked to cancer.

Binding of RBPs near miRNA target sites can regulate miRNA function either directly by affecting miRNA binding or indirectly through a switch in RNA secondary structure. The activity of RBPs is temporally and spatially regulated through changes in transcription rate, post-translational modifications and subcellular localization, and is sometimes deregulated in cancer and other diseases. We speculate that these forms of interplay could hold true for other miRNAs and RBPs as well. Because of this, the challenge now is to connect the mechanisms of action to oncogenesis by applying state-of-the-art genome-wide approaches. In **chapter 4**, we describe an example of miRNA-RBP interplay into more detail. We demonstrate that expression of miRNAs 221 and 222 does not necessarily infer actual binding to target mRNAs followed by translational repression, therefore we suggest the involvement of accessory proteins that control miRNA binding. The RBP Pumilio allows binding of miRNAs 221 and 222 to the p27 mRNA target by resolving an inhibitive secondary RNA structure in the 3' UTR. Therefore, active Pumilio represses p27 expression, which causes quiescent cells to rapidly re-enter the cell cycle.

In **chapter 3**, we focus on the cluster of miRNAs 371, 372 and 373 that is expressed during embryonic development, and is also found in germ cell tumours. MiRNAs 372 and 373 have been found to drive oncogenic transformation. We describe a genomic region upstream of this cluster that is able to promote transcription, and we identify transcription factors (TFs) that contribute to miRNA transcription activation via this region. We find that upon expression of pluripotency-related TFs in differentiated cells, the methylation status of CpG islands in vicinity of the miRNAs 371, 372 and 373 changes and endogenous expression of these miRNAs can be detected. We know that combinations of pluripotency-related TF scan induce reprogramming of differentiated cells into embryonic stem cell-like cells. In the future, this technique may provide a solution for the regeneration of damaged of diseased tissue, but clinical application is hindered by formation of tumours from reprogrammed cells. In **chapter 5**, we investigate the RBP Dead end 1 (DND1) that protects specific mRNAs against the effect of miRNAs. We describe an incomplete protein form of DND1 that is generated as a result of intronic polyadenylation. This DND1 variant binds to mRNA targets, but lacks the ability to prohibit miRNA-mediated repression. Expression of the dysfunctional DND1 variant is found in normal germ cells, but in relatively higher amounts in germ cell tumours. It remains undetermined whether a lack of complete DND1 function contributes to germ cell tumorigenesis.

In order for eukaryotic cells and tissues to function properly, cellular proteins must have the correct spatiotemporal distribution. Transcriptional regulation involves interactions between transcription factors, which enables spatial and temporal differences in gene expression. However, mRNA expression levels are only poor predictors of protein levels. While production of mRNA is dictated by the rate of transcription, the translation of mRNA into protein is regulated by multiple factors either increasing or decreasing translational efficiency. We conclude that microRNAs target many genes concurrently and effect global changes in gene expression, but that their effect is influenced by RBPs or by the secondary (and perhaps tertiary) structure of mRNA. Both intervention by RBPs and conformational changes in RNA secondary structure could provide an explanation for instances where not all present mRNA targets are concurrently repressed by a certain miRNA,

or for instances where a certain mRNA target is not targeted by all miRNAs for which a binding site is present in its 3′ UTR. Interplay between miRNAs and RBPs enables fast anticipation to cellular changes and dynamic and flexible response to a cell's need. Messenger RNA has an essential function in gene expression, however, it is dependent on the tuning of the message on the way.

Cancer comprises more than hundred different diseases. Every case has its origin in a specific cell type, contains tumour cells with a certain drive to proliferate, to spread through the body and to settle in other body parts. Also, the sensitivity of tumours to different treatment strategies varies. The risk of each individual to develop cancer is a combination of hereditary factors and environmental factors throughout life. On the one hand, the risk for cancer increases during life, on the other hand, combinations of new and available therapies enable us to control tumours for an extended period or to eliminate them completely.

NEDERLANDSE SAMENVATTING

Dat het leven begint met een enkele bevruchte eicel die alle erfelijke informatie bevat om een volledig organisme te vormen, is een simpel gegeven. De processen betrokken bij het tot stand komen en in stand houden van een organisme zijn een stuk complexer. Tijdens de ontwikkeling van een embryo wordt bepaald of een cel (embryonale stamcel) zich blijft vermenigvuldigen om bij te dragen aan groei of zich specialiseert in een bepaalde functie. Deze specialisatie wordt differentiatie genoemd en ligt ten grondslag aan alle verschillende types weefsel die samen een lichaam vormen. Samenwerking tussen de verschillende organen en weefsels is uiteindelijk gericht op het ondersteunen van de kiemcellen, de cellen die geslachtscellen vormen met als doel voortplanting.

Een menselijk lichaam bestaat grofweg uit honderdduizend miljard cellen; iedere cel bevat het volledige DNA en daarmee alle genetische informatie. Wanneer een cel zich vermenigvuldigt door deling (celdeling), wordt het DNA verdubbeld en verdeeld over de dochtercellen. Sommige cellen, zoals bloed- of huidcellen, worden vaak vernieuwd, maar de meeste cellen gaan langer mee en stoppen na een aantal celdelingen met delen. Tijdens de celdeling wordt het DNA gecontroleerd op schade, waar het eiwit p53 op verschillende manieren op kan anticiperen. Iedere dag worden er vele schadegevallen opgespoord en hersteld, maar bij ernstige schade wordt de cel op non-actief gezet. Toch kan een cel ook schade aan het DNA oplopen waarmee de cel ontsnapt aan de controle, of kan een overmaat aan groeisignalen de cel stimuleren om door te gaan met delen. Het gevolg is een opeenstapeling van DNA-schade, en veel erger, de beschadigde cellen kunnen ongecontroleerd delen en uitgroeien tot een tumor.

DNA is de erfelijke informatie die opgebouwd is uit vier verschillende bouwstenen (de basen A, C, G en T) en die de genen bevat: verschillende codes voor ieder eiwit dat benodigd is om een organisme te laten functioneren. Kleine delen van het DNA worden gekopieerd (transcriptie) op boodschapper-RNA (mRNA), een tijdelijke en mobiele vorm van informatieopslag. Vervolgens wordt het mRNA naar een plek in de cel gebracht waar eiwit geproduceerd kan worden (translatie). De specifieke functie van een cel hangt samen met die delen van het DNA die in die cel op RNA worden gekopieerd: het mRNA-expressiepatroon.

Recentelijk zijn we meer te weten gekomen over het bestaan van vele andere soorten RNA dan mRNA, de enige soort RNA die codeert voor eiwitten. Deze andere soorten zijn voornamelijk niet-coderende RNAs, dat wil zeggen dat hun functie op het niveau van RNA wordt vervuld. Het feit dat de basevolgorde van deze niet-coderende RNAs gedurende de evolutie behouden is gebleven in het DNA, impliceert dat ze een functie hebben die onmisbaar is voor de voortplanting van een organisme. In dit proefschrift bestuderen we onder andere microRNAs. Deze korte RNAs binden aan een specifieke code van bepaalde mRNAs en dit belemmert de productie van het eiwit waarvoor de mRNA codeert. Er zijn tot nu toe meer dan duizend verschillende microRNAs gevonden in het menselijk lichaam, en iedere microRNA kan veel verschillende mRNAs herkennen. De meeste microRNAs komen in specifieke celtypes voor, echter in tumoren worden vaak afwijkende soorten en hoeveelheden microRNAs gevonden.

In **hoofdstuk 1** lichten we enkele concepten toe die verderop in het proefschrift aan de orde komen. Ook noemen we negen kenmerken van een tumor zoals ze recentelijk gedefinieerd zijn, het vermogen tot ongelimiteerde celdeling en zelfvoorziening van groeisignalen zijn daar twee van. In **hoofdstuk 2** vatten we de tot dusver beschikbare literatuur samen waarin beschreven wordt hoe de productie en functie van microRNAs worden geregeld onder uiteenlopende omstandigheden. Hierbij zijn vaak ook eiwitten betrokken die aan RNA binden en die in een samenspel met de microRNAs het uiteindelijke lot van mRNAs bepalen. In **hoofdstuk 4** beschrijven we hoe het RNA-bindende eiwit Pumilio de binding van microRNAs 221 en 222 regelt, in dit geval de binding aan het mRNA dat codeert voor het eiwit p27. Het eiwit Pumilio wordt geactiveerd in de aanwezigheid van groeisignalen en verandert de structuur van p27 mRNA zodanig dat het beter toegankelijk is voor de microRNAs. De daaropvolgende remming van de productie van p27 eiwit stimuleert de voortgang van een cel in celdeling.

In **hoofdstuk 3** kijken we naar miRNAs (371, 372, en 373) die tot expressie komen tijdens de embryonale ontwikkeling, maar ook voorkomen in kiemceltumoren. We onderzoeken of een gedeelte van het DNA grenzend aan de miRNA-genen een rol speelt bij het activeren van transcriptie. Ook testen we of een aantal eiwitten (transcriptiefactoren) mogelijk betrokken zijn bij de transcriptie van deze miRNAs. Deze transcriptiefactoren spelen een belangrijke rol tijdens de embryonale ontwikkeling en expressie van de eiwitten wordt ook gevonden in kiemceltumoren. Gezamenlijke expressie van deze eiwitten kan zelfs reeds gedifferentieerde cellen weer een expressiepatroon laten aannemen dat sterk lijkt op dat van embryonale stamcellen; dit principe zou in de toekomst klinische oplossingen kunnen bieden voor regeneratie of herstel van ziek of beschadigd weefsel. Een probleem is nog wel dat deze zogenoemde pluripotente cellen ongeremd doorgaan met groeien en bijvoorbeeld kiemceltumoren vormen. In **hoofdstuk 5** richten we ons nogmaals op de kiemcellen en dan specifiek op het RNA-bindende Dead end eiwit dat bepaalde microRNAs kan belemmeren om aan mRNA te binden en daarmee de productie van het eiwit waarborgt. We beschrijven een incomplete vorm van het Dead end mRNA, die slechts codeert voor een deel van het volledige eiwit. Het kleinere eiwit herkent en bindt nog wel mRNA, maar mist de capaciteit om te beschermen tegen het remmend effect van microRNAs. We vinden dit incomplete mRNA onder andere in gezonde kiemcellen, maar in relatief grote hoeveelheden in kiemceltumoren. Het blijft de vraag of het ontbreken van de volledige functie van Dead end werkelijk bijdraagt aan de ontwikkeling van dergelijke tumoren.

We concluderen dat microRNAs weliswaar vele genen tegelijk remmen en het expressiepatroon op grote schaal bijsturen, maar dat de aanwezigheid van een microRNA niet noodzakelijk de maat aangeeft waarmee dit gebeurt. RNA-bindende eiwitten kunnen zowel de productie als de activiteit van microRNAs beïnvloeden, deze invloed kan het effect op genexpressie ofwel bevorderen ofwel verhinderen. Ze kunnen bijvoorbeeld rechtstreeks de plaats van microRNAs bezetten of de structuur van het mRNA zodanig veranderen dat ze beter of slechter toegang bieden aan microRNAs, of een ander regulatiemechanisme exploiteren. Deze eiwitten zijn zelf weer afhankelijk van expressie, lokalisatie en omgevingsfactoren die hun activiteit beïnvloeden. Regulatie door microRNAs en RNA-bindende eiwitten biedt de

mogelijkheid om snel op veranderingen te anticiperen en op een dynamische en flexibele manier in de behoeften van een cel te voorzien. De rol van boodschapper-RNA is een kernonderdeel van genexpressie. Echter, deze rol is afhankelijk van de manier waarop de boodschap onderweg gestuurd wordt.

Kanker is een verzamelnaam voor meer dan honderd verschillende ziekten. In ieder geval is het ontstaan uit een specifiek celtype, hebben de tumorcellen een zekere drang om uit te breiden, om zich te verplaatsen door het lichaam en zich op andere plaatsen te vermenigvuldigen, en zijn tumorcellen meer of minder gevoelig voor verschillende behandelingsmethoden. Het risico voor ieder individu om kanker te ontwikkelen wordt bepaald door een combinatie van erfelijke factoren en omgevingsfactoren tijdens het leven. Enerzijds loopt het risico op kanker gedurende het leven op, anderzijds maken (combinaties van) nieuwe en beschikbare behandelingsmethoden het steeds beter mogelijk om tumoren langere tijd te beheersen of zelfs helemaal te verwijderen.

CURRICULUM VITAE

Marieke van Kouwenhove werd geboren op 20 oktober 1982 te Stadskanaal. In 2000 behaalde zij haar gymnasiumdiploma aan het Dollard College te Winschoten. Datzelfde jaar begon zij de studie Biomedische wetenschappen aan de Universiteit van Utrecht, waar zij haar bachelor diploma behaalde in 2004. In 2003 heeft ze haar studie onderbroken voor een bestuursjaar binnen de studentenvereniging Unitas S.R. te Utrecht. In 2004 begon zij aan de master opleiding Genomics & Bioinformatics aan de Universiteit van Utrecht. Tijdens deze opleiding heeft ze onder begeleiding van Edwin Cuppen in zijn onderzoeksgroep aan het Hubrecht Instituut voor ontwikkelingsbiologie en stamcelonderzoek te Utrecht gewerkt aan een methode om middels het sequencen van polymorfismen (SNPs) in het genoom van *C.elegans* mutaties verantwoordelijk voor fenotypische veranderingen te lokaliseren. Vervolgens heeft ze in de onderzoeksgroep van Marc van de Vijver aan het Nederlands Kanker Instituut te Amsterdam de mRNA expressiepatronen van hormoonreceptor-negatieve borsttumoren bestudeerd onder begeleiding van Hugo Horlings en Bas Kreike. Tijdens deze stage in 2006 heeft zij in de onderzoeksgroep van Thomas Tuschl aan de Rockefeller University in New York, VS, de expressie van microRNAs in hetzelfde type borsttumoren onderzocht. In 2006 behaalde zij haar master diploma, waarna zij in 2007 begon als onderzoeker in opleiding in de groep van Reuven Agami op de afdeling Genregulatie binnen het Nederlands Kanker Instituut te Amsterdam. De resultaten van dat onderzoek staan beschreven in dit proefschrift.

LIST OF PUBLICATIONS

1. Intronic polyadenylation results in a functionally deficient DND1 protein variant. **van Kouwenhove M**, Oude Vrielink JA, Jenal M, Elkon R, van Haaften G, Rooijers K, Kedde M, Urlaub H, Erozenci LA, Gillis AJ, Looijenga LH, Agami R. *Manuscript in preparation*
2. MicroRNA regulation by RNA-binding proteins and its implications for cancer. **van Kouwenhove M**, Kedde M, Agami R. *Nat Rev Cancer* **11(9)**, 644-56 (2011).
3. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Farazi TA, Horlings HM, Ten Hoeve JJ, Mihailovic A, Halfwerk H, Morozov P, Brown M, Hafner M, Reyal F, **van Kouwenhove M**, Kreike B, Sie D, Hovestadt V, Wessels LF, van de Vijver MJ, Tuschl T. *Cancer Res.* **71(13)**, 4443-53 (2011).
4. A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. Kedde M*, **van Kouwenhove M***, Zwart W, Oude Vrielink JA, Elkon R, Agami R. *Nat Cell Biol* **12(10)**, 1014-20 (2010).
5. Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. Turner N, Lambros MB, Horlings HM, Pearson A, Sharpe R, Natrajan R, Geyer FC, **van Kouwenhove M**, Kreike B, Mackay A, Ashworth A, van de Vijver MJ, Reis-Filho JS. *Oncogene* **29(14)**, 2013-23 (2010).
6. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. Kreike B, **van Kouwenhove M**, Horlings H, Weigelt B, Peterse H, Bartelink H, van de Vijver MJ. *Breast Cancer Res.* **9(5),** R65 (2007).

* These authors contributed equally