UNIVERSITAT DE BARCELONA

Epigenetic regulation mediated by antisense non-coding RNAs and its impact on oncogenic pathways: the *HMGA2/RPSAP52 locus*

Cristina Oliveira Mateos

UNIVERSITAT DE BARCELONA

Epigenetic regulation mediated by antisense non-coding RNAs and its impact on oncogenic pathways: the *HMGA2/RPSAP52 locus*

PhD thesis

Cristina Oliveira Mateos

Regulatory RNA and Chromatin Group
Cancer Epigenetics and Biology Program (PEBC)
Bellvitge Biomedical Research Institute (IDIBELL)
Josep Carreras Leukaemia Research Institute (IJC)

Thesis Directors

Dr. Sònia Guil Domènech

Dr. Manel Esteller Badosa

Barcelona, 2020

IDIBELL Institut d'Investigació Biomèdica de Bellvitge

PEBC Cancer Epigenetics and Biology Program Programa d'Epigenètica i Biologia del Càncer Programa de Epigenética y Biología del Cáncer

Josep Carreras LEUKAEMIA Research Institute

Epigenetic regulation mediated by antisense non-coding RNAs and its impact on oncogenic pathways: the *HMGA2/RPSAP52 locus*

This work has been carried out and presented by Cristina Oliveira Mateos to obtain the doctoral degree in the University of Barcelona under the supervision of Dr. Sònia Guil Domènech and Dr. Manel Esteller Badosa

UNIVERSITY OF BARCELONA – FACULTY OF MEDICINE

BIOMEDICINE DOCTORAL PROGRAM 2020

Regulatory RNA and Chromatin Group
Cancer Epigenetics and Biology Program (PEBC)
Bellvitge Biomedical Research Institute (IDIBELL)
Josep Carreras Leukaemia Research Institute (IJC)

Dr. Sònia Guil Domènech
Director

Dr. Manel Esteller Badosa
Director and tutor

Cristina Oliveira Mateos

CONTENTS

CONTENTS

ABBREVIATIONS

#	
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A	
ASO	antisense oligonucleotide
C	
ceRNA	competing endogenous RNA
CGI	CpG island
circRNA	circular RNA
CpG	Cytosine-phosphate-Guanine
CRISPR	clustered regularly interspaced short palindromic repeat
D	
DNMTs	DNA methyltransferase
dsRNA	double-stranded RNA
E	
ENCODE	Encyclopedia Of DNA Elements
eRNA	enhancer RNA
F	
FANTOM	Functional Annotation of the Mammalian Genome
H	
HMGA2	*High Mobility Group A2*
I	
IGF2BP2	*Insulin-like Growth Factor 2 mRNA-Binding Protein 2*
IRES	internal ribosome entry site
K	
KH	K homology domain
L	
let-7	*lethal-7*
LIN28B	*LIN-28 homolog B*
lincRNA	long intergenic non-coding RNA
LNA	locked nucleic acid
lncRNA	long non-coding RNA
M	
miRNA	microRNA
mRNA	messenger RNA

N

NAT	natural antisense transcript
ncRNA	non-coding RNA

P

piRNA	PIWI-interacting RNA
Pol	polymerase
pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA

R

RBP	RNA-binding protein
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNP	ribonucleoprotein
RPSAP52	*Ribosomal Protein SA Pseudogene 52*
RRM	RNA recognition motif
rRNA	ribosomal RNA

S

S/AS	sense/antisense
shRNA	short-hairpin RNA
siRNA	small interfering RNA
sncRNA	small ncRNA
snoRNA	small nucleolar RNA
SNP	single-nucleotide polymorphism
snRNA	small nuclear RNA

T

tRNA	transfer RNA

U

UTR	untranslated region

ABSTRACT

The vast majority of the human genome is transcribed, being many of the RNAs produced non-coding RNAs (ncRNAs) with largely unknown functions. Natural antisense transcripts (NATs) are one of the most abundant types of long non-coding RNAs (lncRNAs), and possess emerging roles in the regulation of the nearby coding genes. This is the case of the oncofetal gene *HMGA2* (*High Mobility Group A2*) and its antisense *RPSAP52* (*Ribosomal Protein SA Pseudogene 52*), whose expression is high and positively correlated in a number of human cancers. The antisense transcript forms an R-loop in the promoter region that changes chromatin conformation, favoring the transcription of the sense gene *HMGA2*.

RPSAP52 exerts additional functions in the cytoplasm through the binding to IGF2BP2 (Insulin-like Growth Factor 2 mRNA-Binding Protein 2) protein, which is a transcriptional target of HMGA2. IGF2BP2 promotes the translation of some genes related with important proliferative pathways, and we demonstrated that it also binds to *LIN28B* (*LIN-28 homolog B*) messenger RNA (mRNA), one of the main negative regulators of *let-7* (*lethal-7*) microRNA (miRNA) maturation. The interaction with *RPSAP52* affects the binding of IGF2BP2 to specific targets such as *LIN28B* mRNA and its recruitment to polysomes. Thus, *RPSAP52* presence increases the translation of *LIN28B* and reduces the levels of the tumor suppressor *let-7*.

The regulation mediated by *RPSAP52* has a severe impact in cancer related pathways, where *RPSAP52* has demonstrated its oncogenic potential. Its depletion impairs tumorigenic features of the cells *in vitro* and decreases tumor progression *in vivo*. Moreover, high *RPSAP52* levels act as a biomarker of worse prognosis in sarcoma.

In summary, the present work proposes a regulatory model mediated by the lncRNA *RPSAP52* with two different levels of action. This antisense transcript promotes the transcriptional activation of *HMGA2* and, in turn, regulates the function of IGF2BP2 protein. Since HMGA2 and IGF2BP2 are in the same proliferative pathway, *RPSAP52* reinforces the function of HMGA2 both on this gene and on its downstream effectors, with the subsequent effect in cancer progression. Due to the important roles performed by *RPSAP52* and its oncogenic properties, this lncRNA could be a potential therapeutic target for the development of new cancer treatments.

INTRODUCTION

1. Cancer

Cancer is a large family of diseases defined by the presence of abnormal cells with an uncontrolled growth that can lead to the formation of tumors in almost any part of the body. Sometimes these cells have the potential to invade or spread to other organs, a process that is known as metastasis, and which is the main cause of death from cancer[1]. The accumulation of genetic alterations in normal cells determine their transformation into malignant cells during the tumorigenesis.

Nowadays, it is considered that there are eight hallmarks that define cancer and that confer tumor capacities to the cells[2,3] **(Fig. 1)**. These are: self-sufficiency in proliferative signals, insensitivity to antigrowth factors, evasion of programmed cell death, limitless replicative potential, sustained angiogenesis, tissue invasion and metastatic capacity, reprogramming of energy metabolism, and immune destruction evasion.

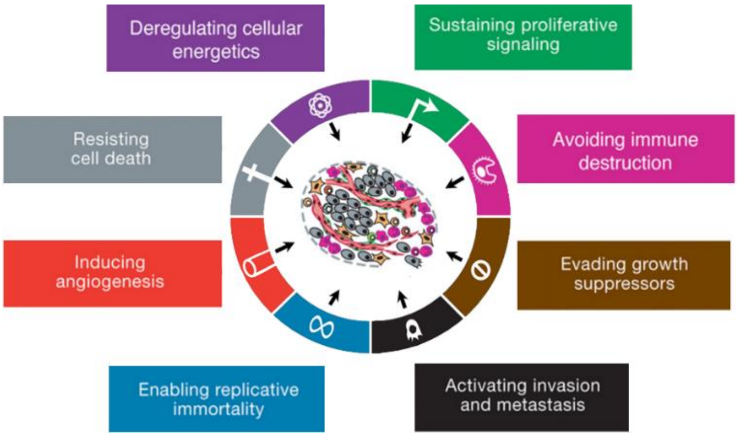

Deregulating cellular energetics
Sustaining proliferative signaling
Resisting cell death
Avoiding immune destruction
Inducing angiogenesis
Evading growth suppressors
Enabling replicative immortality
Activating invasion and metastasis

Figure 1. The hallmarks of cancer. Biological capabilities that distinguish cancer cells from normal cells. These eight alterations are acquired by the cells during malignant transformation and are essential for tumor progression. The majority of cancers, although very different from each other, share these characteristics (Figure modified from Hanahan & Weinberg, 2011[3]).

Cancer is a leading cause of death worldwide[4], and breast cancer is the most frequently diagnosed malignancy in women **(Fig. 2)**, representing 30% of all cases. Its death rate has been stable or decreasing since around 1990 in developed countries thanks to advances in diagnosis and treatment[5]. However, although the survival rate is established around 90%, it is necessary to make an effort to improve the understanding and complete tackling of the disease since the incidence and mortality increase with population ageing, especially in developing countries where they are adopting behaviors that increase cancer risk[4]. Therefore, it remains one of the main causes of cancer death among women[5]. On the other hand, sarcomas are a heterogeneous group of tumors considered rare cancers (incidence of about 6 per 100,000)[6]. Despite the fact that they only represent 1% of adult solid malignancies, they are among the most common solid pediatric cancers (21%)[7], with special relevance of rhabdomyosarcoma in children and Ewing's sarcoma in young adults and adolescents. Their low incidence, together with their differences in origin, location and age of appearance, hinder an early diagnosis and reduce survival expectations[8].

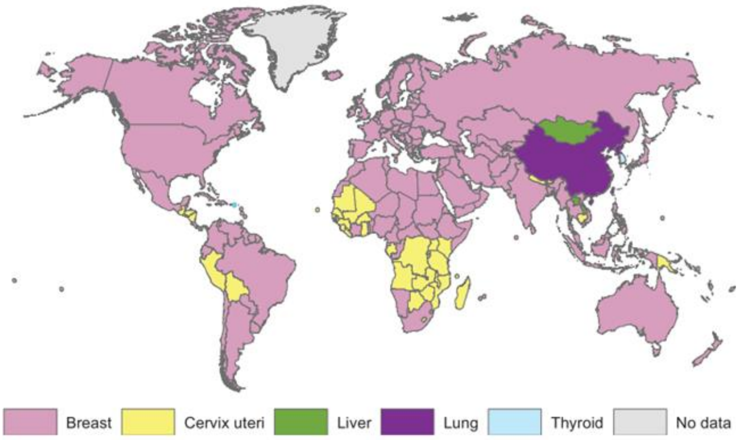


Figure 2. Most commonly diagnosed cancers worldwide in females. Global cancer incidence between women shows breast cancer as the main detected tumor all over the world. This malignancy is only surpassed in a few countries by cervical, lung, liver and thyroid cancers (Figure from Torre *et al*., 2016[4]).

Epithelial cells of the mammary gland give rise to the majority of breast cancers and their origin and invasiveness determine the different types of tumors. Another level of classification is the presence/absence of estrogen and progesterone receptors and of HER2. They are considered triple negative if they do not express any of them, and their aggressiveness and treatment depend on this due to the lack of response to hormone therapy[9]. Soft tissue sarcomas, like rhabdomyosarcoma, appear in muscles, nerves, deep skin tissues, blood vessels…; whereas Ewing's sarcoma affects bones and cartilage[8]. Rhabdomyosarcoma cells are similar to undifferentiated myoblasts because they maintain characteristics of immature skeletal muscle[10]. This sarcoma type is classified in two main subtypes: embryonal and alveolar rhabdomyosarcoma[11]. Embryonal rhabdomyosarcoma is associated with mutations in the RAS pathway[12], while alveolar and Ewing's sarcoma are characterized by chromosomal translocations that generate fusion proteins, PAX3-FOXO1 or PAX7-FOXO1 in the first case and EWS-FLI1 in the second one[13–15].

The huge heterogeneity of cancer makes it impossible to find an effective common treatment for all cases. It is necessary to adjust the therapies to each tumor type according to the specific characteristics that give them an advantage to grow and progress. These characteristics, in turn, are the same that make them vulnerable, and the ones that give an opportunity for the development of new therapies. For this, it is essential to have an in-depth knowledge of the altered pathways and the genetic changes that define each tumor, so that we can make a classification as accurate as possible.

2. Epigenetics

Genetic changes at the DNA sequence or rearrangement at the chromosomal level are not the only cause of cancer development. This higher level of regulation is known as epigenetics. The presence of the prefix “epi” in the word, which means above or beyond, helps to understand the meaning of this term that was defined by Conrad Waddington in 1942[16]. The concept has evolved over time, and the current definition explains the term as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence”[17].

These control mechanisms are required during normal development for the differentiation of cells with the same genotype into the great cellular variety that constitute an organism. This process involves stable activation and repression of certain genes in a cell- and stage-specific manner, but if the wrong genes are affected, pathological processes like cancer can be triggered. The main layers of epigenetic regulation are histone modifications, DNA methylation and ncRNAs.

2.1 Histone modifications

Histones are a family of nuclear proteins that bind to the DNA, conforming the chromatin. Nucleosomes are the fundamental unit of the chromatin and they are composed of an octamer with two histones of each type (H2A, H2B, H3 and H4)[18]. Around the octamer 147 bp of DNA are packaged[19] and about 50 bp of DNA separate one nucleosome from the next one[20]. Histone H1 is located on this linker DNA. All histones have a C-terminal domain and a N-terminal tail subject to a large number of post-translational modifications[21], and the cleavage of part of H3 tail has also been described as a new modification type[22]. Over 60 different residues have been reported to be modified, and the modifications include acetylation and methylation, among others[23].

These epigenetic marks are related with nucleosome positioning and the transcriptional state of the chromatin **(Fig 3)**. Acetylation, for example, is associated with less condensed and transcriptionally active chromatin , whereas deacetylation results in a more compact, transcriptionally silent chromatin[24,25]. Thus, histone deacetylase inhibitors have been used to reexpress silenced genes[26]. On the other hand, the transcriptional effect of histone methylation depends on the modified residues. Another way to modulate chromatin is via incorporation of histone variants, that differ from core histones in their tails, structure and sequence[21]. The final consequence of the epigenetic marks is determined by the crosstalk among all modifications that take place simultaneously in each histone[27].

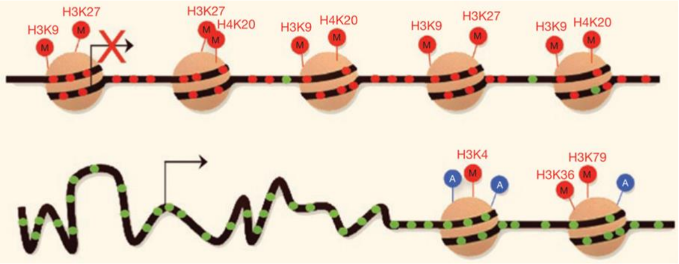


Figure 3. Nucleosome positioning and gene regulation.
The location of the nucleosomes throughout the genome depends on different epigenetic marks. Histone methylation in certain residues (H3K9, H3K27 and H4K20) is associated with high nucleosome occupancy where transcription is not possible (*above*). Acetylation and methylation of H3K4, H3K36 and H3K79 are related with lower nucleosome density. The less condensed chromatin favors the accessibility of the transcriptional machinery to the promoters of active genes (*below*). M: Methylation. A: Acetylation (Figure from Portela & Esteller, 2011[27]).

2.2 DNA methylation

DNA methylation consists in the covalent addition of methyl groups to certain DNA bases. In humans, this reaction occurs mainly to the cytosines of Cytosine-phosphate-Guanine (CpG) dinucleotides, although less prevalent modifications have also been described, such as non-CpG methylation[28,29] or N^6-methyladenine mark[30]. CpGs are usually concentrated in large areas called CpG islands (CGIs)[31], defined as regions of more than 200 bp with a GC content of at least 50% and a ratio above 0.6 of observed CpG dinucleotides versus expected number[32]. They are distributed throughout the genome, located in the promoters of approximately 60% of human genes, particularly housekeeping ones[33]. It should be noted the importance of CGI shores, regions around 2 kb close to CGIs with lower CpG density, which can also be methylated[34] **(Fig. 4a)**.

DNA methyltransferases (DNMTs) are responsible for the transfer of a methyl group from S-adenosyl methionine to C^5 carbon of cytosines in the DNA, giving rise to a 5-methylcytosine (5mC)[27]. DNMT1 binds preferentially to hemimethylated DNA and maintains the methylation state after DNA replication[35]. During embryonic development, *de novo* methylation is established by DNMT3A and DNMT3B[27,36]. On the other hand, DNA demethylation is not fully understood. One option is a passive reaction during DNA replication. Description of 5-hydroxymethylcytosine (5hmC) in brain[37], and the fact that DNMT1 activity is reduced at sites of hemi-5hmC *in vitro*, raises the possibility that hydroxymethylation blocks maintenance methylation and produces a passive demethylation during cell division[38]. An active mechanism has been proposed as a result of the discovery of TET proteins. They catalyze the oxidation of 5mC to 5hmC[39], 5-formylcytosine and 5-carboxylcytosine, and these derivatives can be converted to cytosine[40–42]. The cited marks could be merely intermediaries of the demethylation process. However, some of them are specifically recognized by other proteins[43] and show a particular distribution along the genome[44]. Thus, they can constitute stable epigenetic marks with their own biological meaning[45].

The function for DNA methylation was proposed in 1975, when some studies showed a relationship with gene expression[46,47]. The regulation is achieved through methyl-CpG binding proteins, that recruit chromatin-modifying complexes to the DNA[48], interfering with the binding of transcription factors required for gene induction. Methylation is associated with condensed chromatin and gene silencing, but this is highly related with its localization. The majority of gene promoters remain unmethylated in normal tissues allowing gene expression[49], although some of them become methylated during development and differentiation to give rise to specific cell linages[50] **(Fig. 4b)**. Gene bodies, instead, are deeply methylated even in expressed genes[51], avoiding aberrant initiation of transcription[52] **(Fig. 4c)**. Intergenic regions are also hypermethylated to prevent the harmful expression of repetitive and transposable elements[53] **(Fig. 4d)**.

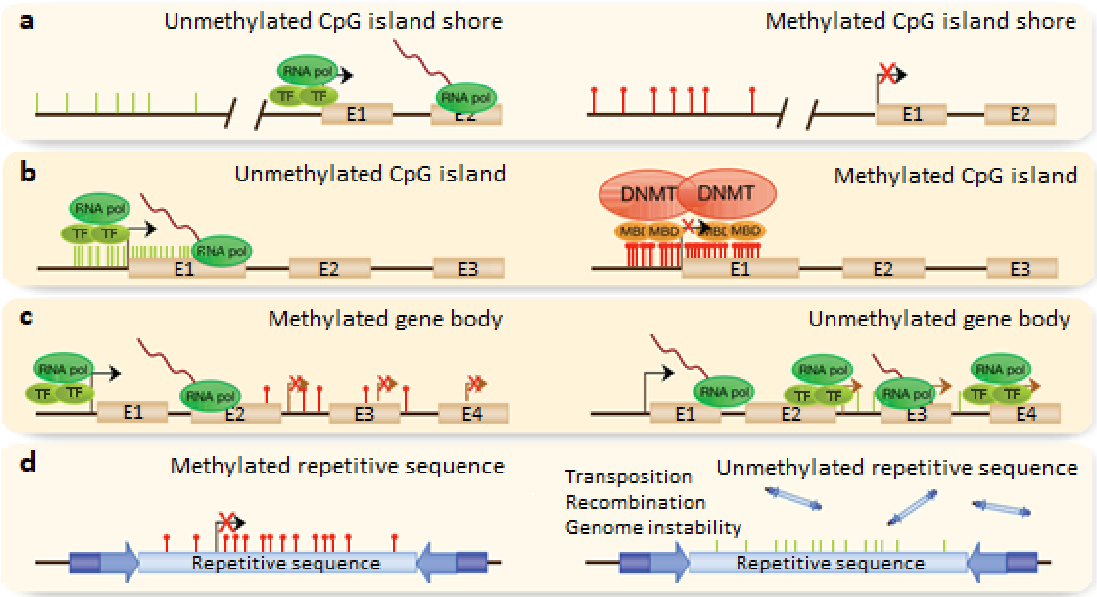


Figure 4. DNA methylation patterns in different regions of the genome.
DNA methylation profile is not the same in physiological (*left*) and pathological (*right*) conditions. **(a)** CGI shores and **(b)** CGIs in gene promoters remain unmethylated in normal tissues allowing gene expression. Their methylation represses transcription and allows differentiation during development, but in many cases hypermethylation is an aberrant condition associated with cancer and other diseases. **(c)** Gene bodies are deeply methylated in actively transcribed genes to avoid the initiation of transcription at incorrect sites. The loss of this methylation is typical in pathologies. **(d)** Methylation in intergenic regions prevents the expression of repetitive and transposable elements. In disease, demethylation of these regions favors chromosomal instability (Figure from Portela and Esteller, 2011[27]).

The correct methylation pattern in each moment and region is essential for normal development, and regulate important physiological processes such as imprinting and X-chromosome inactivation[46,54,55]. Dysregulated methylation is a common cause of tumor suppressor genes silencing and oncogenes activation. There are therapies based on demethylating agents such as decitabine or azacytidine, directed to reverse these changes. Both drugs have been approved for the treatment of myelodysplastic syndromes[26].

RNA can also be modified with epigenetics marks. Until now, there are 163 chemical modifications identified that affect RNA structure and interactions[56], but this is not further addressed here because it falls outside the scope of this thesis.

2.3 Non-coding RNAs

It has been more than a century since the first evidences of the existence of RNA, but for many years its biological importance has been underrated. The central dogma of molecular biology was enunciated in 1958 by Francis Crick[57], according to which the genetic information flows from the nucleic acids to proteins. This placed the RNA molecule as a mere intermediary between DNA and proteins, the essential constituent of living organisms. Since proteins were considered the functional product of genes, the term 'junk DNA' was coined in 1972 for the regions without protein coding potential[58]. It was not until 1981, with the identification of the first ribozyme[59], that the concept of RNA changed. These RNA molecules with catalytic activity showed a much more complex scenario, in which RNA plays relevant biological roles beyond information transmission.

The idea was supported by the discovery that the majority of the genome is transcribed, but not translated. This has been possible thanks to the development of next-generation sequencing techniques that have allowed the in-depth transcriptomic analysis of different organisms and cellular types. The FANTOM (Functional Annotation of the Mammalian Genome) Consortium detected transcribed regions in around 70% of the mouse genome[60]. Similar results were found for humans by the ENCODE (Encyclopedia Of DNA Elements) Project[61], although less than 2% of our genome codes for proteins[62]. This transcribed non-coding portion of the genome constitutes the ncRNAs.

Another evidence of the relevance of these transcripts is that the complexity of the higher organisms cannot be explained by the difference in the number of protein coding genes. It is rather more related with an increase in ncRNAs transcription and the regulatory network established by them[63]. An example of this is that in bacteria, non-coding sequences constitute 5% of the genomic DNA, and this number increases up to 70% and 80% in unicellular eukaryotes and invertebrates, respectively[64]. Moreover, many ncRNAs present specific expression patterns, tightly regulated during differentiation and development[61,65], and they are less conserved than protein coding genes but significantly more than other sequences[66]. This suggests that they are not merely noise or transcriptional byproducts, but they are molecules generated purposedly by organisms for some reason, with the aim of developing specific functions not known in most of the cases. Therefore, the main idea of the central dogma remains essentially true today, but it is necessary to include some modifications. Proteins production is not the only destination of the information contained in the DNA, but it can also be transformed into RNA with regulatory functions[67].

Two different classes are considered within ncRNAs depending on their function, the housekeeping and the regulatory ncRNAs. The first ones are involved in the maintenance of basic cellular functions and include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs). Also, ncRNAs are often arbitrarily divided in two main groups taking into account the length of the transcript: small non-coding RNAs (sncRNAs) with less than 200 nucleotides, and lncRNAs if they have more than 200 nucleotides[68] **(Fig. 5)**.

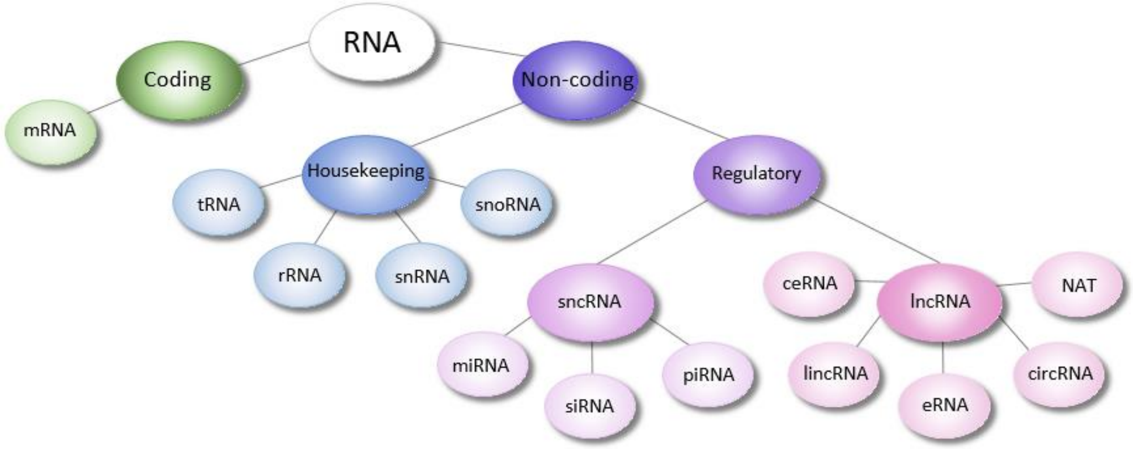


Figure 5. RNAs classification.
The transcribed RNA could be divided in two main groups depending on their capacity to give rise to a protein: coding RNA that produce messenger RNAs (mRNAs), and non-coding RNAs (ncRNAs). The ncRNAs can act as housekeeping or can play regulatory roles. Transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) are part of the housekeeping group of ncRNAs. At the same time, and according to the length, they can be considered small non-coding RNAs (sncRNAs), together with microRNAs (miRNAs), small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs). The group of long non-coding RNAs (lncRNAs) comprises competing endogenous RNAs (ceRNAs), long intergenic non-coding RNAs (lincRNAs), enhancer RNAs (eRNAs), circular RNAs (circRNAs) and natural antisense transcripts (NATs), among others.

2.3.1 Small non-coding RNAs

Several ncRNAs, such as small interfering RNAs (siRNAs) or PIWI-interacting RNAs (piRNAs), are part of this group, but I will focus on miRNAs **(Fig. 5)**. They represent the most extensively studied class to date because they play important roles in diverse biological processes; nevertheless, some of their functions might be still unknown.

2.3.1.1 MicroRNAs

At the beginning of their discovery, it was thought that small RNA fragments were remnants of degradation processes. The first ones for which a function was described were *lin-4* and *let-7*[69,70], but it was not until 2001 when the term miRNA was coined[71–73]. Both were found in *Caenorhabditis elegans*, linked to the regulation of development. Only some years later, more than 2,500 miRNA genes have been identified in humans[74].

The production of miRNAs is a multi-step process. It starts with the transcription of a long RNA precursor in the nucleus that folds in a double-stranded structure, and ends in the cytoplasm with a mature single-stranded form of around 20-24 nucleotides[72,75]. RNA polymerase II (Pol II) or III (Pol III) transcribe the primary miRNA (pri-miRNA)[76,77], normally from an independent transcription unit. This long transcript possesses the appropriate conditions to adopt a stem loop conformation[71]. The RNase III Drosha, together with DGCR8 protein and accessory factors, forms the Microprocessor complex that recognizes the hairpin and cleaves the pri-miRNA, giving rise to the precursor miRNA (pre-miRNA) (~60-70 nucleotides)[78]. These smaller molecules are exported to the cytoplasm by the transport protein Exportin-5[79], where they are cleaved again by another RNase type III called Dicer[80]. The imperfect RNA duplex produced is incorporated into the RNA-Induced Silencing Complex (RISC), where the active or guide strand of the miRNA is selected[81]. The majority of the mature miRNAs reside on the 3′arm of the hairpin[72]. The other arm constitutes the passenger strand, which is released or degraded by Ago proteins, the catalytic components of the complex[81]. Thus, the guide strand loaded into RISC represses mRNAs post-transcriptionally through the binding to miRNA recognition elements. These partially complementary sequences are generally located within the 3' untranslated region (UTR) of the target mRNA[82]. The repression mechanisms are to avoid translation or to mediate mRNA degradation[83] **(Fig. 6)**.

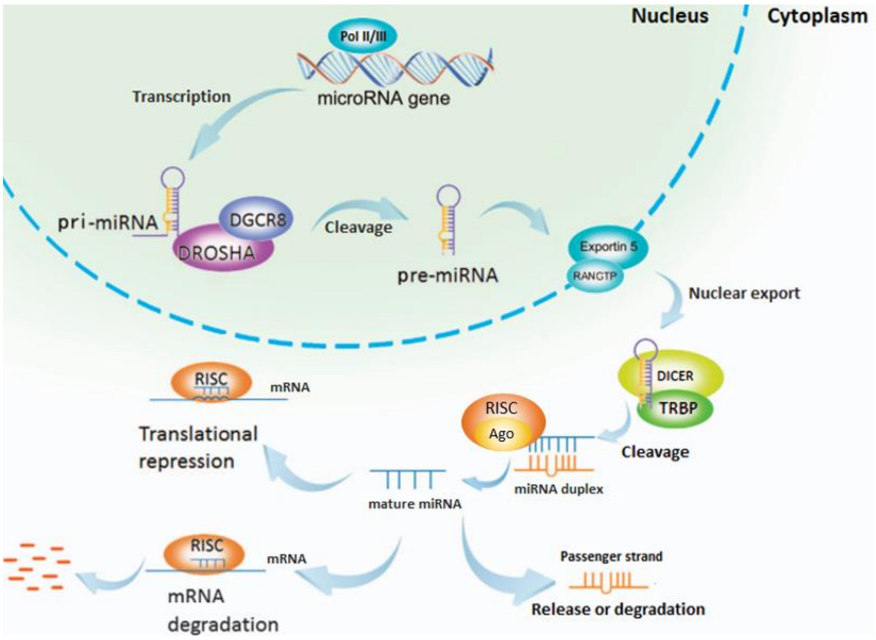


Figure 6. Biogenesis of miRNAs.
Pol II or Pol III mediate the transcription of pri-miRNAs, the primary precursors for miRNAs. These transcripts adopt a double-stranded structure that is recognized by the microprocessor complex. Drosha, the catalytic subunit of the complex, cleaves pri-miRNAs into a shorter stem loop called pre-miRNA. With the help of Exportin-5, pre-miRNAs are exported to the cytoplasm, where they are cut again by Dicer. The RNA duplex produced contains the guide and the passenger strand, and it is incorporated into RISC. Here, the guide strand is selected and the other arm is released or degraded by Ago proteins. The mature single-stranded form loaded into RISC represses mRNAs post-transcriptionally through translational inhibition or mRNA degradation (Modified from Peng & Croce, 2016[84]).

It seems that more than 60% of coding genes could be regulated by miRNAs[85]. Some of the miRNAs are able to target hundreds of genes, while others are much more specific. Likewise, the same target could be regulated by different miRNAs[70]. Thereby, they have an important role in all the biological processes, including development, differentiation, proliferation, apoptosis, regeneration and the immune response.

2.3.1.2 MicroRNAs dysregulation in cancer: let-7 family

The expression of many miRNAs is altered in numerous human cancers. Some of them are described as oncogenes because they are upregulated and target, directly or indirectly, anti-proliferative genes[86,87], while others act as tumor suppressors and their loss leads to tumorigenesis[88,89]. This is the case of *let-7,* the first miRNA detected in humans[90].

Let-7 family is conserved across many species and, in humans, it comprises ten mature sequences (*let-7a*, *let-7b*, *let-7c*, *let-7d*, *let-7e*, *let-7f*, *let-7g*, *let-7i*, *miR-98* and *miR-202*)[91] distributed in many different chromosomes. They share the same seed region (the binding area within the miRNA which normally lies between positions 2 and 8 from the 5′ end and determines mRNA recognition[92]), and because of that, they have a similar repertoire of targets. LIN28 is the main regulator of their maturation[93], and the two related *LIN28* genes that are present in humans act at different levels of regulation. LIN28A prevents the conversion of *pre-let-7* to the mature form by binding to the terminal loop of the precursor. Thus, Dicer processing cannot occur and *pre-let-7* is marked for degradation[94]. On the other hand, LIN28B prevents the Drosha-mediated cleavage of the *pri-let-7*[95] **(Fig. 7)**. According to this, the expression of LIN28 and *let-7* is negatively correlated, and it changes during development. High levels of LIN28 characterize early stages, whereas the expression is reduced in differentiated cells, allowing the increase of *let-7*[93]. A negative double feedback loop is established because *LIN28* transcripts are *let-7* targets.

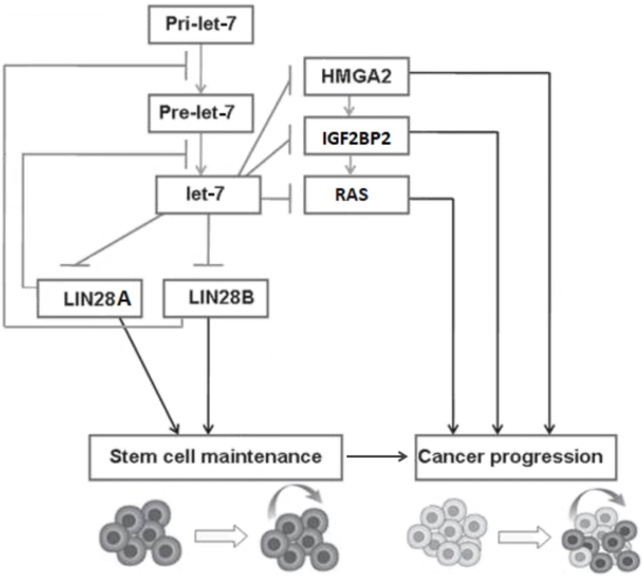


Figure 7. Regulatory pathways related with *let-7* miRNA family.
LIN28 proteins are the main regulators of *let-7* maturation. LIN28B avoids the conversion of *pri-let-7* into *pre-let-7*. LIN28A prevents the transition between the precursor *pre-let-7* and the mature form of the miRNA. At the same time, they are *let-7* targets, together with *HMGA2, IGF2BP2* and *RAS*. HMGA2 protein promotes the transcription of *IGF2BP2*, and IGF2BP2 protein regulates positively *RAS* translation. The expression of these three genes promotes cancer progression, and so does the expression of LIN28 family, because these proteins are associated with the maintenance of stem cell properties. The repression that *let-7* exerts over these oncogenes makes this miRNA an important tumor suppressor (Modified from Barh *et al*., 2010[96]).

Let-7 is highly expressed in most adult tissues because it is related with differentiation processes. There are many examples in which its dysregulation leads to diseases like cancer. *Let-7a* is downregulated in colon carcinomas and breast cancer[95], *let-7b* decreases tumor growth in multiple myeloma[97], *let-7g* suppresses tumorigenesis in non-small cell lung cancer[98], *let-7i* reduces migration, invasion and tumor progression in gastric cancer[99]. It is worth mentioning that some *let-7* targets are known oncogenes such as *RAS*[100], *HMGA2*[101–103] or the *IGF2BP* family[104] **(Fig. 7)**.

RAS was discovered as a *let-7* target because the expression levels of its protein were reduced by the transfection of *let-7* mimics, small molecules designed to reproduce the functions of endogenous miRNAs. This idea was also supported by the increase of *RAS* levels with *let-7* inhibition. *RAS* mRNA does not correlate with protein levels, suggesting that *let-7* regulation is at the level of translation[100].

On the other hand, *HMGA2* truncation by a chromosomal translocation is implicated in many tumors. Although sometimes the open reading frame is unaffected, the lack of the 3′UTR provokes the overexpression of the protein. There are eight putative binding sites for *let-7* in its 3′UTR and at least six are functional[102], so that miRNA cannot exert its function on the truncated form. Thus, only the full-length *HMGA2* increases with *let-7* inhibition and decreases with *let-7* mimics[101,102]. The effect is not only at the protein level because *let-7* promotes the degradation of *HMGA2* mRNA. The expression of *HMGA2* without 3′UTR or with mutated *let-7* sites increases cell growth[102], colony formation and the development of tumors[101].

Regarding the *IGF2BP* family, changes in its expression have been observed with Dicer deletion[104] and LIN28B overexpression[105]. The implication of *let-7* function in the altered expression of *IGF2BPs* has been confirmed through luciferase assays with mutant forms of the regulatory sequences. *IGF2BP2*, in particular, contains two *let-7* binding sites[105]. The dramatic upregulation of this family of proteins, also at the mRNA level, is related with an oncogenic transformation[104], and a higher risk of relapse[105].

2.3.1.3 MicroRNAs related therapies

Nowadays, miRNAs and the machinery involved in their biogenesis are in the focus of translational research as new therapeutic objectives. These molecules have some features as drug targets that make them interesting, such as the conservation across species, which facilitates the transition from preclinical to clinical trials. Moreover, each miRNA normally has many related targets, allowing an easier regulation of entire pathways[106]. At the same time, this is one of the main problems for miRNA based therapies, because the wide range of action may produce unexpected and undesirable effects[107]. Some pharmaceutical companies are working to solve this and other issues, paying special attention to bioavailability, stability and specificity.

There are two different types of strategies **(Fig. 8)**, those that try to avoid miRNA function and those that try to restore them. The first approaches are based on miRNAs sequestration through miRNA sponges or antisense oligonucleotides (ASOs)[85]. Both methods use molecules that bind to the miRNAs by complementarity and compete with the natural targets. There are different kinds of ASOs depending on their chemical modifications, such as locked nucleic acids (LNAs). Another option to inhibit miRNAs function consists in the protection of the binding sites of the target mRNA using molecules similar to miRNAs but non-functional. The binding of these miRNA masks avoids the association of the miRNA with a particular mRNA, limiting the number of affected targets and minimizing the off-target effects[108].

In the case of *let-7*, the therapies that restore miRNA function are the interesting ones. As previously mentioned, dysregulated methylation is a common cause of gene silencing in cancer, including miRNAs. Thus, demethylating agents that derepress expression provide effective treatments. Other possibilities to reestablish normal expression levels include adenoviral vectors that encode the downregulated miRNA, and the use of drugs that regulate the machinery involved in their biogenesis[85]. The enoxacin drug, that promotes miRNA processing[109], and some LIN28 inhibitors[110] have shown promising results. Mimics are the alternative to reexpression therapies[106]. They are double-stranded molecules identical to the miRNA of interest that reproduce miRNAs effects.

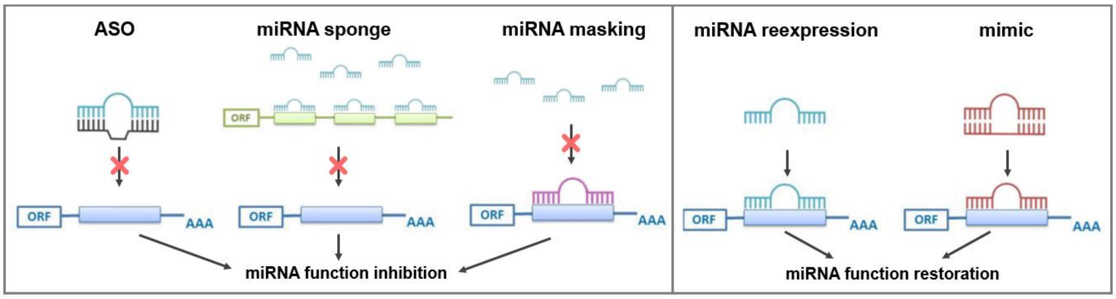


Figure 8. Therapeutic approaches based on miRNA function targeting.
Inhibition strategies of miRNAs functions comprise ASOs, miRNA sponges and miRNA masks. While ASOs and sponges fulfill their function by sequestering miRNAs, miRNA masks block the binding sites of the target mRNA (*left*). Restoration strategies of miRNAs functions are the reexpression of the downregulated genes and the use of mimics. The first ones include demethylating agents, adenoviral vectors and drugs that regulate miRNA biogenesis machinery (*right*) (Modified from Wojciechowska *et al*., 2017[111]).

Pharmaceutical industry is developing miRNAs-based strategies to treat not only cancer, but fibrosis, hepatitis C and cardiovascular diseases, among others. MiRagen Therapeutics (Boulder, Colorado, USA), Mirna Therapeutics (Austin, Texas, USA) and Regulus Therapeutics (San Diego, California, USA) have studies for many miRNAs in preclinical trials. However, very few strategies have reached clinical phases. The most advanced compound is Miravirsen, an ASO against *miR-122* developed by Santaris Pharma (Horsholm, Denmark). The drug, for the treatment of hepatitis C, is in clinical phase II. For instance, Regulus Therapeutics is also testing an ASO for *miR-122*, and Mirna Therapeutics is using a mimic for *miR-34* (MRX34) to treat liver cancer. This company is also working with a *let-7* mimic for cancer treatment[106,107].

2.3.2 Long non-coding RNAs

This group represents the largest portion of the mammalian non-coding transcriptome, and encompasses a heterogeneous set of transcripts much more diverse than sncRNAs. Their classification is still controversial because differences between them hinder the existence of a consensus. The range of sizes, locations, interaction partners and modes of action are very diverse. Consequently, there are different proposals to classify them and, normally, they can be placed in more than one category.

Based on their genomic location, lncRNAs can be intergenic (lincRNAs), when they are located between two protein coding genes; or genic if they share the same *locus* with a coding gene. The latter ones can be subclassified in intronic, exonic or overlapping genes. On the other hand, according to the relative orientation to coding genes, they are sense if they are located in the same strand, or antisense when they are transcribed from the opposite strand[66]. Finally, they act *in cis* if they regulate genes in close proximity to their origin of transcription, or *in trans* if they exert their functions over distant genes[112]. Other division criteria are the association with chromatin elements, the function or the structure. They generate different groups such as enhancer RNAs (eRNAs), competing endogenous RNAs (ceRNAs) and circular RNAs (circRNAs), respectively[113] **(Fig. 5)**.

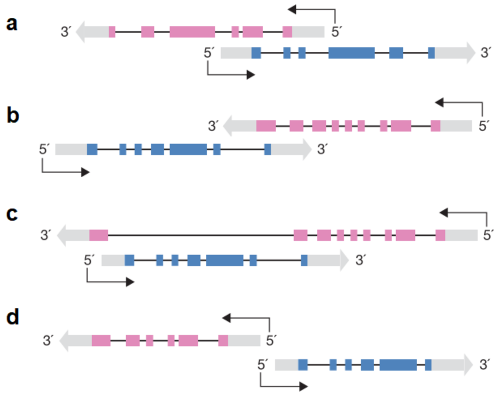
GENCODE v7 collects around 15,000 lncRNAs, a really close number to the 20,687 coding genes annotated[66,114]. On the other hand, a cancer-centric study, MiTranscriptome, analyzed >7,000 RNAseq libraries from human tumors, metastases and benign samples and identified close to 60,000 lncRNAs, with ~8,000 of them being characterized as cancer- and/or lineage-specific[115]. They share many aspects with mRNAs, but differ from them in others. Their transcription, carried out mainly by Pol II[116], gives rise to shorter transcripts that maintain the typical structure of any gene with a reduced number of exons and introns[117]. They can experiment a similar maturation process, with 5′capping, splicing and polyadenylation[60], even post-transcriptional modifications such as editing or methylation. However, there are some lncRNAs that are typically non-polyadenylated, among them some eRNAs, circRNAs and many NATs[118]. As a consequence, and in line with the dynamic functions they perform, many of these transcripts are less stable than mRNAs, although there is a wide range of half-lives[119]. LncRNAs are present both in the nucleus and cytoplasm but, in contrast to coding transcripts, they are predominately localized in the nucleus and chromatin[61,66]. In general, their expression is several orders of magnitude below coding genes and highly tissue-specific[66]. Unlike coding genes and other ncRNAs such as miRNAs, they are poorly conserved across species at the sequence level. In many cases, their secondary structure is the essential feature for the execution of their functions and is the more conserved characteristic[120].

The first lncRNAs identified in humans were *H19*[121] and *XIST*[122] around 1990, and since then, their biological relevance has kept growing with each new discovery. Nowadays, the impact of lncRNA-mediated regulations on normal physiology and development is well established. They are essential in key cellular processes, such as imprinting and X-chromosome inactivation; and they have been described in relation to different pathologies, including cancer. Many regulatory mechanisms have been proposed, most of them based on the induction of epigenetic changes to regulate the expression of proximal genes. Hereafter, I will explain the most important regulatory models described for NATs and pseudogenes, the most relevant lncRNAs for this thesis.

2.3.2.1 Natural antisense transcripts

Given their high abundance, NATs constitute a very important group within lncRNAs[66]. They are endogenous RNAs transcribed from the opposite strand of a sense strand-derived RNA at the same genomic *locus*[123]. The overlap between both transcripts can be partial or total[124] and, as a consequence, they present certain complementarity, forming sense/antisense (S/AS) pairs. Depending on the transcription direction and the overlapping degree, they are divided in three categories: divergent transcription or head-to-head, when they overlap by their 5′ region **(Fig. 9a)**; convergent transcription or tail-to-tail, when they overlap by their 3′ region **(Fig. 9b)** and fully-overlap, when one gene is included in the other[125] **(Fig. 9c)**. As a last possibility, a lincRNA whose transcription starts less than 1.5 kb from the transcription start site of a gene from the opposite strand, but with no overlapping, could be considered a NAT **(Fig. 9d).**

Figure 9. Classification of S/AS pairs according to their overlapping region.
(a) Head-to-head or divergent transcription, gene pairs with an overlap in the 5′UTR. **(b)** Tail-to-tail or convergent transcription, pairs with an overlap in the 3′UTR. **(c)** Fully-overlapping, pairs with one gene included completely within the other. **(d)** Non-overlapping, lincRNAs transcribed from the opposite strand of a sense gene without any overlapping region that can regulate the neighboring gene or a distant one. Blue and pink boxes denote coding exons and grey boxes UTRs (Modified from Lapidot & Pilpel, 2006[126]).



The first antisense transcription phenomenon was described in humans in 1989[127], and nowadays it is known as a much more prevalent mechanism than anticipated. The FANTOM-3 project has annotated NATs for more than 70% of the transcriptional units in mouse[125]. In humans, the numbers are between 20-40% depending on the study, but they could be even higher[128,129]. The pairs can be formed by two non-coding transcripts, two coding genes or a coding and a non-coding transcript, being this last option the predominant[125]. Normally, the sense gene is considered the coding transcript or the first one described of the pair.

Since NATs have sequence complementarity with their sense counterparts and overlap with their promoters and other regulatory regions, they are able to regulate their expression or modulate their post-transcriptional processing. They can interact either with DNA, RNA or proteins in order to affect every level of gene regulation. Many studies have demonstrated this regulatory relationship, showing correlated expression between many S/AS partners and coordinated changes in their levels. They normally exert their function *in cis*, but regulations *in trans* are also possible; and most of them have a repressive role, although they can promote expression as well. The importance of NATs-mediated regulation is explained by the relevance of their target genes, which are implicated, in many cases, in the development of different diseases. The most relevant NATs and their mechanisms of action are explained below but, for further information, there are extensive compilations of the described NATs in the bibliography[130,131].

Chromatin remodeling

The best known and possibly the most frequent mechanism used by NATs is the remodeling of chromatin structure through the recruitment of DNMTs and histone modifying complexes to the promoter region of the sense genes. Thus, they control the transcription of their partners, serving as a scaffold for these proteins. Even very low levels of NATs are enough to exert this function, because only two molecules are necessary to regulate a gene in this way[130].

LUC7L is a protein coding gene expressed by the opposite strand of the globin gene *HBA2*. A rare deletion puts both genes in close proximity and eliminates the termination site of *LUC7L*. Thus, its transcription runs into *HBA2 locus* promoting the methylation of the associated CpG island. This silencing is one of the possible causes of α-thalassemia, an inherited form of anemia[132].

Developmental *HOX* genes are tightly regulated by different lncRNAs that alter chromatin accessibility, mainly through histone modifications. The best-understood example is *HOTAIR*[133,134]. This lincRNA is encoded by the *HOXC locus* but represses the transcription of the *HOXD locus in trans*. The association with PRC2 redirects the repressive complex to the *HOXD locus,* inducing an increase in H3K27me3 **(Fig. 10a)**. Alterations in *HOTAIR* expression are involved in multitude of cancers, with negative implications in invasiveness and metastasis[133,135–137]. There are other NATs associated with *HOX loci* that have been studied, such as *HOXA11-AS*[138] and *HOTTIP*[139]. Unlike *HOTAIR*, these lincRNAs positively regulate gene transcription through the binding to WDR5, which is responsible for the deposit of the active mark H3K4me3.

Some NATs can act by influencing both DNA methylation and histone modifications. *ANRIL*, also known as *p15-AS* or *CDKN2B-AS1,* is upregulated in leukemia and prostate cancer; it acts by mediating the repression of three tumor suppressor genes transcribed from the *locus CDKN2B-CDKN2A.* Through the recruitment of CBX7 protein, a component of PRC1, and PRC2, it induces heterochromatin formation. *ANRIL* also leads to the hypermethylation of the *locus* after cellular differentiation[140,141] **(Fig. 10b)**. Another example is *lincRNA-p21,* antisense to *p21* gene. This NAT forms epigenetic modifier complexes with the collaboration of hnRNP K, and coordinates histone and DNA methylation, resulting in the reduction of *p21* and other pluripotency genes expression[142–144]. An opposite situation is described by Dimitrova and colleagues, where *lincRNA-p21,* together with hnRNP K, promotes the transcription of *p21*[145]. A mechanism of action not related with changes in chromatin conformation is the established between *lincRNA-p21* and the genes CTNNB1 and JUNB. The association to these mRNAs shifts the presence of these transcripts to lighter polysomes, reducing their translation[146].

These regulatory mechanisms are important to control the monoallelic expression of specific genes, clusters or even to inactivate a whole chromosome in mammalian females to compensate the dosage of X-linked genes. This process is regulated by the lncRNA *Xist* and its antisense *Tsix*. For the X-chromosome inactivation, *Xist* recruits histone modifying complexes that induce the mark H3K27me3[147]. *Tsix* is expressed on the other X chromosome and represses *Xist* through CpG methylation and histone modifications, maintaining the transcriptionally active state of the chromosome[148] **(Fig. 10c)**. The expression of imprinted genes, in which only one allele is active depending on the parental origin, is also mediated through the recruitment of chromatin modifiers with NATs participation. *Airn* is a lncRNA expressed from the paternal allele whose antisense transcription across the promoter of the *Igf2r* gene suppresses its expression, together with the expression of the neighboring genes *Slc22a2* and *Slc22a3*. Although *Airn* silences these nearby genes through the usual mechanisms implicated in imprinting, *Igf2r* repression is more related with the antisense transcription itself[149] **(Fig. 10d)**. The involved process is called transcriptional interference, and it will be explained in detail later. Moreover, this lncRNA interacts with the protein IGF2BP2[150], but the role that exerts in this case will be described below. Finally, the production of T-cell receptors and immunoglobulins in lymphocytes also requires monoallelic expression to ensure the presence of a unique antigen receptor in each cell. A recombination process in one of the alleles is indispensable to generate the high variability of immunoglobulins. Antisense transcription favors the recombination through the decondensation of the chromatin. Meanwhile, the other allele is silenced, taking place the allelic exclusion[151].

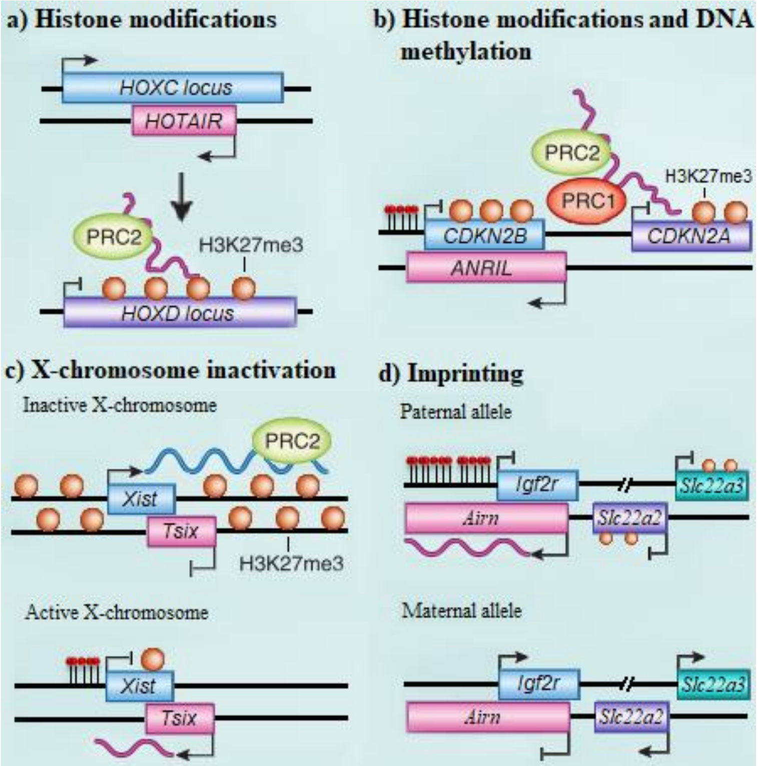


Figure 10. Mechanism used by NATs to regulate gene expression at the transcription level.
(a) *HOTAIR* inhibits the transcription of the *HOXD locus in trans* through the association with PRC2. **(b)** *ANRIL* represses the genes transcribed from the *locus CDKN2B-CDKN2A* combining histone modifications and DNA methylation. **(c)** *Xist* and its antisense *Tsix* regulate X-chromosome inactivation through the recruitment of histone modifying complexes. *Tsix* maintains active the other X chromosome repressing *Xist* through CpG methylation and histone modifications. **(d)** *Airn* is expressed from the paternal allele and regulates the transcription of three imprinted genes. Its antisense transcription leads to *Igf2r* promoter methylation, while the repression of *Slc22a2* and *Slc22a3* genes is mediated through histone modifications. *Airn* is silenced in the maternal allele.

RNA masking/competition

Some NATs can interact by complementarity with the mRNA of their sense gene. The formation of double-stranded RNAs (dsRNAs) could induce genome editing[152] or RNA interference (RNAi) processes[153], but they do not seem very common mechanisms. The duplex formation may mask regulatory segments essential for the binding of other factors.

They are also able to compete with the sense transcript for the binding with other molecules implicated in post-transcriptional regulation. Both mechanisms allow to regulate processes such as splicing, translation, transport, stability and degradation.

Alternative splicing is altered in the thyroid hormone receptor *erbAα* by its NAT *RevErb*. The complementarity between both transcripts leads to a base pairing interaction that affects only the splice sites of one of the *erbAα* isoforms. Hence, the balance between splice variants is controlled by *RevErb*[154]. On the other hand, the translation of *Zeb2* depends on the retention of an intron that contains an internal ribosome entry site (IRES). During epithelial-mesenchymal transition, *Zeb2 NAT* is expressed and avoids the processing of this intron through a binding by complementarity that protects the splice sites. In this situation *Zeb2* is translated correctly[155] **(Fig. 11b)**.

Translation could also be compromised through the competition for the translational machinery. *PU.1* is a transcription factor important in hematopoiesis and closely related to the development of leukemias and lymphomas. Its antisense binds to the translation factor eEF1A, decreasing *PU.1* expression as a result of the impairment of the elongation process during translation[156]. Another example is the antisense of *Uchl1,* a gene involved in neurodegenerative diseases. This NAT controls the association with polysomes of *Uchl1* mRNA to increase its translation. The ability to regulate Uchl1 protein level depends on the binding to the 5′ region of the sense transcript. The presence of a SINEB2 sequence contained within the *Uchl1* antisense transcript is also essential[157] **(Fig. 11a)**.

BACE1 is an important enzyme in neurodegenerative diseases closely linked to Alzheimer. It generates the amyloid-β peptides that aggregate in the neurons leading to their degeneration. The antisense *BACE1-AS* stabilizes *BACE1* because it competes with *miR-485-5p* for the binding sites[158]. Other NATs avoid miRNAs function by direct interaction with them, instead of binding to the mRNAs. They act as ceRNAs, transcripts with multiple recognition elements for specific miRNAs that compete with the target mRNAs for their binding, thereby liberating them from repression. *TUG1* is an example of miRNA sponge for *miR-299* and *miR-34a-5p. VEGFA*, a protein with angiogenic properties, is a known target of these miRNAs. Thus, *TUG1* increases *VEGFA* expression and favors tumor growth in endometrial cancer[159] **(Fig. 11c)**.

NATs can also affect the stability of other transcripts guiding them to degradation or protecting them. The binding of the antisense *HIF1-AS2* to *HIF-1α*, a transcription factor induced by hypoxia, leads to changes in *HIF-1α* mRNA conformation. The structural modification, instead of masking a region, leaves exposed certain AU-rich elements. The exposition of these sequences results in RNA degradation[160]. This lncRNA may also play a role interacting with the RNA-binding protein (RBP) IGF2BP2[161], but the mechanism will be further explained below. In certain cases, the binding between sense and antisense RNAs could serve to stabilize the transcript. The antisense *Wrap53* regulates mRNA and protein levels of the tumor suppressor *p53* through the interaction with a 5′UTR sequence that is determinant for the stability of this mRNA[162].

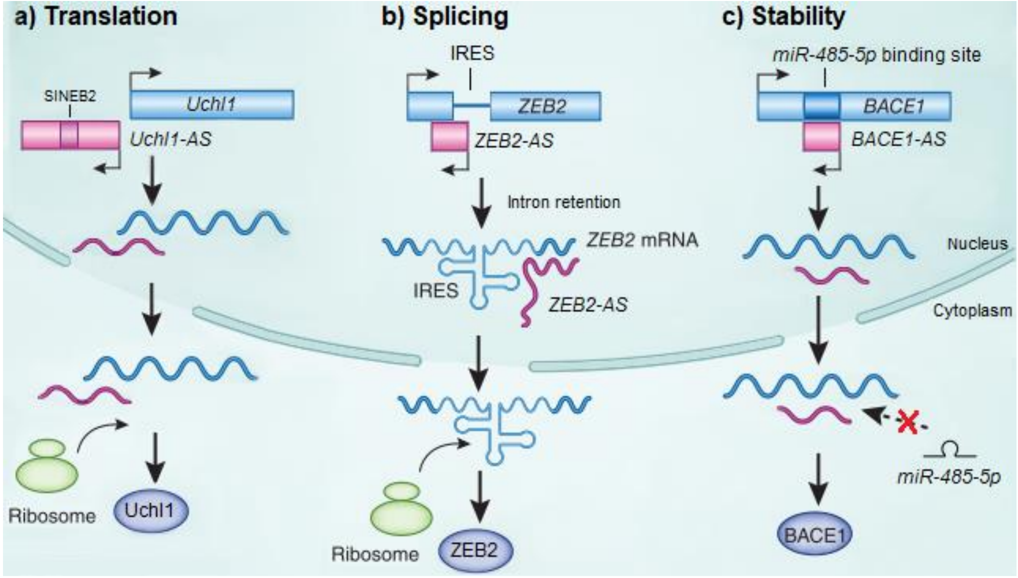


Figure 11. Mechanism used by NATs to regulate gene expression at the post-transcriptional level. **(a)** *Uchl1-AS* controls the translation of *Uchl1*. The association with polysomes of this mRNA depends on the presence of a SINEB2 sequence within the *Uchl1-AS* transcript and its binding to the 5′ region of *Uchl1* mRNA. **(b)** *Zeb2 NAT* regulates the alternative splicing of *Zeb2*. It binds to the splice site of an intron that contains an IRES essential for *Zeb2* translation, and promotes its retention. **(c)** *BACE1-AS* stabilizes *BACE1* competing with *miR-485-5p* for the binding sites (Modified from Guil & Esteller, 2012[124]).

Transcriptional interference

Antisense transcription could be a regulatory mechanism itself, independently of the produced transcript. A transcription process could influence a second transcriptional activity *in cis*. This phenomenon is known as transcriptional interference and takes place by (1) competition: the transcription machinery cannot bind to two promoters in overlapping regions at the same time; (2)"sitting duck interference": a faster polymerase is able to displace another one; (3) occlusion: temporary blockage of the access to a promoter by a polymerase that is already transcribing a gene; (4) collision: two elongating polymerases meet during transcription and block each other[163]. Transcriptional interference mostly suggests an inverse correlation between S/AS expression, although they could be both shutdown through the collision mechanism or expressed if each transcription occurs at different times[126]. It is broadly observed in bacteria[164] or yeasts[165], and one example in mouse is *Airn*[149], as mentioned previously.

2.3.2.2 Pseudogenes as a novel class of long non-coding RNAs

Pseudogenes are genetic elements originating from a parental gene by retrotransposition (processed pseudogenes), duplication (unprocessed pseudogenes) or mutation events (unitary pseudogenes). Normally, reverse transcription and duplication processes are also accompanied by the acquisition of inactivating mutations. Thus, pseudogenes lose the protein coding capacity of the parental gene, reason why they were considered as “junk DNA”[166]. Nevertheless, nowadays, it is demonstrated that many of them are not only transcribed but also translated[167], giving rise to parental derived proteins. In general, there is a single pseudogene for each parental gene, but some housekeepings may have more than 100[168]. The existence of many pseudogenes for a particular parental gene, their similarities and their poor conservation among species, hinder their study. Further research is needed because functions have been described only for a few of them, but they can play roles typically associated to lncRNAs such as miRNA sponging or chromatin remodeling. Thereby, they can contribute to the regulation of parental genes or other genes, having important implications in diseases like cancer[166].

According to GENCODE, there are more than 14,000 pseudogenes in the human genome, and more than 1,000 are transcriptionally active[168]. Different expression patterns have been described for them. The majority of the pseudogenes present a nonspecific expression with an extended presence but high levels only in very few tissues; certain pseudogenes are detected ubiquitously, normally the ones related to housekeeping genes; and only some of them are really tissue-specific. On the other hand, a number of pseudogenes enable the differentiation between normal and cancer samples, and some of them are shared between multitude of cancers[169]. The high retrotransposon activity during carcinogenesis could be related with the formation of new pseudogenes, and it has been reported that sometimes they integrate within coding genes affecting their expression[170]. This not only supports the idea that they have a relevant participation in carcinogenesis, but gives them an important value as biomarkers for cancer diagnosis and prognosis.

Moreover, their participation in cancer development could be related with the fact that many pseudogenes derived from coding genes with relevant functions in tumorigenesis.

RNAs and proteins encoded by a pseudogene can regulate their parental gene competing for the binding of miRNAs and RBPs, among other functions. This regulatory role can spread to other unrelated genes that share the same or similar binding domains. When the sequence is very similar to the parental gene, they can perform the same functions, but processed pseudogenes lose introns and regulatory elements such as enhancers and promoters, which usually implies changes in the expression pattern[171].

One of the best known pseudogenes is *PTENP1*, whose *locus* encodes a sense transcript and two antisense isoforms (*α* and *β*)[172]. The sense transcript increases the levels of the tumor suppressor *PTEN* acting as a ceRNA because they share binding sites for the same miRNAs (*miR-17*, *miR-21*, *miR-214*, *miR-19*, and *miR-26* families)[173] **(Fig. 12a)**. This regulation depends on *PTENP1 β*, because it is the responsible for the stability and cytoplasmic localization of the sense transcript by directly interacting with it **(Fig. 12e)**. On the other hand, *PTENP1 α* has an opposite role silencing *PTEN* through the recruitment of DNMT3a and EZH2 to its promoter[172] **(Fig. 12b)**.

The transcription factor *OCT4* has six pseudogenes[174], some of which are transcribed. *OCT4pg4* competes with the parental gene for the binding of *miRNA-145* **(Fig. 12a)**. Thus, it promotes *OCT4* expression, which leads to poor prognosis in hepatocellular carcinoma[175]. On the contrary, the murine *Oct4pg4* silences the parental gene favoring the deposition of the epigenetic mark H3K9me3 in its promoter through the formation of a complex with the histone methyltransferase SUV39H1, and then it recruits the silencer protein HP1α[176]. The antisense transcribed from the *OCT4pg5 locus* also serves as a scaffold for chromatin modifying complexes such as EZH2. The mark H3K27me3 mediated by EZH2 represses *OCT4*, *OCT4pg4* and *OCT4pg5*[177] **(Fig. 12b)**.

Another gene with this kind of regulation is *HMGA1*. From its seven pseudogenes, *HMGA1p6* and *HMGA1p7* function as ceRNAs not only for the parental gene but also for other genes such as *H19, IGF2* and *EGR1*[178,179] **(Fig. 12a)**. Moreover, *H19* and *IGF2* are the precursors of *miR-675* and *miR-483*, respectively, and the transcription factor *EGR1* activates their expression[179]. *HMGA1p7* adds another level of complexity in the regulation of this genetic network. This pseudogene may also act as a decoy for the protein αCP1, compromising *HMGA1* mRNA stability[180] **(Fig. 12d)**. Thus, *HMGA1* pseudogenes play key roles in cancer progression, showing anti-apoptotic properties and increasing proliferation, cell migration and invasion[181].

A lot of examples corroborate the miRNA sponging activity of pseudogenes and its implications in cancer. However, there are less extended functions that can also be performed by pseudogenes. For instance, *AOC4P* promotes Vimentin degradation by ubiquitination of the protein[182] and *Rps15a-ps4* binds to the transcription factor NF-κB, avoiding the interaction with its targets[183] **(Fig. 12c)**. Finally, antisense RNAs of some pseudogenes can adopt hairpin structures or form dsRNAs with the sense pseudogene or with the parental gene. These molecules can be processed by Dicer to produce endogenous siRNAs that are able to degrade targeted mRNAs[184] **(Fig. 12f)**.

Among the pseudogenes that can be translated, it is worth highlighting the 11 pseudogenes of *NANOG*. Some of them have premature stop codons and give rise to truncated proteins, but *NANOGP7* and *NANOGP8* code for proteins almost identical to NANOG that participate in carcinogenesis[185]. Finally, BRAFP1 peptide is able to interact with BRAF and activate oncogenic pathways such as MAPK in thyroid tumors[186]. Its mRNA can also act as ceRNA for *BRAF* competing for the binding of *miR-30a*, *miR-182* and *miR-876* in lymphoma[187].

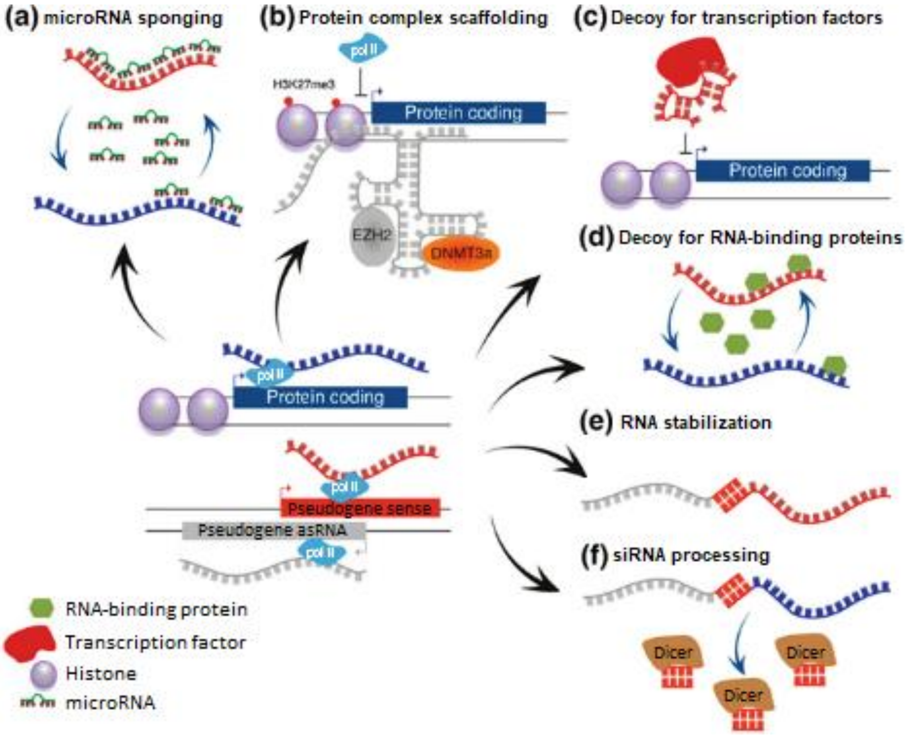


Figure 12. Regulatory mechanisms mediated by pseudogenes.
(a) Pseudogenes can act as ceRNAs and derepress mRNAs that are targeted by miRNAs. **(b)** They serve as a binding platform for chromatin remodeling factors such as DNMT3a and EZH2. Normally, the effect is the repression of the regulated gene. **(c)** They can act as a decoy for certain transcription factors, avoiding its action. **(d)** They compete with the parental gene for the binding to RBPs. **(e)** Antisense pseudogenes can establish RNA:RNA interactions with the sense transcript or with the parental gene, affecting their stability. **(f)** The formation of RNA duplex by the interaction between the antisense pseudogene and the related sense RNAs could trigger Dicer processing and the generation of siRNAs (Modified from Grandér & Johnsson, 2016[166]).

2.3.2.3 Long non-codi ng RNAs related therapies

Few clinical trials are focused on the medical applications of lncRNAs, and the majority of them try to evaluate their use in diagnosis and prognosis as non-invasive biomarkers. Nevertheless, they are considered promising targets for new therapeutic approaches, because some of their features may confer advantages for the development of more specific and effective treatments.

As previously described, lncRNAs are dysregulated in many pathological situations, and they can be transported by extracellular vesicles in body fluids avoiding RNases degradation through the association with RBPs. Thus, these circulating lncRNAs can be measured in saliva, urine or blood and they are useful as indicators of the presence or severity of some diseases[188].

Regarding to their benefits as therapeutic targets, the lower expression of lncRNAs makes it easier to produce changes in their levels with a reduced drug dosage[67], with the consequent decrease in toxicity. Moreover, their tissue-specific expression and their effects normally restricted to a reduced number of genes ensure a much more controlled impact of these therapies. Indeed, the fact that some NATs regulate only their sense gene allows to modify the expression of the gene of interest with less off-target effects, unlike the case of miRNAs-based therapies. Thus, it is possible to get more targeted treatments with less side effects. Furthermore, drugs targeting proteins are more difficult to obtain than RNA complementary oligonucleotides[107,189].

The main strategies focus on the downregulation of lncRNAs with duplex RNAs such as siRNAs and short-hairpin RNAs (shRNAs) **(Fig. 13a)** or ASOs **(Fig. 13b).** Both approaches are based on the cleavage of the targeted RNA, but the second one works better to deplete nuclear RNAs. Besides, its single-stranded structure and its independence on the RNAi processing pathway minimize the toxicity and the off-target effects[190]. ASOs against *Malat1* have probed their value reducing metastasis in mouse models of breast and lung cancer[191,192]. Moreover, NATs inhibition allows the indirect regulation of associated tumor suppressors, oncogenes or other proteins of interest. ASOs that target specifically NATs are known as antagoNATs, and they have been used, for example, to increase *BDNF* expression through the inhibition of its NAT *BDNF-AS*[193]. On the other hand, ASOs are also useful to control splicing. This could modify the functional domains or RNA folding, leading to changes in the binding to its targets[190,194].

CRISPR (clustered regularly interspaced short palindromic repeat)-derived techniques can also be used to repress lncRNAs at both transcriptional and post-transcriptional level. CRISPR-Cas9 is an immunity mechanism of prokaryotic organisms used today for gene editing. CRISPR-interference is based on an inactive Cas9 that can be fused or not to a transcriptional repressor, and that is able to bind to a specific DNA sequence blocking transcription initiation or elongation[195]. On the other hand, CRISPR-Cas13 system targets and cleaves RNA[196] **(Fig. 13c)**.

Ribozymes or Dnazymes are RNA or DNA molecules that display catalytic activity. For instance, hammerhead ribozyme can bind by complementary to an RNA target sequence and catalyze the cleavage of the transcript downstream to a specific site[190] **(Fig. 13d)**.

A strategy not based on RNA cleavage is the steric inhibition by small molecules such as aptamers or morpholinos that could disrupt lncRNAs structure or block the interaction with downstream targets. Aptamers are short oligonucleotides or peptides with a stable tertiary structure **(Fig. 13e)** and morpholinos are ASOs that cannot recruit RNase H[190,194] **(Fig. 13f)**. The advantage of using aptamers is that they bind to their targets with high specificity regardless of the sequence and dependent on the secondary structure[190]. *MALAT1* and *NEAT1* are two examples of lncRNAs with three dimensional structures that can be targeted by these kind of molecules[194].

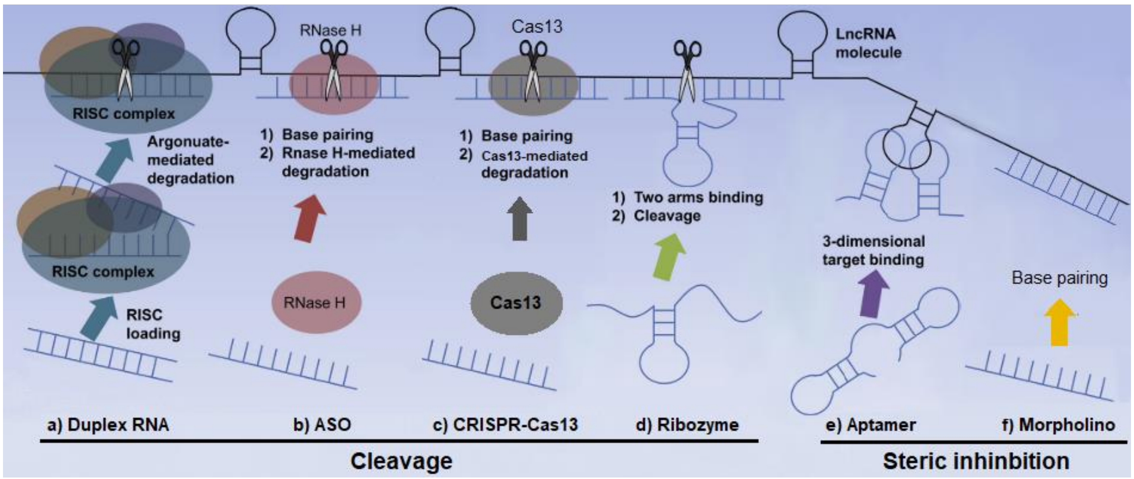


Figure 13. Therapeutic approaches based on the disruption of lncRNAs function.
There are two types of strategies to interrupt lncRNAs function, the ones based on the cleavage of the transcript (*left*) and the ones based on steric inhibition (*right*). **(a)** siRNAs and shRNAs are double-strand molecules that cleave the target RNA through their incorporation in the RNAi processing pathway. **(b)** ASOs are single-stranded structures that bind by complementarity to the target RNA and recruit RNase H for its cleavage. **(c)** CRISPR-Cas13 system contains the RNA-guided ribonuclease Cas13 that cleaves RNA. **(d)** Ribozymes are RNA molecules with catalytic activity that recognize by complementary the RNA sequence of interest and catalyze its cleavage close to a specific site. **(e)** Aptamers are short oligonucleotides or peptides with a stable tertiary structure that bind to their targets dependent on the structure, disrupting the conformation or blocking the interaction with other targets. **(f)** Morpholinos are ASOs that cannot recruit RNase H but could affect lncRNAs structure and interactions (Modified from Li & Chen, 2013[190]).

In some cases it could be interesting to rescue the function of a lncRNA with the use of mimics. On the other hand, the mutated version of a mimic may be useful to compete with an endogenous lncRNA when it plays an oncogenic role, blocking its function. Nevertheless, the size of lncRNA mimics presents a disadvantage for their usage[107].

Some of the unsolved problems of these therapeutic strategies are the poor delivery, the instability of the molecules and the presence of secondary or tertiary structures in RNA targets. To favor the uptake, the size of the molecules has to be reduced, but this can decrease the specificity because even partial complementarity can lead to off-target effects[189]. Bioinformatics tools are used to design and select the most specific drugs. Moreover, chemical modifications are added to the molecules to improve the delivery and avoid degradation[107]. Some examples are the addition of a methyl group to the 2′-OH of the ribose (2′-O-methyl ASOs) or a methylene bridge connecting the 2′ oxygen and the 4′ carbon of the sugar (LNA ASOs)[197]. The most promising molecules are LNA Gapmers, which cleave the target RNA through RNase H action. On the other hand, the transition from the discovery of new transcripts involved in cancer to pre-clinical trials is difficult due to the absence of conservation of the majority of human lncRNAs in mouse[194].

To date, there have only been three ASOs approved by the Food and Drug Administration and a few are in pre-clinical or clinical phases, mainly against coding genes. One example is the drug Nusinersen, prescribed for the treatment of spinal muscular atrophy, that rescues SMN protein levels through the regulation of *SMN2* gene splicing[198]. Companies such as OPKO-CURNA (Miami, Florida, USA) and RaNA Therapeutics (Cambridge, Massachusetts, USA) are working on the development of therapeutic approaches using antagoNATs[189,194]. Meanwhile, a therapy based on lncRNAs properties is in advanced phases of clinical trials in several cancers. *H19* is expressed in 85% of bladder tumors, and this specific presence in cancer cells has allowed to use its promoter to guide the expression of a lethal toxin directly to malignant cells. Specifically, the drug BC-819 consists in a plasmid that expresses diphtheria toxin A under *H19* promoter[199].

2.4 Another layer of regulation: R-loops

R-loops were first characterized in 1976[200], but it took almost 20 years to demonstrate the existence of these structures *in vivo*, in bacteria[201]. It is defined as a stable RNA:DNA hybrid in which a nascent RNA is paired by complementarity with the template strand of the DNA duplex, leaving the opposite chain as a single-strand[200,202] **(Fig. 14)**.

Short hybrids are part of normal processes such as replication or transcription, but it was thought that longer associations that give rise to R-loops were a rare event, only related with bacterial and mitochondrial DNA replication[203,204] and immunoglobulin class-switch recombination[205]. Nevertheless, nowadays, there are around 250,000 predicted places in the genome that are able of forming them, comprising 59% of known genes[206].

Normally, R-loops are formed cotranscriptionally *in cis,* although in yeasts their formation has been found *in trans*[207]. An example of these R-loops *in trans* is observed in the CRISPR-Cas9 system.

There are two different mechanisms to explain how R-loops can be formed: the "extended RNA:DNA hybrid" model and the "thread-back" model[208]. In the first one, the short hybrid formed during transcription remains annealed and it is elongated while polymerase is acting. The second model holds that nascent RNA exits the polymerase and stays as single strand before invading the DNA duplex, hybridizing again with the template strand of the DNA. This model is reinforced by the presence of independent exit channels for RNA and DNA in the polymerase[209]. Whether this invasion can be a spontaneous event is not clear, but several studies have shown the relationship between R-loop formation and proteins such as RecA in bacteria[210], the eukaryotic homolog Rad51 in *Saccharomyces cerevisiae* and human Rad52[207].

They preferentially appear at the promoters[33] and termination sites of genes[211,212] under certain circumstances. One of the features that promotes their formation is the existence of a significant asymmetry in the distribution of GCs in each DNA strand, a property known as GC skew. Hence, a G-rich RNA can be transcribed and can bind to the C-rich strand of the DNA[33,211]. A high GC content is important for stabilization and elongation but it is not sufficient for R-loop initiation if Gs are dispersed. Thus, G clusters in the transcript are necessary[208]. The stability depends on the length and on the base content[213] but, in general, RNA:DNA interactions are more stable than DNA:DNA[214]. This may be due to the conformation they adopt, an intermediate between dsRNA and DNA duplexes[213]; or to the G-quadruplexes that can be formed on the displaced strand of the DNA[215] **(Fig. 14)**. Moreover, it exists the possibility that some factors bind to the region to stabilize the structure, such as AtNDX, a protein described in *Arabidopsis* that associates with single-stranded DNA[216]. An additional characteristic that favors R-loops is the presence of DNA negative supercoiling that, together with GC enrichment, enables the opening of the DNA duplex. Finally, DNA nicks, which are single-strand breaks in the DNA molecule owing to the absence of the phosphodiester bond between adjacent nucleotides of one strand[217], also facilitate the intertwining of RNA with DNA[218].

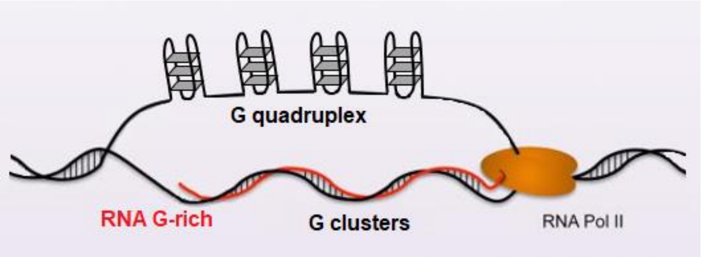


Figure 14. R-loops formation. A G-rich RNA (red) with G clusters in the sequence is transcribed by the RNA Pol II and binds by complementarity with the C-rich strand of the DNA. G-quadruplexes present on the displaced strand of the DNA favor the stability of the structure (Modified from Allison & Wang, 2019[219]).

Since R-loops can have deleterious consequences, there are protection mechanisms against them. Some prevent their apparition, and others disrupt them once formed. Topoisomerases undo the negative supercoiling that favors R-loop formation[201,220]. RBPs implicated in RNA processing or in the formation of ribonucleoproteins (RNPs) reduce the possibility of hybridization with the DNA when they are bound to the RNA[221]. The inhibitors of proteins that favor R-loop formation can also be included in the group of mechanisms that prevent them. An example is Srs2, which acts as an antagonist of Rad51[207]. On the other hand, helicases such as DHX9[222] and SETX disrupt the R-loops in humans, and also RNase H[212], which degrades the RNA of RNA:DNA hybrids.

2.4.1 R-loops functions and human diseases

R-loops play important physiological roles in cells; however, there is a very thin line between the normal and the deleterious effects.

In humans, they participate in the regulation of transcription preventing DNA methylation of CGIs at promoter regions. It is possible that they are not appropriate substrates for DNMTs or, maybe, they contribute to recruit the protective mark H3K4me3 or DNA demethylating complexes that binds to single DNA strands[33,211]. Moreover, their presence at the 3′ end of many genes is related with transcription termination. It seems that R-loops contain functional elements involved in this process but, at the same time, their subsequent resolution is necessary for an efficient termination[212]. They also can play key roles in alternative splicing[206] and in the maintenance of telomeres[223].

On the other hand, R-loops can block the progression of the replication fork[206] and transcription elongation[220,224]. This can lead to genomic instability[202] because single-stranded DNA is susceptible to DNA damage. Therefore, R-loops can be a source of genetic abnormalities linked to pathological conditions. Genetic instability is associated with the expansion of trinucleotide repeats, responsible of disorders like Huntington's disease[225], and R-loops have been found in more than 200 cancer related genes such as *p53*, *BRCA1/BRCA2*, *Myc* and *Kras*[206]. Moreover, it has been shown that some of these genes regulate R-loops presence. The tumor suppressor BRCA2 is recruited to RNA:DNA hybrids and prevents their accumulation[226]. By contrast, EWS-FLI1, the chimeric protein expressed in Ewing's sarcoma, induces their formation[227].

In many cases, R-loops functions involve ncRNAs such as NATs. In *Arabidopsis thaliana*, the expression of *FLC* gene is regulated by the presence of an R-loop in the promoter region of its antisense transcript, named *COOLAIR*. The protein AtNDX inhibits *COOLAIR* expression through the stabilization of the R-loop, and thus, promotes *FLC* expression[216]. The imprinting of the *Ube3a* gene is also determined by the interaction between R-loops and NATs. The transcription of the antisense *Ube3a-ATS*, favored by the formation of an R-loop, silences the paternal *Ube3a* gene. The mutation or deletion of the maternal copy causes Angelman syndrome, and a promising treatment is the use of topotecan. This topoisomerase inhibitor increases R-loop formation, an excess of which terminates transcription and prevents *Ube3a-ATS* expression, with the consequent reactivation of *Ube3a* gene[228].

3. HMGA2-IGF2BP2 pathway

3.1 *HMGA2* gene

HMGA2 gene encodes a small protein with the same name, which is part of the HMGA family together with HMGA1a, HMGA1b and HMGA1c. These non-histone nuclear proteins bind to the minor grooves of DNA helix, specifically to AT-rich sequences, through the three binding motifs known as AT-hooks that they have at their amino terminal end[229]. It is believed that the C-terminal tail is important to modulate protein-protein interactions[230]. HMGA proteins do not have transcriptional activity, but they can alter the structure of the chromatin, facilitating the access of the transcriptional machinery. They are also able to interact directly with transcription factors, change their conformation and increase its affinity to DNA, or even help assemble multiprotein complexes[231]. Thereby, their functions have an important impact in the expression of many genes. Apart from the well-known regulation of *IGF2BP2*, many NF-κB-targets have been identified as genes regulated by HMGA2[232]. Moreover, this protein activates *SNAIL* and *Twist1* transcription in pancreatic, gastric and hepatic cancer[233–235]. Thus, the expression of some epithelial and mesenchymal markers is influenced by *HMGA2* action. In leukemia, it activates Wnt/β-catenin pathway, increases the levels of the antiapoptotic protein Bcl-2[236], and blocks the expression of tumor suppressor *CDKN2A*[237].

The structural similarities and the common mechanism of action derives in an overlap of the targets, but each protein have their own specific functions. In this work, we focus on human HMGA2, whose corresponding gene is located on chromosome 12. It contains 5 exons, and there is a transcript with antisense orientation annotated in the same *locus* in RefSeq, corresponding to the non-coding pseudogene *RPSAP52.* Essential roles have been described for many NATs regarding the expression of their sense gene. Since there is no functional characterization of *RPSAP52* so far, an important part of this thesis will be to unravel the significance of its transcription and its possible effects on *HMGA2* gene.

To date, several mechanisms that regulate *HMGA2* expression have been described. At the transcriptional level, it seems that TGF-β/Smad[238] and Wnt/β-catenin[239] signaling pathways induce its expression; and RUNX1 acts as a negative regulator in human and mouse[240]. Post-transcriptionally, the mRNA of *HMGA2* is affected by miRNA-mediated regulation, through the binding sites present in its 3′UTR. Different studies have shown that it is targeted by *miR-185*[241], *miR-33b*[242] and *miR-93*[243], among others. The overexpression of these miRNAs suppress proliferation and metastasis in breast cancer. Another example is *miR-196a-2*, whose levels are positively regulated by HMGA1. Thus, a regulatory network is established between HMGA family members[244]. Finally, as mentioned before, one of the main miRNAs that regulates *HMGA2* is *let-7*[101,102]. This layer of regulation can be lost in many tumors as a consequence of *HMGA2* truncation and the lack of most of its 3′UTR. At the same time, molecules that regulate the expression of these activators and repressors can affect HMGA2 levels.

The human HMGA2 protein appears in large quantities during development and it is absent in adult tissues[245]. In accordance with its important role in transcriptional regulation, its aberrant reexpression has been observed in a large number of human cancers, including breast[246], lung[247], colorectal[248], pancreatic[249] and leukemia[250], among other types. The oncogenic potential of this protein has been clearly established, and there are many studies that show a positive correlation between its expression and tumor aggressiveness, as well as with the reduction of the survival rate. Moreover, *HMGA2* mRNA could be useful as a prognosis biomarker because it is present in the blood of breast cancer patients, and this is related with a poor outcome[251].

3.2 *IGF2BP2* gene

One identified target of HMGA2 protein is *IGF2BP2*[252–254]. Also known as *VICKZ2* and *IMP2*, this gene is localized on chromosome 3 and has 16 exons, of which the tenth can be skipped, giving rise to a splice variant[255] **(Fig. 15)**. It encodes an RBP which is a member of the IGF2BP family, together with IGF2BP1 and IGF2BP3. Although encoded on different chromosomes, they are highly similar, with an amino acid sequence identity of 56%. IGF2BP2 is the most different protein of the family, and the least understood.

All of them contain a unique combination of six characteristic RNA-binding modules, two RNA recognition motifs (RRM) in the N-terminal portion and four hnRNP K homology domains (KH) in the C-terminal region[256] **(Fig. 15)**. The high similarity of the domains explains some shared roles, but linker regions confer functional diversity to IGF2BPs[257]. The binding mediated by the KH domains form stable complexes between IGF2BP proteins and their mRNA targets[258]. Each KH domain is not enough to generate high affinity interactions by itself, because they recognize short RNA motifs. The presence of multiple copies of these domains increased the affinity and specificity[259]. Moreover, IGF2BPs are able to form homo- and heterodimers once two molecules are bound to different sites of their target mRNAs. This also contributes to the stabilization of the bindings, together with the presence of the RRMs[258]. These interactions may modify transcripts conformation, and create new binding sites for other factors[259]. The consensus RNA motif recognized by IGF2BPs is 5′-CAUH-3′ (H=A, U, or C), and IGF2BP2 binding is enriched in the 3′UTRs of mRNA targets[260], specifically regions with high AT content and miRNA-binding sites[261].

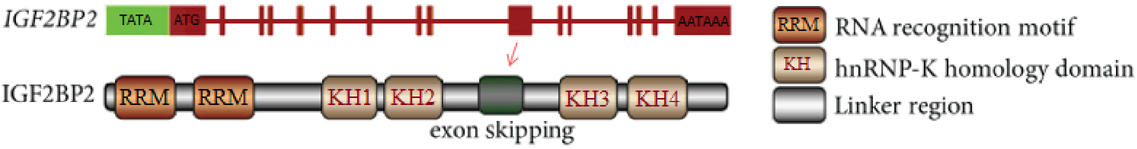


Figure 15. Structure of *IGF2BP2* gene and protein.
Intronic/exonic organization of the gene, with some regulatory sequences such as the TATA box or the polyadenylation signal (AATAA), and the transcription start site (ATG). The gene possesses 16 exons, of which exon 10 could be skipped, giving rise to an alternative isoform. The main RNA-binding domains of the protein are shown, two N-terminal RRMs and four C-terminal KH domains, with the linker region between them (Modified from Cao *et al.*, 2018[257]).

As previously mentioned, *IGF2BP2* is a target of HMGA2. The presence of three AT-rich regions in the DNA encoding the first intron of *IGF2BP2* allows the binding of HMGA2 protein and the induction of *IGF2BP2* transcription[252,253,262]. Although this regulatory mechanism has been observed in mouse, the sequence has 73.2% identity to the human homologous region. The regulation is mediated by NF-κB, thanks to the consensus binding site that is present in the vicinity of the AT-rich regions[253]. Post-transcriptionally, only *miR-216b*[263], *miR-1275*[264] and *let-7*[104] have been described as *IGF2BP2* regulators.

IGF2BPs are considered oncofetal genes because they are highly expressed during development, almost absent in adult tissues, and reexpressed in cancer. The exception is IGF2BP2, whose expression is maintained in many adult tissues[256]. Nevertheless, its overexpression has been associated with multiple cancer types, in which it promotes proliferation, migration, invasion, and metastasis[257]. Particularly interesting for this thesis is its relevance in breast cancer and sarcomas. High levels of IGF2BP2 have been observed in breast cancer patients, with a correlation of the presence of autoantibodies against IGF2BP2 in their blood[265]. Moreover, high expression of IGF2BP2 is related with a decrease in the survival[266]. Thus, the analysis of the autoantibodies can be useful in diagnosis and prognosis. On the other hand, a single-nucleotide polymorphism (SNP) in the second intron of the *IGF2BP2* gene correlates with elevated risk of diabetes[267] and, at the same time, it seems to confer genetic predisposition to breast cancer at least in some populations[268]. This could be explained by the influence that IGF2BP2 has on metabolism and mitochondrial functions[269], directly implicated in diabetes development. In regard to sarcomas, the importance of the HMGA2-IGF2BP2 axis has been widely demonstrated. IGF2BP2 and HMGA2 expression is elevated in myoblasts and regenerating muscle and their depletion impairs muscle cell proliferation and myogenesis[254,262]. Moreover, the overexpression of both proteins in *Hmga2* knockout myoblasts rescues the normal phenotype[254]. Taking into account that many sarcomas are characterized by the impossibility of the cells to complete muscle differentiation, the effect of these two proteins in myogenesis is crucial for muscle cancers development.

3.2.1 Functions and targets of IGF2BP2

The main role of IGF2BPs is the post-transcriptional regulation of mRNAs, and this includes the control of their localization, translation and stability. Considering that IGF2BP2 can bind to more than 2,000 mRNAs[262], its activity has a global impact.

IGF2BPs can form higher-order structures in the cytoplasm through the association with other RBPs. The incorporation of the targeted mRNAs in these RNP complexes may help to transport them along the cytoskeleton[270]. They can also intervene in the nuclear export of some targets because there are two nuclear export signals in their sequence[255,271]. For instance, IGF2BP2 binds to some mRNAs related with mitochondrial activity, allegedly participating in their translocation from the nucleus to the mitochondria[269].

Besides this function, IGF2BPs can positively or negatively regulate the translation of mRNAs. Their inclusion in RNPs can restrict the translation to the right timing[272], or promote translational initiation. This is the mechanism preferentially used by IGF2BP2, and some of its confirmed targets include *c-myc*[262,273] and *IGF1R*[262,264]. Depletion of IGF2BP2 inhibits their translation, with the reduction of the protein level without significant mRNA changes[262]. Moreover, IGF2BP2 binds to the 5′UTR of *IGF2* and promotes translation acting as an IRES *trans*-acting factor[274].

In other cases, IGF2BPs expression affects both mRNA and protein levels, so that the regulation is related with mRNAs stabilization more than with translation control. The degradation of the mRNAs is avoided protecting them from miRNAs action. This can be achieved by two mechanisms. The first one consists in the sequestration of the mRNAs into the RISC-free environment provided by RNPs. That way, although IGF2BP proteins and miRNAs can bind simultaneously to the 3′UTRs of mRNAs, IGF2BP1 and IGF2BP3 protect *let-7* targets such as *HMGA2* and *LIN28B* from degradation[275,276]. The second mechanism implies the competition between the miRNA and the IGF2BP protein for the binding to the mRNA. Unlike the other IGF2BPs, IGF2BP2 is present in processing bodies, together with the proteins involved in miRNA-mediated silencing. Given this, IGF2BP2 avoids the degradation of miRNA targets by the binding to the miRNA response elements of mRNAs[261]. In embryonal rhabdomyosarcoma cells, *N-RAS* mRNA half-life shows an important decrease with IGF2BP2 depletion, with subsequent downregulation of the N-RAS protein level[254]. On the other hand, *let-7* targets expression and tumorigenic capacities were rescued by LIN28B and ASOs against *let-7* in glioblastoma cells depleted of IGF2BP2, indicating that IGF2BP2 exerts its function through *let-7* regulation[261]. Thus, *HMGA1*, *HMGA2*[261,277] and *N-RAS*[254,273] are protected from miRNA degradation.

Post-translational modifications have an essential role in the control of IGF2BPs functions. The release of the mRNAs from RNPs is necessary to induce their translation or degradation, and it involves phosphorylation of the proteins[272]. Furthermore, when mTOR complex 1 phosphorylates IGF2BP2 in the region between RRM2 and KH1, the signal favors its association with IGF2 mRNA and translational initiation[274].

Noteworthy, IGF2BP2 is not only able to interact with mRNAs, but it also binds to ncRNAs. However, rather than being regulated by IGF2BP2, some ncRNAs may actually modulate IGF2BP2 functions. As an example, *LncMyoD* competes with IGF2BP2 mRNA targets for the binding, preventing the translation of the transcripts[273]. Some of the affected genes are *N-Ras*, *c-Myc*, *Igf1r*, *Igf2* and *IGF2BP2* own mRNA. In the presence of *LncMyoD*, low expression levels of these pro-proliferative transcripts permits muscle differentiation **(Fig. 16a)**. The function of IGF2BP2 is also relevant in cardiac muscle, and the regulation mediated by the NAT *Airn* affects cardiomyocyte physiology. However, in this case, the interaction affects positively the expression of some of the IGF2BP2 targets[150]. Another NAT that controls IGF2BP2 function in the same direction is *HIF1A-AS2*. This antisense favors the hypoxic adaptation in glioblastoma increasing the expression of proteins such as *HMGA1*[161] **(Fig. 16b)**.

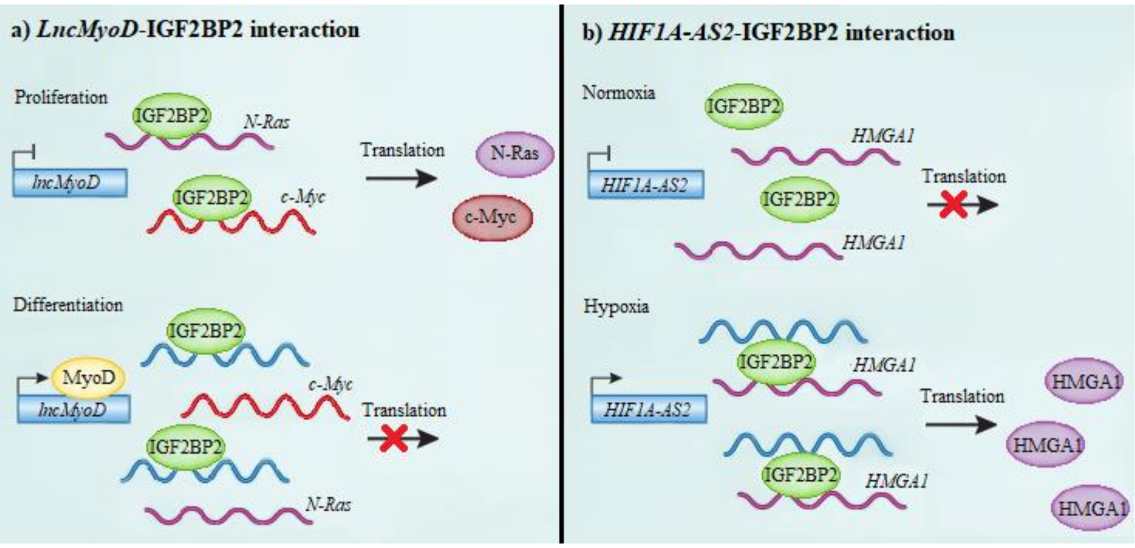


Figure 16. Modulation of IGF2BP2 functions by lncRNAs.
(a) *LncMyoD* is silenced while muscle cells are in proliferation. Thus, IGF2BP2 binds to its pro-proliferative targets, such as *N-Ras* and *c-Myc*, promoting their translation. During differentiation, *lncMyoD* is transcriptionally activated by MyoD and competes with IGF2BP2 mRNA targets for the binding. The presence of the lncRNA regulates negatively the translation of these transcripts. **(b)** In normal oxygen conditions, *HIF1A-AS2* is not expressed and IGF2BP2 cannot bind to its mRNA targets. During hypoxic adaptation, *HIF1A-AS2* is transcribed and its interaction with IGF2BP2 leads to the translation of some IGF2BP2 targets such as *HMGA1*.

AIMS

Many regulations mediated by ncRNAs cause a severe impact on key cellular genes, and they have been described linked to various human diseases, including cancer, which gives an idea of the importance of their understanding. Since the meaning of the large fraction of mammals' transcriptome that does not code for proteins is still an unanswered question, the present project focuses on the emergent roles that NATs, one of the most abundant types of ncRNAs, display in gene expression control. The general purpose of this PhD thesis is to understand the crosstalk between NATs, miRNAs and coding genes, and relate this to diseases like cancer. To finely define the mechanisms underlying the regulation mediated by antisense transcription, we combine biochemical and molecular approaches with transcriptomic and genome-wide techniques. The specific objectives are as follows:

STUDY I

"Head-to-head antisense transcription and R-loop formation promotes transcriptional activation"

We aimed to characterize the impact of antisense transcription on chromatin organization, which is one of the main epigenetic determinants of gene expression programs. We focus on *HMGA2*, a gene relevant in cancer, and the pseudogene expressed from the same *locus* *RPSAP52*. Specific goals:

- To evaluate the role of *RPSAP52* antisense transcription in the regulation of the corresponding sense gene, *HMGA2*.
- To analyze the participation of the lncRNA *RPSAP52* in the formation of regulatory structures such as R-loops.
- To examine the ability of antisense transcription to modify local chromatin accessibility through nucleosomes positioning.

STUDY II

“The transcribed pseudogene RPSAP52 enhances the oncofetal HMGA2-IGF2BP2-RAS axis through LIN28B-dependent and independent let-7 inhibition”

We aimed to describe the regulation exerted by the lncRNA *RPSAP52* on important oncogenic pathways such as the HMGA2-IGF2BP2-RAS axis. Specific goals:

- To characterize the *HMGA2/RPSAP52 locus* and its expression in several tumor types, both cell lines and patients, relating this with its methylation status.
- To identify possible protein partners of *RPSAP52* and to study how the interaction between them, if it exists, affects its activity and the function of the protein.
- To establish the biological relevance of antisense transcription through *in vitro* functional assays and *in vivo* models.
- To transfer the results to other cancer types in which the transcription of the *HMGA2/RPSAP52 locus* may be of special relevance, such as Ewing’s sarcoma and rhabdomyosarcoma.

RESULTS

1. Results summary

There are a number of NATs with regulatory roles in the transcription of the nearby sense genes. Specifically, we study gene pairs with divergent transcription and a GC skew in the promoter region that allows the formation of an R-loop by the NAT. Following these criteria, *HMGA2/RPSAP52* pair was selected for further investigation due to the aberrant expression of HMGA2 in a multitude of cancers. *RPSAP52* depletion led to the decrease of *HMGA2* in MCF10A cell line, consequence of the disruption of the R-loop formed by *RPSAP52*. This R-loop reduces chromatin compaction and increases its accessibility, which has a positive impact on the activation of the sense transcript *HMGA2*.

Both genes are overexpressed in some human cancers, and their expression shows a positive correlation in breast cancer patients and cell lines. In addition, hypermethylation of their promoter correlates with the repression of both transcripts, including an *RPSAP52* isoform not described so far. Although we confirmed the absence of *RPSAP52* coding capacity, its association with polysomes indicates a possible function in translation. Moreover, *RPSAP52* enrichment in the cytoplasm and its polyadenylation also suggest additional roles besides R-loop formation. In order to understand them, we described the binding of *RPSAP52* to IGF2BP2 protein, a transcriptional target of HMGA2 that regulates the translation of some mRNAs such as *IGF1R* and *RAS*. *RPSAP52* depletion resulted in an increase of *let-7* expression that correlated with low levels of the proteins IGF2BP2, IGF1R and RAS, whose mRNAs are *let-7* targets. LIN28B is not expressed in MCF10A cells and LIN28A does not change with *RPSAP52* depletion. Thus, it is not possible to explain the results by alterations in the levels of LIN28 proteins, the main negative regulators of *let-7* biogenesis, in the breast cell lines we have studied. However, IGF2BP2 protein level was rescued by LIN28B overexpression.

The phenotypic impact of *RPSAP52* depletion in breast cell lines includes a decrease in cell proliferation, migration and clonogenicity, either with anchorage-dependent or independent growth. Also, the protein levels of the stemness markers NANOG and OCT4 are reduced in these cells. Similarly, the weight and volume of subcutaneous tumors is lower in immunosuppressed mice injected with *RPSAP52*-depleted cells.

Given the critical role played by the HMGA2-IGF2BP2-NRAS axis in the development of embryonic rhabdomyosarcoma and the elevated expression of *HMGA2* and *RPSAP52* in sarcomas, the study was extended to this cancer type. A673 cell line was selected as a model of Ewing's sarcoma, and stable clones with *RPSAP52* depletion were obtained. As seen in MCF10A cells, an increase in *let-7* expression was observed in the clones, and the interaction between IGF2BP2 and *RPSAP52* was also confirmed. In this case, IGF2BP2 and RAS remain unchanged at the protein level, but the pathway is affected downstream with the decrease of p-ERK. It is noteworthy to mention the important reduction in LIN28B protein in the clones, which is abundantly expressed in A673 cells. *In vivo* consequence of *RPSAP52* depletion was a marked reduction in tumor formation.

Importantly, we described the binding of IGF2BP2 to *LIN28B* mRNA, and this interaction was confirmed using iCLIP-Seq, a technique we used to identify IGF2BP2 RNA targets in a genome-wide manner. *RPSAP52* knockdown caused differences in the binding motif and in the recognition of the 3′UTR, with a specific loss of IGF2BP2 affinity for particular mRNAs. This has no effect in the stability of the targets due to the absence of changes in the half-lives of their mRNAs in *RPSAP52*-depleted cells. Even though *RPSAP52* does not affect the binding of IGF2BP2 to its protein partners, it mediates its recruitment to large polysomes and influences the translational efficiency of *HMGA2* and *LIN28B* mRNAs. This could be the mechanism by which *RPSAP52* controls the expression of LIN28B and, therefore, *let-7* levels. The regulation exerted by *RPSAP52* on IGF2BP2 is independent of LIN28B expression because its depletion does not have an impact on it.

Since *RPSAP52* can globally affect important proliferative pathways, an expression array was performed to study the impact of genome-wide *RPSAP52* silencing. The increase of some tumor suppressor genes was detected in the clones, as well as the decrease in genes usually overexpressed in cancer. In support of the influence that *RPSAP52* has in tumorigenicity, high expression levels imply a worse survival rate in sarcoma patients. It should be noted that its expression correlates positively with *HMGA2*, but only *RPSAP52* levels and methylation of the associated CpG island have prognostic value.

Our findings provide new knowledge about NATs-mediated regulatory mechanisms and highlight their impact on cancer-related genes and on tumor progression itself. According to our results, *RPSAP52* regulates *HMGA2* expression through the formation of an R-loop and IGF2BP2 function through the binding to this protein. We also demonstrated that *LIN28B* mRNA constitutes a new target of IGF2BP2 not described until now, and that is how *RPSAP52* affects LIN28B/*let-7* balance and promotes tumorigenesis. In conclusion, our work establishes *RPSAP52* as a master regulator with oncogenic properties.

2. Directors report

To who may concern, we authenticate that the PhD student CRISTINA OLIVEIRA MATEOS will present her PhD thesis by scientific publications. Her contribution for each publication will be next pointed out:

STUDY I

"Head-to-head antisense transcription and R-loop formation promotes transcriptional activation"

Raquel Boque-Sastre, Marta Soler, **Cristina Oliveira-Mateos**, Anna Portela, Catia Moutinho, Sergi Sayols, Alberto Villanueva, Manel Esteller, and Sonia Guil.

Contribution: In this paper Cristina Oliveira-Mateos was responsible for the experiments related to *HMGA2/RPSAP52 locus*. In addition, she collaborated in bisulfite genomic sequencing, in the nuclear/cytoplasmic fractionation, as well as in cell culturing works. She also participated in data analysis and manuscript revision.

This publication has been used by the first author Raquel Boque-Sastre in her PhD thesis defense.

Journal: PNAS (Proceedings of the National Academy of Sciences of the United States of America). Proc Natl Acad Sci USA. 2015 May 5;112(18):5785-90.
doi:10.1073/pnas.1421197112. Epub 2015 Apr 22.

Impact factor: 9.580 (2018), 9.423 (2015)

STUDY II

"The transcribed pseudogene *RPSAP52* enhances the oncofetal HMGA2-IGF2BP2-RAS axis through LIN28B-dependent and independent *let-7* inhibition"

Cristina Oliveira-Mateos, Anaís Sánchez-Castillo, Marta Soler, Aida Obiols-Guardia, David Piñeyro, Raquel Boque-Sastre, Maria E. Calleja-Cervantes, Manuel Castro de Moura, Anna Martínez-Cardús, Teresa Rubio, Joffrey Pelletier, Maria Martínez-Iniesta, David Herrero-Martín, Oscar M. Tirado, Antonio Gentilella, Alberto Villanueva, Manel Esteller, Lourdes Farré and Sonia Guil.

Contribution: In this paper Cristina Oliveira-Mateos was responsible for the experimental design, and the development of methodology, supervised by Dr. Guil. She also participated in the analysis and interpretation of the generated data together with the manuscript elaboration and revision.

Journal: Nature Communications. Nat. Commun. 2019 Sep 4;10(1):3979.
doi: 10.1038/s41467-019-11910-6.

Impact factor: 11.878 (2018)

For the sake of clarity and higher figure resolution, I next present the published articles in Word format.

Dr. Sònia Guil Domènech, Ph.D.
Regulatory RNA and Chromatin Group, Leader
Josep Carreras Leukaemia Research Institute (IJC)
Ctra de Can Ruti, Camí de les Escoles, s/n
08916 Badalona, Barcelona
+34 93 557 28 00 ext. 4225
sguil@carrerasresearch.org

Dr. Manel Esteller Badosa, M.D, Ph.D.
Cancer Epigenetics Group, Leader
Josep Carreras Leukaemia Research Institute (IJC)
Ctra de Can Ruti, Camí de les Escoles, s/n
08916 Badalona, Barcelona
+34 93 557 28 37
mesteller@carrerasresearch.org

STUDY I

Head-to-head antisense transcription and R-loop formation promotes transcriptional activation

Raquel Boque-Sastre[a], Marta Soler[a], **Cristina Oliveira-Mateos[a]**, Anna Portela[a], Catia Moutinho[a], Sergi Sayols[a,1], Alberto Villanueva[b], Manel Esteller[a,c,d,2], Sonia Guil[a,2]

[a]Cancer Epigenetics and Biology Program, and [b]Translational Research Laboratory, Catalan Institute of Oncology, Bellvitge Biomedical Research Institute, L'Hospitalet, 08908 Barcelona, Catalonia, Spain; [c]Department of Physiological Sciences II, School of Medicine, University of Barcelona, 08907 Barcelona, Catalonia, Spain; and [d]Generalitat de Catalunya, Institucio Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Catalonia, Spain.

[1]*Present address: Institute of Molecular Biology, 55128, Mainz, Germany.*

[2]*To whom correspondence may be addressed. Email: sguil@idibell.cat or mesteller@idibell.cat.*

Edited by Alberto R. Kornblihtt, University of Buenos Aires, Buenos Aires, Argentina, and approved April 2, 2015 (received for review November 6, 2014)

Author contributions: R.B.-S., M.E., and S.G. designed research; R.B.-S., M.S., C.O.-M., A.P., C.M., and S.G. performed research; A.V. contributed new reagents/analytic tools; R.B.-S., M.S., C.O.-M., A.P., S.S., and S.G. analyzed data; and S.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

This article contains supporting information online at

www.pnas.org/lookup/suppl/doi:10.1073/pnas.1421197112/-/DCSupplemental.

Abstract

The mechanisms used by antisense transcripts to regulate their corresponding sense mRNAs are not fully understood. Herein, we have addressed this issue for the vimentin (*VIM*) gene, a member of the intermediate filament family involved in cell and tissue integrity that is deregulated in different types of cancer. *VIM* mRNA levels are positively correlated with the expression of a previously uncharacterized head-to-head antisense transcript, both transcripts being silenced in colon primary tumors concomitant with promoter hypermethylation. Furthermore, antisense transcription promotes formation of an R-loop structure that can be disfavored *in vitro* and *in vivo* by ribonuclease H1 overexpression, resulting in *VIM* downregulation. Antisense knockdown and R-loop destabilization both result in chromatin compaction around the *VIM* promoter and a reduction in the binding of transcriptional activators of the NF-κB pathway. These results are the first examples to our knowledge of R-loop-mediated enhancement of gene expression involving head-to-head antisense transcription at a cancer-related *locus*.

Keywords: vimentin/ antisense transcription/ DNA methylation/ R-loop/ nucleosome occupancy

Significance

The molecular mechanisms used by noncoding RNAs to regulate gene expression are largely unknown. We have discovered a previously unidentified regulatory phenomenon underlying the transcriptional activation of the intermediate filament protein vimentin. This regulation involves the participation of a previously uncharacterized head-to-head antisense transcript that forms part of a hybrid DNA:RNA structure known as the R-loop. R-loops have been the focus of recent research regarding their unexpected involvement in gene expression regulation. Antisense-mediated formation of the R-loop supports a local chromatin environment that ensures the optimal binding of vimentin transcriptional activators. In addition, we describe how hypermethylation of the *locus* in a large panel of colon cancer patients is correlated with antisense silencing and, thereby compromises its regulatory activity.

Introduction

Many well-documented instances of functional long noncoding RNAs (ncRNAs) attest to their multiple roles in regulating transcriptional programs (for a recent review, see ref. 1). The most abundant class of long ncRNAs contains natural antisense transcripts, which partially or totally overlap transcripts originating from the opposite strand. Antisense transcripts may have regulatory effects at different levels, including transcriptional regulation, epigenetic control, imprinting, alternative splicing, translation and RNA editing (reviewed in refs. 2 and 3). Also, recent studies have addressed the role of ncRNAs as spatial regulators of 3D chromatin folding (4). However, we have a far from thorough understanding of the mechanisms underlying antisense-mediated regulation of gene expression. *VIM* is a member of the group of type III intermediate filament genes whose expression increases during the epithelial-to-mesenchymal transition and that are generally associated with an enhanced ability for cell migration and invasion (5). Although the existence of antisense transcription at the *VIM locus* has been reported in rat and is known to influence the epigenetic status of the *locus* (6), it was not known whether a similar mechanism is present in humans. We present data supporting the positive regulation exerted by *VIM* head-to-head antisense transcript on *VIM* mRNA, through the formation of an RNA:DNA hybrid known as the R-loop.

Results

Head-to-Head Antisense Transcription at the *VIM Locus*. The region encompassing the human *VIM* promoter region and transcription start site (TSS) contains an additional, as yet functionally uncharacterized, transcriptional unit corresponding to the antisense strand (Fig. 1*A*), deposited as *VIM-AS1* transcript in the University of California, Santa Cruz (UCSC) data bank (also known as *BC078172* transcript). *VIM-AS1* is a 1.8-kb noncoding RNA transcribed 5′ head-to-head with *VIM*, starting 709 bp downstream from the canonical *VIM* TSS. A minor (expressed at a ~50-fold lower level) alternative *VIM* TSS has been described 993 bp upstream of the canonical TSS (7) (not depicted in Fig. 1*A*). *VIM* is generally scarce in epithelial cells but it can also be expressed in epithelial cell lines as part of the adaptation to *in vitro* culture conditions (8, 9). *VIM* is expressed in normal colon mucosa, mainly in stromal cells and lymphocytes (10). We readily detected both sense and antisense transcripts in end-point PCR by using total RNA from human colon (Fig. 1*A* and *B*) as the template. To confirm strand specificity we carried out primer-specific reverse transcription and PCR (Fig. 1*A*). In addition, oligo-dT-primed and random-primed reverse transcription indicated that the *VIM-AS1* antisense transcript is a polyadenylated RNA (Fig. 1*B*, *Upper*). PolyA$^{+/-}$ partition of total RNA and quantitative RT-PCR (RT-qPCR) confirmed that sense and antisense transcripts are both polyadenylated (Fig. 1*B*, *Lower*). However, analysis of the nuclear and the cytoplasmic fractions showed a clear enrichment of *VIM-AS1* transcript in the nucleus (Fig. 1*C*), suggesting a possible nuclear function. RT-qPCR experiments carried out in this study indicated that the *VIM* mRNA is 2-3 orders of magnitude more abundant than its antisense transcript, consistent with general estimates of the abundance of noncoding antisense transcript relative to its sense partners (ref. 11; see below).

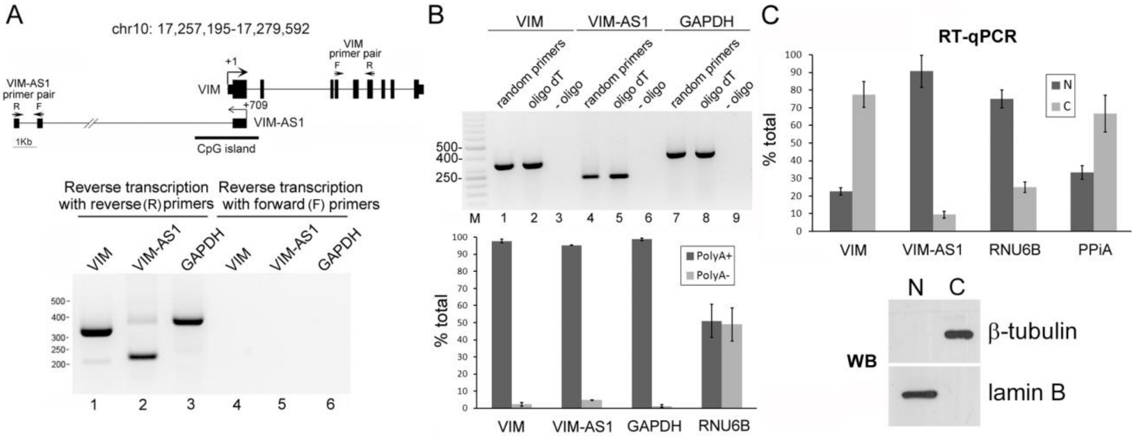


Fig. 1. *VIM-AS1* is a nuclear, polyadenylated transcript running head-to-head with *VIM* transcript. (*A*, *Upper*) Intronic/exonic organization of vimentin (*VIM*) and its antisense *VIM-AS1* transcripts. Coordinates are given relative to the canonical *VIM* TSS and the UCSC Gene data bank (uc001iot.2) for *VIM-AS1* (release hg19). (*A*, *Lower*) End-point RT-PCR from normal colon mucosa total RNA with strand-specific primers. Reverse transcription was carried out either with specific reverse primers ('R', lanes 1-3) or with forward primers ('F', lanes 4-6). (*B*, *Upper*) *VIM-AS1* RNA transcript is polyadenylated. (*B*, *Lower*) PolyA$^+$/polyA$^-$ partition of total RNA from SW480 cells analyzed by RT-qPCR. (*C*) Nuclear/cytoplasmic fractionation of SW480 cells, analyzed by RT-qPCR and Western blot to assess fraction purity.

Hypermethylation in Colon Cancer Is Associated with Sense and Antisense Transcript Silencing. *VIM* promoter has been thoroughly characterized (12-18). In addition, hypermethylation of *VIM* promoter-associated CpG-rich island (CGI) has been reported in colon cancer (10, 19). To investigate the impact of CGI methylation on antisense transcription, we examined DNA samples from 120 normal colon and 120 primary tumors with Illumina's HumanMethylation450 BeadChip. *VIM* promoter region remains largely unmethylated in normal samples, whereas primary tumors display a clearly hypermethylated CpG island (Fig. 2*A* and *B*). Hypermethylation in primary tumors was confirmed by bisulfite sequencing in an independent subset of matched pairs of normal and tumor samples (n=33 per type). In this independent subset, we also observed hypermethylation of tumor samples, allowing us to define a differentially methylated region (DMR, thick red line in Fig. 2*B*) embedded within the CpG island, indicating that hypermethylation of the *VIM* promoter is a hallmark of colon cancer. At the expression level, the two transcripts are positively correlated, both in normal and primary tumors (Fig. 2*C* and *D*). Interestingly, the primary tumors with highest methylation levels display the lowest transcript abundance (Fig. 2*D*). Additionally, 10 out of 12 colon adenocarcinoma cell lines analyzed exhibited CpG island hypermethylation (Fig. 2*E*). *VIM* and *VIM-AS1* transcript levels were also positively correlated in methylated and unmethylated lines (Fig. 2*F*), with 100- to 100,000-fold greater levels of expression in unmethylated compared to methylated lines. Similar correlations were also observed in breast carcinomas and tumor cell lines (Fig. S1*A*-*C*). Methylation levels inversely correlated with the quantities of VIM protein, as shown for HCT116 and HCT116-DKO (hypomorphic for the DNA methyltransferases DNMT1 and DNMT3b) cell lines (Fig. S1*D*). To further estimate the abundance of *VIM-AS1* transcript in comparison with *VIM* mRNA, we performed absolute quantitation of both RNAs in methylated (HCT116) and unmethlyated (DKO, SW480, MCF10A) cell lines (Fig. S2). The results obtained indicate a difference in 2-3 orders of magnitude between *VIM* and *VIM-AS1* levels, and an impact of methylation resulting in a reduction in expression of 2 orders of magnitude for *VIM* mRNA and of 1 order of magnitude for *VIM-AS1* transcript (Fig. S2*C*).

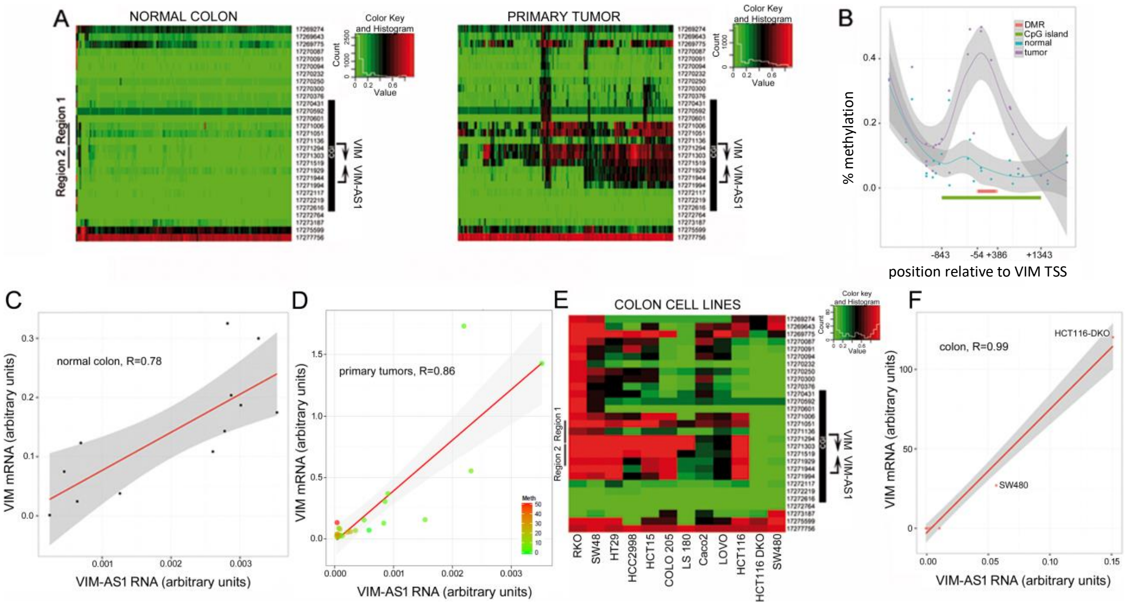


Fig. 2. Sense/antisense transcripts are coordinately expressed in normal and tumor colon samples and inversely correlated with DNA methylation. (*A*) Heatmap representation of a DNA methylation microarray analysis of 120 human normal colon mucosae and 120 tumor samples. (*B*) Percentage methylation levels of individual CpG sites contained in

the 450k array, averaged by class (normal/tumor). The position of the CGI is indicated (green line), and the differentially methylated region defined in *C*. (*C* and *D*) Positive Pearson's correlation coefficients between *VIM* (*y* axis) and *VIM-AS1* (*x* axis) expression for normal and tumor colon samples. For primary tumors, the color code indicates methylation levels assessed by bisulfite sequencing. (*E*) Heatmap representation of a DNA methylation microarray analysis of 12 human colorectal adenocarcinoma cell lines. (*F*) Pearson's correlation coefficients between *VIM* and *VIM-AS1* expression for all colon cell lines shown in *E*.

Antisense Knockdown Results in *VIM* Silencing. Many antisense transcripts correlate positively and are known to act *in cis* to regulate their sense partners (3, 20, 21). To investigate this, we used RNAi to deplete *VIM-AS1* RNA in SW480 cells, which have an unmethylated *VIM* promoter and display high basal levels of *VIM* and *VIM-AS1* transcripts. The shRNAs used target the last exon in *VIM-AS1*, in the region that does not overlap with *VIM* (*SI Materials and Methods*). Two distinct shRNAs were capable of efficiently downregulating *VIM-AS1* levels, concomitant with a two- to three-fold decrease in *VIM* mRNA levels (Fig. 3*A*). This reduction was also detected at the protein level by Western blot and immunofluorescence (Fig. 3*B* and 3*C*; ZsGreen is an indicator of transduced cells). To confirm the specificity of the downregulation, we used two locked nucleic acid (LNA)-based antisense oligonucleotide (ASO) gapmers that target, in a strand-specific manner, the 5′ region of *VIM-AS1* transcript (Fig. 3*D*). Remarkably, both ASOs induced a marked decrease in *VIM-AS1* and *VIM* transcripts, confirming the results obtained with the shRNAs. It is of note that ASO1 was directed against the first intron of *VIM-AS*, suggesting an active functional involvement for this intronic region. We next investigated whether *VIM-AS1* RNA knockdown causes promoter hypermethylation that could account for *VIM* silencing. As seen in Fig. 3*E*, we observed a moderate increase in DNA methylation levels across regions 1 and 2 of the *VIM* promoter, which are the most highly methylated regions in colon tumors and cell lines, therefore indicating that antisense reduction results in a degree of CGI hypermethylation. The same analysis with shRNAs also shows a slight increase in methylation across region 2 of the CpG island (Fig. S3).

Given the suggested involvement of *VIM-AS1* first intron in the regulation, we next designed specific probes to detect by RNA FISH either intron 1 of *VIM* or intron 1 of *VIM-AS1* (Fig. 4*A*). As expected for intronic regions, both probes colocalize at the site of nascent transcription (Fig. 4*B*), with an enrichment of the *VIM-AS1* probes in G2 phase. Remarkably, blocking transcription by treatment with actinomycin D resulted in the loss of *VIM* intron 1 signal, whereas the antisense intron remained localized near the genomic *locus* (Fig. 4*C* and *D* and Fig. S4*A*-*C*, where wider microscope fields with more cells are shown). This FISH signal could be indicative of a special stabilization of the region, possibly due to inefficient splicing. To further analyze the interaction of this intronic RNA region with the local chromatin we used the RNA antisense purification (RAP) method (22), in which a pool of 124-nt-long antisense probes designed against the first intron of *VIM-AS1* was able to specifically retrieve the endogenous transcript (Fig. S4*D*) together with the homologous DNA region (Fig. 4*E*), suggesting that there is a stable RNA:DNA association in this region. Furthermore, RAP signal was maintained even when transcription was arrested, in accordance with RNA-FISH experiments.

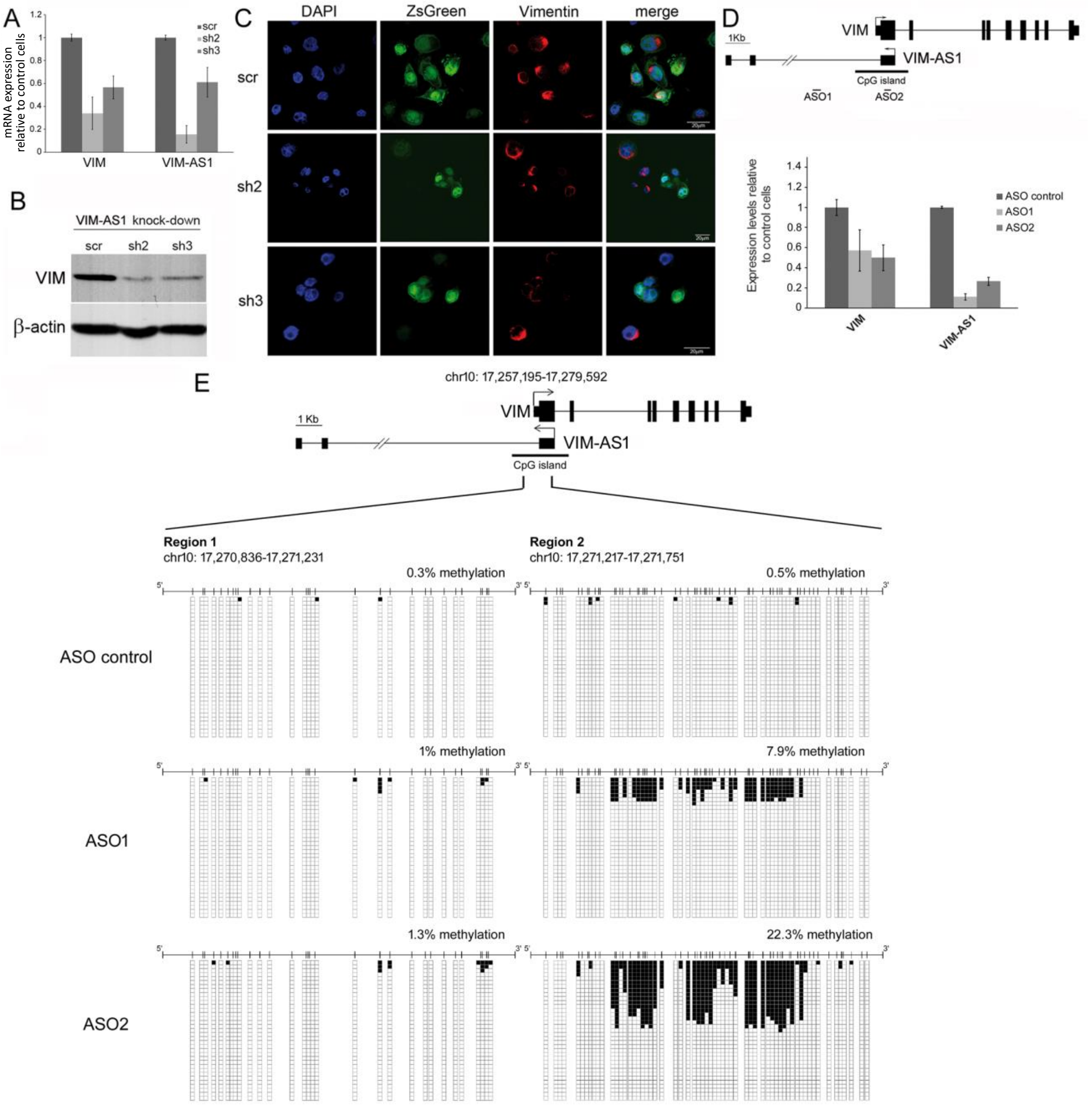
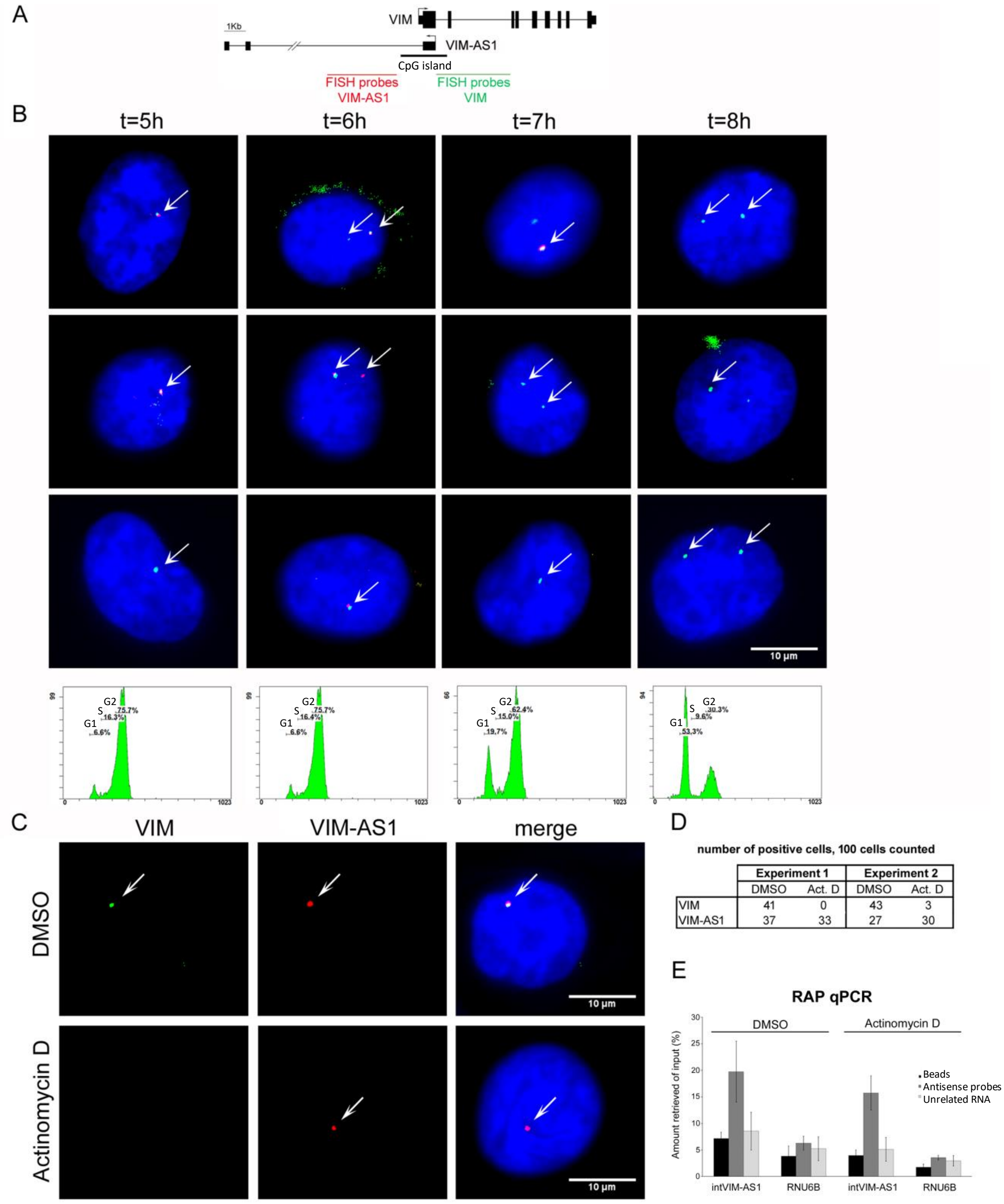


Fig. 3. *VIM-AS1* transcript knockdown results in *VIM* silencing with an effect on promoter CGI methylation. SW480 cells were transduced with lentiviral plasmids overexpressing control shRNA (scr) or shRNAs against *VIM-AS1* RNA (sh2, sh3). (*A*) RT-qPCR analysis of *VIM* and *VIM-AS1* RNAs. (*B*) Western blot to measure vimentin protein levels in the same transduced cells. (*C*) Immunofluorescence detection of endogenous vimentin. (*D*) LNA-based antisense oligonucleotides gapmers (ASOs) targeting intron 1 (ASO1) or exon 1 (ASO2) of *VIM-AS1* transcript were transfected into SW480 cells and expression levels measured by RT-qPCR. (*E*) Bisulfite sequencing of regions 1 and 2 within *VIM* promoter CGI.



number of positive cells, 100 cells counted

	Experiment 1		Experiment 2	
	DMSO	Act. D	DMSO	Act. D
VIM	41	0	43	3
VIM-AS1	37	33	27	30

Fig. 4. RNA FISH detection shows enrichment of antisense transcription during G2 phase and intronic stability following actinomycin D treatment. (*A*) RNA FISH probe design. (*B*) MCF10A cells were synchronized and released, fixed at the indicated times and stained for RNA FISH (*VIM* intron 1 is in green and *VIM-AS1* intron 1 is in red) or analyzed for DNA content by fluorescence-activated cell sorting (FACS) (*Upper* and *Lower*). (*C*) RNA FISH in control (DMSO-treated) or actinomycin D-treated MCF10A cells. (*D*) RNA FISH signal was counted in 100 randomly selected cells. (*E*) Quantitative PCR (qPCR) of the DNA captured in crosslinked MCF10A cells treated as in *C*, using streptavidin beads alone (beads), with antisense probes to VIM-AS1 intron 1 (antisense probes) or against the *LINC00085* RNA (unrelated RNA). Enrichments represent means from two replicate experiments and are relative to the input amount used per pulldown. RNU6B is used as negative control to assess binding specificity.

Antisense Transcript Forms Part of an R-loop Structure and Its Disruption Represses *VIM* Transcription. Further exploration of the genomic region between the two transcription start sites revealed an asymmetric distribution of C and G nucleotides (known as a GC skew) on the plus strand of the DNA along the first half of the CGI and coinciding with the first intron of *VIM-AS1*. As seen in Fig. 5*A*, a fragment of approximately 1 kb between *VIM* and *VIM-AS1* TSS was particularly enriched in C nucleotides in the plus strand (C skew), whereas no enrichment was observed for A or T nucleotides (Fig. S5*A*). This observation points to the potential formation of R-loops throughout this region. R-loops are special three-stranded nucleic acid structures that form *in vivo* as G-rich RNA transcripts invade the DNA duplex and anneal to the template strand to generate an RNA-DNA hybrid (23), leaving the nontemplate, G-rich DNA strand in a largely single-stranded conformation. To explore this possibility, we cloned the DNA region comprising the C skew between two opposing promoters and tested R-loop formation *in vitro*. Transcription from T7 promoter gives rise to an RNA molecule in the direction of *VIM* mRNA, whereas transcription from SP6 promoter originates the antisense *VIM-AS1* RNA (Fig. 5*B*). The formation of extended RNA:DNA hybrid structures results in topological change in the plasmid DNA that can be detected as a lower electrophoretic mobility. Transcription of the region containing the C skew led to a strong shift in DNA migration only when the template strand was the C-rich DNA strand (that is, the RNA produced is *VIM-AS1*). Transcription in the other direction (*VIM* physiological orientation) did not result in such a migration pattern (Fig 5*B*, *Left*). To confirm the involvement of RNA, the reaction was carried out in the presence of radiolabeled [α-^{32}P]-rUTP (Fig. 5*B*, *Right*). The DNA migration shift and radioactive signal are both abolished upon incubation with recombinant RNaseH, which digests RNA:DNA hybrids. Taken together, these properties (slower migration, orientation dependence and sensitivity to RNaseH) indicate the presence of an R-loop structure with involvement of *VIM-AS1* RNA in the vicinity of *VIM* TSS.

To confirm the formation of the R-loop *in vivo*, we used a native bisulfite treatment of SW480 genomic DNA (which converts only accessible cytosines in any DNA template), followed by PCR with primers specific to the first half of the predicted R-loop-forming region, ligation and sequencing of the resulting clones (24) (Fig. 5*C*). This method reveals single-strandedness either of the G-rich or C-rich strand, depending on the type of conversion: C-to-T changes in the sequence of the plus strand are indicative of single-strandedness in the C-rich strand (plus strand), whereas G-to-A changes in the plus strand indicate single-strandedness in the G-rich strand (minus strand). In control-transfected cells, all clones sequenced featured long stretches (>100 bp) of uninterrupted G-to-A conversions (Fig. 5*C*, *Upper*). These changes are a qualitative indications of the existence of an unprotected, single-stranded minus strand, suggesting that R-loop formation occurs endogenously at the *VIM* promoter with the participation of the antisense transcript. Interestingly, ASO gapmers designed against *VIM-AS1* made G-to-A changes less frequent, suggesting the involvement of the antisense *VIM-AS1* transcript in R-loop formation *in vivo* (Fig. 5*C*, *Lower*).

As further proof of the existence of an R-loop structure near *VIM* TSS *in vivo* we performed DNA:RNA immunoprecipitation (DRIP) experiments with the S9.6 antibody (25). Consistent with the previous data, we were able to detect specific R-loop formation by DRIP in an RNaseH-sensitive manner along the C skew region (Fig. 5*D*, *Left*). For comparison and pulldown efficiency estimations, a known amount of *in vitro* generated R-loop was subject to parallel DRIP experiments (Fig. 5*D*, *Right*). Importantly, DRIP signal was decreased in ASO-treated cells, (Fig. 5*E*, *Left*), and in cells transduced with a lentiviral vector encoding human ribonuclease H1 (RNASEH1) (Fig. 5*E*, *Right*), indicating that knockdown of *VIM-AS1* transcripts and overexpression of RNASEH1 both result in R-loop resolution. In addition, immunofluorescence analysis indicates that overexpression of the protein reduced vimentin protein levels in cells expressing the transfected protein (Fig. 5*F*). This result was confirmed by Western blotting after sorting of RNASEH1-overexpressing cells by FACS (Fig. 5*G*, *Left*). Accordingly, both *VIM-AS1* and *VIM* mRNA levels were downregulated under conditions of RNASEH1 overexpression, as detected by RT-qPCR (Fig. 5*G*, *Right*). A further reduction in *VIM* levels was observed in Caco2 cells at the mRNA and protein levels (Fig. S5*B*). Similarly, in the two breast cell lines, MCF7 and MCF10A, *VIM* levels were also sensitive to RNASEH1 overexpression (Fig. S5*C*), indicating that R-loop formation has a generally positive effect on its expression. It is worth noting that *VIM-AS1* RNA levels were also significantly diminished in all cell lines in which RNASEH1 was overexpressed, possibly implying that most of its transcripts are R-loop-associated and direct targets of RNASEH1 digestion. Because R-loop formation has been associated with DNA methylation protection (25), we performed bisulfite sequencing to assess changes in methylation in *VIM* CGI under conditions where R-loop formation is disfavored. A slight increase was observed in Caco2 cell line when we overexpressed RNaseH1 (Fig. S5*D*), whereas no change was seen in SW480 cells (Fig. S5*E*). This difference is probably due to the fact that antisense levels in Caco2 cells are much lower than in SW480 and it might be easier to achieve a more complete resolution of R-loops. Finally, the effect of RNASEH1 overexpression on VIM levels does not result from general changes in expression (Fig. S5*F*).

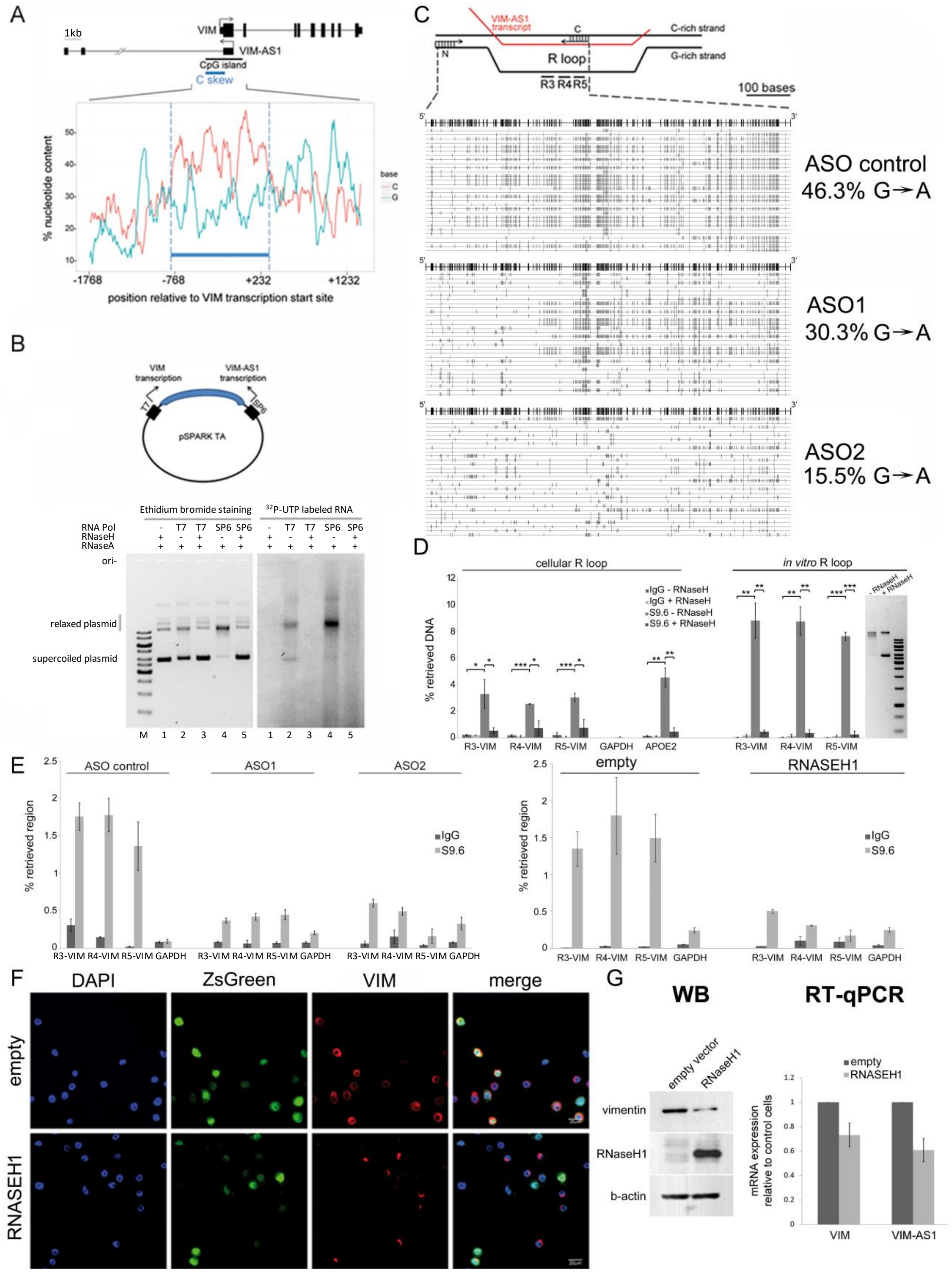


Fig. 5. *VIM-AS1* RNA forms an R-loop structure whose disruption represses *VIM* transcription in SW480 cells. (*A*) Percentage of C and G nucleotides in the *VIM* promoter reveals the presence of a C skew region (thick blue line). For each position on the DNA plus strand, the percentage abundance of each nucleotide within the surrounding 100 nt is counted, with a sliding window of 1 nt. (*B*) *In vitro* R-loop formation assay indicates participation of the *VIM-AS1* transcript. (*C*) *In vivo* detection of R-loop formation within the C skew region. (*Upper*) The diagram depicts the RNA:DNA hybrid and the displaced, single-stranded, minus DNA strand. (*Lower*) PCR amplification and sequencing of 30 clones corresponding to the first half of the C skew-containing region under different ASO treatment. The upper reference line depicts every G position (vertical lines), and every G-to-A change on the plus strand of the sequenced

clones is indicated in light gray by a vertical line. Of the 30 clones represented, 23, 29 and 26 (for ASO control, ASO1 and ASO2, respectively) correspond to unique patterns. (*D*) DRIP with the S9.6 antibody. Signal intensity is presented relative to the input DNA. Three different amplicons (R3, R4, R5, shown in *C*) were measured. *GAPDH* and *APOE* promoters were analyzed as negative and positive controls, respectively. $*P<0.05$;$**P<0.01$; $***P<0.001$. (*E*) DRIP experiments under ASO treatment (*Left*), or overexpression of RNASEH1 (*Right*). The same genomic regions as in *D* were analyzed. (*F*) Immunofluorescence detection of endogenous vimentin (red) in cells overexpressing RNASEH1 (green). (*G, Left*), Western blot of total protein extracts from RNASEH1-positive cells (enriched by FACS). (*Right*) RT-qPCR of total RNA extracted from the pool of transfected cells.

Antisense Transcription and R-loop Structure Support Local Chromatin Decondensation. We next attempted to establish whether R-loop formation had any effect at some other level of chromatin conformation. Nucleosome occupancy is known to be lower in the vicinity of the TSSs of actively transcribed genes (26). In accordance with this premise, histone H3 becomes less prevalent in regions immediately upstream of *VIM* TSS, as revealed by chromatin immunoprecipitation (ChIP) experiments (Fig. 6*A* and *B*). These results might indicate an open conformation coincident with the presence of C skew and R-loop formation. To test the effect of antisense transcription and R-loop formation on the level of chromatin compaction, we isolated native chromatin from control, *VIM-AS1*-depleted and RNASEH1-overexpressing SW480 cells. The graphs in Fig. 6*C* and *D* illustrate the differences in DNA recovery during the first 10 min of micrococcal nuclease digestion, whereby higher values indicate a more thoroughly digested DNA fragment and, thus, greater accessibility to the nuclease, which is associated with more open chromatin and lower nucleosome density. Remarkably, *VIM-AS1* RNA depletion resulted in a three-to five-fold less accessible chromatin conformation in regions 2-6 within the C skew, whereas locations further upstream or downstream did not change significantly (Fig. 6*C*). A similar effect was seen upon overexpression of RNASEH1 (Fig. 6*D*). Accordingly, the quantity of histone H3 increased under conditions of antisense knockdown or RNASEH1 overexpression (Fig. S6). These results indicate that R-loop formation is necessary for maintaining an open chromatin and suggests that transcription of the minus strand relaxes local chromatin and possibly keeps the *VIM* template strand more accessible to the transcriptional machinery.

Antisense Transcription and R-loop Enhance NF-κB binding to the *VIM* Promoter. Enhancer binding sites and negative elements have been characterized for the *VIM* promoter (12, 13, 27). Specifically, binding sites for p65/RelA in the NF-κB pathway are present in regions 4 and 5 of the central region of the C skew (Fig. 6*A*). To determine whether R-loop formation can affect their binding, we performed ChIP experiments on cellular factors of the NF-κB pathway. In accordance with previous studies, binding of p65 following TNF-α stimulation was specifically enriched in regions 4 and 5 in SW480 cells (Fig. 6*E*). The same two regions displayed diminished binding following knockdown of the antisense *VIM-AS1* RNA (Fig. 6*F*) or overexpression of RNASEH1 (Fig. 6*G*). Taken together, our results indicate that transcriptional activation of *VIM* is supported through the cotranscriptional formation of a stable R-loop structure by a head-to-head antisense transcript. This regulatory mechanism could be a general characteristic of GC-rich promoters with divergent sense/antisense transcription and asymmetrically distributed G and C nucleotides. Interestingly, the high-mobility group protein *HMGA2* gene (plus strand) is transcribed head-to-head with the ribosomal protein SA pseudogene *RPSAP52* from a C skew-containing *locus* (Fig. S7*A*). Similar to *VIM-AS1*, *RPSAP52* transcription forms R-loop structures *in vitro* (Fig. S7*B*) and its depletion downregulates the sense *HMGA2* transcript (Fig. S7*C*) concomitantly with an increase in chromatin compaction, as measured by micrococcal nuclease accessibility assays (Fig. S7*D*).

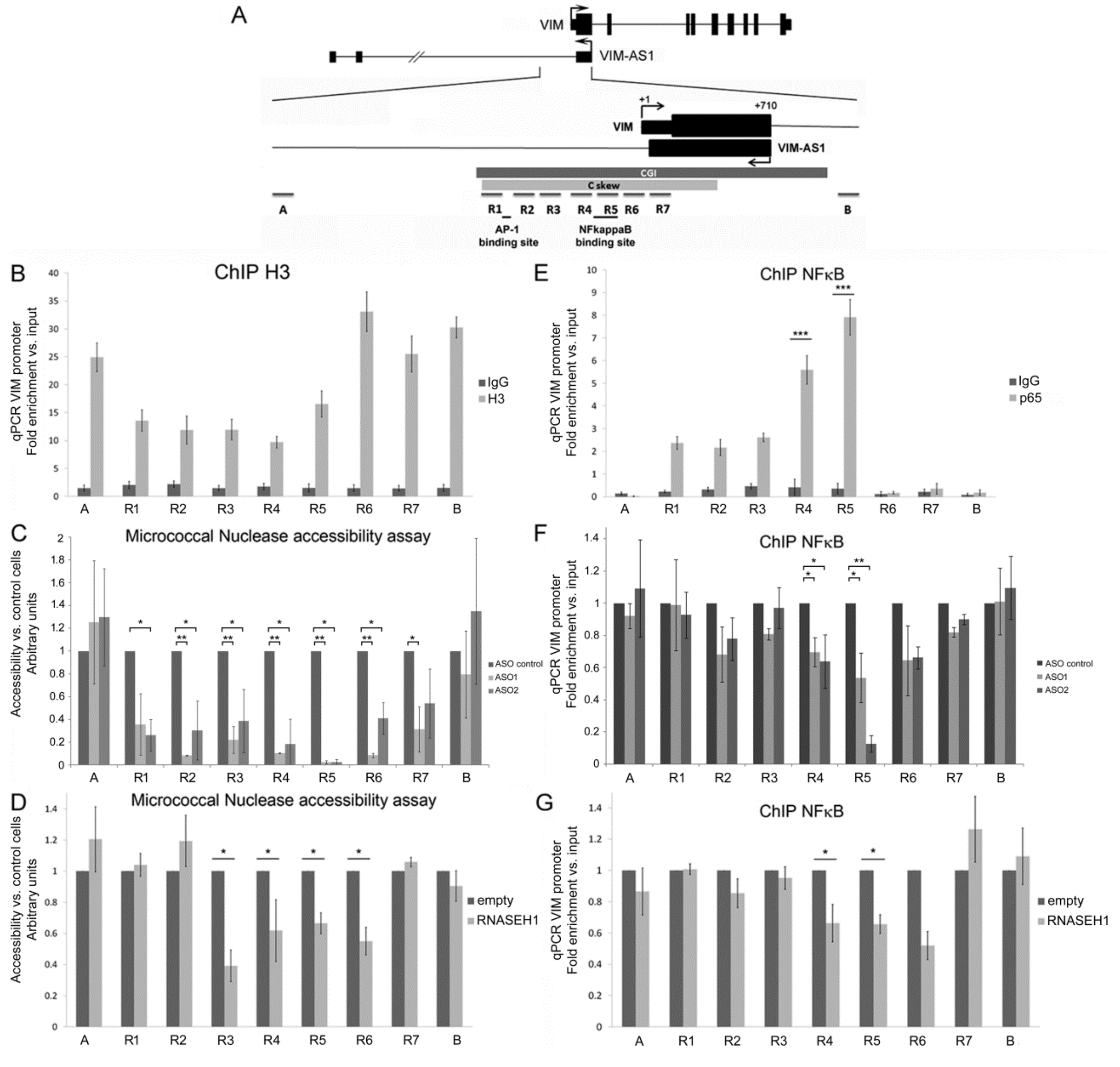


Fig. 6. Disruption of antisense transcription and of R-loop formation results in chromatin compaction and loss of NF-κB binding in the *VIM* gene promoter. (*A*) Fragments analyzed by qPCR in the *VIM* promoter. (*B*) Chromatin immunoprecipitation experiments with histone H3 antibody (H3) and control antibody (IgG) in SW480 cells. (*C*) Micrococcal nuclease accessibility assay on nuclei isolated from ASO-treated SW480 cells. (*D*) As in *C,* but comparing overexpression of RNASEH1 with empty vector. (*E*) Chromatin immunoprecipitation experiments with p65/RelA antibody or control (IgG) antibody in SW480 cells. (*F*) As in *E,* in ASO-treated cells. Levels were calculated relative to control samples. (*G*) As in *F,* but comparing overexpression of RNASEH1 with control-transfected cells. Throughout the figure, **P*<0.05; ***P*<0.01; ****P*<0.005 from Student's *t* test.

Discussion

Our results imply a positive role for R-loop formation by a head-to-head antisense transcript in the regulation of sense transcript expression. Originally considered to be rare transcriptional byproducts, R-loops may have a more general role as a mechanism of gene regulation (28-31). This regulatory mechanism is compatible with generally low levels of antisense RNA, because only two target molecules of DNA are present per cell. We have estimated the absolute abundance of *VIM-AS1* transcripts in different cell lines (Fig. S2) to correspond to a few copies per cell. Importantly, this amount represents the spliced transcript and may not reflect the actual abundance of the functional, intron-containing species. Related to this point, our RNA-FISH data suggests that the antisense region involved in R-loop formation is present as stable RNA in approximately one-third of cells in G2 phase, and in much lower levels in other phases of the cell cycle (Fig. 4). According to our DRIP experiments and taking into account the efficiency of the technique in our hands, we can estimate that, in nonsynchronous cultures, at most 20% of *VIM* promoters form an R-loop. Remarkably, R-loop abundance has been associated to cell cycle progression (32, 33). In this context, it is of note that *VIM* expression has long been known to be cell cycle dependent (34, 35). Further studies are needed to explore the detailed link between cell cycle and R-loop-mediated regulation of gene expression in the *VIM locus*.

Our data indicate that conditions that favor an R-loop lead to decreased nucleosome occupancy and increased binding of transcription factors of the NF-κB pathway, which are known to activate *VIM* expression upon mitogenic stimuli (36). Binding activity [by unknown protein factor(s)] specific to single-stranded DNA present on the minus strand immediately downstream of NF-κB binding elements has also been reported (37), although its regulatory potential is not known. Formation of the R-loop would enhance such binding. Alternatively, we cannot rule out the possibility that this binding indicates the presence of some unknown factor that stabilizes R-loop formation, as has been shown in *Arabidopsis* (38). Either way, the presence of a stable R-loop structure allows the maintenance of an open local chromatin conformation and enhances transcription factor binding to the displaced, single-stranded minus DNA strand. In summary, our results are consistent with a model (Fig. S8) in which an intact R-loop with participation of *VIM-AS1* transcript is essential for the optimal recognition of *VIM* promoter by transcriptional regulators, and specifically indicate activation by the NF-κB pathway, implicating R-loop structures in a previously unidentified positive role in gene transcription at the *loci* of bidirectional sense/antisense transcription.

Materials and Methods

Additional methods are described in the *SI Materials and Methods*.

***In vitro* R-loop formation assay.** R-loop formation was tested *in vitro* essentially as described in ref. 39. Genomic regions of the *VIM* or *HMGA2* promoter were PCR-amplified (with oligos VIMRloop1for and VIMRloop1rev, and HMGA2Rloop1for and HMGA2Rloop1rev, respectively) and cloned into pSPARK TA vector with the antisense strand under SP6 promoter. *In vitro* transcription reactions were carried out in both directions for 45 min at 37°C with either SP6 or T7 RNA polymerases in the presence of 0.15 μCi/μl of α-[^{32}P]-UTP, and further digested with RNaseA and RNaseH as indicated, for 30 min at 37°C. Nucleic acids were phenol-extracted, loaded onto a 1% agarose gel and run in 1x Tris/borate/EDTA. After electrophoresis, the gel was stained with SYBR® Safe DNA Gel Stain (Life Technologies) and UV-visualized. Following picture acquisition, the gel was dried and exposed to an autoradiography film for radiolabel detection.

DRIP. DRIP was performed as described in ref. 39. Genomic DNA was extracted from SW480 cells by SDS/Proteinase K treatment, phenol-chloroform extraction and ethanol precipitation. DNA was then digested with HindIII, EcoRI, XbaI and BamHI restriction enzymes. Samples were then either mock-treated or digested with RNaseH for a further 2 h. After phenol/chloroform extraction and precipitation, samples were resuspended in IP buffer (0.05% Triton X-100 in PBS) and immunoprecipitated with the anti-DNA-RNA hybrid (S9.6) antibody. Retrieved fragments were analyzed by qPCR and compared with appropriate dilution of input DNA. An amplicon from *GAPDH* promoter (lacking target sites for the restriction enzymes above) was used as a negative control.

Chromatin immunoprecipitation. NF-κB (p65) ChIP experiments were done as described (40). See the *SI Materials and Methods* for further details.

ACKNOWLEDGMENTS. We thank Dr. Verónica Dávalos for helpful discussions and A. Aguilera for the gift of the S9.6 antibody. This work was supported by Ministerio de Ciencia e Innovación (MICINN) Grants SAF2011-22895 (to S.G.) and SAF2011-22803 (to M.E.), La Marató de TV3 20131610 (to S.G.), the European Research Council Advanced Grant 268626–EPINORC project (to M.E.), Cellex Foundation, the Health and Science Departments of the Catalan Government (Generalitat de Catalunya), and Fondo de Investigación Sanitaria Grant PI08-1345 (M.E.). S.G. is funded by the Ramón y Cajal Research Program (MICINN). M.E. is an Institució Catalana de Recerca i Estudis Avançats Research Professor. R.B.-S. is a recipient of a Predoctoral Research Grant file BES-2012-058840 from Ministerio de Economía y Competitividad.

References

1. Lee JT (2012) Epigenetic regulation by long noncoding RNAs. *Science* 338(6113):1435–1439.
2. Morris KV & Mattick JS (2014) The rise of regulatory RNA. *Nat Rev Genet* 15(6):423–437.
3. Magistri M, Faghihi MA, St Laurent G 3rd, Wahlestedt C (2012) Regulation of chromatin structure by long noncoding RNAs: Focus on natural antisense transcripts. *Trends Genet.* 28(8):389–396.
4. Lai F, *et al*. (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494(7438):497–501.
5. Kang Y, Massagué J (2004) Epithelial-mesenchymal transitions: Twist in development and metastasis. *Cell* 118(3):277–279.
6. Tomikawa J, *et al*. (2011) Single-stranded noncoding RNAs mediate local epigenetic alterations at gene promoters in rat cell lines. *J Biol Chem* 286(40):34788–34799.
7. Zhou Z, Kahns S, Nielsen AL (2010) Identification of a novel vimentin promoter and mRNA isoform. *Mol Biol Rep* 37(5):2407–2413.
8. Kokkinos MI, *et al.* (2007) Vimentin and epithelial-mesenchymal transition in human breast cancer–observations in vitro and in vivo. *Cells Tissues Organs* 185(1-3):191–203.

9. Pieper FR, *et al*. (1992) Regulation of vimentin expression in cultured epithelial cells. *Eur J Biochem* 210(2):509–519.

10. Chen WD, *et al*. (2005) Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst* 97(15):1124–1132.

11. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW (2008) The antisense transcriptomes of human cells. *Science* 322(5909):1855–1857.

12. Rittling SR, Coutinho L, Amram T, Kolbe M (1989) AP-1/jun binding sites mediate serum inducibility of the human vimentin promoter. *Nucleic Acids Res* 17(4):1619–1633.

13. Lilienbaum A, Paulin D (1993) Activation of the human vimentin gene by the Tax human T-cell leukemia virus. I. Mechanisms of regulation by the NF-kappa B transcription factor. *J Biol Chem* 268(3):2180–2188.

14. Salvetti A, Lilienbaum A, Li Z, Paulin D, Gazzolo L (1993) Identification of a negative element in the human vimentin promoter: Modulation by the human T-cell leukemia virus type I Tax protein. *Mol Cell Biol* 13(1):89–97.

15. Chen JH, *et al*. (1996) PEA3 transactivates vimentin promoter in mammary epithelial and tumor cells. *Oncogene* 13(8):1667–1675.

16. Izmailova ES, Zehner ZE (1999) An antisilencer element is involved in the transcriptional regulation of the human vimentin gene. *Gene* 230(1):111–120.

17. Wu Y, Zhang X, Salmon M, Lin X, Zehner ZE (2007) TGFbeta1 regulation of vimentin gene expression during differentiation of the C2C12 skeletal myogenic cell line requires Smads, AP-1 and Sp1 family members. *Biochim Biophys Acta* 1773(3):427–439.

18. Wu Y, Zhang X, Salmon M, Zehner ZE (2007) The zinc finger repressor, ZBP-89, recruits histone deacetylase 1 to repress vimentin gene expression. *Genes Cells* 12(8):905–918.

19. Li M, *et al*. (2009) Sensitive digital quantification of DNA methylation in clinical samples. *Nat Biotechnol* 27(9):858–863.

20. Guttman M, *et al*. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235):223–227.

21. Cabili MN, *et al*. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 25(18):1915–1927.

22. Engreitz JM, *et al*. (2013) The Xist lncRNA exploits three-dimensional genome

architecture to spread across the X chromosome. *Science* 341(6147):1237973.

23. Aguilera A, García-Muse T (2012) R loops: From transcription byproducts to threats to genome stability. *Mol Cell* 46(2):115–124.
24. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* 4(5):442–451.
25. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* 45(6):814–825.
26. Schones DE, *et al.* (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell* 132(5):887–898.
27. Lilienbaum A, Duc Dodon M, Alexandre C, Gazzolo L, Paulin D (1990) Effect of the human T-cell leukemia virus type I tax protein on activation of the human vimentin gene. *J Virol* 64(1):256–263.
28. Powell WT, *et al.* (2013) R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc Natl Acad Sci USA* 110(34):13938–13943.
29. Ginno PA, Lim YW, Lott PL, Korf I, Chédin F (2013) GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* 23(10):1590–1600.
30. Bhatia V, *et al.* (2014) BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 511(7509):362–365.
31. Chan YA, *et al.* (2014) Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-Chip. *PLoS Genet* 10(4):e1004288.
32. Castellano-Pozo M, *et al.* (2013) R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol Cell* 52(4)583-90.
33. Herrera-Moyano E, Mergui X, García-Rubio ML, Barroso S, Aguilera A (2014) The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription–replication conflicts. *Genes Dev* 28, 735-48.
34. Giese G, Kubbies M, Traub P (1992) Cell cycle-dependent vimentin expression in elutriator-synchronized, TPA-treated MPC-11 mouse plasmacytoma cells. *Exp Cell Res* 200(1):118-25.
35. Giese G, Kubbies M, Traub P (1994) High resolution analysis of cell cycle-correlated vimentin expression in asynchronously grown, TPA-treated MPC-11 cells by the

novel flow cytometric multiparameter BrdU-Hoechst/PI and immunolabeling technique. *J Cell Physiol* 161(2):209-16.

36. Duprey P, Paulin D (1995) What can be learned from intermediate filament gene regulation in the mouse embryo. *Int J Dev Biol* 39(3):443–457.
37. Benazzouz A, Duprey P (1999) The vimentin promoter as a tool to analyze the early events of retinoic acid-induced differentiation of cultured embryonal carcinoma cells. *Differentiation* 65(3):171–180.
38. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C (2013) R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* 340(6132):619–621.
39. Yu K, Roy D, Huang FT, Lieber MR (2006) Detection and structural analysis of R-loops. *Methods Enzymol* 409:316–329.
40. Nowak DE, Tian B, Brasier AR (2005) Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques* 39(5)715–725.

Supporting Information

Boque-Sastre et al. 10.1073/pnas.1421197112

SI Materials and Methods

Cell Culture. HCT116, Caco-2, and SW480 human colon adenocarcinoma cell lines were cultured in DMEM (PAA Laboratories) containing stable glutamine, and supplemented with 10% (vol/vol) [20% (vol/vol) for Caco-2] heat-inactivated FBS (Invitrogen) and 1% penicillin–streptomycin (PAA Laboratories). MCF7 breast adenocarcinoma was cultured under the same conditions, but the medium was supplemented with 0.01 mg/mL human recombinant insulin. Nonmalignant MCF10A breast cells were grown in DMEM/ Ham’s F-12 medium (PAA Laboratories) supplemented with 20 ng/mL EGF (E9644; Sigma), 500 ng/mL hydrocortisone (H0888; Sigma), 10 mg/mL insulin (I0516; Sigma) and 100 ng/mL cholera toxin. All cells were grown at 37 °C in a humidified atmosphere of 5% (vol/vol) CO_2 and 95% (vol/vol) air.

Human Sample Methylation Analysis on Illumina's 450K Bead-Chip Array. DNA preparation, bisulfite conversion, and the analysis of methylation levels were done as described (1). An unpaired-samples Student's *t* test was performed to check for differentially methylated probes between normal and tumor samples. Comparisons with a difference in average beta value greater than 25% and an adjusted value of P (FDR) < 0.01 were considered to be statistically significant. DMR was defined as the region containing CpG sites with a significant corrected P value (FDR < 0.01) from a 2×2 χ^2 contingency test of the association of methylated and unmethylated cytosines with normal and tumor samples.

PolyA$^+$/PolyA$^-$ RNA Selection and Nuclear/Cytoplasmic Fractionation. PolyA$^+$ and polyA$^-$ RNAs were separated by using the Dynabeads mRNA Purification kit (61006; Life Technologies), using three rounds of selection. RNA enrichment in each fraction was then analyzed by RT-qPCR, using *GAPDH* or *U6* RNAs as controls, and normalizing relative to the percentage of RNA from each fraction used in the reverse transcription reaction. Subcellular fractionation was performed with a PARIS kit (Life Technologies; AM1921). Equal amounts of RNA from each fraction were subject to RT-qPCR, and the results were normalized taking into account the total amount of RNA recovered from each fraction. *PPiA* and *U6* RNAs were used as controls for fraction purity.

RNA-FISH and Actinomycin D Treatment. Two pools of 48 probes tiling either the first intron of *VIM-AS1* RNA (starting just downstream of the C skew region to avoid interference with the R-loop region) or the first intron of *VIM* RNA were designed following Stellaris RNA FISH Probe designer (Biosearch Technologies). *VIM-AS1* probes were coupled to TAMRA and VIM probes to FAM reporter dyes. Cell fixation, permeabilization, and probe hybridization were performed by following Stellaris FISH Protocols for adherent cells, with Vectashield Hardset (H1400; Vector Laboratories) as mounting medium. For actinomycin D experiments, synchronized cells were allowed to progress for 6 h into the G_2 phase and treated with either DMSO or 5 μg/mL actinomycin D (Sigma) for 30 min, before cell fixation and RNA-FISH analysis, RAP assays or RT-qPCR experiments.

DNA Methylation Analysis and *in Vivo* R-Loop Detection. DNA methylation was determined by PCR analysis after bisulfite modification of genomic DNA. Region 1 was amplified with primers bsVIM_R1for and bsVIM_R1rev and region 2 with primers bsVIM_R2for and bsVIM-R2rev (Table S1). *In vivo* R-loop was detected in SW480 cells in essentially the same way but with native overnight treatment with sodium bisulfite at 37 °C. The PCR was performed with a forward, native oligonucleotide (N in Fig. 5*C*) outside the C skew-containing region and a reverse, converted oligonucleotide (C in Fig. 5*C*), which takes into account the C-to-U changes (C-to-T after PCR) that occur only on the minus DNA strand following native bisulfite conversion. Following DNA purification, 32 cycles of PCR were carried out with the native forward primer (RLoop_st+_1_F1) and the converted reverse primer (RLoop_st+_1_R1) (Table S1).

Real-Time RT-qPCR. Total RNA from cell lines was extracted by using the TRIzol reagent (Invitrogen) and DNase treated with RQ1 DNase (Promega). For mRNA expression analysis, purified total RNA was reverse-transcribed by using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). Real-time PCRs were performed in triplicate in an Applied Biosystems 7900HT Fast Real-Time PCR system, using 100 ng of cDNA, 6 μL of SYBR Green PCR Master Mix (Applied Biosystems) and 416 nM primers (listed in Table S1) in a final volume of 12 μL for 384-well plates. All data were normalized with respect to two housekeeping genes (*L13* and *PBGD*), with no significant GC skew at their promoters as endogenous control. Relative RNA levels were calculated by using the comparative C_t method ($\Delta\Delta C_t$), considering the PCR efficiency. The order of magnitude of change is equal to $10^{\Delta\Delta Ct/m}$, where m is the average slope of the calibration curves for the gene of interest and the endogenous control.

Western Blot. Cell pellets were resuspended in lysis buffer (10% glycerol, 2% SDS wt/vol, 63 mM Tris·HCl pH 6.8, 0.01% bromophenol blue, 2% 2-mercaptoethanol), sonicated, and boiled for 5 min. Equal amounts of protein extracts were loaded onto Tris-Glycine-SDS gels and transferred to a nitrocellulose membrane (Whatman; GE Healthcare). Primary antibodies were diluted in 5% skimmed milk in TBS and incubated overnight at 4 °C. Final antibody concentrations were 1:5,000 for vimentin (CBL202; Millipore), 1:1,000 for RNaseH1 (H00246243-B01; Abnova), and 1:20,000 for β-actin-HRP (Sigma). After primary antibody incubations, membranes were washed three times (10 min each) with TBS containing 0.05% Tween-20 at RT on a bench-top shaker. Secondary antibodies conjugated to horseradish peroxidase were diluted to a concentration of 1:10,000 in 5% skimmed milk with TBS, containing 0.05% Tween-20.

Membranes were incubated with secondary antibody solutions for 1 h at room temperature (RT) in the dark in a bench-top shaker, washed three times (10 min each) with TBS containing 0.05% Tween-20 at RT, and then briefly rinsed in TBS before detection.

Cell Synchronization. MCF10A cells were synchronized by double thymidine block. Cells were treated with 2 mM thymidine for 14 h in medium supplemented with 10% FCS. After washing twice with PBS, cells were cultured in fresh medium/10% FCS for 10 h and treated again for 14 h with medium/10% FCS containing 2 mM thymidine. After washing cells with PBS, the block was released by the incubation of cells in fresh medium/10% FCS (time 0) and the cells were harvested at the indicated times. Cell cycle progression was detected by flow cytometric analysis.

RAP. The RAP protocol was performed as described by Engreitz et al. (2). Briefly, the 5′ and 3′ ends of the R-loop–forming region on *VIM-AS1* intron were tiled with 10 124-nt antisense RNA probes that had been biotinylated by *in vitro* transcription. The central region of the R-loop was devoid of probes to prevent interference in the RT-qPCR and the qPCR signal. MCF10A cells were synchronized as described above and cross-linked first with 2 mM DSG for 45 min at room temperature and then with 3% formaldehyde for 10 min at 37 °C. For each purification, 100 ng of biotinylated probes were added to the precleared lysates and the mixture was incubated at 45 °C. The probes were then captured by streptavidin beads, and the elutions for the associated RNA and DNA were performed. As a control, the same experiment was carried out in parallel with probes tiling the unrelated *LINC00085* nuclear RNA or with streptavidin beads without any probe. Recovered RNA and DNA samples were analyzed by RT-qPCR together with 1/10 dilution of the input material. Primer sequences for probe construction are available upon request.

Immunofluorescence. Cells were cultured directly on coverslips and fixed with 4% paraformaldehyde in water for 20 min at RT. Cells were permeabilized with 0.1% Triton X-100 in PBS for 5 min and blocked with 1% BSA for 1 h. Cells were then incubated with vimentin primary antibody (1:200, CBL202; Millipore) for 1 h at RT. Finally, 1:1,000 dilution of fluorescent-labeled secondary antibody from Invitrogen (anti-mouse IgG; A21235) was used. The coverslips were mounted on glass slides by using Mowiol with DAPI. Multicolor immunofluorescence imaging was then performed under a Leica SP5 laser scanning confocal spectral microscope (Leica Microsystems) equipped with Argon, DPSS561, HeNe633, and 405 Diode, and using a 63× oil immersion objective lens (N.A. 1.32). Data were analyzed by using the Fiji program.

DRIP. Genomic DNA was extracted from SW480 cells by SDS/proteinase K treatment and phenol-chloroform extraction and ethanol precipitation. DNA was then digested with HindIII, EcoRI, XbaI, and BamHI restriction enzymes. Samples were then either mock-treated or digested with RNaseH for a further 2 h. After phenol/chloroform extraction and precipitation, samples were resuspended in IP buffer (0.05% Triton X-100 in PBS) and immunoprecipitated with the anti-DNA-RNA hybrid (S9.6) antibody. Retrieved fragments were analyzed by qPCR and compared with appropriate dilution of input DNA. An amplicon from *GAPDH* promoter (lacking target sites for the restriction enzymes above) was used as a negative control.

Extraction of Nuclei and the Micrococcal Nuclease Accessibility Assay. Growing cells were trypsinized and washed twice with cold PBS. Cells were then resuspended in 1 mL of ice-cold RSB (Tris·HCl 10 mM pH 7.5, NaCl 10 mM, $MgCl_2$ 3 mM, protease inhibitors) adding Nonidet P-40 to a concentration of 1% and kept on ice for 10 min. After incubation, cells were centrifuged for 5 min at 800 × *g* at 4 °C. The supernatant was discarded, and nuclei were resuspended in RSB plus Nonidet P-40. Samples were centrifuged for 5 min at 800 × *g* at 4 °C. Nuclei were washed with medium salt buffer without Nonidet P-40 and centrifuged for 5 min at 2,300 × *g* at 4 °C. The supernatant was discarded and the nuclei were resuspended in 1× micrococcal nuclease buffer to give 10^6 nuclei per 800 μL. Nuclei from each cell condition were digested in 15 U of micrococcal nuclease S7 restriction enzyme (Roche Applied Science) in a series of increasing incubations at 37 °C: 0, 2, 5, and 10 min. Reactions were stopped by adding 200 μL of stop solution (20 mM Tris·HCl pH 7.5, 0.6 M NaCl, 1% SDS, 10 mM EDTA, and 400 μg/mL proteinase K) and incubating at 37 °C for 2 h. DNA was purified by phenol/chloroform extraction and ethanol precipitation. Amplicons were amplified and quantified by real-time PCR in an Applied Biosystems 7900HT Fast Real-Time PCR System. Results were normalized with respect to a *Sat2* region, which was expected to be extremely compact and, thus, less accessible to micrococcal nuclease. Undigested samples were analyzed relative to 10 min digested samples, because these conditions revealed the maximum differences. Data were then normalized to consider control samples equal to 1.

ChIP. In brief, 5 × 10^6 lentivirus-transfected SW480 cells were seeded on 100-mm dishes. After reattachment, cells were serum-deprived overnight in DMEM supplemented with 0.5% BSA. The following day, cells were stimulated for 30 min with 30 ng/mL TNF-α, washed in PBS and cross-linked twice, first with 2 mM Di (N-succinimidyl) glutarate (DSG; Sigma) for 45 min, and then with 1% formaldehyde for 15 min. Cells were lysed in buffer L1 [50 mM Tris·HCl, pH 8.0, 2 mM EDTA, 0.1% Nonidet P-40 (IGEPAL CA-630, Sigma-Aldrich), 10% glycerol, 1 mM DTT, protease inhibitors] for 15 min on ice, and the nuclei pelleted and resuspended in 500 μL of SDS lysis buffer (50 mM Tris·HCl, pH 8.0, 10 mM EDTA, 1% SDS). Chromatin was sonicated in a Bioruptor (Diagenode) to obtain chromatin fragments of about 150–400 bp. Eighteen A_{260} units of chromatin were used as the input for each immunoprecipitation. Chromatin extracts were precleared overnight with 20 μL of Dynabeads M-280 Sheep Anti-Rabbit IgG (Invitrogen). 4 μg of NfkappaB (p65) rabbit polyclonal antibody (Santa Cruz Biotechnology; sc-109), 1 μg of rabbit polyclonal antibody against H3 (ab1791; Abcam), or 2 μg of normal rabbit IgG control antibody (12–370; Millipore) were coupled overnight to 20 μL of Dynabeads. Precleared extracts were incubated with the Ab–beads complexes for 4 h at 4 °C. After washing, the recovered material was reverse cross-linked with proteinase K, phenol/chloroform extraction, and ethanol precipitation. Immunoprecipitated DNA and 1:50 diluted input sample were analyzed in triplicate by real-time qPCR analyses by using SYBR-Green Master Mix in an ABI 7900 FAST sequence detection system. The primers used are shown in Table S1.

Plasmid Construction and Transfections. Human RNaseH1 lacking the *N*-t mitochondrial localization signal (MLS) was cloned with oligos RNaseHdelMLSEcoRIfor and RNaseHBamHIrev into the EcoRI and BamHI sites of lentiviral expression vector pLVX-IRES-ZsGreen1 (Clontech). shRNA2 and shRNA3 target the 5′ GGTGTACTAGTGAAGTGAT 3′ and 5′ TCCAAATGTGCTACTCAGA 3′ sequences, respectively, of VIM-AS1 mRNA (both located on the last exon), and were expressed by cloning oligos shVIMAS2for and shVIMAS2rev (for sh2) and shVIMAS3for and shVIMAS3rev (for sh3) into the BamHI and EcoRI sites of vector pLVX-shRNA2 (Clontech). Table S1 contains the full list of oligos used for cloning.

For lentivirus-mediated construct overexpression, HEK293T cells were transfected with pLVX-IRES-ZsGreen1-RNaseH1 construct or pLVX-shRNA2-constructs plus packaging plasmids with jetPRIME (Polyplus-transfection) according to the manufacturer’s recommendations. Forty-eight hours after transfection, the supernatant containing viral particles was used to infect target cell lines. ZsGreen1 was used in both cases as a marker to visualize transductants by fluorescence microscopy. In the case of SW480 cells, RNaseH1-transfected cells were enriched by fluorescence-activated cell sorting (FACS) before extract preparation and Western blot analysis.

Transfection with antisense oligonucleotides (LNA GapmeRs, 300600; Exiqon) was carried out as follows: SW480 cells were seeded at 1 × 10^6 cells per well in six-well plates. Transfection mixes were prepared by using HiPerfect (Qiagen) and LNA GapmeRs to a final concentration of 65 nM. Cells were retransfected 48 h later and collected 72 h after the second round of LNA treatment. A control LNA GapmeR (300610; Exiqon) was used as mock transfection.

1. Sandoval J, et al. (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6(6):692–702.

2. Engreitz JM, et al. (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341(6147):1237973.

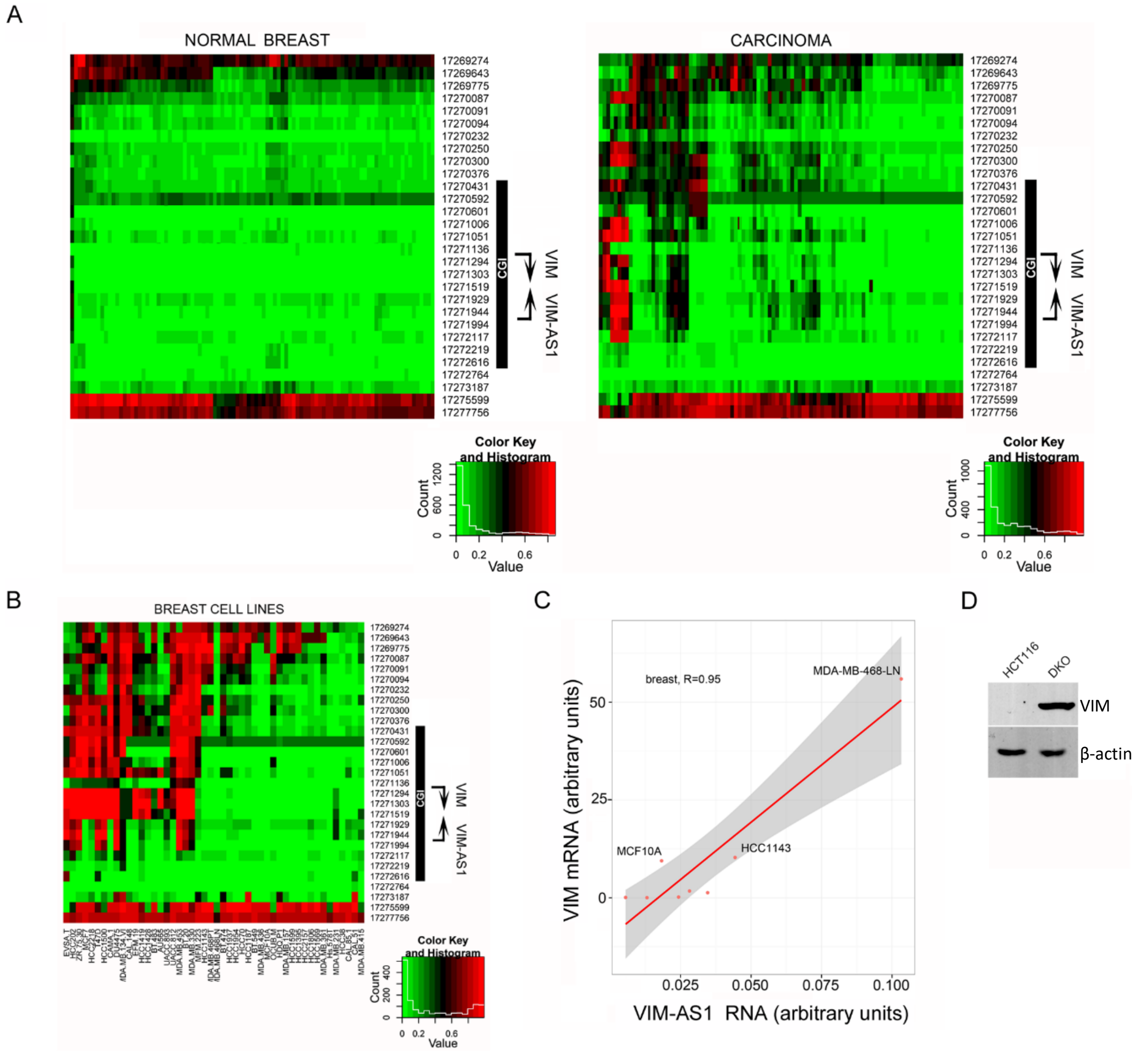
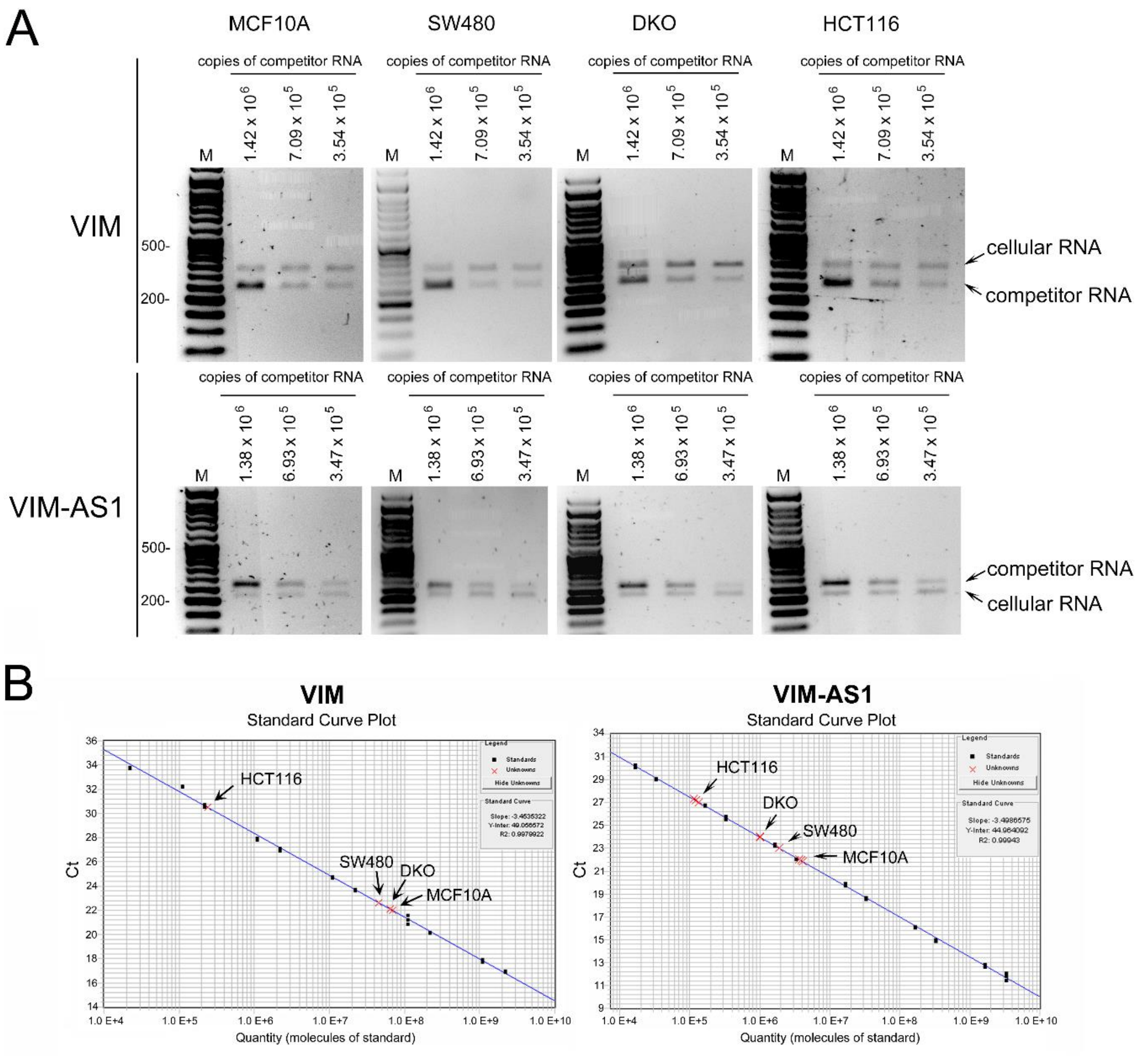


Fig. S1. Sense/antisense transcripts at the vimentin *locus* are positively correlated and reduced in hypermethylated contexts: related to Fig. 2. (*A*) Heatmap representation of a DNA methylation microarray analysis of 97 human normal breast (*Left*) and 97 tumor (*Right*) samples indicates that hypermethylation of *VIM* promoter CGI is a hallmark of cancer. Individual samples are represented along the horizontal axis. CpG sites surrounding *VIM* and *VIM-AS1* transcription start sites are displayed vertically, with the exact CpG coordinate on chromosome 10 indicated. The CGI is represented as a thick line to the right of each plot. (*B*) Heatmap representation of a DNA methylation microarray analysis of 48 human breast adenocarcinoma cell lines (indicated along the *x* axis). (*C*) Positive Pearson's correlation coefficients between *VIM* (*y* axis) and *VIM-AS1* (*x* axis) expression values for eight breast cell lines (HCC1143, MCF10A, MDA-MB-468 LN, MDA-MB-468 PT, MCF7, T47D, MDA-MB-134 VI, BT-474) indicate coordinated expression. The linear trend is displayed. MDA-MB-468-LN, MCF10A, and HCC1143 cell lines, with a hypomethylated CGI, have the highest levels of expression for both transcripts and are highlighted in the plot. (*D*) Promoter CGI hypermethylation is inversely correlated with vimentin protein levels. Western blot analysis of HCT116 colon cell line (hypermethylated at *VIM* promoter CGI) and its derivative DKO (hypomorph of DNMT1 and DNMT3b and hypomethylated at *VIM* promoter CGI).



C

molecules per cell

	VIM	VIM-AS1
HCT116	26-60	0.3-4
DKO	4,000-5,600	1.8-8
SW480	2,300-3,500	3.5-15
MCF10A	15,100-15,900	4.4-13.5

Fig. S2. Absolute quantitation of *VIM* and *VIM-AS1* transcripts. (*A*) Semiquantitative RT-PCR of total RNA from the cell lines indicated in comparison with *in vitro* transcribed competitor RNA. The competitor RNA is identical to the endogenous amplicon but for the inclusion of a 40-nt spacer sequence (in the case of *VIM-AS1*) or a deletion of 75 nt (in the case of *VIM*), resulting in PCR bands of slightly different size that can be resolved in an agarose gel. The amount of competitor RNA included in the PCR is indicated. The amount of total RNA used of each cell line was as follows: for MCF10A, SW480, and DKO, 0.16 ng in *VIM* PCR and 4 ng in *VIM-AS1* PCR. For HCT116, 12 ng in *VIM* PCR and 40 ng in *VIM-AS1* PCR. (*B*) RT-qPCR using *in vitro* transcribed templates as standard for comparison. Increasing amounts (indicated in graphs) of *VIM* or *VIM-AS1* standards and 40 ng of total cellular RNA were retrotranscribed and amplified in parallel by qPCR. Absolute abundance for each transcript was estimated by C_t comparison with the standard curve. C_t values corresponding to each cell line tested are indicated by the red crosses. (*C*) Absolute estimations of number of *VIM* and *VIM-AS1* molecules per cell, derived from both semiquantitative and quantitative RT-PCR.

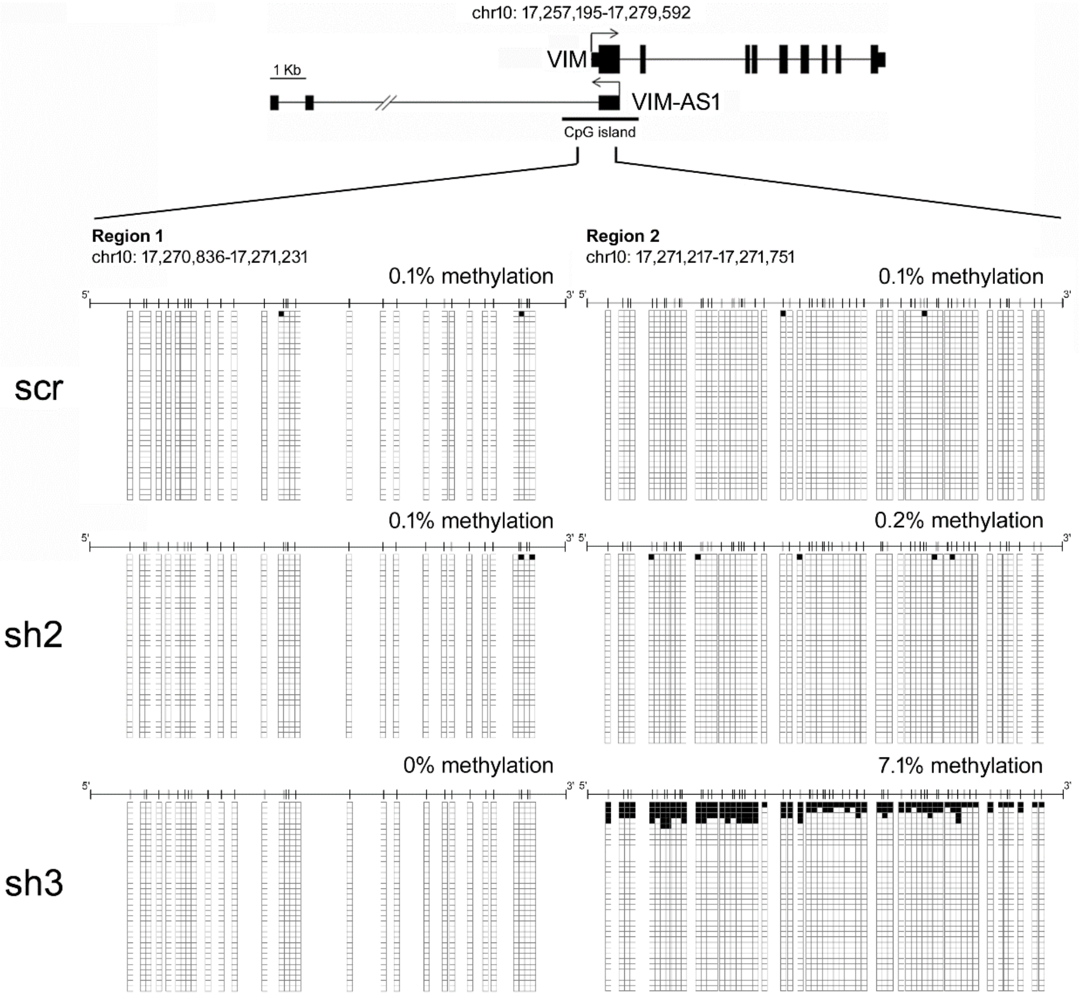


Fig. S3. DNA sequencing following bisulfite treatment of regions 1 and 2 within *VIM* promoter CGI upon shRNA-mediated knockdown.

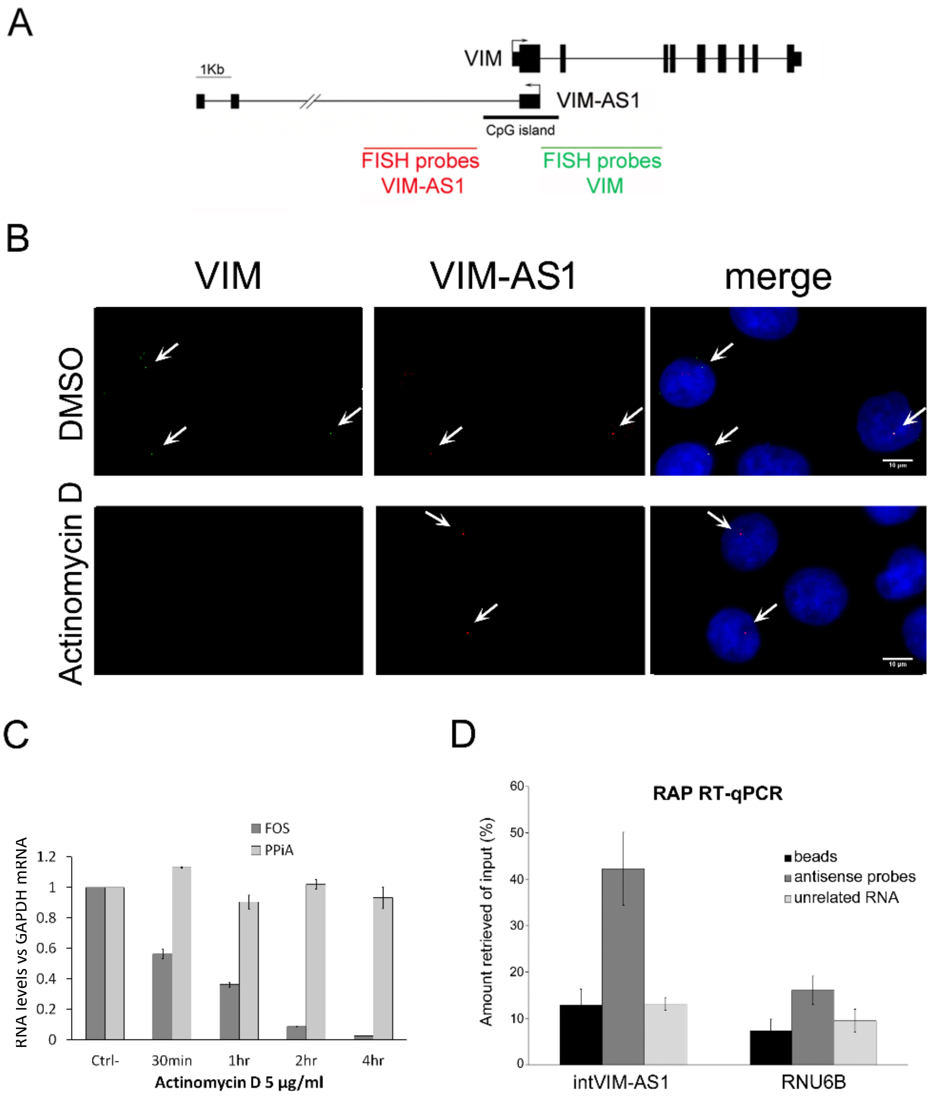


Fig. S4. *VIM-AS1* intron 1 locates to the transcription site even upon Act. D treatment and can be recovered in RAP experiments: related to Fig. 4. (*A* and *B*) RNA FISH with probes targeting *VIM* intron 1 (in green) or *VIM-AS1* intron 1 (in red) in control (DMSO-treated) or actinomycin D-treated MCF10A cells. Cell nuclei were stained with DAPI. (*C* and *D*) Reverse transcription-qPCR of the RNA captured in cross-linked MCF10A cells treated as in *B* by using streptavidin beads alone (beads), with antisense probes to *VIM-AS1* intron 1 (antisense probes) or against the *LINC00085* RNA (unrelated RNA). Enrichments represent means from two replicate experiments and are relative to the input amount used per pulldown. *RNU6B* is used as negative control to assess binding specificity.

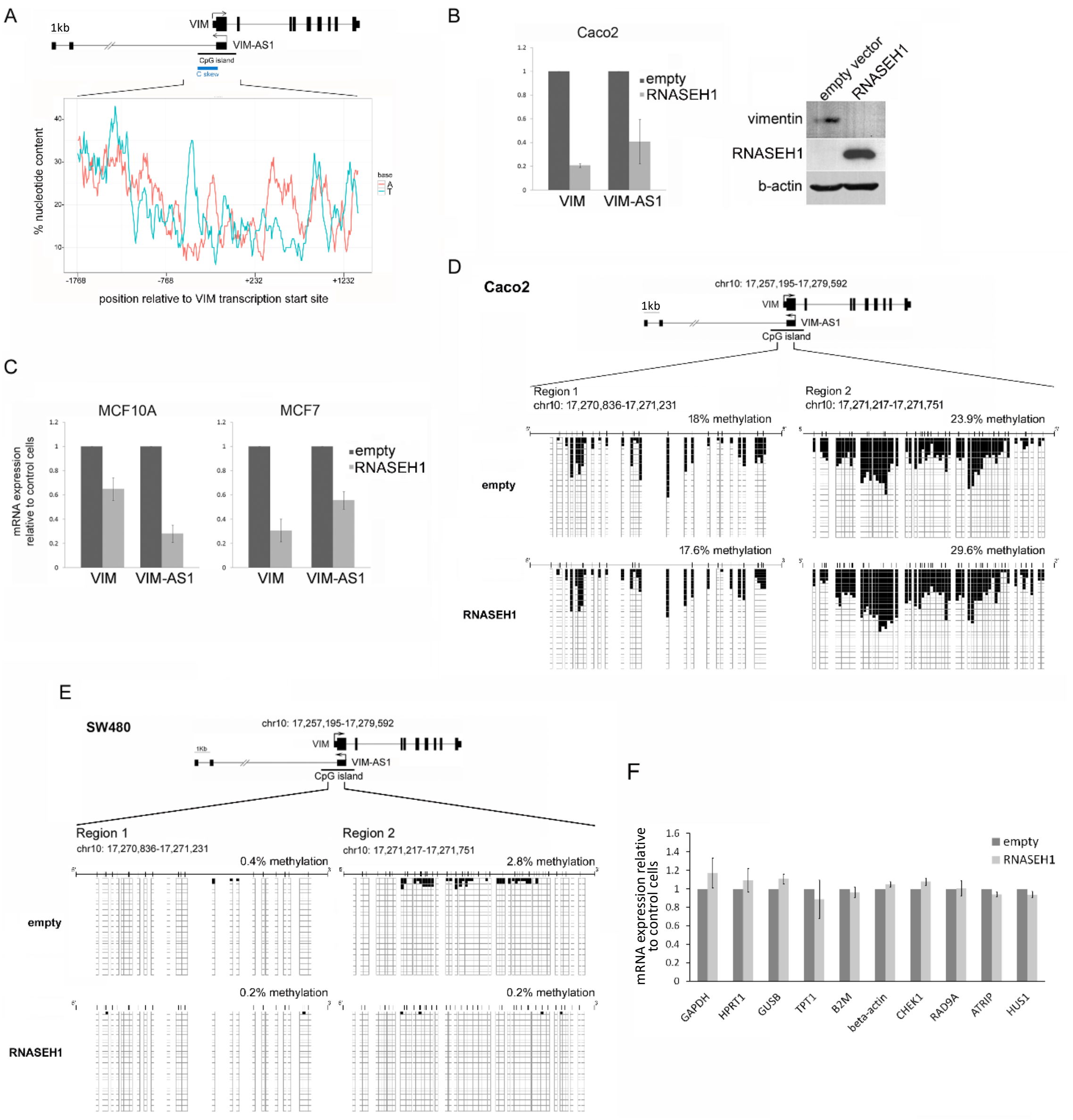


Fig. S5. Overexpression of RNASEH1 reduces *VIM* expression and is accompanied by a slight change in DNA methylation: related to Fig. 5. (*A*) Percentage of A and T nucleotides in the *VIM* promoter. The sequence shown corresponds to the plus strand, and for each position, the percentage abundance of each nucleotide within the surrounding 100 nt is counted, with a sliding window of 1 nt. For clarity, only A (red line) and T (blue line) nucleotides are displayed. The C skew-containing region shown in Fig. 5*A* is indicated by the discontinuous blue line. (*B*) Overexpression of RNASEH1 in Caco2 cells reduces both *VIM* and *VIM-AS1* mRNA expression, as measured by RT-qPCR (*Left*) and vimentin protein levels, as revealed by Western blot (*Right*). Error bars, SDs from three independent experiments. (*C*) Overexpression of RNASEH1 in the nonmalignant breast cell line MCF10A and the breast adenocarcinoma MCF7 reduce *VIM* and *VIM-AS1* RNA expression, as measured by RT-qPCR experiments. Error bars, SDs from three independent experiments. (*D* and *E*) DNA sequencing following bisulfite treatment of regions 1 and 2 within *VIM* promoter CGI reveals a moderate change in methylation upon RNASEH1 overexpression in Caco2 cells but not in SW480 cells. Region 2 overlaps with region 1 so that every CpG dinucleotide is interrogated between coordinates chr10:17,270,836 and 17,271,751 (hg19). Vertical lines represent CpG positions in the whole sequence. Individual clones sequenced are represented horizontally, with empty squares corresponding to unmethylated CpGs and filled squares corresponding to methylated positions. Average methylation levels for each group are indicated. (*F*) RT-qPCR analysis of the mRNA expression levels of the indicated genes in control or RNASEH1-overexpressing SW480 cells.

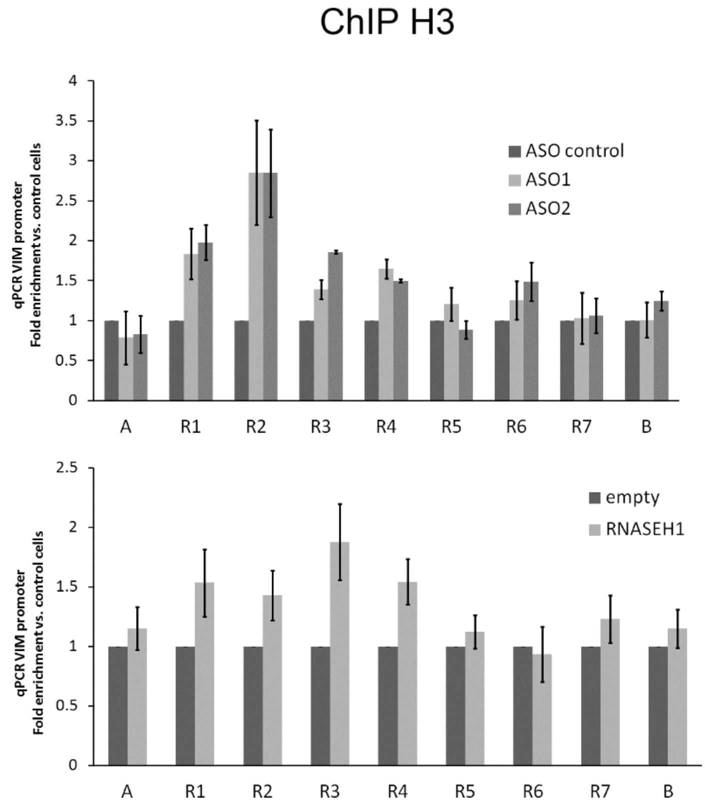


Fig. S6. *VIM-AS1* knockdown or R-loop disruption results in an increase in histone H3 density: related to Fig. 6. (*Upper*) ChIP experiments with histone H3 antibody in control (ASO control) or ASOs against *VIM-AS1*-overexpressing SW480 cells. The pulled-down DNA was measured by qPCR, and the values shown are relative to those of the control sample. Regions analyzed are the same as in Fig. 6. (*Lower*) The same experiment but comparing control or RNASEH1-overexpressing cells. Error bars, SDs from three independent experiments.

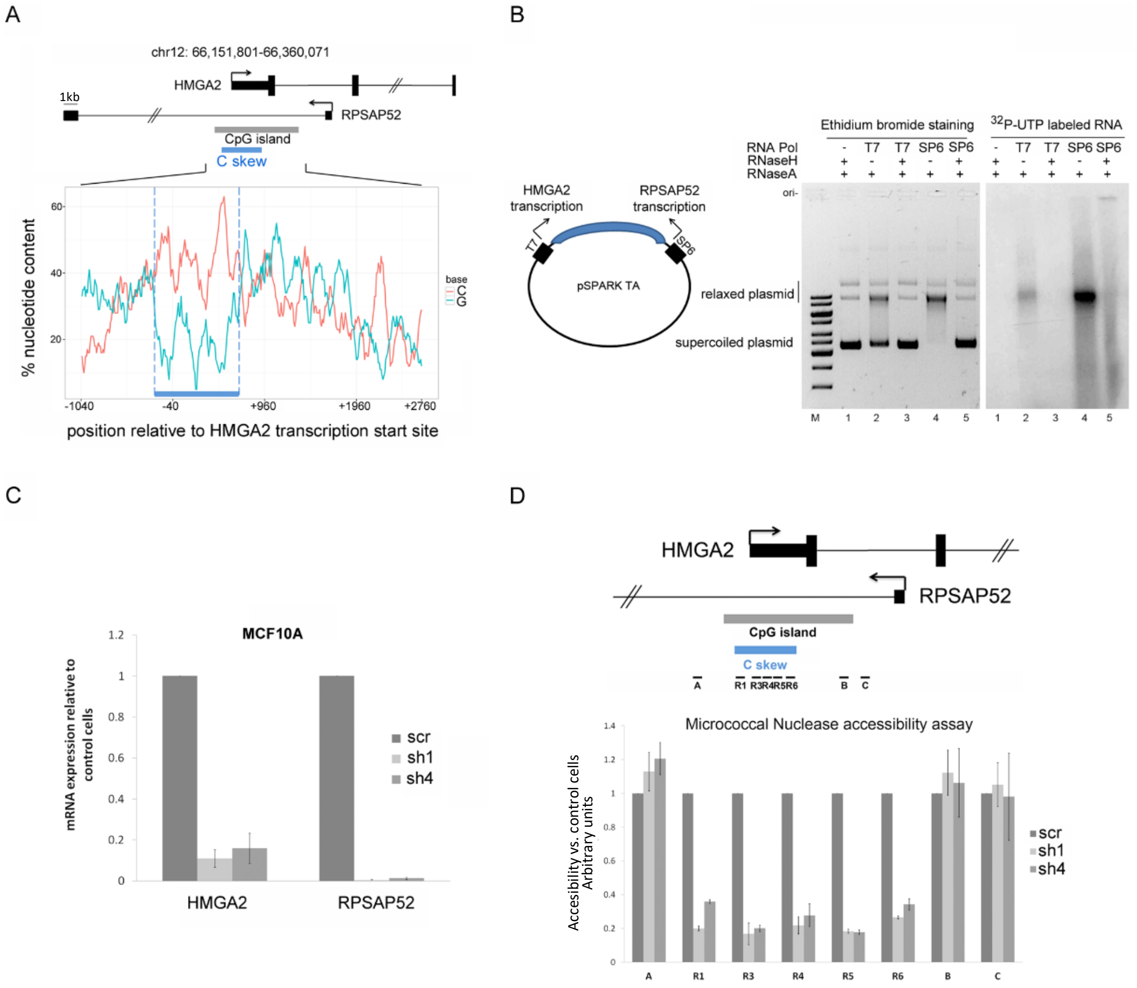


Fig. S7. Antisense-mediated R-loop formation at the *HMGA2 locus* promotes open chromatin conformation and sense transcription. (*A*) Percentage of C and G nucleotides in the *HMGA2* promoter reveals the presence of a 1-kb-long C skew. The upper diagram shows the intronic/exonic organization of *HMGA2* and its antisense transcript, the pseudogene *RPSAP52*. In the lower plot, the sequence represented corresponds to the plus strand, and for each position, the percentage abundance of each nucleotide within the surrounding 100 nt is counted, with a sliding window of 1 nt. For clarity, only C (red line) and G (blue line) nucleotides are displayed. The C skew-containing region is indicated by the thick blue line at the bottom of the diagram. (*B*) The *in vitro* R-loop formation assay indicates the participation of the *RPSAP52* transcript. The region containing the C skew (in blue) was cloned between the T7 and SP6 promoters, and *in vitro* transcription was carried out with either polymerase and in the presence of α-^{32}P-UTP. *HMGA2* transcription corresponds to T7 orientation, whereas antisense *RPSAP52* transcription is under the SP6 promoter (see diagram at left). To reveal RNA:DNA hybrids, the reactions were incubated in the absence (lanes 2 and 4) or presence (lanes 3 and 5) of bacterial recombinant RNaseH. After resolving in a 1% agarose gel, DNA bands were first stained with SYBR Safe (*Left*) and the same gel was then exposed for autoradiography to detect transcribed RNA (*Right*). As a control, a mock reaction without polymerase was also analyzed (lane 1). M, 1-kb DNA ladder. (*C*) MCF10A cells were transduced with lentiviral plasmids overexpressing control shRNA (scr) or shRNAs against *RPSAP52* RNA (sh1, sh4). RT-qPCR analysis of *HMGA2* and *RPSAP52* expression shows a clear reduction of both RNAs. Changes in expression for each case were calculated relative to control cells. Error bars show SDs calculated from two independent experiments. (*D*) Micrococcal nuclease accessibility assay on nuclei isolated from control (scr) or MCF10A cells overexpressing shRNAs against *RPSAP52* (sh1, sh4). The upper diagram indicates the fragments analyzed by qPCR along the *HMGA2* promoter. Regions R1–R6 are within the C skew. A–C fragments are control amplicons outside the C skew region. After chromatin digestion, levels of recovered DNA were estimated by qPCR with primer pairs for each indicated region. Each sample was first normalized against the *Sat2* region (considered constant and highly compacted) to allow comparison between different nuclei preparations. For each primer pair, ΔC$_t$ values were calculated as (0 min digestion) – C$_t$ (10 min digestion), whereby higher values indicate more accessible chromatin. Final levels are presented relative to those of control transfected cells. Error bars, SDs from two independent experiments.

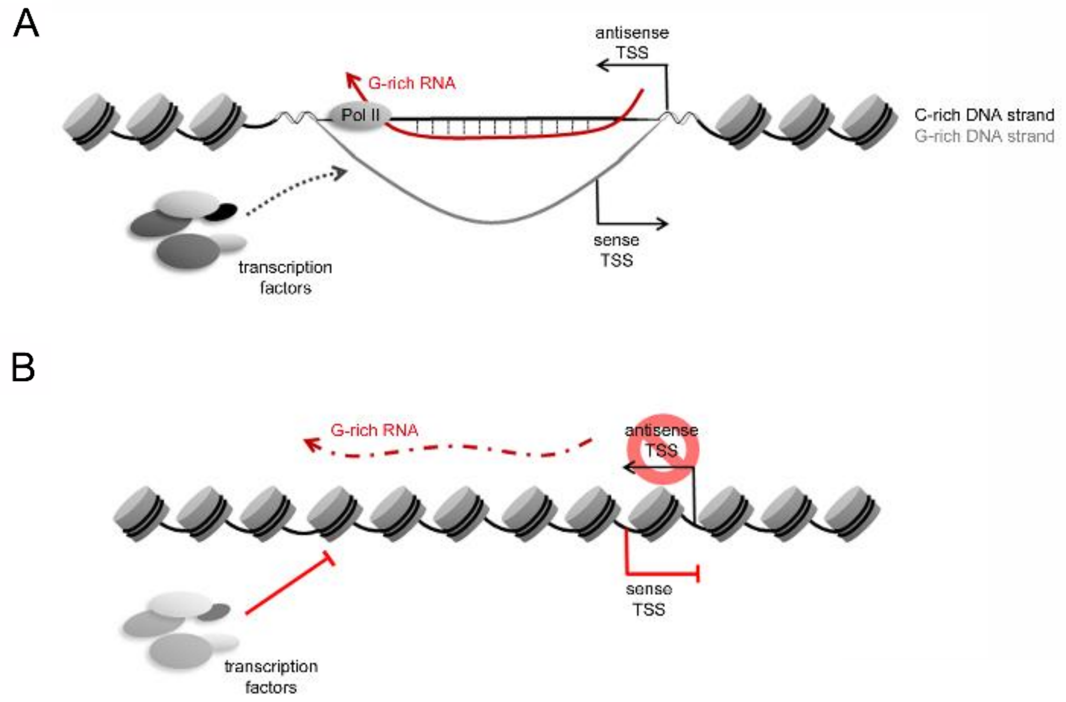


Fig. S8. A model for R-loop involvement in transcriptional activation with the participation of antisense transcripts. (*A*) At the *VIM locus* (and possibly other sites of divergent transcription), an R-loop is formed between the nascent, G-rich antisense transcript and the C-rich DNA strand. This structure maintains a local open chromatin conformation that allows for efficient recognition and binding of transcription factors, resulting in the sustained activation of nearby sense transcription. (*B*) Upon decrease of antisense transcription (for example, under hypermethylated conditions), R-loop formation is prevented and the region remains compacted, inhibiting sense transcription.

Table S1. Primer sequences used in this work

Oligo name	Sequence 5'-3'
RNaseHdelMLSEcoRIfor	CTGGAATTCATGTTCTATGCCGTGAGGAGGG
RNaseHBamHIrev	CTGGGATCCCTAGTCTTCCGATTGTTTAGCTCC
shVIMAS2for	GATCCGGTGTACTAGTGAAGTGATTTCAAGAGAATCACTTCACTAGTACACCTTTTTTACGCGTG
shVIMAS2rev	AATTCACGCGTAAAAAAGGTGTACTAGTGAAGTGATTCTCTTGAAATCACTTCACTAGTACACCG
shVIMAS3for	GATCCGTCCAAATGTGCTACTCAGATTCAAGAGATCTGAGTAGCACATTTGGATTTTTTACGCGTG
shVIMAS3rev	AATTCACGCGTAAAAAATCCAAATGTGCTACTCAGATCTCTTGAATCTGAGTAGCACATTTGGACG
VIM_Afor	GCCTAAAAGAGGCTTGTCCA
VIM_Arev	CAGGGGGTACTGCAGGTTACT
VIM_Bfor	CGAAAACACCCTGCAATCTT
VIM_Brev	AATTGCTCGTGGGTTGTGTT
VIM_R1for	GGCCCAGCTGTAAGTTGGTA
VIM_R1rev	AGGGGAAACCGTTAGACCAG
VIM_R2for	GGACTGAGCCCGTTAGGTC
VIM_R2rev	CCTCTGTCCATCGACTTGC
VIM_R3for	CAATCTCAGGCGCTCTTTGT
VIM_R3rev	GAGCGGGAAGAGGAAAGAGT
VIM_R4for	ACCGGACCCCTCTGGTTC
VIM_R4rev	ACCCTGGGGTGCTGAAAA
VIM_R5for	GAAAGCCCCCAAAAGTCC
VIM_R5rev	CCTCGAGCCTTCCTGCTC
VIM_R6for	GAGGGGACCCTCTTTCCTAA
VIM_R6rev	GGAGCGAGAGTGGCAGAG
VIM_R7for	CCTCCTACCGCAGGATGTT
VIM_R7rev	GGTGGACGTAGTCACGTAGC
GAPDH_DRIPfor	AGAGAAACCCGGGAGGCTA
GAPDH_DRIPrev	TGACTCCGACCTTCACCTTC
qVIMfor	GGCTCAGATTCAGGAACAGC
qVIMrev	GCTTCAACGGCAAAGTTCTC
qVIM-AS1for	CAAAGCTCCCTTTGGATGAC
qVIM-AS1rev	ACTAGTACACCCCCGACGTG
VIMRloop1 for	TCTCCAAAGGCTGCAGAAGT
VIMRloop1rev	ATGATGTCCTCGGCCAGGTT
HMGA2Rloop1for	AGACGCTTCCTGCAAAGTGT
HMGA2Rloop1rev	TGGAGGTAGCAAGAGGAGGA
RLoop_st+_1_F1	AGACAGGCTTTAGCGAGTTATT
RLoop_st+_1_R1	AATAGGGATTTAGTGAGAAGTG
bsVIM_R1for	GATTTGAGGGATTTTTTATTTTTTT
bsVIM_R1rev	AAAAAATCCCCTCCCACT
bsVIM_R2for	GGGAGGGGATTTTTTTTTTTA
bsVIM-R2rev	CAACTCCTACAACTCCACCTTC
HMGA2_Afor	GGGATGGAGGCTCTCTCTCT
HMGA2_Arev	CACTTTGCTGCACGTTGAGT
HMGA2_Bfor	TTGAGTAGGGGACGATCGAG
HMGA2_Brev	GCACGCTTAATTGGTTGCAT
HMGA2_Cfor	ATTTAGACTGGAGGCCATGC
HMGA2_Crev	TGGGAGGTTTTGCTTGAATC
HMGA2_1for	CTCCGGGACAGTCACGTT
HMGA2_1rev	CTAGCTCCACCCGCCTCT
HMGA2_3for	CACGATTAGAGGTGGGCACT
HMGA2_3rev	TGTGAGTGTGAGTGTGTGTGG
HMGA2_4for	GAATCTTGGGGCAGGAACTC
HMGA2_4rev	GGCTGCTAGCTCCTGAGTCTT
HMGA2_5for	GGTGCCACCCACTACTCTGT
HMGA2_5rev	CAAAGGAGGATGGGGAGACT
HMGA2_6for	GCAACTCCTGATCCCAACC
HMGA2_6rev	TGGAGGTAGCAAGAGGAGGA

STUDY II

The transcribed pseudogene *RPSAP52* enhances the oncofetal HMGA2-IGF2BP2-RAS axis through LIN28B-dependent and independent *let-7* inhibition

Cristina Oliveira-Mateos[1], Anaís Sánchez-Castillo[1,2], Marta Soler[1], Aida Obiols-Guardia[1], David Piñeyro[1], Raquel Boque-Sastre[1,3], Maria E. Calleja-Cervantes[1], Manuel Castro de Moura[1], Anna Martínez-Cardús[1], Teresa Rubio[4], Joffrey Pelletier[4], Maria Martínez-Iniesta[5], David Herrero-Martín[6], Oscar M. Tirado[2,6], Antonio Gentilella[4,7], Alberto Villanueva[5], Manel Esteller[1,2,8,9,10], Lourdes Farré[5,11] & Sonia Guil[1,10]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain. [2]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Carlos III Institute of Health (ISCIII), Madrid, Spain. [3]Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3AX Wales, UK. [4]Laboratory of Cancer Metabolism, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain. [5]Program Against Cancer Therapeutic Resistance (ProCURE), ICO, IDIBELL, L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain. [6]Sarcoma Research Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain. [7]Department of Biochemistry and Physiology, Faculty of Pharmacy, University of Barcelona (UB), Barcelona, Catalonia, Spain. [8]Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain. [9]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain. [10]Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Catalonia, Spain. [11]Laboratory of Experimental Pathology (LAPEX), Gonçalo Moniz Research Center, Oswaldo Cruz Foundation (CPQGM/FIOCRUZ), Salvador, Bahia, Brazil. Correspondence and requests for materials should be addressed to L.F. (email: mfarre@idibell.cat) or to S.G. (email: sguil@carrerasresearch.org)

Abstract

One largely unknown question in cell biology is the discrimination between inconsequential and functional transcriptional events with relevant regulatory functions. Here, we find that the oncofetal *HMGA2* gene is aberrantly reexpressed in many tumor types together with its antisense transcribed pseudogene *RPSAP52*. *RPSAP52* is abundantly present in the cytoplasm, where it interacts with the RNA binding protein IGF2BP2/IMP2, facilitating its binding to mRNA targets, promoting their translation by mediating their recruitment on polysomes and enhancing proliferative and self-renewal pathways. Notably, downregulation of *RPSAP52* impairs the balance between the oncogene LIN28B and the tumor suppressor *let-7* family of miRNAs, inhibits cellular proliferation and migration *in vitro* and slows down tumor growth *in vivo*. In addition, high levels of *RPSAP52* in patient samples associate with a worse prognosis in sarcomas. Overall, we reveal the roles of a transcribed pseudogene that may display properties of an oncofetal master regulator in human cancers.

Introduction

The largest part of the mammalian genome is transcribed into RNA species with little or no coding potential, known as noncoding RNAs (ncRNAs)[1]. Although their biological roles are still largely unknown, a growing number of long noncoding RNAs (lncRNAs, a label arbitrarily assigned to transcripts longer than 200 nucleotides) display regulatory properties by acting at all levels in gene expression control (from epigenetic modifications and chromatin dynamics to the control of post-transcriptional messenger RNA stability and translation)[2,3]. In some cases, their key functions in normal homeostasis and development links the dysregulation of their expression with causal roles in cancer[4], and there are instances of lncRNAs involved in each of the cancer hallmarks, including sustained proliferative signaling and growth (e.g., *ANRIL*[5], *lincRNA-p21*[6], *MEG3*[7]), invasion and metastasis (e.g., *HULC*[8], *MALAT1*[9], *HOTAIR*[10]), resistance to cell death (e.g., *PCGEM1*[11]), and replicative immortality (e.g., *TERC*[12], *TERRA*[13]). Mechanisms of action include the interaction with other nucleic acids and/or protein factors, which confers the ability to function as scaffolds, guides, decoys, or allosteric regulators of several nuclear or cytoplasmic processes[14,15]. In a growing number of examples, their roles intertwine with that of the better studied miRNAs[16], either by cooperating in their function[17] or by impairment of the miRNA-mediated regulation[18]. The latest annotation in GENCODE estimates that up to 16,000 genes in the human genome correspond to lncRNAs, and a similar number is given to pseudogenes (https://www.gencodegenes.org/stats/current.html#). Although some pseudogenes do code for proteins, the majority are thought to be lncRNAs owing to the accumulation of mutations in the definition of the open reading frames, and as such their biological functions include the ability to regulate gene expression similarly to lncRNAs[19], and are thereby also involved in growth-regulatory roles in cancer[20].

RPSAP52 is a pseudogene-transcribed RNA that runs antisense to the oncofetal gene *HMGA2*, a transcriptional co-regulator that is expressed at high levels during embryonic development, silenced in virtually all adult tissues and re-expressed in several human cancers, where its levels are generally associated with the presence of metastases and poor prognosis[21,22]. Our previous results indicate that *RPSAP52* positively regulates *HMGA2* expression through the formation of an R-loop structure[23]. Herein we further study the role of this transcribed pseudogene in breast and sarcoma tumors, and uncover its role as a pro-growth factor through the regulation of the IGF2BP2/IGF1R/RAS axis and the balance between LIN28B and *let-7* levels.

Results

***RPSAP52* impacts on IGF2BP2 and *let-7* in breast cancer cells.** We have previously uncovered the positive impact of the expression of the pseudogene *RPSAP52* on its sense, protein-coding gene *HMGA2* (Fig. 1a)[23]. Both genes are generally expressed at low levels in differentiated normal tissues and overexpressed in a number of human cancers, including breast cancer, concomitant with a hypomethylation of the associated CpG island (Fig. 1b). In breast cancer patients, a positive correlation between the expression of both genes is observed (Fig. 1c), as is also seen in the NCI60 panel of cell lines (Supplementary Fig. 1a). Other studies have reported that high *HMGA2* expression predicts poor outcome in breast cancer patients[24]. Since our observations indicate that knockdown of *RPSAP52* results in a reduction in *HMGA2* expression[23], we decided to look further into the molecular mechanism of *RPSAP52*-mediated regulation of the *locus*. A panel of breast cancer cell lines was used to confirm the presence of *RPSAP52* transcript by semi-quantitative PCR (Fig. 1d). Surprisingly, most cell lines expressed the annotated *RPSAP52* transcript (Refseq NR_026825.2) together with an additional species that corresponds to the inclusion of a 104-nucleotides-long internal exon (Fig. 1d and Supplementary Fig. 1b). Evidence as to the presence of this alternative exon in the spliced transcript can also be found in the MiTranscriptome database[25], which catalogs long polyadenylated RNA transcripts (www.mitranscriptome.org, with reference G018828|T081486). The quantitative measurement of expression levels indicates that *HMGA2* mRNA and the two isoforms of *RPSAP52* are 2-3 orders of magnitude overexpressed when there is hypomethylation of the promoter-associated CpG island, as shown with Illumina’s HumanMethylation450 BeadChip analysis (Fig. 1e) and was confirmed by bisulfite sequencing at the nucleotide level (Supplementary Fig. 1c). Altogether, these observations confirm the coordinate expression of both genes and their silencing in hypermethylated conditions. *RPSAP52* is annotated as a noncoding RNA in Refseq, but is labeled as coding in some coding potential calculator tools. Pseudogenes are more likely to give false positive results in programs such as PhyloCSF (since they are similar to their parental protein-coding, and PhyloCSF evaluates conservation to predict coding capacity). We thus conducted *in vitro* transcription/translation assays, which confirmed the absence of *RPSAP52* coding potential (Supplementary Fig. 2a). However, analysis of RNA presence along sucrose gradients from MCF10A cells showed the presence of *RPSAP52* transcripts in polysomal fractions, indicating a role in translation. Interestingly, a strong correlation in co-sedimentation of *HMGA2* mRNA and *RPSAP52* transcript was observed (Supplementary Fig. 2b–e). Indeed, further characterization of *RPSAP52* transcripts showed that they are enriched in the cytoplasm (Fig. 1f) and polyadenylated (Fig. 1g), suggesting additional roles besides the ability to regulate *HMGA2* transcription in the nucleus. In order to identify protein partners of *RPSAP52* that could help characterize its activity, we performed RNA pull-down assays combined with mass spectrometry (MS). *In vitro* synthesized full-length *RPSAP52* RNA was incubated in the presence of MCF10A extracts and the retrieved proteins were analyzed by SDS-PAGE. As shown in Fig.2a, a protein band of ~70 kDa is specifically pulled-down by *RPSAP52* RNA, but not by its antisense sequence or another unrelated RNA. This band was characterized by MS, which identified two proteins within the isolated fragment: the insulin-like growth factor 2 mRNA-binding protein 2 (IGF2BP2), also known as IMP2 (from which seven peptides were identified), and the heterogeneous nuclear ribonucleoprotein Q (HNRNPQ), also known as SYNCRIP (identified with six peptides) (Supplementary Fig. 3a). IGF2BP2, together with IGF2BP1 and IGF2BP3 partakes of a family of RNA binding proteins that have been implicated in post-transcriptional control, including the regulation of mRNA localization, stability, and translation[26]. Similar to HMGA2, although the expression of IGF2BPs is normally restricted to embryonic stages, they are re-expressed upon malignant transformation, playing roles in the maintenance of cancer stem cells and the promotion of tumor growth[27]. Western blot with specific antibodies confirmed that both IGF2BP2 and HNRNPQ are enriched in the *RPSAP52* pull-down, and analysis of *RPSAP52* truncates indicate that the two isoforms are able to bind to these two factors (Fig. 2b). In accordance with the pull-down results, the previously reported consensus binding site for IGF2BP2, the CAUH (H= A, C, U) motif[28], is abundant along the two constitutive *RPSAP52* exons, but absent on the alternative exon (Supplementary Fig. 3b), suggesting that the alternative splicing event does not impact on the affinity of the binding.

IGF2BP2 is a direct transcriptional target of HMGA2[29] and both proteins partake of a pro-proliferative axis[30,31] that interweaves with the function of *let-7* family of miRNAs. Both *HMGA2* and *IGF2BP2* mRNAs are direct targets of *let-7*, but IGF2BP proteins have been suggested to modulate *let-7* action via the formation of cytoplasmic mRNPs that would protect certain mRNAs from *let-7* binding and repression[32,33]. The most abundantly expressed members of the family are *let-7a/b/e* in MCF10A cells, and *let-7a/d/f/g/i* in Hs578T cells (Supplementary Fig. 3c). Interestingly, the levels of the mature form of these miRNAs were upregulated in *RPSAP52*-depleted cells, both in MCF10A and Hs578T clones stably expressing shRNAs (Fig. 2c and Supplementary Fig. 3g) and in cells transiently expressing three different locked nucleic acid (LNA)-based antisense oligonucleotides (ASOs) gapmers (Fig. 2e). *RPSAP52* has a region of homology with other *RPSA* pseudogenes, but none was affected by our depletion strategy (Supplementary Fig. 3d). Also, given the possibility that some pseudogenes regulate parental gene expression, we analyzed the levels of RPSA protein in the *RPSAP52*-depleted cells, but no quantitative change was found (Supplementary Fig. 3e). Of note, gapmer-mediated depletion of *HMGA2* increased both *RPSAP52* isoforms and resulted in a decrease in *let-7* levels, suggesting that the negative regulation exerted by *RPSAP52* on the miRNAs is not through HMGA2 pathway (Fig. 2e). *Let-7* regulates IGF2BP2 mRNA and other members of IGF1 signaling pathway, among others *IGF1R* and *RAS*. Downregulation of *RPSAP52* with shRNAs or gapmers reduces the amount of these proteins, in accordance with the increased *let-7* levels (Fig. 2d, f and Supplementary Fig. 3g, h). LIN28A and LIN28B are the main negative regulators of *let-7* biogenesis, through direct binding to either *pre-let-7* and/or *pri-let-7*[34,35], but often only one of the two proteins is found expressed in human cancer cell lines[36]. We could not detect LIN28B protein expression in MCF10A cells, and LIN28A was not altered upon *RPSAP52* knockdown (Supplementary Fig. 3f), suggesting that changes in *let-7* in MCF10A cells were not consequences of impaired biogenesis at the level of regulation by LIN28. However, *RPSAP52*-mediated regulation of IGF2BP2 protein levels is reverted by overexpression of LIN28B, indicating a convergence on the same regulatory network (Fig. 2g).

Next, the phenotypic impact of the altered control of the IGF2BP2/IGF1R/RAS pathway by *RPSAP52* was tested both *in vitro* and *in vivo*.

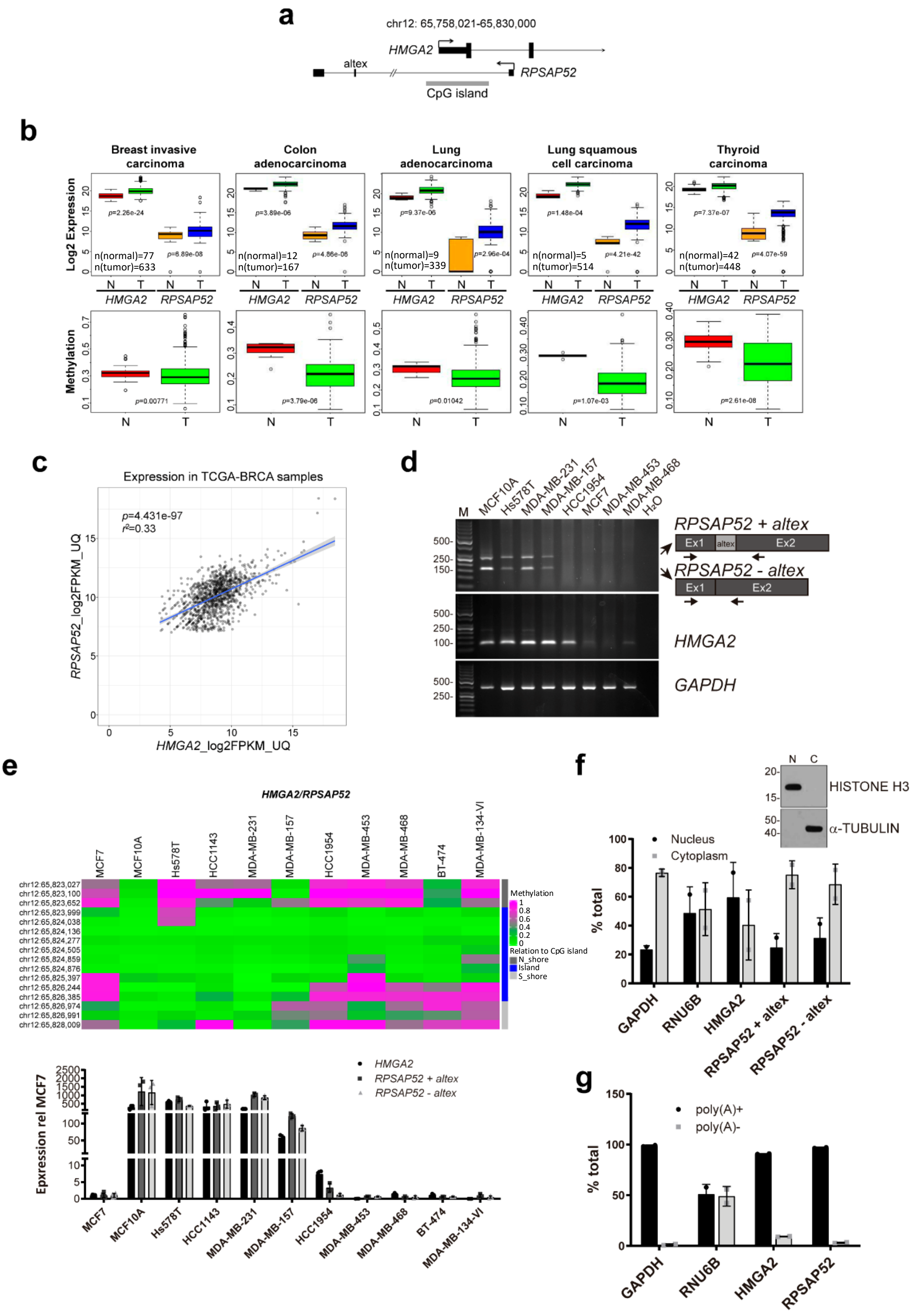


Fig. 1 Characterization of *RPSAP52* expression in breast cancer. **a** Intronic/exonic organization of sense/antisense transcripts in *HMGA2 locus*. Coordinates are referred to the UCSC Genome Browser (GRCh38/hg38 release). Only the 5′ regions of *HMGA2* transcripts are included. **b** Box plot representations show both *HMGA2* and *RPSAP52* transcripts are overexpressed in a variety of human cancers compared with normal controls (TCGA dataset) (upper panels), concomitant with hypomethylation of the associated CpG island (lower panels). On each box plot, the central mark indicates the median, and the bottom and top edges indicate the interquartile range (IQR). The box plot whiskers represent either 1.5 times the IQR or the maximum/minimum data point if they are within 1.5 times the IQR. *P*-values

are according to the two-tailed Wilcoxon signed-rank test. **c** Pearson correlation between *HMGA2* and *RPSAP52* transcripts in breast cancer primary tumors (all stages included). Normalized values of RNA-seq data from TCGA are represented. **d** Semiquantitative RT-PCR to detect the expression of *HMGA2* and *RPSAP52* in a panel of breast cancer cell lines. Detection of *RPSAP52* transcripts was done with primers that detect both isoforms. **e** Upper panel: heatmap representation of the DNA methylation profile for the CpG island-containing promoter at the *HMGA2 locus*, as analyzed with the 450K DNA methylation microarray. Single CpG methylation levels are shown. Green, unmethylated; magenta, methylated. Data from 11 breast cancer cell lines are shown. Lower panel: the expression levels of both *RPSAP52* (including or excluding the alternative exon) and *HMGA2* were analyzed by RT-qPCR and represented relative to MCF7 cell line. Graphs represent the means of three replicates from different RNA extractions ±SD. **f** Nuclear/cytoplasmic fractionation of MCF10A cells, analyzed by RT-qPCR and western blot to assess fraction purity. Graphs represent the mean ±SD of two replicates of fractionation. **g** Poly(A)+/poly(A)− partition of total RNA from MCF10A cells and analysis by RT-qPCR. Primers that detect the *RPSAP52+altex* isoform were used. Graphs represent the mean ±SD of two replicates of poly(A) selection. Source data are provided as a Source Data file.

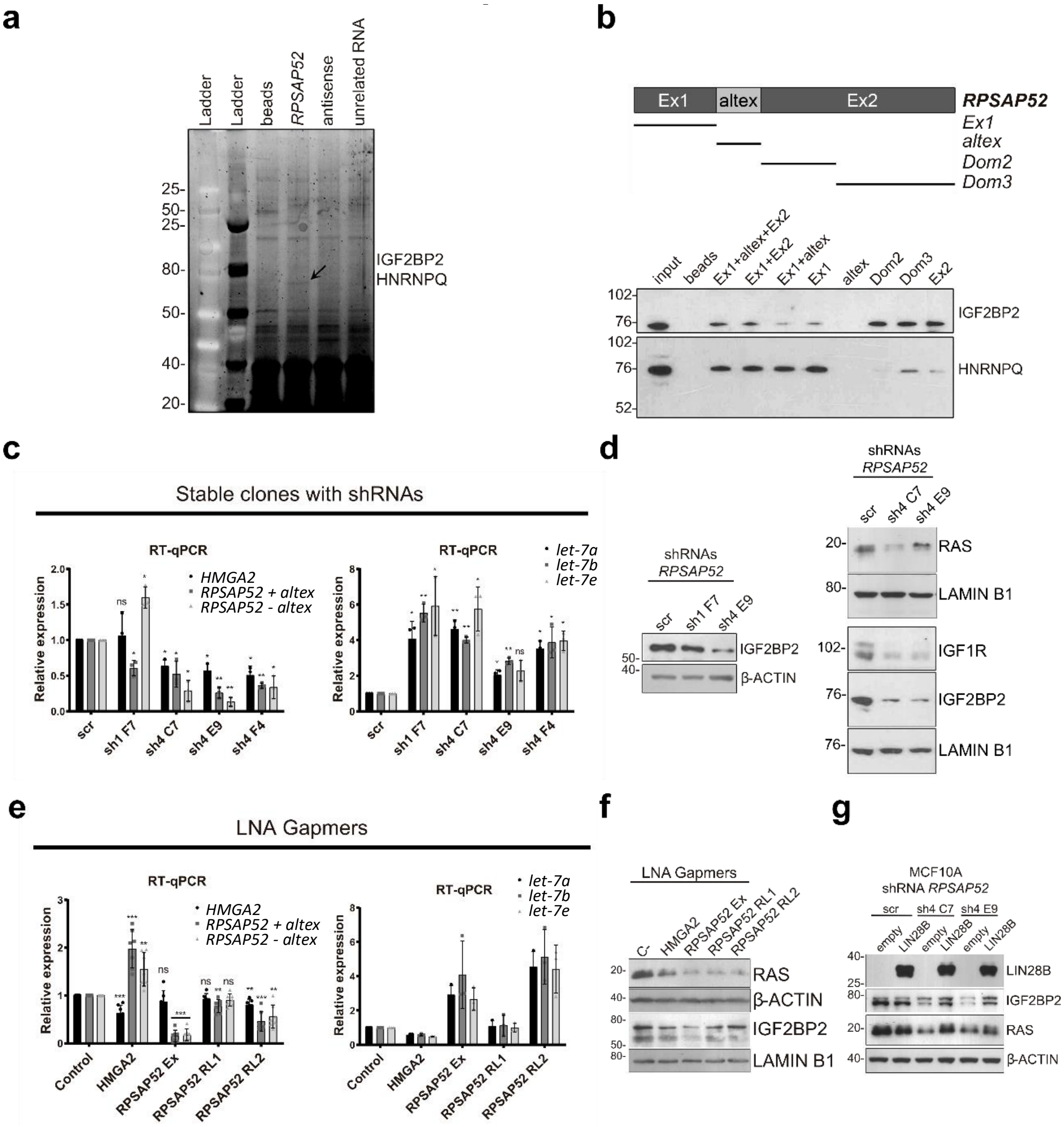


Fig. 2 *RPSAP52* interacts with IGF2BP2 and HNRNPQ and influences proliferative pathways in MCF10A cells. **a** RNA pull-down assay to detect *RPSAP52*-associated proteins. *In vitro* synthesized full-length *RPSAP52* transcript (including the alternative exon) or control sequences (the antisense transcript and the unrelated *Uc.160+* RNA) were tested. The proteins retrieved were analyzed by SDS-PAGE and the band of ~70 kDa indicated by the arrow was identified by MS as containing IGF2BP2 and HNRNPQ. **b** Western blot showing the association between *RPSAP52* RNA and IGF2BP2 and HNRNPQ proteins. Different truncated fragments of *RPSAP52* RNA (as shown in the upper

diagram) were incubated in the presence of MCF10A total protein extracts and the pulled-down material was subject to western blot with specific antibodies. Total extract from the MCF10A cell lines was used as input control, and a reaction without RNA (beads) as negative control. **c** Stable knockdown of *RPSAP52* results in upregulation of *let-7* family of miRNAs. Total RNA fromMCF10A clones constitutively expressing two different shRNAs against *RPSAP52* (sh1 or sh4) was analyzed by RT-qPCR to assess *HMGA2* mRNA, *RPSAP52* transcripts and *let-7* miRNAs levels. Graphs represent the mean ±SD of three independent RNA extractions. Two-tailed student *t*-test were used (**P* < 0.05, ***P* < 0.01, ns = not significant). **d** Western blot to analyze IGF2BP2, IGF1R, and RAS protein levels upon stable knockdown of *RPSAP52* transcripts. **e** Transient transfection of MCF10A cells with locked nucleic acid (LNA)-based antisense oligonucleotides (ASOs) gapmers targeting *HMGA2* mRNA (HMGA2), exon1 (RPSAP52 Ex) or the first intron (RPSAP52 RL1 and RL2) of *RPSAP52* transcript. Expression levels of *HMGA2*, *RPSAP52* and *let-7* were measured by RT-qPCR. Graphs represent the mean ±SD of seven independent replicates (for *HMGA2* and *RPSAP52*) or three replicates (for *let-7*). Two-tailed student *t*-test were used (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). **f** Western blot to analyze IGF2BP2 and RAS protein levels upon transient knockdown of *RPSAP52* transcripts. **g** Western blot to analyze IGF2BP2 and RAS protein levels upon transient overexpression of LIN28B protein in the background of *RPSAP52* depletion. Source data are provided as a Source Data file.

***RPSAP52* has oncogenic-like features *in vitro* and *in vivo*.** Upon *RPSAP52* knockdown, all three breast cell lines tested (the non-transformed MCF10A and the tumorigenic Hs578T and HCC1143 cells) proved to be significantly less proliferative in the sulforhodamine B (SRB) assay (Fig. 3a), and had a significantly lower percentage colony formation density than control cells (Fig. 3b). Interestingly, *RPSAP52* depletion was also associated with a decreased migration potential (Fig. 3c). High levels of *let-7* miRNAs often correlate with a lower capacity for self-renewal and pluripotency. Given the observed reduction in proliferation and migration following *RPSAP52* knockdown, we next assessed the levels of markers of cell stemness (Fig. 3d). NANOG and OCT4 protein levels were decreased in *RPSAP52*-depleted cells, suggesting this lncRNA promotes features of cancer stem cells. This was further confirmed in soft-agar colony formation experiments, in which the measure of the anchorage-independent growth of the cells showed a significant decrease upon depletion of *RPSAP52* (Fig. 3e). For the *in vivo* approach, we next used tumor formation assays in nude mice. MCF10A and Hs578T cells stably expressing either scrambled shRNAs or shRNAs against *RPSAP52* were subcutaneously injected into mice, and the tumor formation and volume was monitored. Tumors originating from *RPSAP52* knockdown cells had a significantly lower volume and weight at endpoint than control tumors, both for the non-tumorigenic and the tumorigenic cells (Fig. 3f, g). Importantly, the amount of RAS and IGF2BP2 proteins were markedly reduced in the excised tumors at end point, indicating that the *in vitro* findings were maintained in the *in vivo* context (Fig. 3f, g).

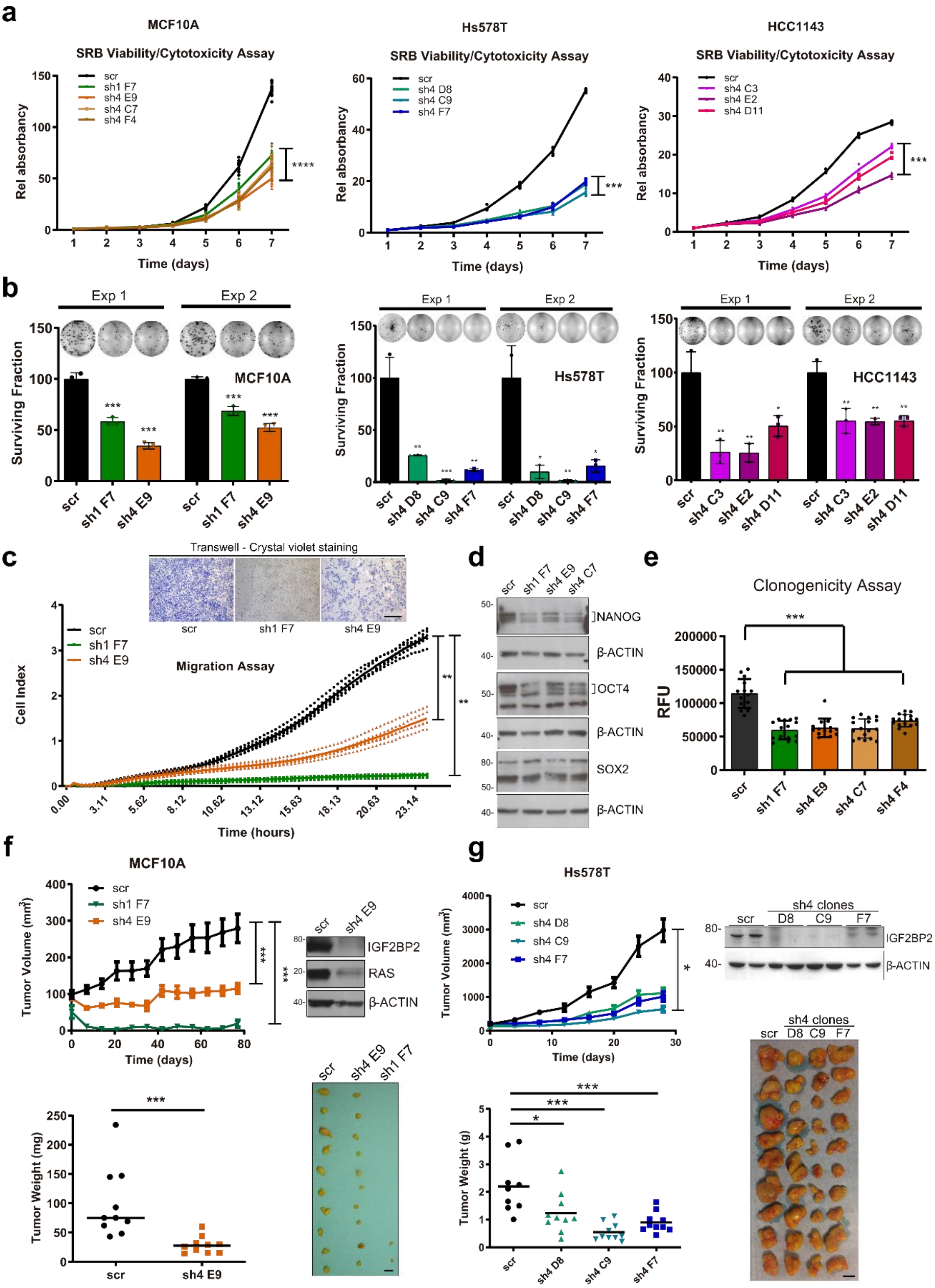


Fig. 3 *RPSAP52* displays oncogenic features in breast cancer cells. **a** Viability/cytotoxicity assays in MCF10A, Hs578T and HCC1143 clones. The experiment was performed three times and one representative graph is shown for each cell line. Values are mean ±SD of $n \geq 6$ measurements. One-way ANOVA was used (***P < 0.001, ****P < 0.0001). **b** Effect of *RPSAP52* silencing on colony formation ability. Representative plates are shown. Colonies were counted from three replicate plates and two independent experiments. Values are mean ±SD. Two-tailed unpaired *t*-test were used (*P < 0.05, **P < 0.01, ***P <0.001, ****P < 0.0001). **c** Migration capacity of *RPSAP52*-depleted clones was monitored over 24 h (n = 5 replicates per condition), with higher cell index indicating higher migration. Values are mean ±SD. A two-tailed Mann–Whitney U test of data at end point was used (**P < 0.01). Inset: migration was also

assessed with transwells. Scale bar = 100 μm. **d** Western blot analysis of NANOG, OCT4, and SOX2 in *RPSAP52*-depleted cells. **e** The clonogenic ability was assessed with at least *n* = 12 replicates per condition. Values are mean ±SD. A two-tailed Mann–Whitney U test was used (****P* < 0.001). **f** Growth inhibitory effect of *RPSAP52* knockdown in MCF10A mice xenografts. Upper graph: tumor volume (*n* =10) was monitored over time. Mean values are shown ±SEM. Lower graph: tumors were excised and weighed at 77 days (****P* < 0.001, two-tailed Mann–Whitney U test). Western blot was carried out from sh4 tumors since no material could be recovered from sh1 tumors, and the levels of RAS and IGF2BP2 proteins were analyzed. The photograph shows the relative size of all tumors extracted. Scale bar =10 mm. **g** Growth-inhibitory effect of *RPSAP52* knockdown in Hs578T mice xenografts. Upper graph: tumor volume (*n* =9 for scr and *n* = 10 for sh4 clones) was monitored over time. Mean values are shown ±SEM. Lower graph: tumors were excised and weighed at 28 days (**P* < 0.05, ****P* < 0.001, two-tailed Mann–Whitney U test). Western blot was carried out from tumors at end point and the levels of IGF2BP2 protein were analyzed. The photograph shows the relative size of all tumors extracted. Scale bar = 10 mm. Source data are provided as a Source Data file.

***RPSAP52* regulates IGF2BP2/LIN28B/*let-7* axis in sarcoma.** We next attempted to determine whether *RPSAP52*-mediated regulation of proliferative pathways occurred in other cancer types. The analysis of the collection of human cancers available from The Cancer Genome Atlas (TCGA) indicates *HMGA2* and *RPSAP52* expression is specially increased in adrenocortical carcinoma, mesothelioma and in the sarcoma samples available (as measured by Z-score, Supplementary Fig. 4a). TCGA RNA expression and DNA methylation data showed that *RPSAP52* promoter hypermethylation was associated with transcript downregulation across sarcoma samples (Fig. 4a, upper panel; Pearson correlation, r^2 = 0.264, *P*-value = 4.764e–10). Of note, *HMGA2* expression in the same samples shows a poorer correlation (Supplementary Fig. 4b, r^2 = 0.131, *P*-value = 2.582e–05). Both genes maintain a positive expression correlation (Fig. 4a, lower panel) and a difference of ~1–2 orders of magnitude in their relative expression (Supplementary Fig. 4c). This is in agreement with absolute quantification of *RPSAP52* and *HMGA2* transcripts in MCF10A and A673 cell lines, in which *HMGA2* mRNA is 1–2 orders of magnitude at higher levels (Supplementary Fig. 4d). Since the HMGA2-IGF2BP2-RAS pathway has been previously involved in the pathogenesis of embryonic rhabdomyosarcoma[31], we then assessed *HMGA2* and *RPSAP52* expression in a panel of cell lines derived from rhabdomyosarcoma and also Ewing’s sarcoma (Fig. 4b). Both *RPSAP52* isoforms were abundantly expressed in most rhabdomyosarcoma cell lines, with just one order of magnitude higher *HMGA2* expression in Rh28, Rh41, or CW9019 cells. In Ewing’s sarcoma, *RPSAP52* was generally lowly expressed with the exception of A673 cell line. We thus focused on A673 cells to further characterize the molecular function of this lncRNA. As seen in MCF10A cells, RNA pull-downs confirmed the ability of *RPSAP52* to interact with IGF2BP2 and SYNCRIP/HNRNPQ (Fig. 4c). Importantly, stable clones expressing two different shRNAs against both *RPSAP52* isoforms resulted in a strong increase in *let-7* family members (Fig. 4d, upper panel), even when *HMGA2* levels were only moderately reduced (Fig. 4d, lower panel). In this case, *RPSAP52* knockdown did not correlate with IGF2BP2 decrease, but with a marked reduction in LIN28B protein levels, which in contrast to breast cell lines, is abundantly expressed in A673 cells (Fig. 4e and Supplementary Fig. 4e). Also, while RAS levels were only partially reduced, downstream signaling was impaired, as observed by the decrease in p-ERK levels (Fig. 4e). In an *in vivo* setting, and similarly to the observations in breast cell lines, this results in a marked reduction in tumor formation when mice are subcutaneously injected with *RPSAP52*-depleted A673 cells (Fig. 4f). Interestingly, LIN28B protein reduction is only partially explained by a decrease in *LIN28B* mRNA levels (Supplementary Fig. 4f). This discrepancy, together with the interaction detected between *RPSAP52* and IGF2BP2, and the presence of *RPSAP52* along sucrose gradient's heavy fractions, which correspond to translating poly-ribosomes (see text above and Supplementary Fig. 2d), prompted us to investigate the possibility that LIN28B levels were regulated by the lncRNA at the translational level.

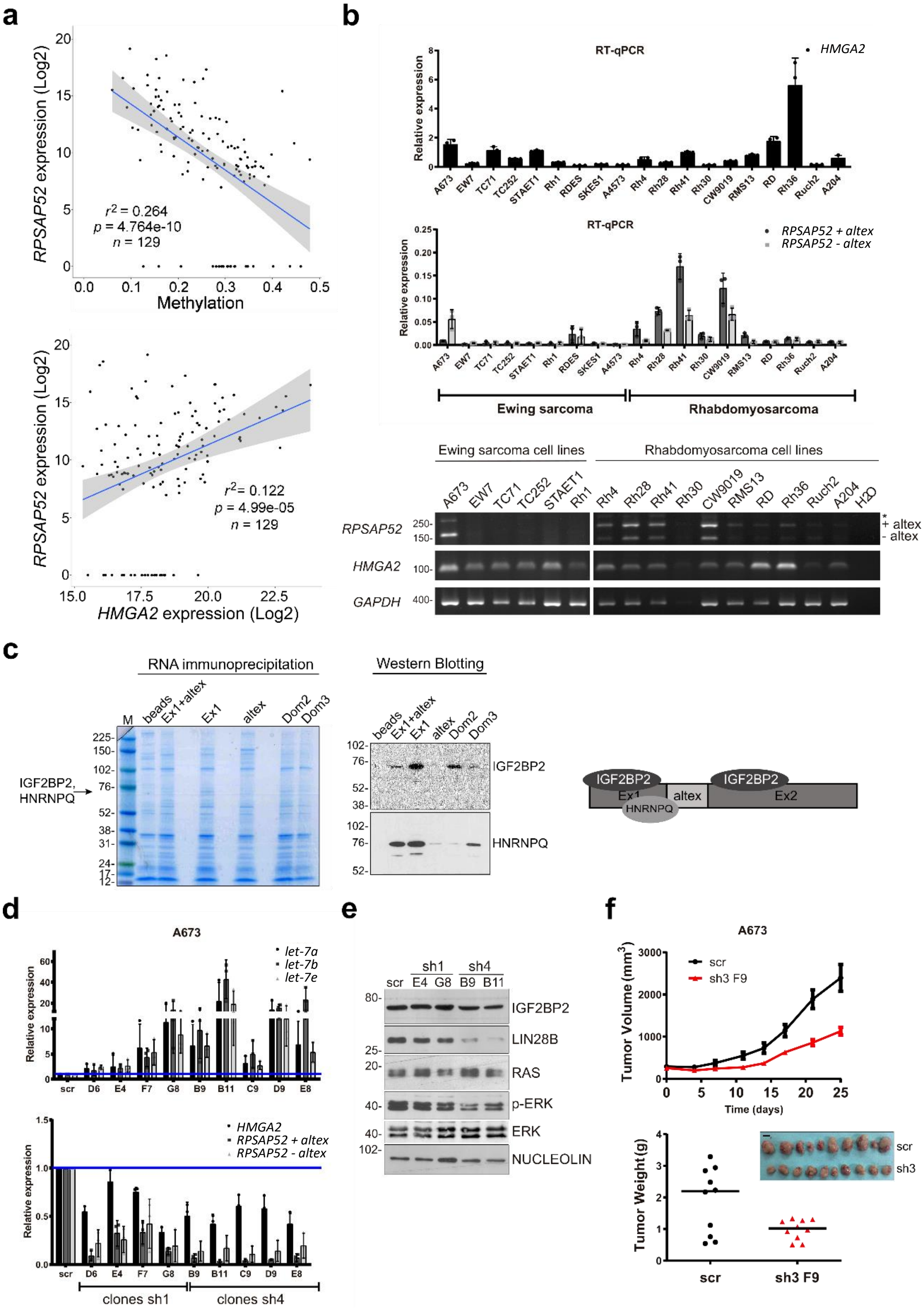


Fig. 4 *RPSAP52* is abundantly expressed in sarcoma and regulates the LIN28B/*let-7* balance. **a** Upper graph: Pearson coefficient between *RPSAP52* expression levels and CGI methylation in the TCGA sarcoma cohort indicates a negative correlation. Lower graph: Pearson's index indicates a weaker association between *RPSAP52* and *HMGA2* expression levels in the same cohort. **b** Upper graphs: RT-qPCR analysis to estimate *HMGA2* and *RPSAP52* expression levels in a panel of Ewing's sarcoma and rhabdomyosarcoma cell lines. Expression is relative to *GUSB* mRNA levels. Graphs represent the mean ±SD of three independent RNA extractions. Lower panel: semi-quantitative RT-PCR analysis of expression in the same cell lines. The two *RPSAP52* isoforms are indicated, and the higher migrating band depicted by

an asterisk contains an additional exonic sequence (encompassing coordinates chr12:66,169,917–66,170,002 (hg19)), detected in those cells lines with the highest expression of *RPSAP52*. **c** RNA pull-down assays confirm the interaction of *RPSAP52* with IGF2BP2 and HNRNPQ in A673 cell extracts. Different truncated fragments of *RPSAP52* were assayed as indicated, and the band identified by MS and corresponding to IGF2BP2 and HNRNPQ is indicated on the protein gel (left). Western blot to test the association between *RPSAP52* RNA and IGF2BP2 and HNRNPQ proteins (middle panel). The drawing summarizes the data obtained from the pull-downs (right). **d** Total RNA from A673 stable clones constitutively expressing sh1 or sh4 shRNA sequences was analyzed by RT-qPCR to assess *HMGA2* and *RPSAP52* transcripts levels (lower graph) or *let-7* miRNAs levels (upper graph). Graphs represent the mean ±SD of three independent replicates. **e** Western blot on A673 clones to analyze protein levels upon stable knockdown of *RPSAP52* transcripts. **f** Growth-inhibitory effect of *RPSAP52* knockdown in A673 mice tumor xenografts. Upper graph: tumor volume (*n*=10) was monitored over time. Mean values are shown ±SEM. Lower graph: tumors were excised and weighed at 25 days. The photograph shows the relative size of all tumors extracted. Scale bar=10mm. Source data are provided as a Source Data file.

***RPSAP52* modulates IGF2BP2 binding to its mRNA targets.** IGF2BP2 is a mRNA stability and translational regulator with some well-described targets, such as *IGF2*[37], *NRAS*[31], or *HMGA1*[38]. The closely related IGF2BP1 protein has been shown to interact with *LIN28B* mRNA and increase LIN28B protein levels in ES-2 cells[32]. In order to assay the interaction of IGF2BP2 with *LIN28B* mRNA in A673 cells, we carried out protein immunoprecipitation followed by RT-qPCR of the pulled-down RNA. We could confirm the interaction of IGF2BP2 with both *RPSAP52* isoforms, with *IGF2BP2* and *NRAS* mRNA, and importantly, with *LIN28B* mRNA. Of note, even though the *RPSAP52* isoform lacking the alternative exon is more abundant in A673 cells, both transcripts were recovered in comparable amounts in IGF2BP2 immunoprecipitate, with a ~10-fold higher affinity of IGF2BP2 for *RPSAP52+altex* RNA (Fig. 5a, see RT-qPCR). Further, LIN28B protein was not co-immunoprecipitated with IGF2BP2 protein (Fig. 5b), suggesting its putative regulation by IGF2BP2 is at the level of transcript. We next wanted to test the possibility that this binding is regulated by *RPSAP52* presence. Interestingly, whereas binding to *IGF1R* and *IGF2BP2* mRNAs was not altered, binding of IGF2BP2 to *LIN28B* mRNA was reduced upon stable knockdown of *RPSAP52* (Fig. 5c). This suggests that this lncRNA might regulate *LIN28B* post-transcriptionally through modulation of IGF2BP2 function. In view of this, we decided to characterize in a transcriptome-wide manner the IGF2BP2-RNA interactions with individual nucleotide resolution (iCLIP-seq) under control or *RPSAP52*-knockdown conditions. We identified 290,060 and 131,729 iCLIP-tags in control and *RPSAP52*-depleted A673 cells, corresponding to 3075 and 1639 peak regions, respectively (Supplementary Fig. 5a, b). As has been shown before, IGF2BP2 iCLIP tags were enriched in 3′UTRs[28,33], with ~60% of the iCLIP peaks falling within 3′UTRs in control cells. Remarkably, knockdown of *RPSAP52* resulted in a specific decrease in the number of 3′UTR peaks revealed by iCLIP and an increase in intronic regions (Fig. 5d). Motif enrichment analysis around the crosslinking-induced truncation sites (CITS) indicate that the previously described CAUH (H = A, C, U) consensus binding site[28] also ranks high in our iCLIP experiments, but is more enriched in the control samples (Supplementary Fig. 5c, d). We found 1775 and 810 peaks with the CAUH motif in the control and depleted sample, respectively, representing a statistically significant difference in occurrence (Fisher's exact test *P*-value = 5.286e–08). In addition, the number of peaks with more than one CAUH motif was higher in the control cells (average number of motifs was 1.81 for control and 1.64 for depleted cells; Mann–Whitney U test, *P*-value = 1.786e–05). The full list of statistically significant iCLIP-seq peaks and CITS can be found in Supplementary Data 1. Differences in motif binding and the reduction in 3′UTR recognition results in a shortlist of 34 transcripts with differential IGF2BP2 iCLIP counts along their 3′UTR (Fig. 5e). GO enrichment analysis shows these genes belong to categories that may relate to IGF2BP2 involvement in cancer invasion and metastasis, including cell-substrate adhesion, spreading, and wound healing, as well as the canonical function for IGF2BP2 pathway, cellular glucose homeostasis (Fig.5e). Of note, previous CLIP experiments for IGF2BPs in pluripotent stem cells have revealed that cell adhesion is also the most significant GO category for CLIP-enriched 3′UTRs for IGF2BP1[39]. This suggests that the levels of *RPSAP52* have a dramatic impact on IGF2BP2 global role. In addition, top ten GO categories for genes with significant iCLIP peaks present on their 3′UTRs in the control sample correspond to signaling pathways and cell cycle progression, whereas none of these categories are enriched in the *RPSAP52*-depleted sample (Fig. 5f).

Previous CLIP-seq studies with IGF2BP2 had revealed binding sites on the 3′UTR of *LIN28B* mRNA in HEK293T cells[28], and we detected similar sites in our experimental setting and a tendency to decrease upon *RPSAP52* depletion, although without any statistical power (Fig. 5g). For other validated IGF2BP2 targets, such as *HMGA2*, we also detected abundant iCLIP signal corresponding to direct binding of IGF2BP2 to its 3′UTR. In this case, binding is dramatically lost upon *RPSAP52* knockdown (Fig. 5h), and a corresponding decrease in HMGA2 protein level is observed (Supplementary Fig. 5e). This is not a general phenomenon for all IGF2BP2 targets, since other well-characterized mRNA partners, such as *HMGA1*, *NRAS*, and *IGF1R* maintain comparable iCLIP signals in both control and *RPSAP52*-depleted conditions (Supplementary Fig. 5f). Our results thus suggest a specific loss of IGF2BP2 affinity for particular mRNA targets. Several of the best characterized IGF2BP2 targets are regulated by *let-7* (e.g., *RAS*, *HMGA2*…) but, interestingly, we could not find a differential presence of *let-7* miRNA recognition motifs along the 3′UTRs of the immunoprecipitated mRNAs in control or depleted samples (Fisher's exact test *P*-value = 0.8974). Also, depletion of LIN28B does not impact on IGF2BP2 levels or its binding to mRNA targets, indicating that the regulation of *RPSAP52* on IGF2BP2 does not proceed through LIN28B (Supplementary Fig. 5h–j). Both *HMGA2* and *LIN28B* mRNAs are present at high levels upon *RPSAP52* depletion (Fig. 4d and Supplementary Fig. 4f), and their half-lives are not substantially altered (Supplementary Fig. 5g), pointing to a decrease in their translation as a consequence of a diminished binding to IGF2BP2.

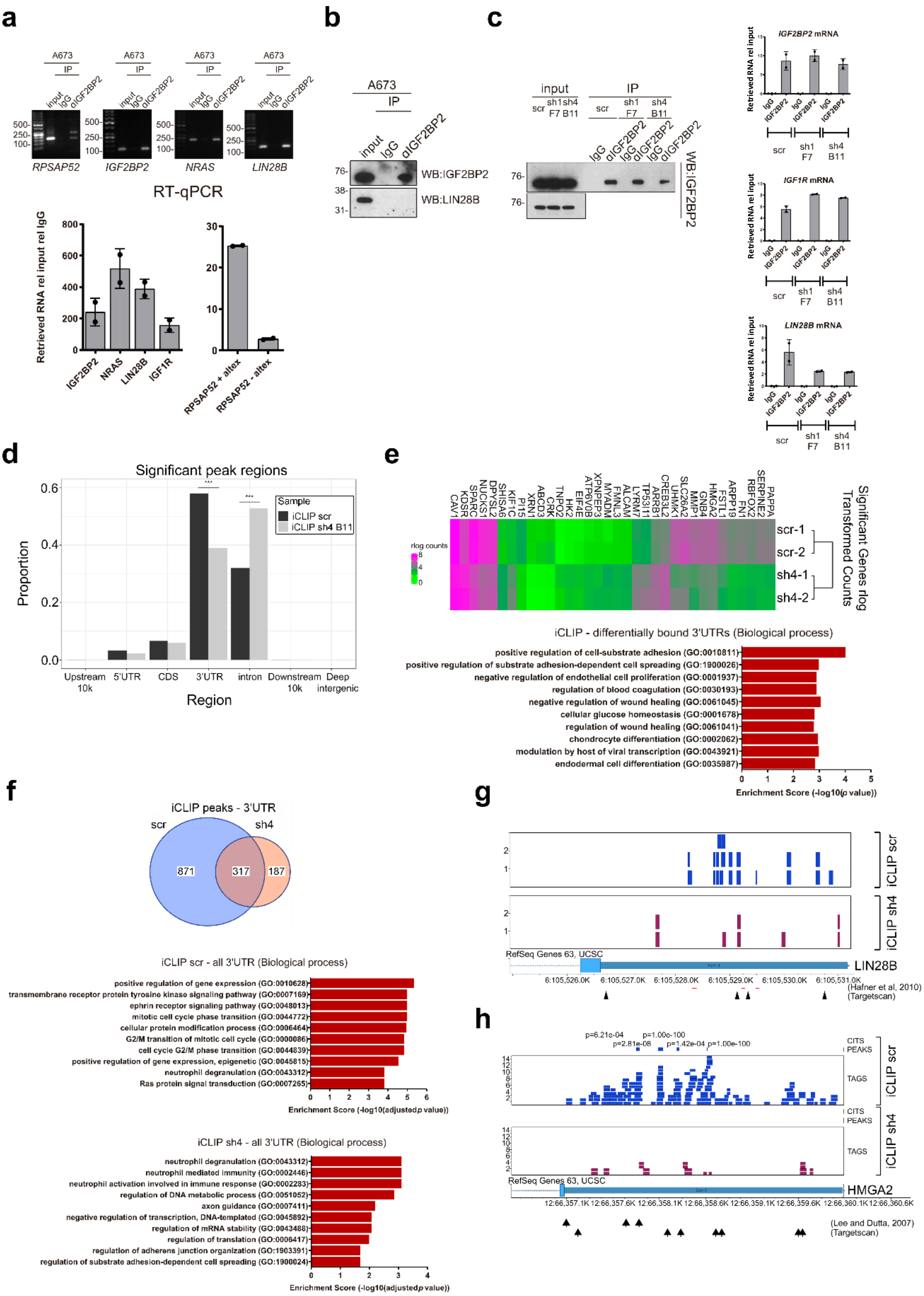


Fig. 5 Binding of IGF2BP2 to its mRNA targets is affected by *RPSAP52* knockdown. **a** IGF2BP2 was immunoprecipitated from A673 extracts and the pulled-down RNA was analyzed by RT-PCR. The *IGF2BP2* and *NRAS* mRNAs were used as positive controls. Gel images represent semi-quantitative RT-PCR, whereas data on graphs represent means of two independent RT-qPCR analysis ±SD. **b** Immunoprecipitation of IGF2BP2 from A673 extracts followed by western blot of retrieved proteins. 10% of total extract prior to IP was loaded as control (input). **c** IGF2BP2 was immunoprecipitated in control or *RPSAP52*-depleted A673 cells, and the retrieved proteins and RNAs were isolated and analyzed by western blot (left) or RT-qPCR (right), respectively. Graphs correspond to means from two replicates ±SD. **d** Analysis of IGF2BP2 binding targets from iCLIP-seq experiments in control (scr) or depleted cells

(sh4 B11). The absolute number of peaks mapping to 3′UTR regions were 1762 (scr) and 622 (sh4), and to intronic regions were 973 (scr) and 846 (sh4). Asterisks correspond to *P*-values < 2.2e–16 (two-tailed Fisher's tests). **e** Above: heatmap of genes with differential iCLIP counts on their 3′UTR. Results from two experiments are shown. Below: differential enrichment of these genes according to GO biological process categories (top ten are shown). **f** Above: Venn diagram showing the relation between genes with significant iCLIP peaks present on their 3′UTR regions in the control (scr) or *RPSAP52*-depleted (sh4) sample. Below: top ten GO enrichment categories (Biological process) for the same genes in the scr or sh4 sample. **g** UCSC GenomeBrowser view of *LIN28B* 3′UTR with the read coverage from IGF2BP2 iCLIP experiment. Previous IGF2BP2-CLIP data positions are shown in red, and predicted *let-7* binding sites are indicated by the arrows. **h** UCSC Genome Browser view of *HMGA2* 3′UTR with the read coverage from IGF2BP2 iCLIP experiment. Position of significant peaks and CITS are shown above the profiles, and predicted *let-7* binding sites are indicated by the arrows. Source data are provided as a Source Data file.

***RPSAP52* controls IGF2BP2 and mRNA distribution on polysomes.** To obtain direct evidence of the changes in translation efficiency for specific IGF2BP2 targets, we analyzed the distribution of mRNAs across sucrose gradients in control or *RPSAP52*-depleted A673 cells. We observed no major changes in the polysome profiles of cells depleted of *RPSAP52* when compared with control cells, indicating that *RPSAP52* knockdown does not alter the global translational output of the cell (see gradient profiles in Fig. 6a–d and Supplementary Fig. 6a). Surprisingly, we detected a remarkable decrease in the amount of *HMGA2* and *LIN28B* mRNAs associated with polysomes (Fig. 6a, b), indicating a selective regulation in their translation as a function of *RPSAP52* expression. This was not observed for *NRAS* and *GAPDH* mRNAs (Fig. 6c, d). Analysis of total *HMGA2*, *LIN28B*, and *NRAS* mRNAs under the same conditions does not justify the specific redistribution of *HMGA2* and *LIN28B* across the gradients (Fig. 6e). These results indicate that the loss of binding to IGF2BP2 previously observed in iCLIP experiments correlates with lower translational efficiency for individual mRNAs, and we next asked whether IGF2BP2 protein itself is redistributed across the gradient. Importantly, even though IGF2BP2 co-immunoprecipitates with the same protein partners in pull-down experiments (Supplementary Fig. 6b, c), its co-sedimentation with translating poly-ribosomes is markedly reduced upon *RPSAP52* knockdown (Fig. 6f), demonstrating that *RPSAP52* expression mediates the recruitment of IGF2BP2 on polysomes. Taken together, the results suggest that the absence of the pseudogene decreases the recruitment of IGF2BP2 to large polysomes, thereby impacting on the translation of specific mRNAs.

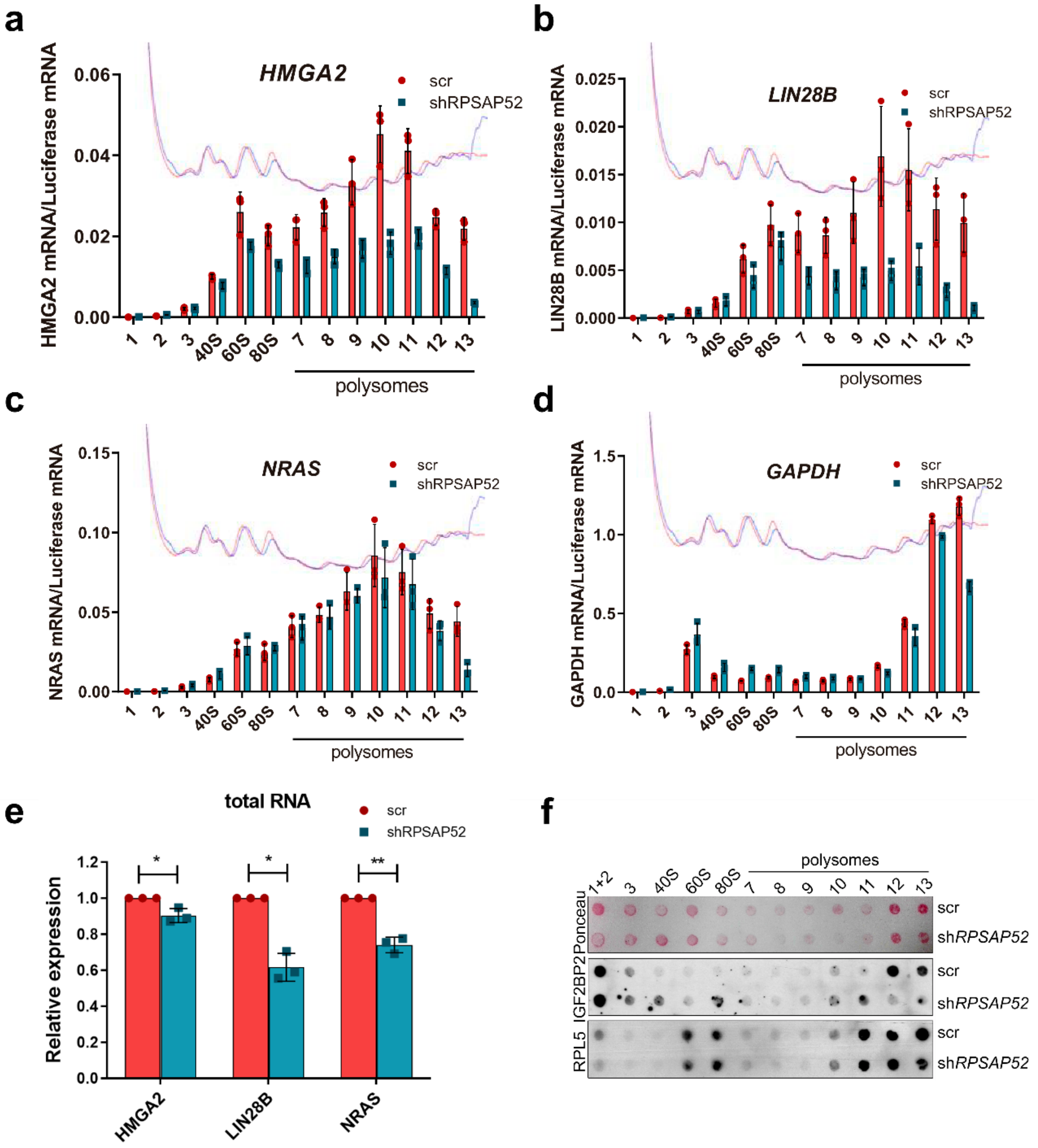


Fig. 6 IGF2BP2 is redistributed on polysome gradients upon *RPSAP52* depletion. **a–d** Polysome profiles of A673 cells stably depleted of *RPSAP52* (shRPSAP52, corresponding to sh4 B11 clone) or control cells (scr). *HMGA2* mRNA (**a**), *LIN28B* mRNA (**b**), or *NRAS* mRNA (**c**) distribution across the gradient was evaluated in each fraction by RT-qPCR. For comparison, *GAPDH* mRNA distribution was also assessed (**d**). Graphs represent the mean ±SD of three replicates. purpoThe red and blue lines indicate absorbance at 260 nm for each fraction in control or depleted cells, respectively. **e** Total RNA from the same cells ($n = 3$) was analyzed by RT-qPCR. Mean values are shown ±SD. A two-tailed student *t*-test was used (*$P < 0.05$, **$P < 0.01$). **f** Protein extracted from the 20% of the polysome profile fractions shown in (**a–d**) were subjected to dot blot analysis with an anti-IGF2BP2 antibody (middle panel) or with anti-RPL5 antibody as control (lower panel). Proteins from 10% of fractions 1 and 2 were loaded together. Membranes were previously stained with Ponceau S (top panel) for loading control. Source data are provided as a Source Data file.

***RPSAP52* alters key pathways and is a biomarker in sarcoma.** The influence of *RPSAP52* on IGF2BP2 binding affinity to its multiple mRNA targets might reflect the impact of this pseudogene on the control of several cellular processes. To identify such processes we interrogated general gene expression with an expression microarray platform under conditions of *RPSAP52* knockdown by shRNAs. As shown in Fig. 7a, 1%

of the ~30,000 interrogated Entrez Gene RNAs were downregulated following knockdown, and 0.7% of the transcripts were upregulated. In agreement with a regulation mainly at the level of translation, none of the genes with differential IGF2BP2 binding along the 3′UTR, as seen by iCLIP-seq, is deregulated in the expression array analysis, indicating that (i) the differential binding observed is not a consequence of altered transcript expression, and (ii) none of the IGF2BP2 targets whose interaction with this protein is influenced by *RPSAP52* see their stability significantly altered as a consequence. However, they might participate in similar cellular pathways, since GO terms analysis indicated that the subset of downregulated genes was enriched in components of response to stimulus and signaling, whereas upregulated genes appeared more involved in development (Fig. 7a). The full list of altered transcripts (fold change > 2, unpaired *t*-test *P*-value < 0.05) can be found in Supplementary Data 2. Among the downregulated genes, molecular functions that were overrepresented included genes involved in receptor binding and growth factor activity (e.g., *TIAM1*, *STYK1*, *AREG*, *MICB*) and sulfur compound binding (*CYR61* and *MGST1*). Of note, MGST1 is involved in the glutathione metabolism pathway and a marker of Ewing's sarcoma prognosis[40], high levels of NPY (a direct target of the EWS-FLI1 fusion) promotes the metastasis of Ewing's sarcoma models *in vivo*[41], CRABP1 and CPT1C favor tumor malignancy[42,43], and CD109 and PTPRZ1 are highly expressed in several cancer types (including sarcoma cell lines[44]) and promote stem cell-like properties[45]. Among the upregulated genes, genes involved in cytoskeletal protein binding were enriched and included MTSS1, a regulator of actin dynamics whose loss increases metastatic potential in a number of cancer types[46,47] (Supplementary Fig. 7a). The results from the expression arrays were validated by RT-qPCR under conditions of depletion of *RPSAP52* where *HMGA2* levels are largely unaffected (probably because the R-loop forming, nuclear *RPSAP52* is not effectively depleted by shRNAs), and with two different shRNAs that target distant regions on *RPSAP52* transcripts (Fig. 7b and Supplementary Fig. 7b). These results indicate that *RPSAP52* depletion implies a decrease in proliferative and self-renewal programs and suggests its potential as a biomarker in human samples. In support of this, patients with high *RPSAP52* expression levels had poorer prognosis than cases with low expression in the sarcoma patients cohort from TCGA database, whereas *HMGA2* expression did not show any prognostic effect in the same cohort (Fig. 7c). Since *RPSAP52* expression negatively correlates with hypermethylation of the associated CpG island (Fig. 4a), methylation itself is also a marker of better prognosis (Fig. 7d), reinforcing the relevance of considering lncRNA expression regulation in translational medicine.

Taken together, our results suggest that the pseudogene *RPSAP52* controls the HMGA2/IGF2BP2/LIN28B axis through a double mechanism that involves, in the nucleus, the positive transcriptional regulation of *HMGA2*, and in the cytoplasm, the regulation of the function of IGF2BP2 protein as a translational co-regulator (among others, of *LIN28B* and *HMGA2* mRNAs),which in turn results in a downregulation of *let-7* miRNAs and derepression of their pro-proliferative targets (see diagram depicting our working model in Fig. 7e). *RPSAP52* thus displays characteristics of an oncogenic gene whose dysregulation might contribute to the progression of a number of human cancers.

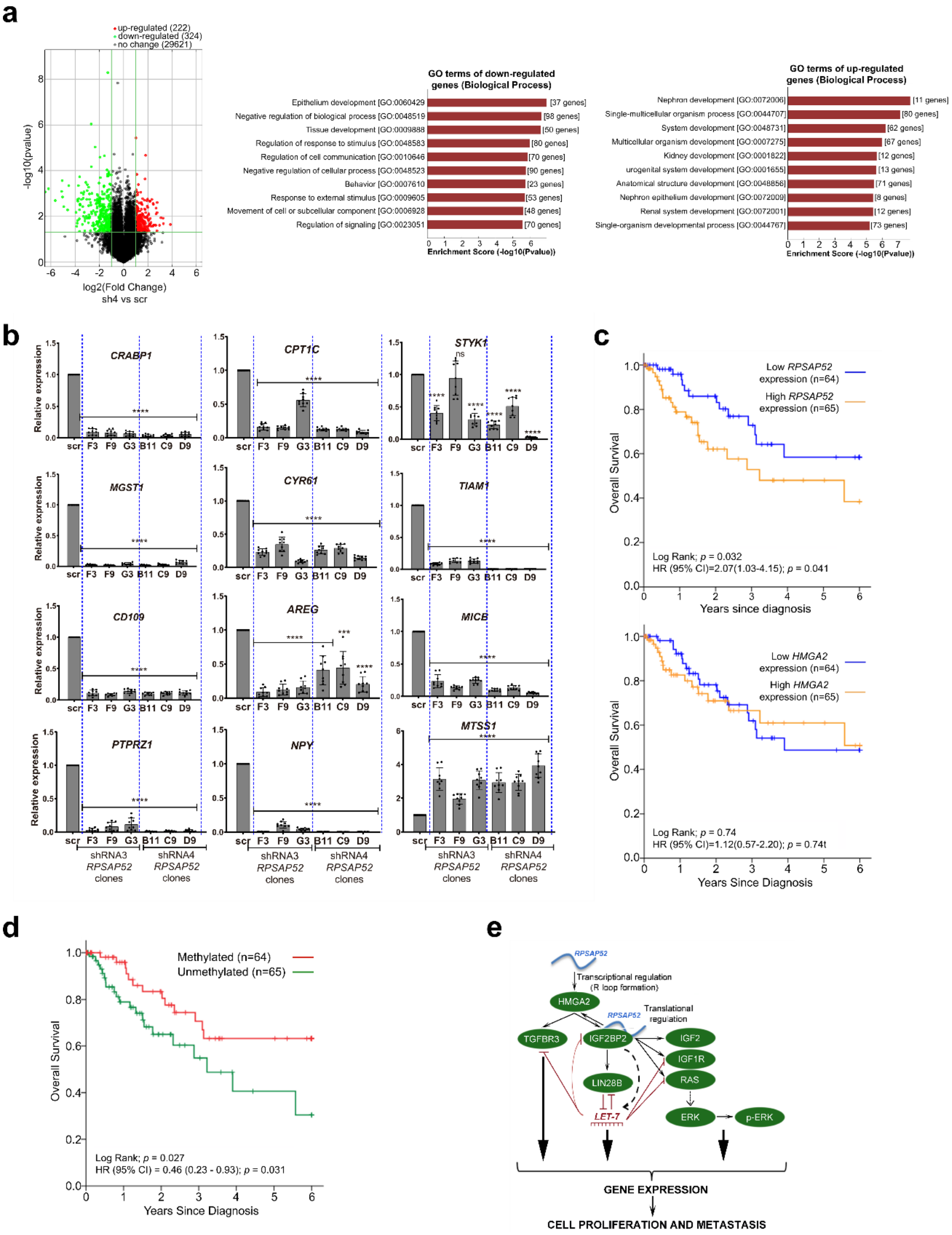


Fig. 7 *RPSAP52* expression influences proliferative cellular programs and is a prognosis factor. **a** Left: Volcano plot indicating differential expression (green= down, red= up) between control (scr) and *RPSAP52*-depleted (sh4) A673 cells. The vertical green lines correspond to 2.0-fold up and down, respectively, and the horizontal green line represents a *P*-value of 0.05 (two-tailed unpaired *t*-test). Right: enriched GO terms for shRNA-*RPSAP52*-affected genes. The *y* axis shows GO terms and the *x* axis shows statistical significance (two-tailed Fisher's exact test). **b** RT-qPCR analysis of candidate genes altered in *RPSAP52*-depleted A673 cells. Clones stably expressing two different shRNAs were analyzed. Graphs represent the mean ±SD of three replicates (two-tailed unpaired student *t*-test, ****P* < 0.001, *****P* < 0.0001). **c** Above: in the TCGA sarcoma cohort, Kaplan–Meier analysis of overall survival indicates that patients with high *RPSAP52* expression levels have poorer prognosis than cases with low expression. Below: *HMGA2* expression has no prognostic value in the same cohort. Significance of the log-rank test is shown. **d** Kaplan–Meier analysis of overall survival in the sarcoma cohort from TCGA, indicating that patients with a hypermethylated *HMGA2*/*RPSAP52* promoter display better prognosis. **e** Summary of the results in the context of

HMGA2/IGF2BP2/*let-7* axis. *RPSAP52* positively regulates *HMGA2* expression through both transcriptional and post-transcriptional mechanisms. Binding of *RPSAP52* to IGF2BP2 in the cytoplasm might promote downregulation of *let-7* levels by LIN28B-dependent and independent mechanisms. This binding could also modulate the formation of mRNPs for a number of IGF2BP2 mRNA targets, thereby directing their translation efficiency. Source data are provided as a Source Data file.

Discussion

The detailed mechanism by which lncRNAs may contribute to altering the output of signal transduction pathways is largely unexplored. Our previous work had shown that the transcribed pseudogene *RPSAP52* enhances *HMGA2* transcription through the formation of an R-loop structure[23]. We have further explored the impact of *RPSAP52* expression in cell physiology and propose a mechanism of action that also influences post-transcriptional regulation in the cytoplasm through the interaction with the RNA binding protein IGF2BP2. Regulation of IGF2BP2 expression or function by lncRNAs appears as a common theme in a number of lineage commitment programs, including adipocyte, cardiac or muscle differentiation[48–50]. However, while other studies have reported lncRNAs that interact with IGF2BP2 and compete for its binding to target mRNAs (e.g., *LncMyoD* promotes muscle differentiation by outcompeting *c-Myc* and *N-Ras* mRNAs for IGF2BP2 binding[48]), in the cancer setting that we are studying reexpression of *RPSAP52* facilitates IGF2BP2 binding to a subset of mRNA targets, prominently *HMGA2* and *LIN28B* mRNAs. Our data indicate that this is achieved through modulation of the binding affinity that IGF2BP2 has for particular 3′UTRs and its distribution in large polysomes. This is reminiscent of the mechanism of action of *HIF1A-AS2* in glioblastoma cell lines, where binding of an antisense transcript to IGF2BP2 and DHX9 stimulates expression of their target mRNAs and promotes adaption to hypoxic stress[51]. Thus, our working model is that by forming ternary complexes (IGF2BP2-*RPSAP52*-other mRNAs), *RPSAP52* may influence the recruitment into ribonucleoprotein particles that dictate mRNA fate, and in particular enhance the translation of mRNAs that would otherwise be repressed by miRNAs. Binding by IGF2BP3 (another member of the IGF2BP family), for instance, has been associated with resistance to miRNA-dependent destabilization for many oncogenes, including *HMGA2* and *LIN28B*[52]. We hereby describe a similar scenario for IGF2BP2, and IGF2BPs are thus emerging as key nodes that integrate lncRNA-mediated post-transcriptional regulation of gene expression and pro-proliferative and self-renewal axis.

An important added layer of regulation exerted by *RPSAP52* is the influence on *let-7* levels, which may be a consequence of the control on LIN28B translation efficiency, or (in those cells where LIN28B is absent, such as MCF10A), may derive from the altered levels observed in IGF2BP2 protein itself upon depletion of the pseudogene. In fact, in glioblastoma cells lacking LIN28, *let-7* targets have been observed to be protected from miRNA dependent silencing by the binding of IGF2BP2 to *let-7* miRNA responsive elements[27]. This in turn may indirectly cause a decrease of *let-7* levels, since miRNA turnover might also depend on binding to mRNA targets, with some previous evidence suggesting that target availability prevents miRNA decay[53,54]. The greater effect of directly inhibiting biogenesis versus indirectly influencing the turnover might explain why *let-7* levels increase moderately in MCF10A upon RPSAP52 depletion (where LIN28B is not expressed and LIN28A is not altered, but IGF2BP2 levels decrease), and, by contrast, increase by almost one order of magnitude more in A673 cells (where expression of LIN28B protein is reduced) (compare Figs. 2c and 4d). Taken as a whole, *RPSAP52* is a pseudogene with an important impact on a major tumor suppressor miRNA. While silencing of *HMGA2* expression by *let-7* has been reported before[55], this is the first time that regulation of *let-7* levels by transcripts originating from *HMGA2 locus* is proposed. Importantly, this effect does not proceed through *HMGA2* itself, since depletion of *HMGA2* expression with gapmers actually increases *RPSAP52* levels and consequently results in a decrease in *let-7* family (Fig. 2e). This adds further complexity to the regulatory network, one hypothesis being that the tumorigenic cell activates an alternative pathway (increase of *RPSAP52*) to compensate for the loss of HMGA2 function.

Consistent with their convergent roles in the same pathway, low expression of *HMGA2/RPSAP52* in differentiated cells and reexpression in cancer mirrors LIN28 levels, which is one of the key players in maintenance of the pluripotent state. *Let-7* levels are maintained low in embryonic stem cells and certain primary tumors due to inhibition by LIN28 proteins, which are present at characteristically high levels in undifferentiated cells[56,57]. Of all tumor suppressor miRNAs, *let-7* is the one whose loss is most frequently correlated with poor prognosis in meta-analysis reports[58]. Accordingly, LIN28A/B high expression is a marker of poor prognosis and more aggressive tumors in a variety of cancers, and their levels have also been associated with metastatic and drug-resistant cases[59]. Thus, regulation of LIN28B/*let-7* balance is one important driver in cancer development.

An important aspect of this LIN28B/*let-7* balance is their counteracting action on the stemness characteristics of cancer cells. Interestingly, LIN28B/*let-7* signaling has been shown to regulate endogenous Oct4 and Sox2 expression by using ARID3B and HMGA2 as downstream effectors, and thereby regulate stemness properties in oral squamous cancer[60]. Also, the role of *let-7* in antagonizing self-renewal and promoting differentiation has been established via targeting of Myc, Ras, and HMGA2 pathways[61,62]. In accordance with *let-7* anti-pluripotency properties, we observe a decrease in NANOG and OCT4 levels as well as in clonogenicity upon *RPSAP52* depletion (Fig. 3d, e), suggesting that *RPSAP52* is an enhancer of stem cell characteristics. To date, few lncRNAs have been thoroughly described regarding their involvement in stemness, among them *H19* (whose downregulation reduces NANOG, OCT4, and SOX2 in glioma and breast cancer[63]) and *lncRNA ROR* (which inhibits proliferation of glioma stem cells by negatively regulating KLF4[64]). In particular, *H19* has recently been proposed to facilitate tumorigenesis through sponging of *let-7*[65]. The regulatory mechanism used by *RPSAP52*, by contrast, targets *let-7* family through modulation of LIN28B and/or target availability.

Taken together, we have observed that *RPSAP52* (1) stimulates proliferative and self-renewal axes together with a reduction of *let-7* levels, (2) promotes tumorigenic behavior *in vitro* and *in vivo*, and (3) is overexpressed in a number of human cancers and its expression is associated with worse outcome. This, together with its virtual absence in normal differentiated cells and embryonic expression pattern allows us to propose that *RPSAP52* is an oncofetal pseudogene that enhances proliferative and survival programs across several tumor types and whose expression in cancer can have important clinical implications. Indeed, *RPSAP52* levels are more useful as biomarkers in sarcoma than *HMGA2* mRNA levels, which do not seem to correlate well with protein levels (as suggested by our work and others[66]), probably due to the complex post-transcriptional regulation of HMGA2. The potential use of this pseudogene as an effective therapeutic target in human cancer will thus be the focus of future studies.

Methods

DNA methylation analysis. Genome-wide DNA methylation analysis was performed with the 450K DNA methylation microarray from Illumina (Infinium HumanMethylation450 BeadChip). Bisulfite-treated DNA from the indicated breast cancer cell lines was hybridized onto the array. A three-step normalization procedure was performed using the lumi package v2.30.0 (available for Bioconductor, within the R v3.4.3 statistical environment), consisting of color bias and background level adjustment and quantile normalization across arrays. The methylation level (*β*-value) of CpG sites was calculated as the ratio of methylated signal divided by the sum of methylated and unmethylated signals.

Bisulfite genomic sequencing. The Methyl Primer Express v1.0 software (Applied Biosystems) was used to design specific primers for the methylation analysis of *HMGA2/RPSAP52* island (Supplementary Table 1). Genomic DNA (1 µg) was subjected to sodium bisulfite treatment using the EZ DNA Methylation-Gold kit (Zymo Research). For bisulfite genomic sequencing, 300–500 bp fragments were amplified using 1–2 µl of bisulfite-converted DNA with Immolase Taq polymerase (Bioline) for 42 cycles. The resulting PCR products were gel-purified with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) and then cloned into the pSpark® TA vector (Canvas) according to the manufacturer's protocol. For all samples, 10 colonies were randomly chosen, the DNA was purified using NucleoSpin® 96 Plasmid (Macherey-Nagel) and sequenced by the 3730 DNA Analyzer (Applied Biosystems). After sequencing analysis with BioEdit v7.2.5 software, C nucleotides that remained unaltered were transformed into percentages of CpGs showing methylation.

Western blotting. Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8, 5 mM EDTA, 350 mM NaCl, 0.5% NP40, 10% glycerol, 0.1% SDS and phosphatase inhibitors), sonicated and centrifuged to recover the supernatant. The concentration was determined with the Pierce BCA Protein Assay Kit (#23227, ThermoFisher). Proteins were boiled for 5 min with Laemmli buffer (2% SDS, 10% glycerol, 60 mM Tris-Cl pH 6.8, 0.01% bromophenol blue) plus 2% 2-mercaptoethanol as a loading buffer, and equal amounts of extracts were loaded onto Tris-Glycine-SDS gels. Proteins were transferred to a nitrocellulose membrane (Whatman, GE Healthcare) and incubated overnight at 4 °C with primary antibodies diluted in 5% skimmed milk in PBS containing 0.1% Tween-20. The detected proteins were IGF2BP2 (H00010644-M01, Abnova, 1:500), RAS (ab55391, Abcam, 1:500, which recognizes all RAS proteins), LIN28B (ab71415, Abcam, 1:1000), IGF1R (#3027, Cell Signaling, 1:1000), ERK (#4695, Cell Signaling, 1:1000), p-ERK (#9101, Cell Signaling, 1:1000), LAMIN B1 (ab16048, Abcam, 1:4000), LIN28A (#8641, Cell Signaling, 1:750), α-TUBULIN HRP (ab40742, Abcam, 1:5000), HNRNPQ (ab184946, Abcam, 1:10,000), β-ACTIN HRP (a3854, Sigma, 1:20,000), NANOG (#4903, Cell Signaling, 1:2000), OCT4 (#2750, Cell Signaling, 1:1000), SOX2 (#4195, Cell Signaling, 1:1000), NUCLEOLIN (#8031, SantaCruz, 1:2000), HMGA2 (ab97276, Abcam, 1:1000), HISTONE H3 (ab1791, Abcam, 1:5000), RPSA (ab133645, Abcam, 1:1000), and RPL5 (A303-933A, Company Bethyl, 1:1000). After three washes with PBS containing 0.1% Tween-20, membranes were incubated for 1 h at room temperature in a bench-top shaker with the secondary antibodies conjugated to horseradish peroxidase anti-rabbit IgG (A0545, Sigma, 1:10,000) or anti-mouse IgG (NA9310, GE HealthCare, 1:5000). ECL reagents (Luminata-HRP; Merck-Milllipore) were used to visualize the proteins.

Nuclear/cytoplasmic fractionation and poly(A) selection. Subcellular fractionation was performed with PARIS™ kit (#AM1921, Life Technologies). Equal amounts of RNA from each fraction were subject to RT-qPCR and the results were normalized taking into account the total quantity of RNA recovered from each fraction. To verify the nuclear and cytoplasmic fractionation of the mRNA, *RNU6B* and *GAPDH* were used as controls, respectively. The separation was confirmed at the protein level by western blot with HISTONE H3 (ab1791, Abcam, 1:5000) and α-TUBULIN HRP (ab40742, Abcam, 1:5000). Poly(A)+ and poly(A)− RNAs were separated using the Dynabeads® mRNA Purification kit (#61006, Life Technologies),using three rounds of selection. RNA enrichment in each fraction was then analyzed by RT-qPCR, using *GAPDH* and *RNU6B* as controls.

RNA-biotin pull-down. Full-length *RPSAP52* (including alternative exon) or truncated fragments, as well as the antisense version or the sequence corresponding to the unrelated *Uc.160* + RNA, were biotin-labeled by standard *in vitro* transcription reactions and gel-purified. DNA templates for transcription were prepared by PCR with oligos described in Supplementary Table 1. The pull-downs were carried out with 10 pmol of each biotinylated RNA and 1 mg of total MCF10A or A673 protein extracts. Following incubation with the extract, each RNA was retrieved with 25 µl of Dynabeads® M-270 Streptavidin beads (#65305, Invitrogen) and washed in RIP buffer (150 mM KCl, 25 mM Hepes at pH 7.9, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 1 × protease inhibitor cocktail (Roche)). Binding proteins were released through boiling in SDS loading buffer and samples were run on a 4–12% gradient pre-cast Bis-Tris protein gels (Invitrogen) in MOPS buffer. After electrophoresis, the gels were either stained with SYPRO Ruby (Invitrogen) for band visualization and MS analysis or transferred to nitrocellulose membranes for western blotting.

In-gel digestion and LC-MS/MS analysis. Gel bands were manually excised and digested with trypsin overnight. Bands were then washed with water, 50 mM ammonium bicarbonate and 50% acetonitrile. Samples were subsequently reduced with 10 mM DTT and alkylated with 35 mM iodoacetamide. Extracted peptides were analyzed on a Ion Trap Amazon Speed ETD (Bruker Daltonics, Bremen,Germany) fitted with a captive spray source (Bruker, Daltonisc) following separation with Easy-nLCII apparatus (Proxeon). Peptides were separated in a reverse phase chromatography using a nano-capillary analytical c18 column. Peptide masses were analyzed at full scan MS, and then at MS/MS fragmentation for the most intense peaks. Data were analyzed using the Mascot search engine and the SwissProt human database.

Reverse pull-down. Endogenous IGF2BP2 was immunoprecipitated from A673 cell extracts. One milligram of total protein was incubated overnight with 2 µg of anti-IGF2BP2 polyclonal antibody (#H00010644-M01, Abnova) or control mouse IgG antibody (#12-371, Millipore) and 40 µl of Dynabeads® M-280 anti-mouse IgG beads (#11202D, ThermoFisher) in 1 ml of RIP buffer (150 mM KCl, 25 mM Hepes at pH 7.9, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 1× protease inhibitor cocktail (Roche)). Beads were then washed three times with RIP buffer and 10% of the volume was boiled in the presence of Laemmli buffer for western blot analysis. The pulled-down RNA in the remaining 90% of beads was extracted by adding 1 ml of TRIzol® Reagent (15596-018, ThermoFisher). After phenol extraction and isopropanol precipitation, the final pellet was resuspended in 10 µl of H_2O and retrotranscribed. cDNA was analyzed by either 30 cycles of semi-quantitative RT-PCR or by RT-qPCR (Applied Biosystems 7900HT Fast Real-Time PCR System). Two micrograms of input RNA was processed in parallel to estimate pull-down efficiency.

Cell culture. MCF7, MCF10A, HCC1143, and A673 cell lines were purchased from ATCC. The remaining breast and sarcoma cell lines were obtained from Dr. Esteller and Dr. Tirado’s labs, respectively. Authenticity of the cell lines was routinely confirmed by

STR profiling analysis done at qGenomics SL (Esplugues de Llobregat, Barcelona, Spain). All cell lines were routinely checked for mycoplasma contamination. Non-malignant MCF10A breast cells were grown in DMEM/Ham's F-12 medium (#L0093-500, Biowest) supplemented with 20 ng ml^{-1} EGF (#SRP3027, Sigma), 500 ng ml^{-1} hydrocortisone (#H0888, Sigma), and 10 µg ml^{-1} insulin (#I9278, Sigma). Ewing's sarcoma A673 cells and breast cancer HCC1143 cell line were grown in RPMI-1640 medium with GlutaMAX (#61870-010, Gibco). HEK293T cells, used for the production of lentiviral particles, and breast cancer Hs578T cell line were cultivated in DMEM with GlutaMAX (#31966-021, Gibco). All the media were supplemented with 10% fetal bovine serum (FBS) (#10270,Gibco), and the cells were grown at 37 °C in a humidified atmosphere of 5% CO_2 and 95% air.

Plasmid construction and transfections. Stable knockdown of *RPSAP52* was achieved with the following sequences: shRNA1 and shRNA4 target, respectively, the 5′ TCCTTAAGCTCCTTGCAGT 3′ and 5′ CACGGACTCTTAAGCAACA 3′ sequences of *RPSAP52* mRNA (both located on the last exon), whereas shRNA3 targets the 5′ GTGCAAGACTCAGGAGCTA 3′ sequence of *RPSAP52* (on the first intron, which is inefficiently spliced). These shRNAs were expressed by cloning oligos shRPSAP52-1for and shRPSAP52-1rev (shRNA1), shRPSAP52-4for and shRPSAP52-4rev (shRNA4), and shRPSAP52-3for and shRPSAP52-3rev (shRNA3) into the BamHI and EcoRI sites of the vector pLVX-shRNA2 (Clontech). A scramble (scr) sequence was used as a control. For lentivirus-mediated depletion, HEK293T cells were transfected with pLVX-shRNA2-constructs plus packaging plasmids with jetPRIME® (Polyplus-transfection) according to the manufacturer's recommendations. The target cell line was infected with the supernatant containing viral particles 48 h post-transfection. ZsGreen1 was used as a marker to visualize transductants by fluorescence microscopy, and these cells were selected by fluorescence-activated cell sorting (FACS) and plated to obtain stable clones. Antisense oligonucleotides (LNA™ GapmeRs, #300600, Exiqon) targeting *HMGA2* mRNA (HMGA2), exon1 (RPSAP52 Ex) or the first intron (RPSAP52 RL1 and RL2) of *RPSAP52* transcript were transfected to a final concentration of 65 nM using HiPerfect (Qiagen). Cells were retransfected 48 h later and collected 72 h after the second round of LNA treatment. A control LNA GapmeR (#300610, Exiqon) was used as mock transfection. For siRNA-mediated knockdown of LIN28B, cells were transfected with a 1:1 mix of two different siRNAs against LIN28B (#216387-216388, Ambion) and a negative control (C-) (#AM4611, Ambion), using Lipofectamine™ RNAiMAX Transfection Reagent (#13778, Invitrogen) according to the manufacturer’s recommendations. The overexpression of LIN28B was achieved with pcDNA3-FLAG-Lin28B (#51373, Addgene), and we used pcDNA3.1 + (Invitrogen) as a control (empty) and jet-PRIME® (Polyplus-transfection) as the transfection reagent.

Transcription/translation assay. *RPSAP52* full-length transcript (from RefSeq NR_026825 annotation) was amplified by PCR from A673 cDNA with oligos T7-RPSAP52for-TnT and RPSAP52rev-TnT, producing two isoforms that include or exclude the alternative exon downstream of the T7 bacteriophage promoter. After gel purification, the PCR product was used directly in coupled transcription and translation reactions in reticulocyte extracts (TNT® Quick Coupled Transcription/Translation Systems, Promega), following the manufacturer’s indications and by labeling the reactions with ^{35}S. The translation products were separated by SDS-PAGE, the gel was then vacuum-dried and exposed overnight with an autoradiography film.

RNA isolation and RT-qPCR analysis. Total RNA, including miRNAs, was extracted using the Maxwell® RSC instrument with the Maxwell® RSC miRNA Tissue kit (Promega) according to the manufacturer’s recommendations. For mRNA expression analysis, total RNA was reverse transcribed using the Super-Script™ III Reverse Transcriptase (#18080, Invitrogen). Real-time PCR reactions were performed in triplicate in an Applied Biosystems 7900HT Fast Real-Time PCR system, using 30–100 ng cDNA, 6 µl SYBR® Green PCR Master Mix (Applied Biosystems), and 416 nM primers in a final volume of 12 µl for 384-well plates. All data were acquired and analyzed with the QuantStudio Design & Analysis Software v1.3.1 and normalized with respect to *GUSB* as endogenous control. Relative RNA levels were calculated using the comparative C_t method ($\Delta\Delta C_t$). For miRNA expression analysis, miRCURY LNA™ Universal RT microRNA PCR System (Exiqon) was used, according to the manufacturer’s recommendations, with the Universal cDNA Synthesis Kit II (#203301) for the RNA retrotranscription and the ExiLENT SYBR® Green master mix (#203421) for the RT-qPCR in the LightCycler® 480 (Roche) with the LightCycler® 480 Software v1.5.0 SP4. To normalize the data, *RNU6B* or *miR-195* were used as endogenous control. Actinomycin D treatment and RNA stability analysis. Control or *RPSAP52*-depleted A673 clones were treated with either 0.5% DMSO or 5 µg ml^{-1} Actinomycin D (Sigma) for 9 h. Pellets of each condition and treatment were harvested at different times and RNA was extracted for the RT-qPCR experiments. All data were normalized with respect to *GUSB* as endogenous control and gene expression fold-changes induced by Actinomycin D were calculated relative to the control (DMSO) cells of each condition and time point. *c-FOS* and *GAPDH* were used as controls of the experiment due to their short and long half-life, respectively.

SRB assay. Cell viability and proliferation were determined by the sulforhodamine B (SRB) assay. Cells were seeded in 96-well microplates in medium with 10% FBS, and the experiment started after 24 h of incubation at 37 °C and 5% CO_2. The optimal cell number (100 cells per well for MCF10A and 2000 cells per well for Hs578T and HCC1143) was determined to ensure that the cells were in growth phase at the end of the assay. During 7 consecutive days, at least 6 wells per condition were processed as follows. The medium was removed and the cells were fixed with 10% trichloroacetic acid for 1 h at 4 °C. Then, two washes with 1% acetic acid were performed and the viable cells were stained with 0.057% SRB in 1% acetic acid. Following 30 min of incubation at RT, the SRB was removed by washing twice with 1% acetic acid. The wells were air-dried completely and the SRB bound to the viable cells was dissolved with 100 µl of Tris-HCl 10 mM (pH 10.0). Absorbance at 540 nm was determined on an automatized microtiter plate reader PowerWave XS (BioTek).

Colony formation assay. Cells were seeded into 35 mm dishes with three triplicates per condition at a density of 200 cells per plate for MCF10A and HCC1143 and 500 cells per plate for Hs578T. They were maintained for 8–15 days in a humidified incubator with 5% CO_2 at 37 °C. Cells were then fixed in 4% paraformaldehyde and stained with 0.5% crystal violet for 30 min. Digital images were obtained using GBox (Syngene) and colonies containing more than 50 cells were counted manually using ImageJ v1.50 software. Plating efficiency and survival fractions were determined by using the following formulas:

$$\text{Plating efficiency} = \frac{\text{number of colonies obtained}}{\text{number of cells seeded}}$$

$$\text{Surviving fraction} = \frac{\text{plating efficiency}}{\text{number of colonies obtained in the control condition}} \times 100$$

Real-time migration assay. The xCELLigence Real-Time system (ACEA Biosciences) was used with CIM-16 plates of 8 µm pore membranes. The lower chamber wells were filled with 160 µl of medium containing 10% FBS and the top chamber wells with 40 µl of serum-free medium. The two chambers were assembled together and allowed to equilibrate for 1 h at 37 °C and 5% CO_2. Cells were incubated for 24 h in serum-free medium, rinsed with PBS, trypsinized and resuspended in medium supplemented with 10% FBS to inactivate the trypsin, followed by centrifugation and resuspension in serum-free medium. A total of 8×10^4 cells were seeded onto the top chamber of CIM-16 plates and placed into the xCELLigence system for data collection after background measurement. The software RTCA 2.0 was set to collect impedance data every 15 min. The cell index represents the capacity for cell migration, whereas the slope of the curve can be related to the cell invasion ability.

Transwell migration assay. Transwell® Permeable Supports (#3422, Cultek) with 8 µm pore polycarbonate membranes in 24-well plates were used to measure cell migration. Cells were incubated for 24 h in serum-free medium, rinsed with PBS, trypsinized, and resuspended in 10% FBS-containing medium to inactivate the trypsin, followed by centrifugation and resuspension in serum-free medium. A total of 1×10^5 cells were seeded onto each transwell with 150 µl of serum-free medium and the transwells were placed in the wells of a 24-well plate with 500 µl of10% FBS-containing medium. The chemoattractant promoted the migration of the cells from the upper part of the transwell to the lower part. After 24 h of incubation at 37 °C and 5% CO_2, the cells in the upper part of the membrane were removed with a cotton swab and several washes with 1× PBS. Cells in the lower part were fixed for 10 min with ice-cold 100% methanol. For the staining, cells were covered with 0.5% crystal violet in 25% methanol for 10 min. Transwells were washed several times with 1× PBS and air-dried. Membranes were then mounted on a slide for image acquisition.

Clonogenicity assay. The clonogenicity of *RPSAP52*-depleted MCF10A clones was tested in soft agar by using the CytoSelect 96-well Cell Transformation Assay Kit (Cell Biolabs, #CBA-130), following the manufacturer’s instructions. Briefly, a base agar layer was prepared by mixing equal volumes of 1.2% agar solution and 2× DMEM/20% FBS medium in each well of a 96-well flat-bottom microplate. In total, 5000 cells per well were seeded in a top layer by mixing equal volumes of the cell suspension, 1.2% agar solution and 2× DMEM/20% FBS (1:1:1), and incubated for 6 days after covering the solidified cell agar layer with 100 μl of DMEM-F12 medium plus supplements. The CyQuant GR dye was used to detect the lysed colonies and the proportional fluorescence to the number and size of colonies was read using a PerkinElmer's VICTOR X5 multilabel plate reader with a 485/535 filter set and 1 s of measurement time. The data are expressed in relative fluorescence units (RFU).

***In vivo* xenograft.** Athymic nude female mice (Charles River, Inc (USA), strain Crl:NU(NCr)-Fox1nu) were subcutaneously injected at 7–8 weeks of age in one flank with a total of 10×10^6 MCF10A, 7×10^6 A673, and 5×10^5 Hs578T cells from clones expressing either scrambled or *RPSAP52*-shRNAs, soaked in 100 μl of Matrigel (BD Biosciences). Tumor growth was monitored every 7 days for MCF10A, 3–4 days for A673, and 4 days for Hs578T by measuring tumor width (W, mm) and length (L, mm) until mice were killed at the indicated days post-injection. After allowing to grow for several weeks, tumor volume (V, mm^3) was estimated from the formula $V = \frac{\pi \times L \times W^2}{6}$ and tumor weight (g) measured. Animal tests complied with ethical regulations. All the mouse experiments were approved by IDIBELL's Committee for Animal Experimentation.

Absolute quantification. Estimations of the absolute amounts of RNAs were obtained by comparison with in vitro transcribed RNA standards of known concentration. These RNA standards correspond to the sequences amplified in RT-qPCR in the analysis of *RPSAP52 + altex*, *RPSAP52 – altex*, *HMGA2* and *LIN28B* transcript expression in all figures, and were generated by introducing the T7 RNA Polymerase promoter upstream of the amplicon by PCR and subsequent *in vitro* transcription. Serial dilutions of the synthesized RNA standards were used as spike-ins in total RNA extractions from MCF7 cells (which do not express any of the transcripts of interest) and processed in parallel in RT-qPCR with RNA extractions from a known number of MCF10A or A673 cells, so that transcript copy number per cell could be measured. Similarly, for determination of *let-7* copy number, we generated a standard curve using synthetic *let-7a, b* and *e* purified RNA oligonucleotides (Sigma), corresponding to the sequences *hsa-let-7a-5p* (5′-rUrGrArGr-GrUrArGrUrArGrGrUrUrGrUrArUrArGrUrU-3′), *hsa-let-7b-5p* (5′-rUrGrArGrGrUr-ArGrUrArGrUrUrGrUrGrUrGrGrUrU-3′), and *hsa-let-7e-5p* (5′-rUrGrArGrGrUrArGr-GrArGrGrUrUrGrUrArUrArGrUrU-3′). For protein quantification, total extracts from a recorded number of cells was analyzed by western blot in parallel with known amounts of the following recombinant proteins: LIN28B (ab134596, Abcam), IGF2BP2 (ab153107, Abcam), HNRNPQ (ab153089, Abcam). Western blot was performed with the following antibodies: anti-LIN28B (ab71415, Abcam, 1:1000), anti-IGF2BP2 (H00010644-M01, Abnova, 1:500), anti-HNRNPQ (NBP1-57197, Novus Biologicals, 1:1000). Band intensity was measured by densitometry with an iBright™ CL1000 Imaging System (ThermoFisher).

iCLIP-seq. iCLIP-seq was performed on stable A673 clones. Approximately 8 × 10^6 A673 cells stably expressing scrambled shRNA (scr) or shRNA-4 against *RPSAP52* (clone B11) were crosslinked with 150 mJ cm^{-2} total 254-nm irradiation in a Stratalinker 2400. The same amount of non-crosslinked cells were used as controls. Cell lysates were treated with different concentrations (2 or 0.4 U μl^{-1}) of RNaseI (#AM2294, ThermoFisher) and 4 U of Turbo DNase (#AM2238, ThermoFisher) in a final volume of 1 ml. Lysates were then cleared and immunoprecipitated overnight at 4 °C with 10 µg of anti-IGF2BP2 antibody (#RN008P, MBL) preincubated for 1 h at room temperature with 60 µl anti-rabbit IgG Dynabeads (#11204D,ThermoFisher). After two washes in high-salt buffer (50 mM Tris-HCl pH 4.4, 1 M NaCl, 1 mM EDTA, 1% Igepal CA-630, 01% SDS, 0.5% sodium deoxycholate) and one wash in PNK buffer (20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 0.2% Tween-20), RNA 3′end was dephosphorylated with PNK for 20 min at 37 °C. Beads were then washed once with PNK buffer, once with high-salt buffer and twice with PNK buffer. L3 adapter was then ligated overnight at 16 °C in a 20 µl reaction containing 1.5 µM pre-adenylated L3-App adapter (rAppAGATCGGAAGAGCGGTTC-AG//ddC/), 4 µl PEG400 and 10 U T4 RNA Ligase1 (#M0204, New England Biolabs). Beads were then washed twice with high-salt buffer and twice with PNK buffer, and 20% of beads were radioactively labeled with γ-[^{32}P]-ATP and 0.5 U μl^{-1} PNK (#M0201, New England Biolabs) for 5 min at 37 °C, added to the remaining cold beads and incubated in 20 µl 1x NuPAGE buffer for 5 min at 70 °C prior to loading the supernatant on a 4–12% NuPAGE Bis-Tris gel (#NP0341BOX, ThermoFisher). The gel was run at 180 V for 50 min, and the protein–RNA complexes were transferred to a nitrocellulose membrane at 30 V for 1 h. The membrane was then autoradiographed through exposure to a film at −80 °C for 1 h. Regions of interest containing the IGF2BP2-RNA crosslinked products were cut out of the membrane and the RNA fragments isolated, reverse transcribed, purified and circularized by incubating with CircLigase II (Epicentre) for 1 h at 60 °C, followed by annealing to Cut_oligo (5′-GTTCAGGATCCACGACGCTCTTCAAAA-3′) and digestion with BamHI. iCLIP libraries were amplified for 27 cycles with P3/P5 Solexa primers, and the appropriate size of products were confirmed by gel electrophoresis. Sequencing of the libraries was performed on a MiSeq instrument following standard manufacturer's procedures, using MiSeq Reagent Kit v3 reagents with single-read, 151-bases read profile. Two independent experiments were performed for each condition. The four libraries were sequenced in two separated pools (#31 & #38) and data acquired with MiSeq Reporter v2.6.3.2. Raw data can be downloaded from https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA484688.

iCLIP computational analysis. Read quality was assessed using FastQC (v0.11.7) software (available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc). After sequencing, PhiX sequences were removed using BWA (Burrows-Wheeler Aligner, v0.7.17)[67]. All the pre-processing steps, peak calling, CITS calling, and annotations were performed using CTK (CLIP Tool Kit) (v1.0.9) software[68], following the recommendations found in https://zhanglab.c2b2.columbia.edu/index.php/ICLIP_data_ analysis_using_CTK. After pre-processing steps, final tags from the two biological replicates of the same condition were merged to proceed with peak and CITS calling steps. Peak calling was statistically assessed using a Bonferroni adjusted P-value < 0.05 as a significance threshold. For CITS calling, all tags presenting substitutions were excluded from the analysis, since the abnormally high frequency of substitutions observed could be due to reverse transcriptase read through. CITS were considered significant with a *P*-value < 0.001. No proximity clustering was applied in either of the analysis. Adapter trimming, sequence alignments, and alignment manipulations were performed using Cutadapt (v1.16), BWA and samtools (v1.8), respectively. All genome alignments and annotations used hg19 human genome (GCA_000001405.1) as a reference. *De novo* motif discovery from CITS was performed using HOMER (v4.10) software[69], for which a window of CITS +/−10 nucleotides was taken. Additional python scripts were used for specific CAUH motif enrichment analysis. Genome Browser images were generated using Golden Helix GenomeBrowse® v3.0.0 software (available from http://www.goldenhelix.com). *Let-7a/b/e* predicted target genes, used for *let-7* enrichment analysis in the obtained peaks, were downloaded from miRDB (http://www.mirdb.org/). For differential binding analysis, 3′UTR tags obtained after preprocessing steps by CTK software were used to generate 3′UTR tag counts for each biological replicate of the two conditions. Using these counts, differential binding analysis was performed using DESeq2 bioconductor package v1.18.1[70]. Gene enrichment analyses from significant peaks were conducted using Enrichr v2.0 software (http://amp.pharm.mssm.edu/Enrichr/)[71,72]. Additional graphics and statistical analysis were performed using R v3.4.3 programming language (https://www.R-project.org/).

Primary tumors expression and methylation analysis. RNA expression and DNA methylation data from different tumor types was collected from The Cancer Genome Atlas (TCGA) Data Portal (https://tcga-data.nci.nih.gov/tcga). TCGA data were downloaded using TCGAbiolinks v2.9.2[73]. Bioconductor package from the current GDC (Genomic Data Commons) harmonized database aligned against hg38 genome. Box plots analysis represent normalized expression and methylation values corresponding to TCGA COAD, LUSC, LUAD, THCA, and BRCA projects from the GDC data portal (https://portal.gdc.cancer.gov/). We use the Wilcoxon signed-rank test to compare differences between groups. The association between *HMGA2* and *RPSAP52* expression in primary tumors and cell lines was estimated with a Pearson's correlation. All the statistical analysis and graphical representations were performed using R v3.4.3. For primary samples, an average of *HMGA2/RPSAP52* promoter methylation > 0.26 (median of the population) was considered as hypermethylated. NCI60 cell lines expression data were downloaded from cBioPortal database (http://www.cbioportal.org/).

Polysome profile analysis. MCF10A cells or A673 stable clones were plated in 150 mm dishes and treated with 100 µg ml^{-1} cycloheximide (CHX) at 37 °C for 5 min. Cells were washed twice with cold PBS supplemented with CHX, pelleted and resuspended in 250 µl of hypotonic lysis buffer (1.5 mM KCl, 2.5 mM $MgCl_2$, 5 mM Tris-HCl pH 7.4, 1 mM DTT, 1% sodium deoxycholate, 1% Triton X-100, 100 µg ml^{-1} CHX) supplemented with mammalian protease inhibitors (SIGMA) and RNase inhibitor (NEB) at a concentration of 100 U ml^{-1}, and left on ice for 5 min. Cell lysates were cleared of debris and nuclei by centrifugation for 5 min at 17,000 × *g*. Protein concentrations were determined by BCA assay and 500 µg of lysate were loaded on 10–50% sucrose linear gradients containing 80 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.4, 1 mM DTT, 10 U ml^{-1} RNase inhibitor with a BIOCOMP gradient master. Gradients were centrifuged on a SW40 rotor for 3.5 h at 217,290 × *g*. Gradients were analyzed on a BIOCOMP gradient station and collected in 11 (MCF10A) or 13 (A673) fractions ranging from light to heavy sucrose.

Fractions were supplemented with SDS at a final concentration of 1% and placed for 10 min at 65 °C. To each fraction was added 1 ng of firefly luciferase mRNA, followed by phenol–chloroform extraction and precipitation with isopropanol. Purified RNAs from each fraction were retrotranscribed and subjected to qPCR. mRNA quantification was normalized to firefly mRNA.

IGF2BP2 co-immunoprecipitation and mass spectrometry analysis.

Immunoprecipitation and sample digestion for mass spectrometry analysis: 1 mg of precleared protein extract from three replicates of control cells (scr) and cells depleted for *RPSAP52* (sh4 B11 clone) were immunoprecipitated overnight at 4 °C using 5 µg of anti-IGF2BP2 antibody (#H00010644-M01, Abnova) and 40 µl of Dynabeads® M-280 anti-mouse IgG beads (#11202D, ThermoFisher) in 1 ml RIP buffer (150 mM KCl, 25 mM Hepes at pH 7.9, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 1× protease inhibitor cocktail (Roche)). After three washes with RIP buffer, the resulting material was in-bead digested with trypsin. Briefly, the beads were washed three times with 500 ml of 200 mM Ammonium Bicarbonate (ABC) and resuspended in 60 ml of 6 M Urea 200 mM^{-1} ABC. The samples were then reduced with 10 ml of 10mM DTT (1 h, 30 °C) and alkylated with 10 ml of 20 mM Iodoacetamide (30 min, room temperature and darkness). After that, the samples were diluted with 280 ml of 200 mM ABC and digested with 5 ml of 0.2 mg ml^{-1} Trypsin for 16 h at 37 °C. The beads were finally pulled-down (5 min at 5000 g), the supernatant transferred to new, cleaned tubes and acidified with 20 ml of 100% Formic acid. The resulting peptides mixtures were desalted using C18 stage tips (UltraMicroSpin Column, The Nest Group, Inc., MA) and dried in a SpeedVac.

Mass spectrometry analysis (LC-MS/MS): the dried-down peptide mixtures were analyzed in a nanoAcquity liquid chromatographer (Waters) coupled to aLTQ-OrbitrapVelos (ThermoScientific) mass spectrometer. The tryptic digests were resuspended in 10 µl 1% FA solution and an aliquot of 3 µl of each sample was injected for chromatographic separation. Peptides were trapped on a Symmetry C18TM trap column (5 µm, 180 µm × 20 mm; Waters), and were separated using a C18 reverse phase capillary column (ACQUITY UPLC BEH column; 130 Å, 1.7 µm, 75 µm × 250 mm, Waters). Eluted peptides were subjected to electrospray ionization in an emitter needle (PicoTipTM, New Objective) with an applied voltage of 2000 V. Peptide masses (300–1700 *m/z*) were analyzed in data dependent mode, where a full scan MS was acquired in the Orbitrap with a resolution of 60,000 FWHM at 400 *m/z*. Up to the 15th most abundant peptides (minimum intensity of 500 counts) were selected from each MS scan and then fragmented in the linear ion trap using CID (38% normalized collision energy) with helium as the collision gas. The scan time settings were Full MS: 250 ms (1 microscan) and MSn:120 ms. Generated .raw data files were collected with *ThermoXcalibur* (v2.2).

Data analysis: the .raw files were analyzed with the MaxQuant (v.1.6.2.6a) software using the built-in search engine Andromeda to search against the Swissprot Human database downloaded from UniprotKB website in March 26, 2018. The search parameters were set as follow: the enzyme was trypsin with a maximum of two allowed missed cleavages. Oxidation in methionines as well as Acetylation at protein N-terminal was set as variable modifications while carbamidomethylation in cysteines was set as fixed modification. The mass tolerances for the first and main search were set at 20 and 4.5 ppm, respectively. In addition, only peptides with more than 6 and up to 25 aminoacids were considered. The final list of identified peptides and proteins were filtered by using a 5% false discovery rate (FDR) both at peptide and protein level. To enhance the identification of proteins the match between runs option was selected.

Protein–protein interaction analysis: the statistical analysis of the protein–protein interactions found in our experimental conditions was performed with the help of the Significance Analysis of the INTeractome (SAINT) algorithm which was implemented in the http://statsms.crg.es/ site.

Expression arrays. Total RNA from each sample was quantified using the NanoDrop ND-1000 and RNA integrity was assessed by standard denaturing agarose gel electrophoresis. For microarray analysis, Agilent Array platform was employed. The sample preparation and microarray hybridization were performed based on the manufacturer's standard protocols. Briefly, total RNA from each sample was amplified and transcribed into fluorescent cRNA with using the manufacturer's Agilent's Quick Amp Labeling protocol (version 5.7, Agilent Technologies). The labeled cRNAs were hybridized onto the Whole Human Genome Oligo Microarray (4 × 44 K, Agilent Technologies). After having washed the slides, the arrays were scanned by the Agilent Scanner G2505C. Agilent Feature Extraction software (version 11.0.1.1) was used to analyze acquired array images. Quantile normalization and subsequent data processing were performed using the GeneSpring GX v12.1 software (Agilent Technologies).

Differentially expressed genes were identified through Fold Change filtering and Volcano filtering. Pathway analysis and GO Analysis were applied to determine the roles of these differentially expressed genes.

Statistical analysis. Bar graphics and statistical comparisons were obtained with the GraphPad Prism 8.1.2 software. Comparative analyses between different experimental groups were performed using *t*-student test and one-way ANOVA with Bonferroni's or Dunnet's *post hoc* tests for intergroup comparisons. For cancer patients' samples, we used the Kaplan–Meier method for survival analysis and the log-rank test was used to analyze the differences between the groups. Cox regression method was used to analyze the independent prognostic importance of expression or methylation. Results of the univariate Cox regression analysis are represented by the hazards ratio (HR) and 95% confidence interval (CI). Results were considered significant if the *P*-value was < 0.05 (*), < 0.01 (**), < 0.001 (***), or < 0.0001 (****). Unless otherwise stated, data are presented as the mean ±SD.

Unprocessed scans. All unprocessed and uncropped scans and images can be found in the source data file.

Data availability

All relevant data are available from the authors. Raw data for the iCLIP-seq experiment have been deposited under the accession code PRJNA484688 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA484688). iCLIP-seq peaks and CITS are provided in the Oliveira-Mateos et al_Supplementary Data 1. The list of significant altered transcripts from the microarray expression analysis is provided in the Oliveira-Mateos et al_Supplementary Data 2. Numerical source data for Figs. 1e–g, 2c, e, 3a–c, e–g, 4b, d, f, 5a, c, 6a–e, 7b; Supplementary Figs. 1c, 2b–e, 3c, d, f–h, 4d–f, 5g, j, 6a, c and 7b and all unprocessed images and scans for Figs. 1d, f, 2a, b, d, f, g, 3d, f, g, 4b, c, e, 5a–c, 6f; Supplementary Figs. 2a, 3e–h, 4d, 5a, e, h, i, and 6b can found in the source data file.

Received: 29 August 2018
Accepted: 8 August 2019
Published online: 04 September 2019

References

1. Djebali, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
2. Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell. Biol.* **17**, 756–770 (2016).
3. Délas, M. J. & Hannon, G. J. LncRNAs in development and disease: from functions to mechanisms. *Open Biol.* **7**, 170121 (2017).
4. Di Gesualdo, F., Capaccioli, S. & Lulli, M. A pathophysiological view of the long non-coding RNA world. *Oncotarget* **5**, 10976–10996 (2014).
5. Holdt, L. M. et al. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. *PloS Genet* **9**, e1003588 (2013).
6. Huarte, M. et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
7. Ying, L. et al. Downregulated MEG3 activates autophagy and increases cell proliferation in bladder cancer. *Mol. Biosyst.* **9**, 407–411 (2013).
8. Panzitt, K. et al. Characterization of HULC, a novel gene with striking upregulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology* **132**, 330–342 (2007).
9. Arun, G. et al. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev.* **30**, 34–51 (2016).
10. Gupta, R. A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
11. Fu, X., Ravindranath, L., Tran, N., Petrovics, G. & Srivastava, S. Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. *DNA Cell Biol.* **25**, 135–141 (2006).
12. Venteicher, A. S. et al. A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644–648 (2009).
13. Yu, T. Y., Kao, Y. W. & Lin, J. J. Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proc. Natl Acad. Sci. USA* **111**, 3377–3382 (2014).

14. Achour, C. & Aguilo, F. Long non-coding RNA and Polycomb: an intricate partnership in cancer biology. *Front. Biosci.* **23**, 2106–2132 (2018).
15. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018).
16. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-seq data. *Nucleic Acids Res.* **42**, D92–D97 (2014).
17. Wu, Q. et al. Analysis of the miRNA-mRNA-lncRNA networks in ER+ and ER− breast cancer cell lines. *J. Cell. Mol. Med.* **19**, 2874–2887 (2015).
18. Arun, K. et al. Comprehensive analysis of aberrantly expressed lncRNAs and construction of ceRNA network in gastric cancer. *Oncotarget* **9**, 18386–18399 (2018).
19. Johnsson, P., Morris, K. V. & Grandér, D. Pseudogenes: a novel source of trans-acting antisense RNAs. *Methods Mol. Biol.* **1167**, 213–226 (2014).
20. Poliseno, L. et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
21. Wend, P. et al. WNT10B/β-catenin signalling induces HMGA2 and proliferation in metastatic triple-negative breast cancer. *EMBO Mol. Med.* **5**, 264–279 (2013).
22. Mahajan, A. et al. HMGA2: a biomarker significantly overexpressed in high grade ovarian serous carcinoma. *Mod. Pathol.* **23**, 673–681 (2010).
23. Boque-Sastre, R. et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl Acad. Sci. USA* **112**, 5785–5790 (2015).
24. Wu, J. et al. Elevated HMGA2 expression is associated with cancer aggressiveness and predicts poor outcome in breast cancer. *Cancer Lett.* **376**, 284–292 (2016).
25. Iyer, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
26. Yisraeli, J. K. VICKZ proteins: a multi-talented family of regulatory RNA binding proteins. *Biol. Cell.* **97**, 87–96 (2005).
27. Degrauwe, N., Suvà, M. L., Janiszewska, M., Riggi, N. & Stamenkovic, I. IMPs: an RNA-binding protein family that provides a link between stem cell maintenance in normal development and cancer. *Genes Dev.* **30**, 2459–2474 (2016).
28. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

29. Cleynen, I. et al. HMGA2 regulates transcription of the Imp2 gene via an intronic regulatory element in cooperation with nuclear factor-kappaB. *Mol. Cancer Res.* **5**, 363–372 (2007).
30. Li, Z. et al. An HMGA2-IGF2BP2 axis regulates myoblast proliferation and myogenesis. *Dev. Cell* **23**, 1176–1188 (2012).
31. Li, Z. et al. Oncogenic NRAS, required for the pathogenesis of embryonic rhabdomyosarcoma, relies upon the HMGA2-IGF2BP2 pathway. *Cancer Res.* **73**, 3041–3050 (2013).
32. Busch, B. et al. The oncogenic triangle of HMGA2, LIN28B and IGF2BP1 antagonizes tumor-suppressive actions of the let-7 family. *Nucleic Acids Res.* **44**, 3845–3864 (2016).
33. Degrauwe, N. et al. The RNA binding protein IMP2 preserves glioblastoma stem cells by preventing let-7 target gene silencing. *Cell Rep.* **15**, 1634–1647 (2016).
34. Heo, I. et al. TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696–708 (2009).
35. Piskounova, E. et al. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **147**, 1066–1079 (2011).
36. Viswanathan, S. R. et al. Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.* **41**, 843–848 (2009).
37. Nielsen, J. et al. A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol. Cell. Biol.* **19**, 1262–1270 (1999).
38. Dai, N. et al. IGF2 mRNA binding protein-2 is a tumor promoter that drives cancer proliferation through its client mRNAs IGF2 and HMGA1. *eLife* **6**,e27155 (2017).
39. Gong, C. et al. A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev. Cell.* **34**, 181–191 (2015).
40. Conway, A. E. et al. Enhanced CLIP uncovers IMP protein-RNA targets in human pluripotent stem cells important for cell adhesion and survival. *Cell Rep.* **15**, 666–679 (2016).
41. Scotlandi, K. et al. Overcoming resistance to conventional drugs in Ewing sarcoma and identification of molecular predictors of outcome. *J. Clin. Oncol.* **27**, 2209–2216 (2009).

42. Hong, S. H. et al. High neuropeptide Y release associates with Ewing sarcoma bone dissemination—in vivo model of site-specific metastases. *Oncotarget* **6**, 7151–7165 (2015).
43. Kainov, Y. et al. CRABP1 provides high malignancy of transformed mesenchymal cells and contributes to the pathogenesis of mesenchymal and neuroendocrine tumors. *Cell Cycle* **13**, 1530–1539 (2014).
44. Zaugg, K. et al. Carnitine palmitoyl transferase 1C promotes cell survival and tumor growth under conditions of metabolic stress. *Genes Dev.* **25**, 1041–1051 (2011).
45. Emori, M. et al. High expression of CD109 antigen regulates the phenotype of cancer stem-like cells/cancer-initiating cells in the novel epithelioid sarcoma cell line ESX and is related to poor prognosis of soft tissue sarcoma. *PloS One* **8**, e84187 (2013).
46. Fujikawa, A. et al. Targeting PTPRZ inhibits stem cell-like properties and tumorigenicity in glioblastoma cells. *Sci. Rep.* **7**, 5609 (2017).
47. Zeleniak, A. E., Huang, W., Brinkman, M. K., Fishel, M. L. & Hill, R. Loss of MTSS1 results in increased metastatic potential in pancreatic cancer. *Oncotarget* **8**, 16473–16487 (2017).
48. Agarwal, E. et al. Role of Akt2 in regulation of metastasis suppressor 1 expression and colorectal cancer metastasis. *Oncogene* **36**, 3104–3118 (2017).
49. Zhang, X. et al. Interrogation of nonconserved human adipose lincRNAs identifies a regulatory role of linc-ADAL in adipocyte metabolism. *Sci. Transl. Med.* **10**, eaar5987 (2018).
50. Hosen, M. R. et al. Airn regulates Igf2bp2 translation in cardiomyocytes. *Circ. Res.* **122**, 1347–1353 (2018).
51. Mineo, M. et al. The long non-coding RNA HIF1A-AS2 facilitates the maintenance of mesenchymal glioblastoma stem-like cells in hypoxic niches. *Cell Rep.* **15**, 2500–2509 (2016).
52. Jønson, L. et al. IMP3 RNP safe houses prevent miRNA-directed HMGA2 mRNA decay in cancer and development. *Cell Rep.* **7**, 539–551 (2014).
53. Chatterjee, S., Fasler, M., Büssing, I. & Grosshans, H. Target-mediated protection of endogenous microRNAs in C. elegans. *Dev. Cell.* **20**, 388–396 (2011).
54. Pitchiaya, S., Heinicke, L. A., Park, J. I., Cameron, E. L. & Walter, N. G. Resolving subcellular miRNA trafficking and turnover at single-molecule resolution. *Cell Rep.* **19**, 630–642 (2017).

55. Boyerinas, B. et al. Identification of let-7-regulated oncofetal genes. *Cancer Res.* **68**, 2587–2591 (2008).

56. Rybak, A. et al. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell. Biol.* **10**, 987–993 (2008).

57. Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin28. *Science* **320**, 97–100 (2008).

58. Nair, V. S., Maeda, L. S. & Ioannidis, J. P. Clinical outcome prediction by microRNAs in human cancer: a systematic review. *J. Natl. Cancer Inst.* **104**, 528–540 (2012).

59. Kugel, S. et al. SIRT6 suppresses pancreatic cancer through control of Lin28b. *Cell* **165**, 1401–1415 (2016).

60. Chien, C. S. et al. Lin28B/Let-7 regulates expression of Oct4 and Sox2 and reprograms oral squamous cell carcinoma cells to a stem-like state. *Cancer Res.* **75**, 2553–2565 (2015).

61. Akao, Y., Nakagawa, Y. & Naoe, T. Let-7 microRNA functions as a potential growth suppressor in human colon cancer cells. *Biol. Pharm. Bull.* **29**, 903–906 (2006).

62. Damanakis, A. I. et al. MicroRNAs let7 expression in thyroid cancer: correlation with their deputed targets HMGA2 and SLC5A5. J. *Cancer Res. Clin. Oncol.* **142**, 1213–1220 (2016).

63. Peng, F. et al. H19/let-7/LIN28 reciprocal negative regulatory circuit promotes breast cancer stem cell maintenance. *Cell Death Dis.* **8**, e2569 (2017).

64. Feng, S. et al. Expression and functional role of reprogramming-related long noncoding RNA (lincRNA-ROR) in glioma. *J. Mol. Neurosci.* **56**, 623–630 (2015).

65. Kallen, A. N. et al. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol. Cell* **52**, 101–112 (2013).

66. Sgarra, R. et al. High mobility group A (HMGA) proteins: molecular instigators of breast cancer onset and progression. *Biochim. Biophys. Acta* **1869**, 216–229 (2018).

67. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

68. Shah, A., Qian, Y., Weyn-Vanhentenryck, S. M. & Zhang, C. CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics* **33**, 566–567 (2017).

69. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

71. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinforma*. **14**, 128 (2013).

72. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

73. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).

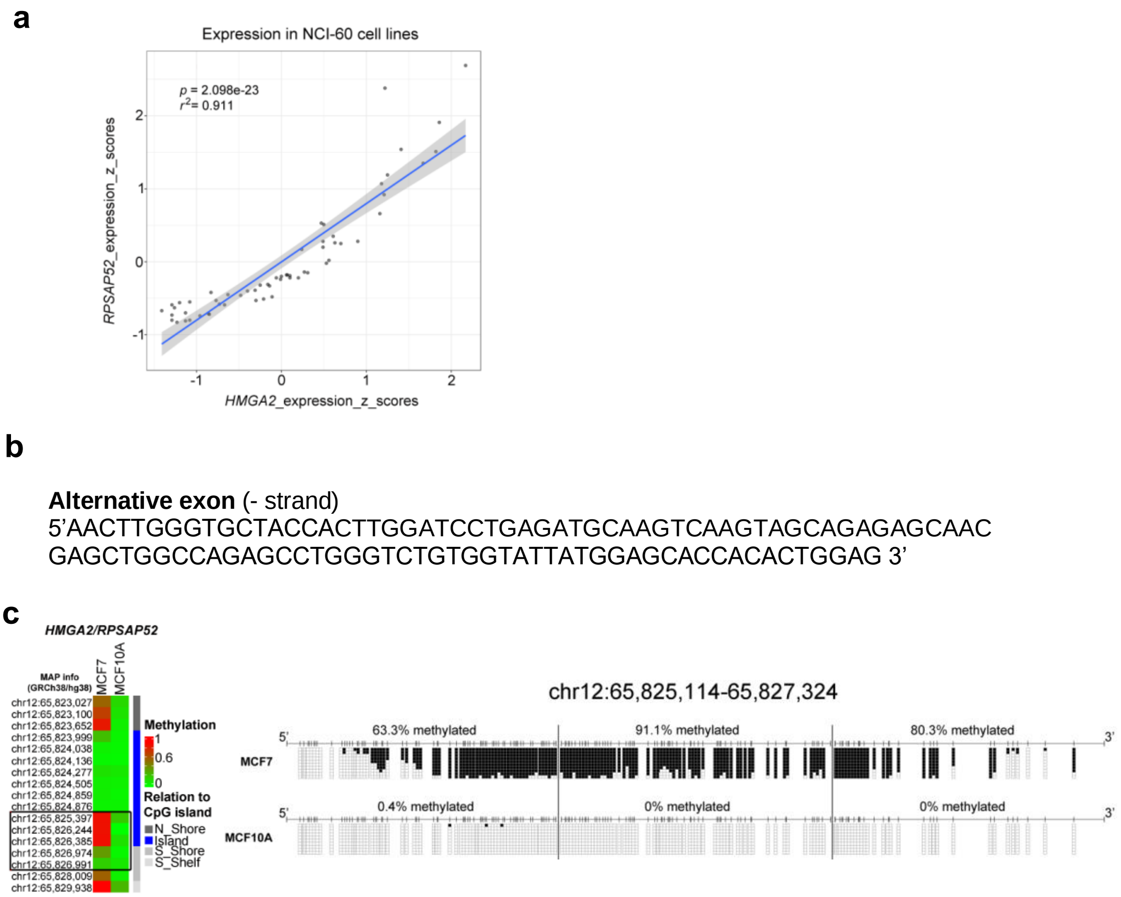
Acknowledgements

We thank CERCA Program/Generalitat de Catalunya for their institutional support. This work was also supported by the Fundació La Marató de TV3, grant number #20131610 (S.G.), the AECC-Junta de Barcelona (S.G.), the Fundación Científica de la AECC undergrant GCB13131578DEÁ (O.M.T.), the Ministerio de Economía y Competitividad (MINECO) and the Instituto de Salud Carlos III (ISCIII), co-financed by the European Development Regional Fund, 'A way to achieve Europe' ERDF, under grant numbers SAF2014-56894-R (S.G.), SAF2014-55000-R (M.E.), CB16/12/00312 (M.E.) and PI16/01898 (A.V.), the Health and Science Departments of the Catalan Government (Generalitat de Catalunya), and the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie, grant agreement number 799850 (A.V., L.F.). C.O.-M. is a pre-doctoral fellow funded by the Basque Government (PRE_2013_1_1009). A.O.-G. is a pre-doctoral fellow funded by MINECO (BES-2015-071452). M.E. is an ICREA Research Professor.

Author contributions

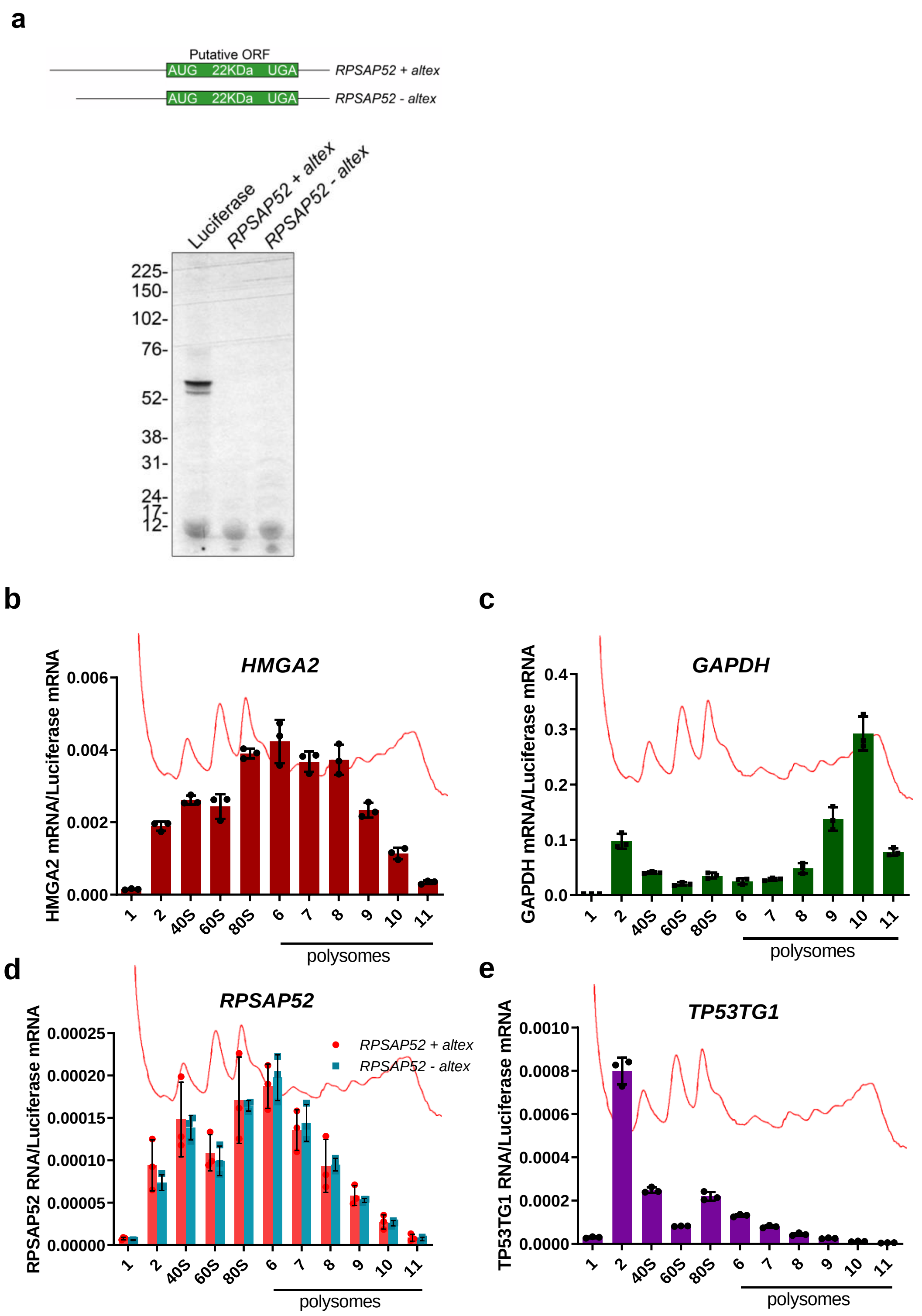
Conception and design: S.G., C.O.-M. and L.F. Development of methodology: C.O.-M., A.S.-C., A.O.-G., R.B.-S., M.S., T.R., J.P., A.G., M.M.-I., D.H.-M., L.F. and S.G. Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc): O.M.T., A.V., and M.E. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D.P., M.E.C.-C., M.C.de M., A.M.-C. and A.G. Writing, review and/or revision of the paper: All authors. Administrative, technical or material support (i.e., reporting or organizing data, constructing databases): D.P., M.E.C.-C. and M.C.deM. Study supervision: S.G.

Supplementary Information

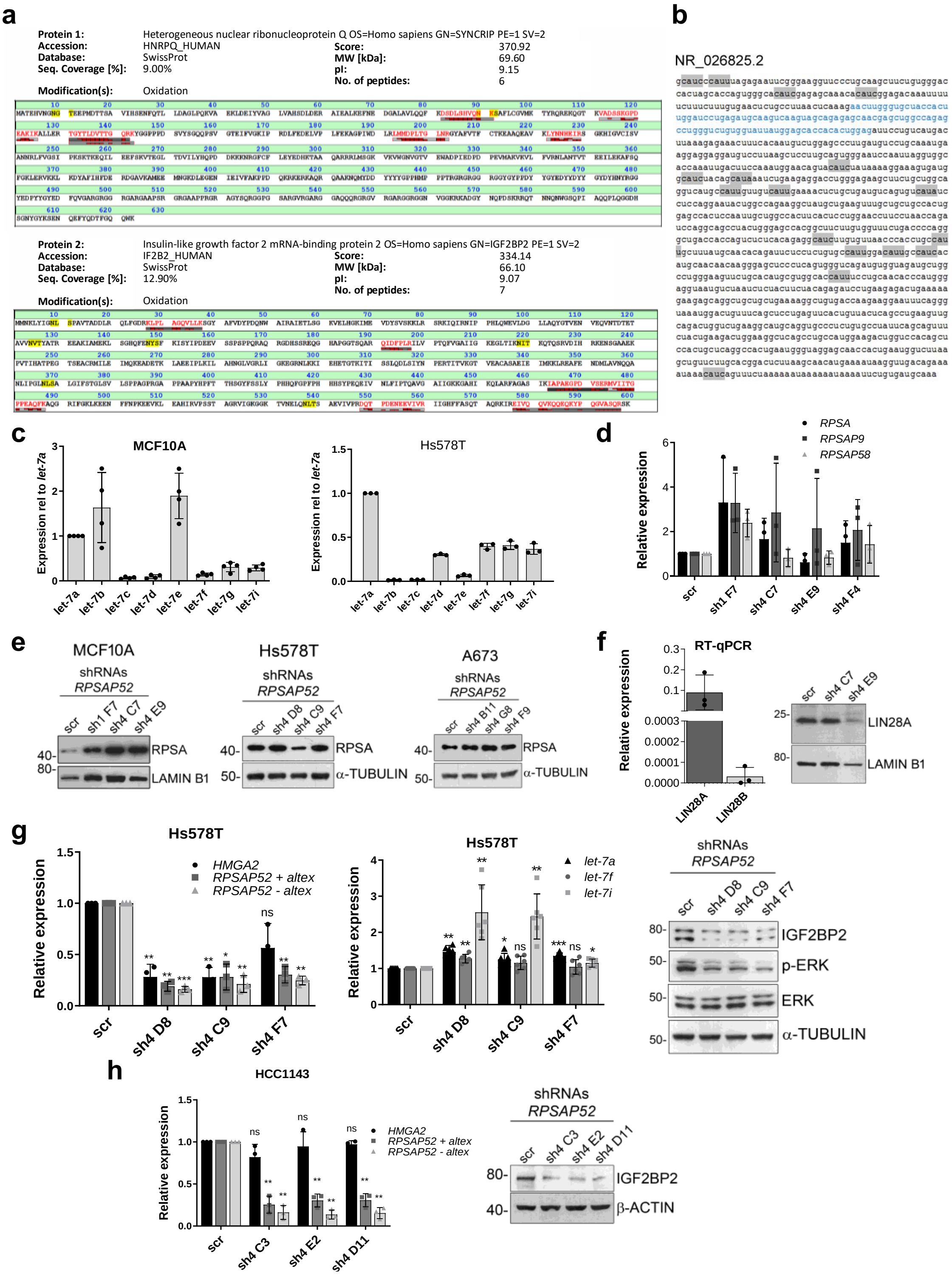


Alternative exon (- strand)
5’AACTTGGGTGCTACCACTTGGATCCTGAGATGCAAGTCAAGTAGCAGAGAGCAAC
GAGCTGGCCAGAGCCTGGGTCTGTGGTATTATGGAGCACCACACTGGAG 3’

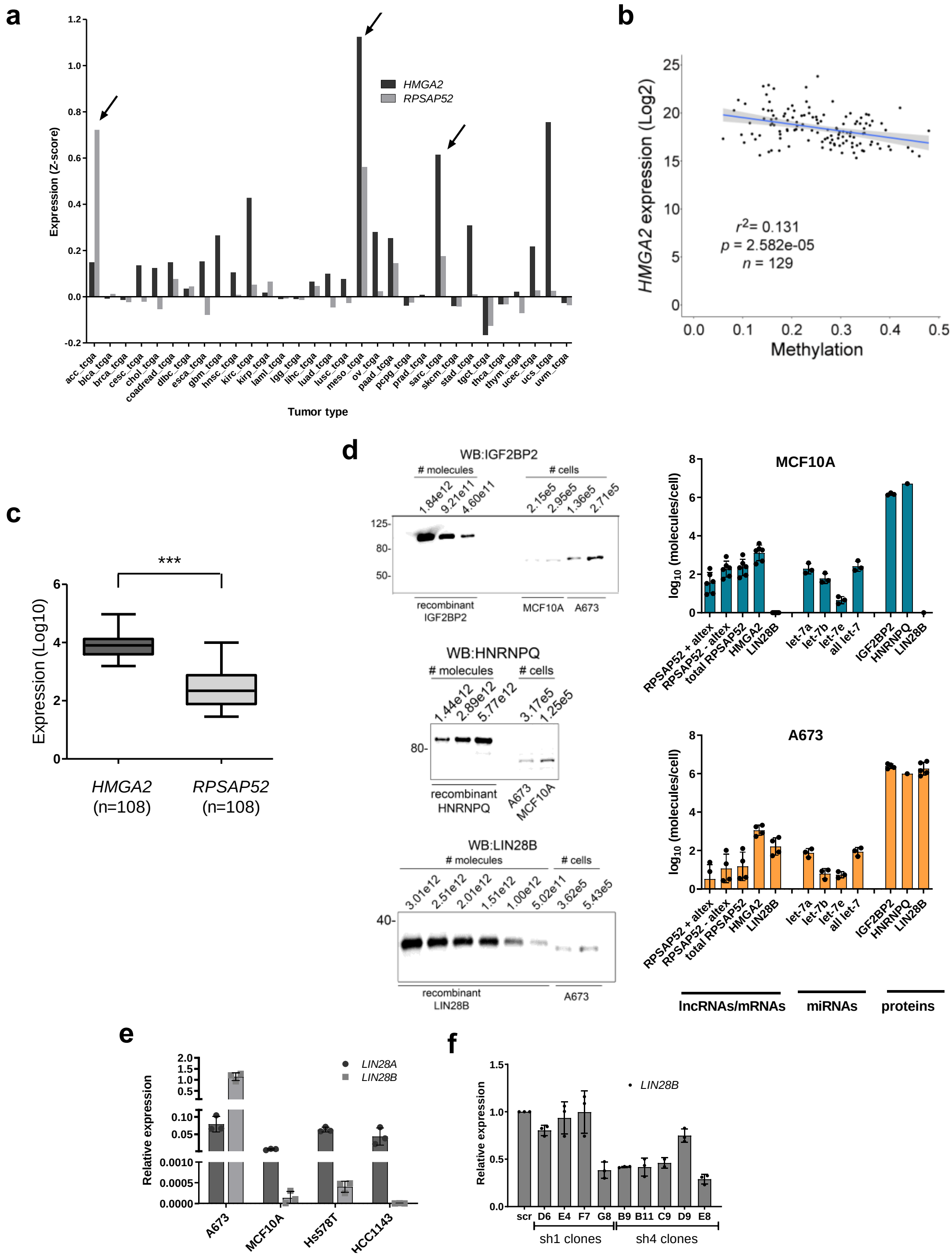
Supplementary Fig. 1. Expression and methylation of the *HMGA2/RPSAP52 locus*. a Correlation between *HMGA2* and *RPSAP52* in the NCI60 panel of cancer cell lines. Normalized expression array data (Z-score) are represented. R squared of Pearson correlation coefficient is shown. **b** The 104 nucleotides of *RPSAP52* alternative exon are indicated. **c** *Left*, heatmap representing methylation levels in MCF7 and MCF10A cell lines. The black square indicates the DNA region that was subject to sequencing following bisulfite treatment (*right*). 3 overlapping DNA fragments were analyzed, so that every CpG is interrogated between coordinates chr12:65,825,114 and chr12:65,827,324 (hg38). Vertical lines represent CpG positions along the whole sequence. Individual clones sequenced are represented horizontally, with empty squares corresponding to unmethylated CpGs and filled squares corresponding to methylated positions. Average methylation levels for each fragment are indicated. Source data for **c** are in Oliveira-Mateos et al_Source Data 1.



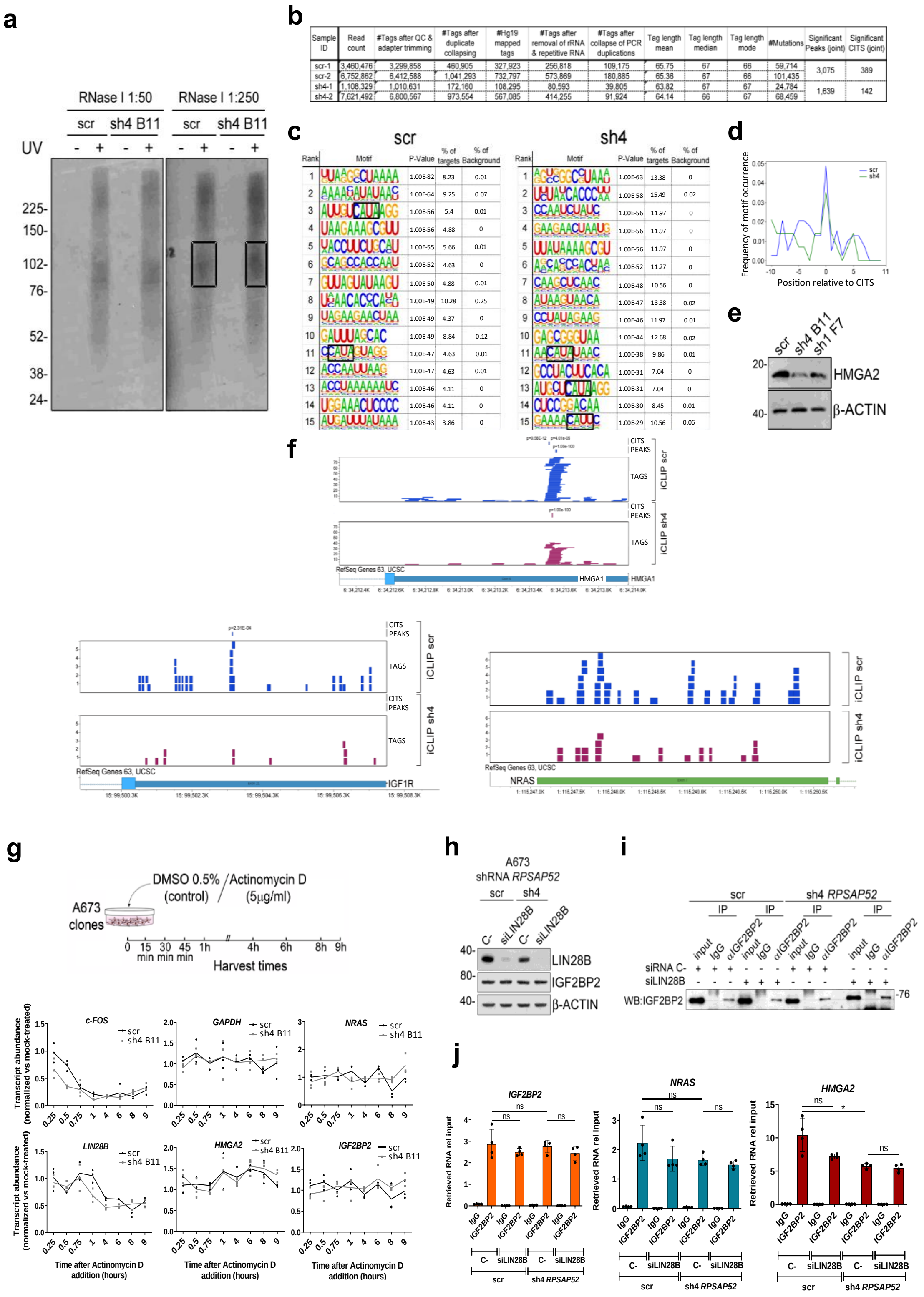
Supplementary Fig. 2. *RPSAP52* transcripts are associated with polysomes in MCF10A cells. **a** Transcription/Translation assay to test the coding potential of *RPSAP52*. A DNA fragment corresponding to the Luciferase open reading frame was used as positive control. The transcripts corresponding to *RPSAP52 + altex* and *RPSAP52 – altex* isoforms were assayed (both have a predicted encoded protein of 22 kDa, as indicated in the upper diagram). The molecular weight for Luciferase is 61 kDa. **b-e** Polysome distribution on a 10%-50% sucrose gradient from MCF10A wild-type cells. The presence of the indicated transcripts in each fraction was analyzed by RT-qPCR. *GAPDH* and *TP53TG1* were used for comparison since they represent an actively translated and a non-protein coding transcript, respectively. Data are means ±SD, and error bars represent 3 replicates of RT-qPCR from each fraction. The red line indicates absorbance at 260 nm for each fraction. Source data for **b-e** are in Oliveira-Mateos et al_Source Data 1. Unprocessed scans are available in Oliveira-Mateos et al_Source Data 2.



Supplementary Fig. 3. *RPSAP52* loss impacts on IGF2BP2/*let-7* pathway in breast cancer cell lines. a Mass spectrometry (MS) identification of the band isolated in the *RPSAP52* RNA pull-down. The specific band was cut out from the gel and trypsin digested for MS analysis, which detected 6 peptides for HNRNPQ protein and 7 peptides for IGF2BP2 protein. **b** *RPSAP52* pseudogene contains several CAU(H) motifs (shaded in grey). The sequence of the alternative exon is highlighted in blue. **c** RT-qPCR analysis of the expression of *let-7* family members in MCF10A cells (left) and Hs578T (right). Abundance is expressed relative to *let-7a*. Data are means of at least three independent RNA extractions ±SD. **d** Expression of the *RPSA* mRNA and the pseudogenes *RPSAP9* and *RPSAP58* upon knockdown of *RPSAP52*. Total RNA from MCF10A clones stably expressing shRNAs against *RPSAP52* was analyzed by RT-qPCR. Data are means of three independent RNA extractions ±SD. **e** Western Blot to analyze RPSA levels upon *RPSAP52* depletion in MCF10A, Hs578T and A673 cell lines. **f** *LIN28A/B* expression in MCF10A cells. *Left*, mRNA relative expression as measured by RT-qPCR from 3 different RNA extractions. Data are means ±SD; *right*, Western Blot of LIN28A protein. LIN28B protein is undetectable in MCF10A cells. **g** RNA and protein analysis of Hs578T clones stably expressing shRNA4 against *RPSAP52*. *Left*, RT-qPCR analysis of *HMGA2* and *RPSAP52* expression. Three different total RNA extractions were analyzed, and two-tailed student *t*-tests were used (**P*<0.05, ***P*<0.01, ****P*<0.001, ns=not significant). Data are means ±SD. *Middle*, RT-qPCR to assess *let-7* miRNAs levels. Six RT-qPCR analysis were performed from three different total RNA extractions, and two-tailed student *t*-tests were used (**P*<0.05, ***P*<0.01, ****P*<0.001, ns=not significant). Data are means ±SD. *Right* (*image*), Western Blot to analyze IGF2BP2, total ERK and phosphorylated ERK (p-ERK) protein levels in the same clones. **h** RNA and protein analysis of HCC1143 clones stably expressing shRNA4 against *RPSAP52*. *Left*, RT-qPCR analysis of *HMGA2* and *RPSAP52* expression. Three different total RNA extractions were analyzed, and two-tailed student *t*-tests were used (**P*<0.05, ***P*<0.01, ns=not significant). Data are means ±SD. *Right* (*image*), Western Blot to analyze IGF2BP2 protein levels. Source data for **c**, **d**, and **f-h** are in Oliveira-Mateos et al_Source Data 1. Unprocessed scans are available in Oliveira-Mateos et al_Source Data 2.



Supplementary Fig. 4. Evaluation of relative and absolute abundance of the RNAs and proteins involved in the regulatory network. a Z-score values for the expression of *HMGA2* and *RPSAP52* transcripts in all tumor types available at the TCGA database. Tumor types with the highest *RPSAP52* expression are indicated by an arrow. **b** *HMGA2* expression in the TCGA sarcoma cohort displays a weak negative correlation with the methylation of its associated CpG island, as measured by Pearson's coefficient. **c** Box plots of *HMGA2* and *RPSAP52* relative expression levels in the TCGA cohort of sarcomas. Only patients with a recorded (non-zero) value for *RPSAP52* expression were considered. (****P*<0.001, two-tailed Mann-Whitney U test). The central mark of the box plot indicates the median, and the bottom and top edges of the box indicate the interquartile range (IQR). The box plot whiskers represent the minimum and maximum of all of the data. **d** Absolute estimations of number of transcripts and proteins per cell. Recombinant proteins or *in vitro* transcribed templates of known amounts were used for comparison. *Left images*, increasing amounts of each recombinant protein were loaded onto SDS/PAGE gels together with total extracts from an exact number of MCF10A or A673 cells, as indicated, and blotted with the corresponding antibodies. Densitometry analysis of the Western Blot signal was used for absolute estimation. *Right graphs*, the number of molecules per cell (log10) in each cell line is shown for RNAs (*RPSAP52*, *HMGA2* and *LIN28B*), *let-7* miRNAs and proteins (IGF2BP2, HNRNPQ and LIN28B). Data are means ±SD, and error bars represent results from at least 2 different experiments (1 for HNRNPQ). **e** *LIN28A/B* mRNAs expression in the cell lines indicated. Relative expression was measured by RT-qPCR taking as a reference *LIN28B* levels in A673 cells. Data are means ±SD, and error bars represent 3 replicates of RT-qPCR from different RNA extractions. **f** RT-qPCR analysis of *LIN28B* mRNA levels upon *RPSAP52* knockdown in A673 cells. Data are means ±SD, and error bars represent 3 replicates of RT-qPCR. Source data for **d-f** are in Oliveira-Mateos et al_Source Data 1. Unprocessed scans are available in Oliveira-Mateos et al_Source Data 2.



Sample ID	Read count	#Tags after QC & adapter trimming	#Tags after duplicate collapsing	#Hg19 mapped tags	#Tags after removal of rRNA & repetitive RNA	#Tags after collapse of PCR duplications	Tag length mean	Tag length median	Tag length mode	#Mutations	Significant Peaks (joint)	Significant CITS (joint)
scr-1	3,460,476	3,299,858	460,905	327,923	256,818	109,175	65.75	67	66	59,714	3,075	389
scr-2	6,752,862	6,412,588	1,041,293	732,797	573,869	180,885	65.36	67	66	101,435		
sh4-1	1,108,329	1,010,631	172,160	108,295	80,593	39,805	63.82	67	67	24,784	1,639	142
sh4-2	7,621,492	6,800,567	973,554	567,085	414,255	91,924	64.14	66	67	68,459		

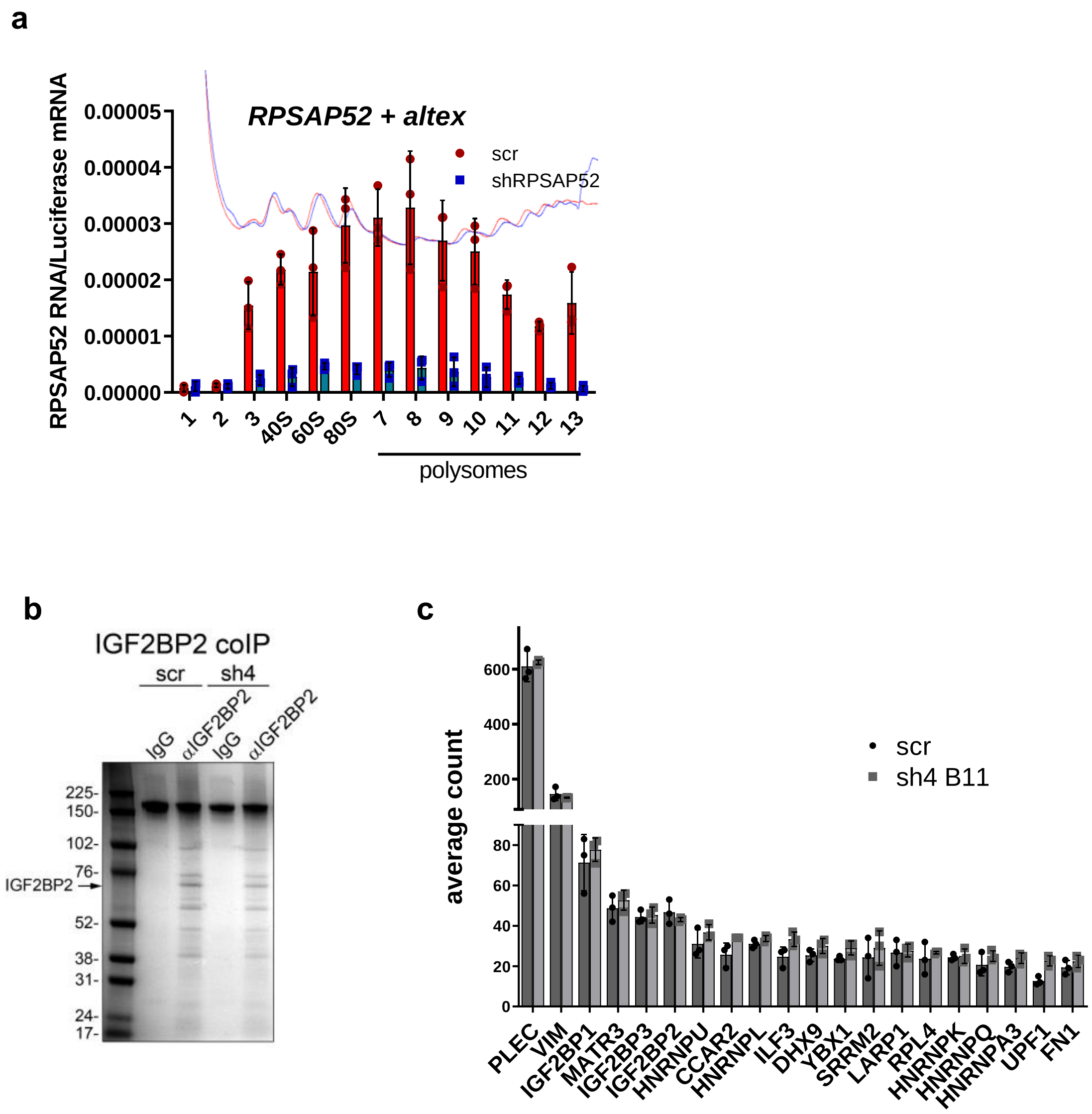
scr

Rank	Motif	P-Value	% of targets	% of Background
1	UUAAGGCUAAAA	1.00E-82	8.23	0.01
2	AAAAGAUAUAAC	1.00E-64	9.25	0.07
3	AUUGUCAUAAGG	1.00E-56	5.4	0.01
4	UAAGAAAGCGUU	1.00E-56	4.88	0
5	UACCUUCUGCAU	1.00E-55	5.66	0.01
6	GCAGCCACCAAU	1.00E-52	4.63	0
7	GUUAGUAUAAGU	1.00E-50	4.88	0.01
8	UGAACACCCAUA	1.00E-49	10.28	0.25
9	UAGAAGAACUAA	1.00E-49	4.37	0
10	GAUUUAGCAC	1.00E-49	8.84	0.12
11	CCAUAGUAGG	1.00E-47	4.63	0.01
12	ACCAAUUAAG	1.00E-47	4.63	0.01
13	ACCUAAAAAAUC	1.00E-46	4.11	0
14	UGGAAACUCCCC	1.00E-46	4.11	0
15	AUGAUUUAUAAA	1.00E-43	3.86	0

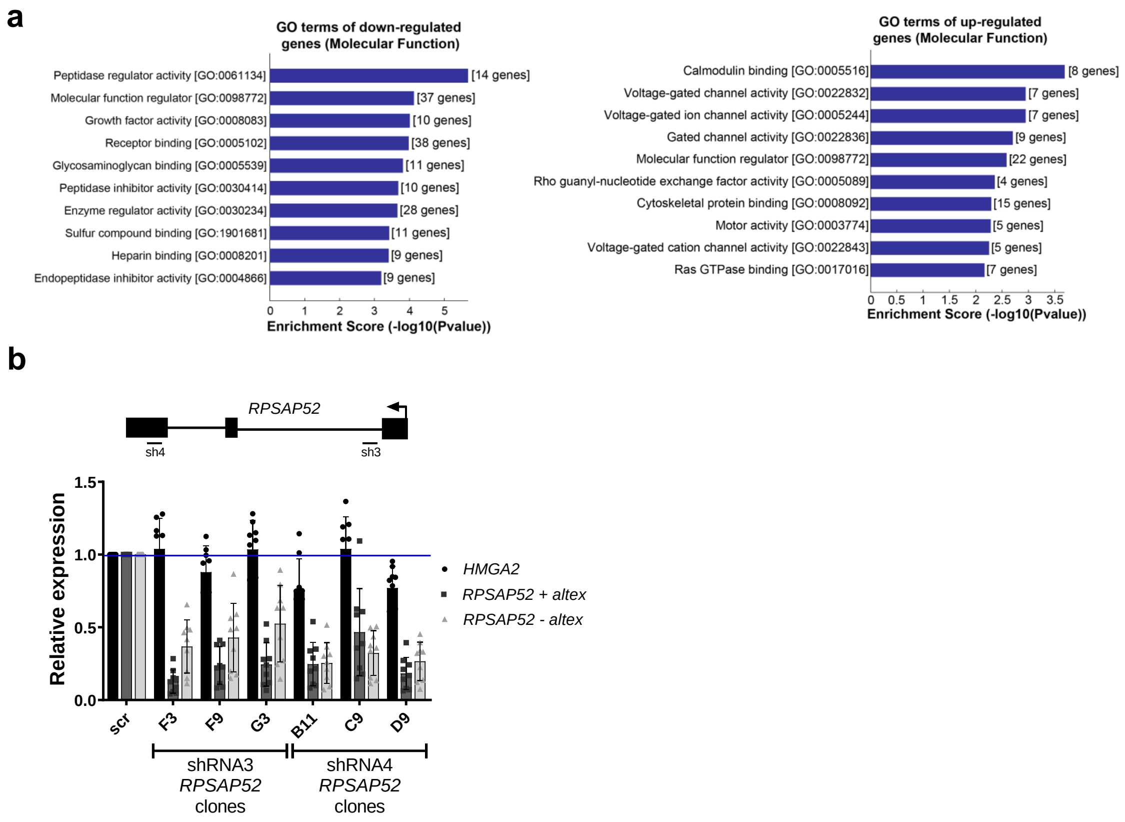
sh4

Rank	Motif	P-Value	% of targets	% of Background
1	AGUCGGCCUAAA	1.00E-63	13.38	0
2	UUUAUCACCCUU	1.00E-58	15.49	0.02
3	CCAAUCUAUC	1.00E-56	11.97	0
4	GAAGAACUAAUG	1.00E-56	11.97	0
5	UUAUAAAAGCGU	1.00E-56	11.97	0
6	AGCAGCCACUAA	1.00E-52	11.27	0
7	CAAGCUCAAC	1.00E-48	10.56	0
8	AUAAGUAACA	1.00E-47	13.38	0.02
9	CCUAUAGAAG	1.00E-46	11.97	0.01
10	GAGCGGGUAA	1.00E-44	12.68	0.02
11	AACAUAUAAC	1.00E-38	9.86	0.01
12	GCCUACUUCACA	1.00E-31	7.04	0
13	AUGCUCAUAAGG	1.00E-31	7.04	0
14	CUCCGGACAA	1.00E-30	8.45	0.01
15	GAAAACAUUC	1.00E-29	10.56	0.06

Supplementary Fig. 5. iCLIP-seq experiment of IGF2BP2 in the context of *RPSAP52* depletion. a Autoradiograph of IGF2BP2 iCLIP experiment in A673 cells expressing control shRNAs (scr) or the sh4 sequence against *RPSAP52* (sh4 B11). Two different concentrations of RNaseI were tested, and for each condition a –UV control was included. The excised RNA-protein bands are marked by the black squares. **b** iCLIP-seq experiment statistics. Results for each of the two experimental replicates for each condition are indicated. **c** Sequence logos of the IGF2BP2 RNA binding motif for each condition, generated by Homer analysis of all significant CITS positions (+/-10nts). The CAUH motif is highlighted by the black squares. **d** Enrichment analysis of the CAUH motif within the iCLIP CITS identified for control (scr) or *RPSAP52*-depleted cells (sh4). **e** Western Blot to assess HMGA2 protein levels in control (scr) or *RPSAP52*-depleted (sh4 and sh1) A673 cells. **f** UCSC Genome Browser view of *HMGA1*, *IGF1R* and *NRAS* 3′UTR with the read coverage from IGF2BP2 iCLIP experiment. Results from control (scr) or *RPSAP52* (sh4) samples are shown. The position of statistically significant CITS and peaks are indicated. **g** RNA stability of IGF2BP2 targets upon *RPSAP52* knockdown. A673 control (scr) or *RPSAP52*-depleted cells (sh4 B11) were treated with 5μg/ml Actinomycin D or DMSO for the times indicated before harvesting, as indicated in the drawing. mRNA levels for each gene at each time-point were then assessed by RT-qPCR (graphs). Data are means ±SD, and error bars represent data from 3 independent Actinomycin D treatments. **h** Western Blot analysis of IGF2BP2 protein levels upon LIN28B depletion. Both control and *RPSAP52*-depleted A673 cells were subject to LIN28B depletion by means of siRNAs, as indicated. **i**, IGF2BP2 immunoprecipitation from the cells in (**h**) followed by Western Blot to assess IGF2BP2 pull-down. **j** RNA from 90% of the pull-down in (**i**) was extracted and analyzed by RT-qPCR. Identity of the genes analyzed are indicated in each graph. Data are means ±SD, and error bars represent the results from 4 replicates of RT-qPCR analysis (**P*<0.05, ns=not significant, two-tailed student *t*-test). Source data for **g** and **j** are in Oliveira-Mateos et al_Source Data 1. Unprocessed scans are available in Oliveira-Mateos et al_Source Data 2.



Supplementary Fig. 6. *RPSAP52* depletion does not change neither IGF2BP2 affinity for protein binding partners nor global translation efficiency. a *RPSAP52 + altex* distribution across a polysome gradient in control (scr) or depleted (shRPSAP52) A673 cells. The presence of RNA in each fraction was analyzed by RT-qPCR. Data are means ±SD, and error bars represent the results from 3 replicates of the RT-qPCR reaction. The red and blue lines indicate absorbance at 260 nm for each fraction in control or depleted cells, respectively. **b** Coomassie staining of a IGF2BP2 coimmunoprecipitation experiment in control and *RPSAP52*-depleted A673 cells (sh4). Mouse IgG was used as a negative control. **c** Average counts from the peptides eluted in 3 IGF2BP2 coimmunoprecitation experiments. The first top 20 proteins with highest counts are represented. No statistical differences between conditions were found among the interactors with highest counts (BFDR=1 in all cases). Data are means ±SD. Source data for **a** and **c** are in Oliveira-Mateos et al_Source Data 1. Unprocessed scans are available in Oliveira-Mateos et al_Source Data 2.



Supplementary Fig. 7. Impact of *RPSAP52* and LIN28B depletion in A673 cells. a Enriched GO terms for shRNA-*RPSAP52*-affected genes (two-tailed Fisher's exact test). The *y* axis shows Molecular Function terms and the *x* axis shows statistical significance. Enrichments for the down-regulated (left) or up-regulated (right) genes are shown. **b** *HMGA2* levels are not quantitatively altered upon *RPSAP52* knockdown with the sh3 and sh4 shRNA sequences. Location of the targeted regions on *RPSAP52* gene is indicated in the upper drawing. RT-qPCR assessment of *HMGA2* mRNA and *RPSAP52* transcripts is shown below. Data are means ±SD from 3 RT-qPCR replicates. Source data for **b** are in Oliveira-Mateos et al_Source Data 1.

Supplementary Table 3. Oligos used in this work

Name	Sequence 5' - 3'	Experiment
HMGA2for	CCTAAGAGACCCAGGGGAAG	RT-qPCR/RT-PCR
HMGA2rev	TCCAGTGGCTTCTGCTTTCT	RT-qPCR/RT-PCR
RPSAP52for	GAGCAAACACATCGGAGACA	RT-qPCR/RT-PCR
RPSAP52rev	AATTGGATTCCCACTGCAAG	RT-PCR
RPSAP+altexrev	CAGCTCGTTGCTCTCTGCTA	RT-qPCR
RPSAP52for2	ACTAGCACCAGTGGGCACAT	RT-qPCR
RPSAP-altexrev	CATGACAGGAATCTTTGAGTTAAG	RT-qPCR
GUSBfor	TGGTTGGAGAGCTCATTTGGA	RT-qPCR
GUSBrev	GCACTCTCGTCGGTGACTGTT	RT-qPCR
GAPDHfor	TGCACCACCAACTGCTTAGC	RT-qPCR/RT-PCR
GAPDHrev	GGCATGGACTGTGGTCATGAG	RT-qPCR/RT-PCR
GAPDHfor2	TGCACCACCAACTGCTTAGC	RT-qPCR
GAPDHrev2	GGCATGGACTGTGGTCATGAG	RT-qPCR
RNU6Bfor	CTCGCTTCGGCAGCACA	RT-qPCR
RNU6Brev	AACGCTTCACGAATTTGCGT	RT-qPCR
c-FOSfor	CCGGGGATAGCCTCTCTTACT	RT-qPCR
c-FOSrev	CCAGGTCCGTGCAGAAGTC	RT-qPCR
IGF2BP2for	AGCCTGTCACCATCCATGC	RT-qPCR/RT-PCR
IGF2BP2rev	CTTCGGCTAGTTTGGTCTCATC	RT-qPCR/RT-PCR
NRASfor	ATGACTGAGTACAAACTGGTGGT	RT-qPCR/RT-PCR
NRASrev	CATGTATTGGTCTCTCATGGCAC	RT-qPCR/RT-PCR
IGF1Rfor	GGAATGAAGTCTGGCTCCG	RT-qPCR
IGF1Rrev	CAGCTGCTGATAGTCGTTGC	RT-qPCR
LIN28Afor	CTTTGTGCACCAGAGTAAGC	RT-qPCR
LIN28Arev	GACCCTTGGCTGACTTCTTA	RT-qPCR
LIN28Bfor	CATCTCCATGATAAACCGAGAGG	RT-qPCR/RT-PCR
LIN28Brev	GTTACCCGTATTGACTCAAGGC	RT-qPCR/RT-PCR
β-ACTINfor	CATCCGCAAAGACCTGTACG	RT-qPCR
β-ACTINrev	CCTGCTTGCTGATCCACATC	RT-qPCR
FLucfor	ACAGATGCACATATCGAGGTG	RT-qPCR
FLucrev	GATTTGTATTCAGCCCATATCG	RT-qPCR
TP53TG1for	CTTTCCTTTAATCTTCGGAGGC	RT-qPCR
TP53TG1rev	TGCCAGCTCTCAGAGTCCTT	RT-qPCR
MGST1for	ATTCATGGCTTTTGCATCC	RT-qPCR
MGST1rev	CTGCGTACACGTTCTACTCTGTC	RT-qPCR
CRABP1for	AAAACCTACTGGACCCGTGA	RT-qPCR
CRABP1rev	GAAAGTAGGAGCAAGCCAGC	RT-qPCR
CYR61for	AACGAGGACTGCAGCAAA	RT-qPCR
CYR61rev	CCCGTTTTGGTAGATTCTGG	RT-qPCR
CD109for	CCAAGATGCTTCAGTGTCC	RT-qPCR
CD109rev	CACAGGAGGACAGCTTCAC	RT-qPCR
PTPRZ1for	TACTGGCCAAATAAAGATGAGC	RT-qPCR
PTPRZ1rev	TGCCTCACTTCAAGTACATAATCA	RT-qPCR
STYK1for	CCTGTGGCTTTTTATCAGAGA	RT-qPCR
STYK1rev	AGGTCCCTAGGTGGAGGA	RT-qPCR
AREGfor	AGCCGACTATGACTACTCAGAAGA	RT-qPCR
AREGrev	CACTTTCCGTCTTGTTTTGG	RT-qPCR
CPT1Cfor	TCAAAGAGTTGCTGCCTGA	RT-qPCR
CPT1Crev	CAGCCGTGGTAGGACAGA	RT-qPCR
NPYfor	CCTCATCACCAGGCAGAG	RT-qPCR
NPYrev	TGGGAACATTTTCTGTGCTT	RT-qPCR
MTSS1for	ACCATCATCAGCGACATGA	RT-qPCR
MTSS1rev	GCCATGTCAGCCACTTTCT	RT-qPCR
TIAM1for	TGGAGGCAAAAGATTGTGTG	RT-qPCR
TIAM1rev	CCTCCTCCTCCCAAGAGACT	RT-qPCR
MICBfor	AAGAAAACATCAGCGGCAG	RT-qPCR
MICBrev	CATCCCTGTGGTCTCCTGT	RT-qPCR
RPSAfor	TCATTTCCTGCCGCCTGT	RT-qPCR
RPSArev	CATCCTCCTCCTTCATTTGC	RT-qPCR
RPSAP9for	ACCCCAATCCATTTTTACCC	RT-qPCR
RPSAP9rev	GGTCTTTTTGTGGCTTGATAGC	RT-qPCR
RPSAP58for	TCTGGAGCGAGAAAAAGAGC	RT-qPCR
RPSAP58rev	GGGTTCATCCACCATCTCAT	RT-qPCR
shRPSAP52_1for	gatccGTCCTTAAGCTCCTTGCAGTTTCAAGAGAACTGCAAGGAGCTTAAGGATTTTTTACGCGTg	Gene silencing
shRPSAP52_1rev	aattcACGCGTAAAAAATCCTTAAGCTCCTTGCAGTTCTCTTGAAACTGCAAGGAGCTTAAGGACg	Gene silencing
shRPSAP52_3for	gatccGTGCAAGACTCAGGAGCTATTCAAGAGATAGCTCCTGAGTCTTGCACTTTTTTACGCGTg	Gene silencing
shRPSAP52_3rev	aattcACGCGTAAAAAAGTGCAAGACTCAGGAGCTATCTCTTGAATAGCTCCTGAGTCTTGCACg	Gene silencing
shRPSAP52_4for	gatccGCACGGACTCTTAAGCAACATTCAAGAGATGTTGCTTAAGAGTCCGTGTTTTTTACGCGTg	Gene silencing
shRPSAP52_4rev	aattcACGCGTAAAAAACACGGACTCTTAAGCAACATCTCTTGAATGTTGCTTAAGAGTCCGTGCg	Gene silencing
bHMGA2for1	GGTAGTTTAAGTAATAGTAG	Methylation analysis
bHMGA2rev1	AAATAAACTAATACCCCCAC	Methylation analysis
bHMGA2for2	GTGGGGGTATTAGTTTATTT	Methylation analysis
bHMGA2rev2	ACCCCCAAAACTCTAACCCC	Methylation analysis
bHMGA2for3	GGGGTTAGAGTTTTGGGGGT	Methylation analysis
bHMGA2rev3	CAAACAAAACCCTCCACTCC	Methylation analysis
bHMGA2for4	GGAGTGGAGGGTTTTGTTTG	Methylation analysis
bHMGA2rev4	AAACTCAAAAACCTCTAAATC	Methylation analysis
bHMGA2for5	TAGAGGTTTTTGAGTTTTTT	Methylation analysis
bHMGA2rev5	ATTAACTTAAAACCCATAAA	Methylation analysis
bHMGA2for6	AATTAGTTTTATTTAATTAT	Methylation analysis
bHMGA2rev6	TAAAAAATTTTACTTAAATC	Methylation analysis
T7-RPSAP52for	GAAATTAATACGACTCACTATAGGGGCATCCCATTTAGAGAAT	*In vitro* biotin-transcription
RPSAP52-flrev	ATCGATCGCTCGAGTTTGCATCACAGAATTTT	*In vitro* biotin-transcription
T7-antiRPSAP52for	GAAATTAATACGACTCACTATAGGGTTTGCATCACAGAATTTT	*In vitro* biotin-transcription
RPSAP52-flfor	ATCGATCGCTCGAGGCATCCCATTTAGAGAAT	*In vitro* biotin-transcription
T7-RPSAP52altexfor	GAAATTAATACGACTCACTATAGGGAACTTGGGTGCTACCACTTGGATCC	*In vitro* biotin-transcription
RPSAP52dom1rev	CTTTAAGTCATGACAGGAATCT	*In vitro* biotin-transcription
T7-RPSAP52dom2for	GAAATTAATACGACTCACTATAGGGAGAAACTTTCACAATGTCTGG	*In vitro* biotin-transcription
RPSAP52dom2rev	GTGATGGCAATGTCCAATGG	*In vitro* biotin-transcription
T7-RPSAP52dom3for	GAAATTAATACGACTCACTATAGGGATGCAACAACAAGGGAGCTCCC	*In vitro* biotin-transcription
RPSAP52dom3rev	TTTGCATCACAGAATTTTATTTTTTA	*In vitro* biotin-transcription
T7-RPSAP52for-TnT	GAAATTAATACGACTCACTATAGGGaaggttccctgcaagcttct	TnT assay
RPSAP52rev-TnT	TTGCTTAAGAGTCCGTGCAA	TnT assay

DISCUSSION

HMGA2 regulation through *RPSAP52*-mediated R-loop formation

A strand-specific RNA-Seq and the Illumina's HumanMethylation450 Bead Chip array allowed the comparison of normal and breast cancer cell lines transcriptome, as well as their methylation profiles. We were interested specifically in S/AS gene pairs with differential expression between normal and tumor condition dependent on changes in methylation. It was also taken into account the presence of a theoretical GC skew. The selected pairs for further study were *VIM/VIM-AS1* and *HMGA2/RPSAP52*, because in addition to the mentioned criteria, there are numerous bibliographical references that show VIM and HMGA2 oncogenic potential and their aberrant expression in a multitude of cancers[246–250,278].

The expression of the *HMGA2* gene is positively correlated with the levels of *RPSAP52* pseudogene, and both are overexpressed in a variety of human cancers, as it was described previously for oral carcinoma[279]. Divergent transcripts often share regulatory elements that control the transcriptional state of both genes at the same time. The presence of a bidirectional promoter could help explain the coordinated expression of this S/AS pair[280]. On the other hand, the well-defined association between DNA methylation and gene expression[281–283] is demonstrated once again in this case, where the hypomethylation of the promoter region leads to the overexpression of both genes. The lack of methylation in certain genomic regions such as oncogene promoters is a typical circumstance in tumors[284,285]. It has been proved that CGI shore methylation has a relevant role in the regulation of gene expression[34], being involved in differentiation and cancer development[34,286], and it seems the critical region that determines the transcriptional state of this gene pair. The numerous examples that show the regulation exerted by NATs on their respective sense genes[132,140,193], along with the fact that *RPSAP52* depletion by shRNAs results in a substantial decrease of *HMGA2* mRNA, suggest the existence of a regulation mediated by the antisense transcript.

The presence of RNA:DNA hybrids has been related with antisense transcription, pointing to a possible implication of NATs in their formation[216,287]. Moreover, the GC skew present in the promoter region of the studied genes enables the formation of an R-loop with the participation of the G-rich mRNA[33], in this case, the antisense transcript *RPSAP52*. We confirmed the existence of an R-loop that is affected by RNase H, as expected given its ability to degrade the RNA moiety in RNA:DNA hybrids. The *in vitro*

R-loop formation assay shows that its production depends on the direction of transcription, from where it can be deduced that the responsible of its formation is *RPSAP52*, and not *HMGA2* mRNA. Different experiments indicated that the *RPSAP52* region involved in the R-loop is an intron not efficiently processed by splicing that remains in the transcription start site. This could be linked with the described implication of the loss of certain splicing factors in R-loops formation[221]. Native bisulfite sequencing, RNA-FISH and DRIP experiments shows that the structure is maintained *in vivo*, at least on the *VIM* promoter, where *VIM-AS1* exerts the same function as *RPSAP52*.

R-loops effects on chromatin conformation and transcription

Local chromatin structures such as R-loops are related to nucleosome positioning and DNA methylation of the nearby region[33,228], and antisense transcription is also associated with both processes[132,288]. Changes at the methylation level were expected upon inhibition of *RPSAP52*; however, no differences in this regard have been found, although the existence of alterations in the unstudied regions of the CGI cannot be ruled out.

Not having observed methylation variations in the absence of *RPSAP52* led us to believe that this NAT might perform its regulatory functions through changes in nucleosome occupancy in the promoter region, an essential issue for gene expression[289]. Micrococcal nuclease assay shows less accessibility in the GC skew area in cells infected with shRNAs against *RPSAP52*, compared to controls. The DNA within nucleosomes is partially protected from enzyme digestion[290], so that this lower accessibility implies a higher nucleosome density. This suggests that the antisense transcript affects their positioning. According to this, the presence of *RPSAP52* is responsible for lower chromatin compaction along the promoter region, with the consequent increase in accessibility that favors the activation of the sense transcript. These chromatin regions with low nucleosome density seems to facilitate the binding of proteins that stabilize and protect R-loops[216,291], as well as the access of transcription factors[292,293]. As a consequence, a decrease in nucleosome occupancy is usually detected where functional binding sites are present. An example is observed in the *VIM* promoter, where the region implicated in the formation of the R-loop contains binding sites for NF-κB. The grade of chromatin compaction determine the accessibility of this transcription factor, and the absence of the antisense transcript that allows the formation of the R-loop reduces its interaction with the DNA. Something similar may be happening at the promoter of the *HMGA2* gene, but additional experiments would be necessary to corroborate it.

A similar mechanism has been described for the *COOLAIR* gene of *Arabidopsis thaliana*[216]. In this example, however, the R-loop is located in the promoter region of the antisense transcript but in the 3′UTR of the sense gene. This implies that mechanistically the regulation can be very different. Indeed, the presence of this structure impairs *COOLAIR* transcription and favors the expression of the sense *FLC* gene, establishing a negative correlation between S/AS expression levels. The same happens in the *Ube3a-ATS locus*, but in this case the transcription of the antisense is favored by the formation of the R-loop in the paternal allele, and the sense *Ube3a* gene results silenced[228]. The R-loop present in the termination region of the gene upstream *Ube3a-ATS* causes chromatin decondensation and allows transcriptional elongation through the *locus* of the antisense transcript. The transcription of *Ube3a-ATS* beyond the *Ube3a* promoter region inhibits the expression of the sense gene, probably by the transcriptional interference mechanism[294].

R-loops prevalence and its implications is diseases

There are numerous characterized R-loops, and many of them are involved in human pathologies. For example, different studies have demonstrated the presence of R-loops in the genes responsible of neurological disorders, autoimmune diseases and cancer[295]. Both a dysregulated increase in the global formation of R-loops and the loss of specific regulatory structures can have a detrimental role for the cell. As a consequence, depending on the context, therapeutic strategies may involve their global targeting (e.g., in cancer) as well as their specific stabilization (e.g., some neurological diseases).

The expression of some proteins that interact with RNA:DNA hybrids correlates with survival in different cancers. These proteins and R-loops itself are potential molecular targets for the development of new therapies against cancer and other diseases[296]. For instance, CX-3543 and CX-5461 are drugs in advanced phases of clinical trials that produce DNA damage through G-quadruplex stabilization[297], a typical structure on the DNA strand displaced from the R-loop. A non-cancer related example is topotecan, which could be used to treat Angelman syndrome through the increase of R-loop formation and the reactivation of *Ube3a* gene expression[228].

Although the link between R-loops and cancer-associated *loci* is well established, this has not been previously related with divergent transcription. Our work connects these three aspects for the first time and shows how the R-loop formation with the participation of an antisense transcript activates the expression of an oncogenic gene. This mechanism is not restricted to *HMGA2/RPSAP52* pair, because our group has also identified it for *VIM* gene. Moreover, the case of *GATA3/GATA3-AS1* pair has been recently described with a similar mechanism[298]. These genes do not overlap but they are encoded in close proximity with divergent transcription. The antisense transcript promotes the formation of an R-loop within the central intron of *GATA3-AS1* gene and thus, it favors the recruitment of methyltransferases that maintain the active status of the *locus*. The existence of the same kind of regulation in several *loci* with similar characteristics indicates that it could be a much more general mechanism than expected, being extensible to other genes with oncogenic potential. Several methods have been recently developed to detect R-loops genome-wide[299–301], but only a study in yeasts recognized the relationship between RNA:DNA hybrids and divergent transcription across the genome[287]. Therefore, it would be interesting to carry out a genome-wide characterization of the presence of R-loops associated with S/AS transcription in human cell lines.

The results obtained in this work open the possibility of new therapeutic strategies based on the implication of antisense transcripts in the formation of R-loops in oncogenic *loci*.

***RPSAP52* impact on IGF2BP2 function**

The abundant presence of *RPSAP52* in the cytoplasm suggests that its functions are not restricted to the activation of *HMGA2* transcription through the formation of an R-loop in the nucleus. Here, we demonstrated that *RPSAP52* interacts with IGF2BP2 protein and, thereby, it exerts regulatory control of important oncogenic pathways.

LIN28B reduced expression in A673 clones with *RPSAP52* depletion could be explained by the decrease in the binding of IGF2BP2 to *LIN28B* mRNA. Previous evidences had shown that *LIN28B* mRNA is a direct target of IGF2BP1[275], but with our data, it arises as a new target of IGF2BP2 not described so far, whose efficient translation depends on the interaction between *RPSAP52* and IGF2BP2. Thanks to the immunoprecipitation of this protein and the iCLIP-Seq experiment, we observed changes in the affinity of IGF2BP2 for certain 3′UTRs, among them *LIN28B* mRNA. The same mechanism affects the translation of *HMGA2*, a known IGF2BP2 target[261], while others such as *RAS*, *IGF1R* or the own mRNA of *IGF2BP2* do not suffer affinity changes. As a consequence, the expression of these genes do not experiment differences at the protein level in the clones with respect to the control cells.

Surprisingly, IGF2BP2 protein levels are not altered in spite of the variations promoted by *RPSAP52* depletion in the expression of its transcriptional activator HMGA2 and its negative regulator *let-7*. Since the absence of *RPSAP52* does not affect the binding between IGF2BP2 and its own mRNA, protection against *let-7* degradation is maintained. Another possibility is that these changes are not enough to affect the high expression levels of IGF2BP2 in A673 cell line. However, in MCF10A, where basal levels are lower, the expression decreases under *RPSAP52* silencing.

Several lncRNAs, such as *Airn*[150] and *HIF1A-AS2*[161], have shown similar functions as partners of this RBP, promoting the expression of some IGF2BP2 targets. Another example of this type of regulation has been recently defined for the lncRNA *THOR*, but in this case the binding protein is IGF2BP1, another member of the family[302]. The specific effect only over certain mRNA targets could be related with different aspects. *RPSAP52* might not affect the bindings with high affinity that involve several KH or RRM domains of IGF2BP2 or more than one IGF2BP forming dimers, but it could be essential to the stabilization of weaker bindings. Moreover, *RPSAP52* might promote the encounter between molecules expressed with different proportions or might favor proper structural conformations for the interaction. Another relevant factor could be the implication of additional molecules that ensure the association with polysomes without *RPSAP52* presence. Conversely, interactions with some lncRNAs impair IGF2BP2 function because they compete for the binding with target mRNAs. For instance, *lncMyoD* promotes muscle differentiation regulating negatively the translation of *c-Myc* and *N-Ras*[273]; and *linc-ADAL* modulates adipogenesis related genes expression through the same mechanism[303]. These lncRNAs might compete with the target mRNAs for the same binding domains, while *RPSAP52* binds in a compatible manner with the interaction of other RNAs. This could explain the differences between the positive regulation exerted by *RPSAP52* and the interference of the function mediated by *lncMyoD* and *linc-ADAL*. Moreover, since *RPSAP52* does not appear in the iCLIP-seq, it cannot be ruled out the possibility that it does not bind IGF2BP2 directly. RNA immunoprecipitation experiments might be identifying a secondary interaction mediated by an unknown protein that connects the two molecules.

Consequences of *RPSAP52* cytoplasmic functions on *let-7* levels

HMGA2 mRNA has several functional binding sites for *let-7* in its 3′UTR[102], so that it might control this miRNA availability acting as a ceRNA. However, in our models, the regulation takes place at the expression level, and through an *HMGA2* independent mechanism because differences in *let-7* family are observed even without changes in the mRNA of the coding gene. Moreover, the specific depletion of *HMGA2* with LNA gapmers decreases *let-7* levels, concomitant with an increase in *RPSAP52* that occurs through an unknown mechanism, indicating that the main regulator of *let-7* in the contexts we have studied is not HMGA2 but *RPSAP52*.

Many pseudogenes also modulate miRNA-mediated repression by acting as ceRNAs or masking the miRNA binding sites of the parental mRNA[173,175], but the effect that *RPSAP52* exerts over *let-7* is determined by changes in the miRNA expression. Since LIN28B is one of the main negative regulators of *let-7* maturation, and its translation depends on the control of the interaction between *LIN28B* and IGF2BP2 by *RPSAP52*, this could be the mechanism by which *RPSAP52* regulates *let-7* levels in A673 cell line.

On the contrary, alterations of *let-7* levels in MCF10A cannot be justified by LIN28B changes, since no expression is detected in this line and the presence of the related protein LIN28A is residual. A possible explanation could be related with the ability of IGF2BP proteins to protect some mRNAs from miRNA degradation[275,276]. The most likely mechanism in this case seems the masking of miRNA response elements present in mRNA targets, as previously described for IGF2BP3[304] and for IGF2BP2 itself[261]. Under *RPSAP52* silencing, the decrease of IGF2BP2 protein levels impairs the translation of *IGF1R* and *RAS* mRNAs and hinders their protection from miRNA action, with the consequent protein reduction. The higher availability of *let-7* mRNA targets might explain the increase in *let-7* levels without the intervention of LIN28B, since miRNA stability is influenced by the binding to the targets[305,306]. Thus, *RPSAP52* established an indirect negative regulation over *let-7* expression, in which the presence of the lncRNA results in the degradation of *let-7* through the regulation of IGF2BP2 function.

The magnitude of the change in *let-7* levels might be determined by the different mechanisms that control its expression in each cell line. The modulation through LIN28B regulation in A673 cells is, most likely, stronger than the control of the turnover through target availability in MCF10A. Thus, the increase of *let-7* expression upon *RPSAP52* depletion is much more pronounced in A673 cell line.

Regulation of IGF2BP2 distribution on polysomes by *RPSAP52*

The known functions of IGF2BP proteins that affect the expression levels of their targets are the control of mRNA translation[262] and stability[254]. Half-lives of IGF2BP2 targets are not significantly altered in this study, indicating that their stability is not disturbed. An additional fact that points to an effect at the level of translation is that *RPSAP52* does not follow the typical distribution across a polysome gradient of a ncRNA. Given that the retention of coding potential is typical in many pseudogenes[185,186], this could indicate that the lncRNA *RPSAP52* is, indeed, translated. Nevertheless, according to our *in vitro* transcription/translation assays, this possibility seems unlikely.

Taking into account both results, the most likely explanation is that *RPSAP52* plays a role in the translational regulation of other mRNAs. The correct association of mRNAs with polysomes has been linked previously with lncRNAs functions, such as *lincRNA-p21*[146] and the antisense of *Uchl1*[157]. Our results show that the localization of IGF2BP2 on heavy polysomes is determined by its interaction with *RPSAP52*. Thus, the association of this complex with mRNA targets not only protects them from *let-7*-mediated repression, but it favors the appropriate recruitment into polysomes and, thereby, its translation.

LIN28B effect on IGF2BP2 expression and function

Overexpression of LIN28B rescues the decreased levels of IGF2BP2 protein in MCF10A clones depleted of *RPSAP52*. This is due to the ability of the exogenous protein to downregulate *let-7*. Although IGF2BP2 cannot properly drive the translation of its own mRNA in the absence of *RPSAP52*, LIN28B overexpression liberates *IGF2BP2* mRNA from *let-7* repression. Thus, *IGF2BP2* mRNA could be accumulated and increase its translation which would help to explain the recovery of basal levels. Indeed, alternative functions have been described for LIN28 proteins that relate them with the translation increase of some mRNAs through the control of their association with polysomes[307,308].

There are LIN28 binding sites in thousands of mRNAs[309,310], and some confirmed targets are their own mRNA, *HMGA1*, *HMGA2* and *IGF2BPs*[310–312], which could also explain the rescue of IGF2BP2 levels upon LIN28B overexpression.

On the other hand, LIN28B depletion in A673 clones does not have any influence on IGF2BP2 levels and function. The discrepancy between the impact that LIN28B overexpression has over IGF2BP2 in MCF10A and the absence of effects of its depletion in A673 could be related with the fact that *RPSAP52* silencing does not affect the binding between IGF2BP2 and its own mRNA. Hence, even in circumstances that promote the upregulation of *let-7*, IGF2BP2 maintains the protection of its mRNA from miRNA degradation. However, it cannot be ruled out the existence of additional levels of regulation for IGF2BP2 that may differ between the two systems.

RPSAP52 as a master regulator with oncogenic properties

The functional effects of *RPSAP52* observed in breast cancer and Ewing's sarcoma cell lines both *in vitro* and *in vivo*, reveals the biological importance of this lncRNA in this two cancer types. Moreover, the results from an expression array showed that *RPSAP52* is able to promote the oncogenic properties of the cells. This NAT decreases the levels of some tumor suppressors, such as *MTSS1*[313], and favors the upregulation of genes related to proliferative pathways and metastasis, *CD109*[314] and *CRABP1*[315] among others.

Its relevance promoting stemness and self-renewal characteristics of the cells is evident considering that it reduces the levels of the important tumor suppressor *let-7* and positively regulates the expression and/or function of the oncogenic proteins HMGA2, IGF2BP2 and LIN28B. In that respect, some lncRNAs and NATs have been described in relation to the maintenance of cells pluripotency. One strategy to protect the undifferentiated state consists in repressing lineage-specific genes. For instance, *ANCR* blocks the expression of essential genes for epidermis differentiation in humans[316], and *HOTAIR* suppresses the transcription of developmental genes[317]. Conversely, it seems that *RPSAP52* acts in another way preserving the expression of the stemness factors OCT4 and NANOG. It is known that LIN28B favors *OCT4* and *SOX2* expression through the regulation of transcription factors targeted by *let-7*. Indeed, HMGA2 is one of these factors and acts as a positive regulator of *SOX2* transcription[318]. The regulation of the axis LIN28B/HMGA2/*let-7* could be one of the mechanisms used by *RPSAP52* to maintain pluripotency. On the other hand, *linc-RoR* has been described as a sponge for miRNAs that degrade the stemness genes *OCT4*, *NANOG* and *SOX2* in embryonic stem cells[319]. The sequence of *RPSAP52* contains potential binding sites for *miR-299*[320] and *miR-150*[321], miRNAs that target *OCT4* and *NANOG*, respectively. Thus, another possibility would be that *RPSAP52* regulates the pluripotency factors by acting as a ceRNA for these miRNAs, but additional experiments are needed to confirm this.

Implications of *RPSAP52* regulatory pathway in other systems

The analysis of *HMGA2* and *RPSAP52* expression in TCGA samples, and the high levels detected in some of them, indicates that *RPSAP52*-mediated regulation could be relevant in other types of human tumors. Embryonic rhabdomyosarcoma would be a case of special relevance that warrants further exploration due to the critical role played by the HMGA2-IGF2BP2-NRAS axis in its development and in the myogenesis process[254,262,273]. Moreover, Lin28 has also been linked to myogenic differentiation. Although this protein does not seem involved in the origin of rhabdomyosarcomas, it is worth noting its effect in the association of mRNAs such as *MyoD* to polysomes in mouse myoblasts[308].

In that respect, the expression profile of *RPSAP52* and *HMGA2* was analyzed during muscle differentiation as part of a different project related to this thesis. We hypothesize that *RPSAP52* could be a barrier against differentiation in human myoblasts because it decreases greatly and rapidly during this process, together with *HMGA2*. Thus, its oncogenic nature could be essential to allow cells to remain in a proliferative state. It is worth highlighting the relevance that this may have not only at the level of normal differentiation processes throughout development, but also regarding to pathological circumstances such as rhabdomyosarcomas. The cells that make up this type of tumor have characteristics of immature skeletal muscle, which means that they are similar to undifferentiated myoblasts. Contrary to expectations, preliminary results from the comparison of proliferating myoblasts and under differentiation conditions show a decrease in differentiation markers in the absence of *RPSAP52*. Different experiments are currently being carried out in the laboratory to analyze possible changes in IGF2BP2 binding to its mRNA targets that help explain these results. Therefore, the precise role of *RPSAP52* participation in the mentioned process still remains enigmatic.

RPSAP52 as a potential therapeutic target

The important functions performed by NATs, as well as their relationship with the development of diseases such as cancer, make them a potential source of new therapeutic strategies. In this particular case, the overexpression of *HMGA2* oncogene could be decreased silencing the corresponding NAT, due to the positive correlation of their expression levels. Likewise, IGF2BP2 function could be impaired in the absence of *RPSAP52*, avoiding the activation of proliferative pathways. Thus, NATs-related treatments could be an alternative to traditional chemotherapy, since by limiting their regulatory roles to a restricted number of genes, they have reduced side effects and lower toxicity. In addition, oncofetal lncRNAs such as *RPSAP52* that are highly expressed in tumors and absent in normal tissues, offer the opportunity to attack cancer cells without damaging healthy tissues. These expression differences act, at the same time, as a biomarker for sarcoma prognosis, unlike *HMGA2* levels, which do not show correlation with survival neither in the studied TCGA cohort nor in previous works[322]. Interestingly, methylation of the promoter of the *locus* itself act as a marker due to the correlation established with *RPSAP52* expression. Since the results have been obtained from a cohort of patients with different types of sarcomas, this suggests an important role of *RPSAP52* not only in Ewing's sarcoma, but also in other types of soft tissue cancers.

The lack of conservation of lncRNAs in mouse also affects *RPSAP52*, which impedes the study of its silencing in animal models. Nevertheless, *Hmga2* and *Igf2bp2* knockout mice could help us understand the *in vivo* consequences of *RPSAP52*-mediated regulation. Both models show a marked effect on the insulin metabolism and on mice growth. Disruption of *Hmga2* derives in smaller mice with a decrease of the adipose content[323]. The appearance of growth disorders in humans as a result of *HMGA2* deletions or specific SNPs reinforce the observed phenotype in mouse[324–326]. *Igf2bp2* knockout mice show the same features and higher insulin sensitivity[327]. Moreover, their implication in the insulin pathway is supported by the correlation between a SNP of *IGF2BP2* and the risk of diabetes[267], as well as by the presence of the disease in humans and mice in the absence of Hmga1, a protein that may have some overlapping functions with HMGA2[328]. Another characteristics of these mice are the defect in self-renewal capacities of the cells[329] and the longer lifespan due to the absence of tumors comparing with wild type mice for *Igf2bp2*, showing the consequences of both genes in tumorigenesis[327].

These *in vivo* models could be useful to understand *RPSAP52* functions in humans because the effect in the insulin signaling pathway is in agreement with the typical alterations observed in rhabdomyosarcomas. In addition to RAS cascade, it represents one of the main altered molecular routes in this cancer. Mutations in tyrosine kinase receptors are commonly observed, being the insulin receptor of this type. Moreover, IGF1R and its ligand IGF2 are overexpressed in rhabdomyosarcomas[330,331]. Some compounds are in late stages of clinical trials such as an antibody against IGF1R (R1507)[332] and the kinase inhibitor sorafenib[333], but the most interesting option seems a therapeutic strategy that combined the blockade of this and RAS pathway.

In conclusion, NATs represent a set of biomolecules of great importance due to their potential therapeutic and diagnostic use that opens up novel possibilities for the design of strategies against different diseases. This thesis contributes to the comprehension of the regulatory mechanisms mediated by antisense transcripts through the study of *HMGA2/RPSAP52 locus*. The NAT *RPSAP52* activates the transcription of *HMGA2* through the formation of an R-loop in the promoter region, with the corresponding impact in the expression of its target *IGF2BP2*. Beyond the effects over the sense gene, it regulates IGF2BP2 function, modulating its binding to specific mRNAs such as LIN28B and its distribution on polysomes. Moreover, *RPSAP52* has an indirect impact on the levels of the tumor suppressor *let-7*. According to our results, we define *RPSAP52* as an oncofetal pseudogene with properties of a master regulator and with important implications in a number of human cancers.

CONCLUSIONS

Based on the findings of this PhD thesis, we can conclude that:

The new knowledge obtained about the regulatory mechanisms mediated by NATs reveals the impact that these have on the expression of sense genes related to cancer and on the tumor progression itself. Given the high number of transcripts in whose regulation NATs could be involved, their understanding is extremely important to know the activation potential of a large number of genes. The full comprehension of the underlying regulatory mechanisms is still far from being a reality, but this thesis proposes a new model of regulation mediated by the antisense transcription of the lncRNA *RPSAP52*.

STUDY I:

1. There are pairs of S/AS genes that are expressed in a coordinated manner. The expression of the *HMGA2* sense gene is positively correlated with the levels of the *RPSAP52* pseudogene, and it is regulated by this antisense transcript, since its depletion leads to a clear reduction of both RNAs.
2. The presence of a GC skew in the surroundings of *HMGA2* and *RPSAP52* transcription start site allows the formation of an R-loop structure at least *in vitro* with the participation of the *RPSAP52* antisense transcript.
3. Antisense transcription modifies chromatin compaction. *RPSAP52* presence promotes open chromatin conformation through changes in nucleosome occupancy. This more accessible chromatin seems to facilitate the binding of transcriptional factors and favors *HMGA2* sense gene expression.
4. This positive regulation of gene expression, which involves the formation of an R-loop through antisense transcription, could be broadly extended in the genome as a general mechanism of transcriptional activation.

STUDY II:

1. *HMGA2/RPSAP52 locus* is overexpressed in many types of human cancers and these high expression levels correlate with aberrant hypomethylation of the CGI associated with their promoter region.

2. *RPSAP52* interacts with the RNA binding protein IGF2BP2 and facilitates its binding to a subset of mRNA targets, promoting their translation. The binding affinity of IGF2BP2 for some 3′UTRs and its recruitment on polysomes depends on this interaction. This is especially relevant for *HMGA2* and *LIN28B* mRNAs.

3. *RPSAP52* reduces the expression of some members of *let-7* family, the major tumor suppressor miRNA. This could be consequence of a direct regulation through the control of *LIN28B* translation, the main regulator of *let-7* maturation, or indirect, influencing miRNAs turnover through the availability of *let-7* targets such as *IGF2BP2* and *HMGA2*.

4. *RPSAP52* antisense transcript is involved in the upregulation of important oncogenic pathways, which results in an increase of migration, proliferation and self-renewal capacities of breast cancer and sarcoma cells. Moreover, it promotes tumorigenic progression *in vivo*, which makes it an important master regulator with oncogenic properties.

5. High *RPSAP52* expression and hypomethylation of its promoter are associated with poor prognosis in sarcomas, and this correlates better with survival than *HMGA2* expression. According to this, *RPSAP52* can have significant clinical implications in cancer, because it can be useful as a biomarker and it could be a new therapeutic target.

REFERENCES

1. WHO. Cancer, Fact sheet N°297, World Health Organization. 2018.
2. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57–70.
3. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144(5):646–74.
4. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends–An update. Cancer Epidemiol Biomarkers Prev. 2016;25(1):16–27.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7–30.
6. Stiller CA, Trama A, Serraino D, Rossi S, Navarro C, Chirlaque MD, et al. Descriptive epidemiology of sarcomas in Europe: Report from the RARECARE project. Eur J Cancer. 2013;49(3):684–95.
7. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer. cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Research Data, Nov 2010 Sub (1973–2008) <Katrina/Rita Population Adjustment> − Linked To County Attributes - Total U.S., 1969–200.
8. Burningham Z, Hashibe M, Spector L, Schiffman JD. The epidemiology of sarcoma. Clin Sarcoma Res. 2012;2(1):14.
9. Sgarra R, Pegoraro S, Ros G, Penzo C, Chiefari E, Foti D, et al. High Mobility Group A (HMGA) proteins: Molecular instigators of breast cancer onset and progression. Biochim Biophys Acta - Rev Cancer. 2018;1869(2):216–29.
10. Bizer LS. Rhabdomyosarcoma. Am J Surg. 1980;140:687–91.
11. Davicioni E, Anderson MJ, Finckenstein FG, Lynch JC, Qualman SJ, Shimada H, et al. Molecular classification of rhabdomyosarcoma - Genotypic and phenotypic determinants of diagnosis: A report from the Children’s Oncology Group. Am J Pathol. 2009;174(2):550–64.
12. Martinelli S, McDowell HP, Vigne SD, Kokai G, Uccini S, Tartaglia M, et al. RAS signaling dysregulation in human embryonal rhabdomyosarcoma. Genes Chromosomes Cancer. 2009;48:975–82.
13. Xia SJ, Pressey JG, Barr FG. Molecular pathogenesis of rhabdomyosarcoma. Cancer Biol Ther. 2002;1(2):97–104.
14. May WA, Gishizky ML, Lessnick SL, Lunsford LB, Lewis BC, Delattre O, et al. Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. Proc Natl Acad Sci USA. 1993;90(12):5752–6.
15. May WA, Lessnick SL, Braun BS, Klemsz M, Lewis BC, Lunsford LB, et al. The Ewing’s sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. Mol Cell Biol. 1993;13(12):7393–8.
16. Waddington CH. The epigenotype. Endeavour. 1942;1:18–20.
17. Riggs AD, Porter TN. Overview of epigenetic mechanims. In: Riggs AD, Martienssen RA, Russo VEA (eds). Epigenetic mechanisms of gene regulation. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 1996. 32:29-45.
18. Simpson RT, Bustin M. Histone composition of chromatin subunits studied by immunosedimentation. Biochemistry. 1976;15(19):4305–12.
19. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389:251–60.
20. van Holde KE. Chromatin. Springer-Verlag, New York. 1989.
21. Kamakaka RT, Biggins S. Histone variants: Deviants? Genes Dev. 2005;19(3):295–310.

22. Santos-Rosa H, Kirmizis A, Nelson C, Bartke T, Saksouk N, Cote J, et al. Histone H3 tail clipping regulates gene expression. Nat Struct Mol Biol. 2009;16(1):17–22.
23. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693–705.
24. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA. 1964;51(1938):786–94.
25. Grant PA. A tale of histone modifications. Genome Biol. 2001;2(4):1–6.
26. Galm O, Herman JG, Baylin SB. The fundamental role of epigenetics in hematopoietic malignancies. Blood Rev. 2006;20(1):1–13.
27. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28(10):1057–68.
28. Woodcock DM, Lawler CB, Linsenmeyer ME, Doherty JP, Warren WD. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. J Biol Chem. 1997;272(12):7810–6.
29. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–22.
30. Achwal CW, Iyer CA, Chandra HS. Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, Drosophila and mealybug DNA. FEBS Lett. 1983;158(2):353–8.
31. Kulis M, Esteller M. DNA Methylation and Cancer. In: Herceg Z, Ushijima T (eds). Advances in Genetics. Epigenetics and Cancer, Part A. San Diego, CA: Academic Press. 2010. 70:27-56.
32. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol. 1987;196:261–82.
33. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell. 2012;45(6):814–25.
34. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009;41(2):178–86.
35. Pradhan S, Bacolla A, Wells RD, Roberts RJ. Recombinant human DNA (cytosine-5) methyltransferase. J Biol Chem. 1999;274(46):33002–10.
36. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99:247–57.
37. Kriaucionis S, Heintz N. The nuclear DNA base, 5-hydroxymethylcytosine is present in brain and Purkinje neurons. Science. 2009;324(5929):929–30.
38. Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. 2007;67(3):946–50.
39. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2010;930(2009).
40. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300–3.
41. He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-

carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303–7.

42. Pfaffeneder T, Hackner B, Truß M, Münzel M, Müller M, Deiml CA, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chemie - Int Ed. 2011;50(31):7008–12.
43. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell. 2013;152(5):1146–59.
44. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011;473(7347):394–397.
45. Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem. 2014;6(12):1049–55.
46. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975;14:9–25.
47. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187:226–232.
48. Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol. 1998;18(11):6538–47.
49. Illingworth RS, Bird AP. CpG islands - "A rough guide." FEBS Lett. 2009;583(11):1713–20.
50. Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, et al. Developmental programming of CpG island methylation profiles in the human genome. Nat Struct Mol Biol. 2009;16(5):564–71.
51. Ball MP, Li JB, Gao Y, Lee J-H, LeProust E, Park I, et al. Targeted and genome-scale methylomics reveals gene body signatures in human cell lines. Nat Biotechnol. 2009;27(4):361–8.
52. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, et al. Intragenic DNA methylation prevents spurious transcription initiation. Nature. 2017;543(7643):72–7.
53. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 1997;13(8):335–40.
54. Reik W, Lewis A. Co-evolution of X-chromosome inactivation and imprinting in mammals. Nat Rev Genet. 2005;6(5):403–10.
55. Kacem S, Feil R. Chromatin mechanisms in genomic imprinting. Mamm Genome. 2009;20(9–10):544–56.
56. Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, et al. MODOMICS: A database of RNA modification pathways. 2017 update. Nucleic Acids Res. 2018;46(D1):303–7.
57. Crick F. On protein synthesis. Symp Soc Exp Biol. 1958;12:138–63.
58. Ohno S. So much 'junk' DNA in our genome. Brookhaven Symp Biol. 1972;23:366–370.
59. Cech TR, Zaug AJ, Grabowsk PJ. In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell. 1981;27(3):487–96.
60. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309:1559–63.
61. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi AM, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101–8.

62. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. Nat Rev Genet. 2010;11(8):559–71.
63. Mattick JS. Non-coding RNAs: The architects of eukaryotic complexity. EMBO Rep. 2001;2(11):986–91.
64. Szymanski M, Erdmann VA, Barciszewski J. Noncoding RNAs database (ncRNAdb). Nucleic Acids Res. 2007;35:162–4.
65. Ward M, McEwan C, Mills JD, Janitz M. Conservation and tissue-specific transcription patterns of long noncoding RNAs. J Hum Transcr. 2015;1(1):2–9.
66. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775–89.
67. Wahlestedt C. Targeting long non-coding RNA to therapeutically upregulate gene expression. Nat Rev Drug Discov. 2013;12(6):433–46.
68. Losko M, Kotlinowski J, Jura J. Long noncoding RNAs in metabolic syndrome related disorders. Mediators Inflamm. 2016;2016(5365209):1–12.
69. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75:843–54.
70. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403:901–6.
71. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science. 2001;294:853–8.
72. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science. 2001;294:858–62.
73. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Science. 2001;294:862–4.
74. Griffiths-Jones S, Grocock RJ, Dongen S van, Bateman A, Enright AJ. miRBase: MicroRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34:140–4.
75. Lee Y, Jeon K, Lee J-T, Kim S, Kim VN. MicroRNA maturation: Stepwise processing and subcellular localization. EMBO JournalEMBO J. 2002;21(17):4663–70.
76. Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23(20):4051–60.
77. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 2006;13(12):1097–101.
78. Gregory RI, Yan K-P, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. Nature. 2004;432(7014):235–40.
79. Bohnsack MT, Czaplinski K, Görlich D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA. 2004;10(2):185–91.
80. Hutvágner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 2001;293(2001):834–9.
81. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509–24.
82. Bartel DP. MicroRNA target recognition and regulatory functions. Cell. 2009;136(2):215–33.

83. Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, et al. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell. 2014;56(1):104–15.
84. Peng Y, Croce CM. The role of microRNAs in human cancer. Signal Transduct Target Ther. 2016;1(15004):1–9.
85. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861–74.
86. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043):828–33.
87. Yan L-X, Huang X-F, Shao Q, Huang M-Y, Deng L, Wu Q-L, et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. RNA. 2008;14(11):2348–60.
88. Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci USA. 2007;104(40):15805–10.
89. He L, He X, Lim LP, Stanchina E de, Xuan Z, Liang Y, et al. A microRNA component of the p53 tumour suppressor network. Nature. 2007;447(7148):1130–4.
90. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature. 2000;408:86–9.
91. Roush S, Slack FJ. The let-7 family of microRNAs. Trends Cell Biol. 2008;18(10):505–16.
92. Lewis BP, Shih I-H, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell. 2003;115:787–98.
93. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin-28. Science. 2008;320(5872):97–100.
94. Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. Lin28 mediates the terminal uridylation of let-7 precursor microRNA. Mol Cell. 2008;32(2):276–84.
95. Piskounova E, Polytarchou C, Thornton JE, LaPierre RJ, Pothoulakis C, Hagan JP, et al. Lin28A and Lin28B inhibit let-7 MicroRNA biogenesis by distinct mechanisms. Cell. 2011;147(5):1066–79.
96. Barh D, Malhotra R, Ravi B, Sindhurani P. MicroRNA let-7: An emerging next-generation cancer therapeutic. Drug Dev Contemp Oncol. 2010;17(1):70–80.
97. Manier S, Powers JT, Sacco A, Glavey SV, Huynh D, Reagan MR, et al. The LIN28B/let-7 axis is a novel therapeutic pathway in multiple myeloma. Leukemia. 2017;31(4):853–60.
98. Kumar MS, Erkeland SJ, Pester RE, Chen CY, Ebert MS, Sharp PA, et al. Suppression of non-small cell lung tumor development by the let-7 microRNA family. Proc Natl Acad Sci. 2008;105(10):3903–8.
99. Shi Y, Duan Z, Zhang X, Zhang X, Wang G, Li F. Down-regulation of the let-7i facilitates gastric cancer invasion and metastasis by targeting COL1A1. Protein Cell. 2019;10(2):143–8.
100. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. Cell. 2005;120(5):635–47.
101. Mayr C, Hemann MT, Bartel DP. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. Science. 2007;315:1576–80.
102. Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2

oncogene. Genes Dev. 2007;21:1025–30.

103. Boyerinas B, Park S-M, Shomron N, Hedegaard MM, Vinther J, Andersen JS, et al. Identification of let-7-regulated oncofetal genes. Cancer Res. 2008;68(8):2587–91.

104. JnBaptiste CK, Gurtan AM, Thai KK, Lu V, Bhutkar A, Su M-J, et al. Dicer loss and recovery induce an oncogenic switch driven by transcriptional activation of the oncofetal Imp1-3 family. Genes Dev. 2017;31(7):674–87.

105. Alajez NM, Shi W, Wong D, Lenarduzzi M, Waldron J, Weinreb I, et al. Lin28b promotes head and neck cancer progression via modulation of the Insulin-Like Growth Factor survival pathway. Oncotarget. 2012;3(12):1641–52.

106. van Rooij E, Kauppinen S. Development of microRNA therapeutics is coming of age. EMBO Mol Med. 2014;6(7):851–64.

107. Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nat Rev Drug Discov. 2013;12(11):847–65.

108. Beyer S, Fleming J, Meng W, Singh R, Haque SJ, Chakravarti A. The role of miRNAs in angiogenesis, invasion and metabolism and their therapeutic implications in gliomas. Cancers. 2017;9(85):1–21.

109. Shan G, Li Y, Zhang J, Li W, Szulwach KE, Duan R, et al. A small molecule enhances RNA interference and promotes microRNA processing. Nat Biotechnol. 2008;26(8):933–40.

110. Balzeau J, Menezes MR, Cao S, Hagan JP. The LIN28/let-7 pathway in cancer. Front Genet. 2017;8(31):1–16.

111. Wojciechowska A, Braniewska A, Kozar-Kamińska K. MicroRNA in cardiovascular biology and disease. Adv Clin Exp Med. 2017;26(5):865–74.

112. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013;10(6):924–33.

113. St.Laurent G, Wahlestedt C, Kapranov P. The landscape of long noncoding RNA classification. Trends Genet. 2015;31(5):239–51.

114. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for the ENCODE project. Genome Res. 2012;22(9):1760–74.

115. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015;47(3):199–208.

116. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223–7.

117. Clark BS, Blackshaw S. Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. Front Genet. 2014;5(164):1–19.

118. Zhang Y, Yang L, Chen LL. Life without A tail: New formats of long noncoding RNAs. Int J Biochem Cell Biol. 2014;54:338–49.

119. Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, et al. Genome-wide analysis of long noncoding RNA stability. Genome Res. 2012;22:885–98.

120. Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long noncoding RNAs; sequence, structure, function. Biochim Biophys Acta. 2014;1840(3):1063–71.

121. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. Mol Cell Biol. 1990;10(1):28–36.

122. Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, et al.

The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell. 1992;71(3):527–42.

123. Gomes AQ, Nolasco S, Soares H. Non-coding RNAs: Multi-tasking molecules in the cell. Int J Mol Sci. 2013;14(8):16010–39.

124. Guil S, Esteller M. Cis-acting noncoding RNAs: Friends and foes. Nat Struct Mol Biol. 2012;19(11):1068–75.

125. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. Science. 2005;309(5740):1564–6.

126. Lapidot M, Pilpel Y. Genome-wide natural antisense transcription: Coupling its regulation to its different regulatory mechanisms. EMBO Rep. 2006;7(12):1216–22.

127. van Duin M, van Den Tol J, Hoeijmakers JH, Bootsma D, Rupp IP, Reynolds P, et al. Conserved pattern of antisense overlapping transcription in the homologous human ERCC-1 and yeast RAD10 DNA repair gene regions. Mol Cell Biol. 1989;9(4):1794–8.

128. Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, et al. Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res. 2004;32(16):4812–20.

129. Balbin OA, Malik R, Dhanasekaran SM, Prensner JR, Cao X, Wu YM, et al. The landscape of antisense gene expression in human cancers. Genome Res. 2015;25(7):1068–79.

130. Faghihi M, Wahlestedt C. Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol. 2009;10(9):637–43.

131. Khorkova O, Myers AJ, Hsiao J, Wahlestedt C. Natural antisense transcripts. Hum Mol Genet. 2014;23(R1):R54–63.

132. Tufarelli C, Stanley JAS, Garrick D, Sharpe JA, Ayyub H, Wood WG, et al. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet. 2003;34(2):157–65.

133. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong J, et al. Long noncoding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–6.

134. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by non-coding RNAs. Cell. 2007;129(7):1311–23.

135. Kim K, Jutooru I, Chadalapaka G, Johnson G, Frank J, Burghardt R, et al. HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. Oncogene. 2013;32(13):1616–25.

136. Yang Z, Zhou L, Wu L-M, Lai M-C, Xie H-Y, Zhang F, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol. 2011;18(5):1243–50.

137. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 2011;71(20):6320–6.

138. Liu Z, Chen Z, Fan R, Jiang B, Chen X, Chen Q, et al. Over-expressed long noncoding RNA HOXA11-AS promotes cell cycle progression and metastasis in gastric cancer. Mol Cancer. 2017;16(1):1–9.

139. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation. Nature. 2011;472(7341):120–4.

140. Yu W, Gius D, Onyango P, Muldoon-jacobs K, Karp J, Feinberg AP, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature. 2008;451(7175):202–6.

141. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, et al. Molecular interplay of the non-coding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell. 2010;38(5):662–74.

142. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-broz D, et al. A large intergenic non-coding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010;142(3):409–19.

143. Morris KV, Santoso S, Turner AM, Pastori C, Hawkins PG. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet. 2008;4(11):1–9.

144. Bao X, Wu H, Zhu X, Guo X, Hutchins AP, Luo Z, et al. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. Cell Res. 2015;25:80–92.

145. Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, et al. LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. Mol Cell. 2014;54(5):777–90.

146. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, et al. LincRNA-p21 suppresses target mRNA translation. Mol Cell. 2012;47(4):648–55.

147. Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, et al. Role of histone H3 lysine 27 methylation in X inactivation. Science. 2003;300:131–5.

148. Ohhata T, Hoki Y, Sasaki H, Sado T. Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. Development. 2008;135(2):227–35.

149. Latos PA, Pauler FM, Koerner M V., Şenergin HB, Hudso QJ, Stocsits RR, et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science. 2012;336(14):1469–73.

150. Hosen MR, Militello G, Weirick T, Ponomareva Y, Dassanayaka S, IV JBM, et al. Airn regulates Igf2bp2 translation in cardiomyocytes. Circ Res. 2018;122(10):1347–1353.

151. Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, et al. Antisense intergenic transcription in V(D)J recombination. Nat Immunol. 2004;5(6):630–7.

152. Peters NT, Rohrbach JA, Zalewski BA, Byrkett CM, Vaughn JC. RNA editing and regulation of Drosophila 4f-rnp expression by sas-10 antisense readthrough mRNA transcripts. RNA. 2003;9:698–710.

153. Katsutomo Okamura, Balla S, Martin R, Liu N, Lai EC. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol. 2008;15(6):581–90.

154. Hastings ML, Milcarek C, Martincic K, Peterson ML, Munroe SH. Expression of the thyroid hormone receptor gene, erbAα, in B lymphocytes: Alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. Nucleic Acids Res. 1997;25(21):4296–300.

155. Beltran M, Puig I, Peña C, García JM, Álvarez AB, Peña R, et al. A natural

antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev. 2008;22(6):756–69.

156. Ebralidze AK, Guibal FC, Steidl U, Zhang P, Lee S, Bartholdy B, et al. PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. Genes Dev. 2008;22(15):2085–92.

157. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature. 2012;491(7424):454–7.

158. Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol. 2010;11(R56):1–13.

159. Liu L, Chen X, Zhang Y, Hu Y, Shen X, Zhu W. Long non-coding RNA TUG1 promotes endometrial cancer development via inhibiting miR-299 and miR-34a-5p. Oncotarget. 2017;8(19):31386–94.

160. Uchida T, Rossignol F, Matthay MA, Mounier R, Couette S, Clottes E, et al. Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1α and HIF-2α expression in lung epithelial cells. J Biol Chem. 2004;279(15):14871–8.

161. Mineo M, Ricklefs F, Rooj AK, Lyons SM, Ivanov P, Ansari KI, et al. The long non-coding RNA HIF1A-AS2 facilitates the maintenance of mesenchymal glioblastoma stem-like cells in hypoxic niches. Cell Rep. 2016;15(11):2500–9.

162. Mahmoudi S, Henriksson S, Corcoran M, Méndez-Vidal C, Wiman KG, Farnebo M. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol Cell. 2009;33(4):462–71.

163. Rosikiewicz W, Makałowska I. Biological functions of natural antisense transcripts. Acta Biochim Pol. 2016;63(4):665–73.

164. Giangrossi M, Prosseda G, Tran CN, Brandi A, Colonna B, Falconi M. A novel antisense RNA regulates at transcriptional level the virulence gene icsA of Shigella flexneri. Nucleic Acids Res. 2010;38(10):3362–75.

165. Prescott EM, Proudfoot NJ. Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci U S A. 2002;99(13):8796–801.

166. Grandér D, Johnsson P. Pseudogene-expressed RNAs: Emerging roles in gene regulation and disease. In: Morris, KV (ed). Current topics in microbiology and immunology. Long non-coding RNAs in human disease. Springer; 2016. 394:111-126.

167. Chan JJ, Tay Y. Noncoding RNA:RNA regulatory networks in cancer. Int J Mol Sci. 2018;19(5):1–26.

168. Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, et al. The GENCODE pseudogene resource. Genome Biol. 2012;13(R51):1–26.

169. Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu Y-M, Cao X, et al. Expressed pseudogenes in the transcriptional landscape of human cancers. Cell. 2012;149(7):1622–34.

170. Cooke SL, Shlien A, Marshall J, Pipinikas CP, Martincorena I, Tubio JMC, et al. Processed pseudogenes acquired somatically during cancer development. Nat Commun. 2014;5(3644):1–9.

171. Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X. Pseudogene in cancer: Real functions and promising signature. J Med Genet. 2015;52(1):17–24.

172. Johnsson P, Ackley A, Vidarsdottir L, Lui W-O, Corcoran M, Grandér D, et al. A pseudogene long noncoding RNA network regulates PTEN transcription and translation in human cells. Nat Struct Mol Biol. 2013;20(4):440–6.

173. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-

independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465(7301):1033–8.

174. Suo G, Han J, Wang X, Zhang J, Zhao Y, Zhao Y, et al. Oct4 pseudogenes are transcribed in cancers. Biochem Biophys Res Commun. 2005;337:1047–51.

175. Wang L, Guo ZY, Zhang R, Xin B, Chen R, Zhao J, et al. Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma. Carcinogenesis. 2013;34(8):1773–81.

176. Scarola M, Comisso E, Pascolo R, Chiaradia R, Maria Marion R, Schneider C, et al. Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. Nat Commun. 2015;6(7631):1–13.

177. Hawkins PG, Morris KV. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. Transcription. 2010;1(3):165–75.

178. De Martino M, Forzati F, Marfella M, Pellecchia S, Arra C, Terracciano L, et al. HMGA1P7-pseudogene regulates H19 and Igf2 expression by a competitive endogenous RNA mechanism. Sci Rep. 2016;6(37622):1–10.

179. De Martino M, Palma G, Azzariti A, Arra C, Fusco A, Esposito F. The HMGA1 pseudogene 7 induces miR-483 and miR-675 upregulation by activating Egr1 through a ceRNA mechanism. Genes. 2017;8(330):1–10.

180. Chiefari E, Iiritano S, Paonessa F, Le Pera I, Arcidiacono B, Filocamo M, et al. Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. Nat Commun. 2010;1(40):1–7.

181. Esposito F, de Martino M, Petti MG, Forzati F, Tornincasa M, Federico A, et al. HMGA1 pseudogenes as candidate proto-oncogenic competitive endogenous RNAs. Oncotarget. 2014;5(18):8341–54.

182. Wang TH, Lin YS, Chen Y, Yeh CT, Huang Y lin, Hsieh TH, et al. Long non-coding RNA AOC4P suppresses hepatocellular carcinoma metastasis by enhancing vimentin degradation and inhibiting epithelial-mesenchymal transition. Oncotarget. 2015;6(27):23342–57.

183. Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and antiinflammatory therapeutics. Elife. 2013;2(e00762):1–16.

184. Chan WL, Yuo CY, Yang WK, Hung SY, Chang YS, Chiu CC, et al. Transcribed pseudogene ψpPM1K generates endogenous siRNA to suppress oncogenic cell growth in hepatocellular carcinoma. Nucleic Acids Res. 2013;41(6):3734–47.

185. Gawlik-Rzemieniewska N, Bednarek I. The role of NANOG transcriptional factor in the development of malignant phenotype of cancer cells. Cancer Biol Ther. 2016;17(1):1–10.

186. Zou M, Baitei EY, Alzahrani AS, Al-Mohanna F, Farid NR, Meyer B, et al. Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. Neoplasia. 2009;11(1):57–65.

187. Karreth FA, Reschke M, Ruocco A, Ng C, Chapuy B, Léopold V, et al. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. Cell. 2015;161(2):319–32.

188. Pardini B, Sabo AA, Birolo G, Calin GA. Noncoding RNAs in extracellular fluids as cancer biomarkers: The new frontier of liquid biopsies. Cancers. 2019;11(1170):1–52.

189. Wanowska E, Kubiak MR, Rosikiewicz W, Makałowska I, Szcześniak MW. Natural antisense transcripts in diseases: From modes of action to targeted therapies. Wiley Interdiscip Rev RNA. 2018;9(2):1–16.

190. Li CH, Chen Y. Targeting long non-coding RNAs in cancers: Progress and prospects. Int J Biochem Cell Biol. 2013;45(8):1895–910.
191. Arun G, Diermeier S, Akerman M, Chang KC, Wilkinson JE, Hearn S, et al. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. Genes Dev. 2016;30(1):34–51.
192. Gutschner T, Hämmerle M, Eißmann M, Hsu J, Kim Y, Revenko A, et al. The non-coding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res. 2013;73(3):1180–9.
193. Modarresi F, Faghihi MA, Lopez-Toledano MA, Fatemi RP, Magistri M, Brothers SP, et al. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. Nat Biotechnol. 2012;30(5):453–9.
194. Arun G, Diermeier SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. Trends Mol Med. 2018;24(3):257–77.
195. Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar K, Kabadi AM, et al. Highly specific epigenome editing by CRISPR/Cas9 repressors for silencing of distal regulatory elements. Nat Methods. 2015;12(12):1143–9.
196. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, et al. RNA targeting with CRISPR-Cas13a. Nature. 2017;550(7675):280–4.
197. Adams BD, Parsons C, Walker L, Zhang WC, Slack FJ. Targeting noncoding RNAs in disease: Challenges and opportunities. J Clin Invest. 2017;127(3):761–71.
198. Scoles DR, Minikel E V., Pulst SM. Antisense oligonucleotides: A primer. Neurol Genet. 2019;5(2):1–8.
199. Gofrit ON, Benjamin S, Halachmi S, Leibovitch I, Dotan Z, Lamm DL, et al. DNA based therapy with diphtheria toxin-A BC-819: A phase 2b marker lesion trial in patients with intermediate risk nonmuscle invasive bladder cancer. J Urol. 2014;191(6):1697–702.
200. Thomas M, White RL, Davis RW. Hybridization of RNA to double-stranded DNA: Formation of R-loops. Proc Natl Acad Sci USA. 1976;73(7):2294–8.
201. Drolet M, Phoenix P, Menzel R, Massé E, Liu LF, Crouch RJ. Overexpression of RNase H partially complements the growth defect of an Escherichia coli ΔtopA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proc Natl Acad Sci USA. 1995;92(8):3526–30.
202. Aguilera A, García-Muse T. R loops: From transcription byproducts to threats to genome stability. Mol Cell. 2012;46(2):115–24.
203. Masukata H, Tomizawa J. Effects of point mutations on formation and structure of the RNA primer for ColE1 DNA replication. Cell. 1984;36:513–22.
204. Xu B, Clayton DA. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: An implication for RNA-DNA hybrids serving as primers. EMBO J. 1996;15(12):3135–43.
205. Yu K, Chedin F, Hsieh C-L, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol. 2003;4(5):442–51.
206. Wongsurawat T, Jenjaroenpun P, Kwoh CK, Kuznetsov V. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. Nucleic Acids Res. 2012;40(2):1–15.
207. Wahba L, Gore SK, Koshland D. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. Elife. 2013;2013(2):1–20.
208. Roy D, Yu K, Lieber MR. Mechanism of R-loop formation at immunoglobulin

class switch sequences. Mol Cell Biol. 2008;28(1):50–60.

209. Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: Separation of RNA from DNA by RNA polymerase II. Science. 2004;303:1014–6.

210. Kasahara M, Clikeman JA, Bates DB, Kogoma T. RecA protein-dependent R-loop formation in vitro. Genes Dev. 2000;14:360–5.

211. Ginno PA, Lim YW, Lott PL, Korf I, Chédin F. GC skew at the 5’ and 3’ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome Res. 2013;23(10):1590–600.

212. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human Senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell. 2011;42(6):794–805.

213. Shaw NN, Arya DP. Recognition of the unique structure of DNA:RNA hybrids. Biochimie. 2008;90(7):1026–39.

214. Roberts RW, Crothers DM. Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. Science. 1992;258(5087):1463–6.

215. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. 2004;18:1618–1629.

216. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science. 2013;340(6132):619–21.

217. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993;362:709–15.

218. Roy D, Zhang Z, Lu Z, Hsieh C-L, Lieber MR. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: A nick can serve as a strong R-loop initiation site. Mol Cell Biol. 2010;30(1):146–59.

219. Allison DF, Wang GG. R-loops: formation, function, and relevance to cell stress. Cell Stress. 2019;3(2):38–46.

220. El Hage A, French SL, Beyer AL, Tollervey D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. 2010;24(14):1546–58.

221. Li X, Manley JL. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell. 2005;122(3):365–78.

222. Chakraborty P, Grosse F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. DNA Repair. 2011;10(6):654–65.

223. Toubiana S, Selig S. DNA:RNA hybrids at telomeres - when it’s better to be out of the (R) loop. FEBS J. 2018;1–15.

224. Belotserkovskii BP, Shin JHS, Hanawalt PC. Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. Nucleic Acids Res. 2017;45(11):6589–99.

225. Reddy K, Tam M, Bowater RP, Barber M, Tomlinson M, Nichol Edamura K, et al. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic Acids Res. 2011;39(5):1749–62.

226. Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA

export factor PCID2. Nature. 2014;511(7509):362–5.

227. Gorthi A, Romero JC, Loranc E, Cao L, Lawrence LA, Goodale E, et al. EWS–FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. Nature. 2018;555(7696):387–91.

228. Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, et al. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc Natl Acad Sci U S A. 2013;110(34):13938–43.

229. Reeves R, Nissen MS. The A·T-DNA-binding domain of mammalian High Mobility Group I chromosomal proteins. J Biol Chem. 1990;265(15):8573–82.

230. Fedele M, Battista S, Manfioletti G, Croce CM, Giancotti V, Fusco A. Role of the high mobility group A proteins in human lipomas. Carcinogenesis. 2001;22(10):1583–91.

231. Fusco A, Fedele M. Roles of HMGA proteins in cancer. Nat Rev Cancer. 2007;7(12):899–910.

232. Henriksen J, Stabell M, Meza-Zepeda LA, Lauvrak SAU, Kassem M, Myklebost O. Identification of target genes for wild type and truncated HMGA2 in mesenchymal stem-like cells. BMC Cancer. 2010;10(329):1–14.

233. Watanabe S, Ueda Y, Akaboshi SI, Hino Y, Sekita Y, Nakao M. HMGA2 maintains oncogenic RAS-induced epithelial-mesenchymal transition in human pancreatic cancer cells. Am J Pathol. 2009;174(3):854–68.

234. Luo Y, Li W, Liao H. HMGA2 induces epithelial-to-mesenchymal transition in human hepatocellular carcinoma cells. Oncol Lett. 2013;5(4):1353–6.

235. Sun J, Sun B, Sun R, Zhu D, Zhao X, Zhang Y, et al. HMGA2 promotes vasculogenic mimicry and tumor aggressiveness by upregulating Twist1 in gastric carcinoma. Sci Rep. 2017;7(2229):1–13.

236. Yang S, Gu Y, Wang G, Hu Q, Chen S, Wang Y, et al. HMGA2 regulates acute myeloid leukemia progression and sensitivity to daunorubicin via Wnt/β-catenin signaling. Int J Mol Med. 2019;44:427–36.

237. Eguchi-Ishimae M, Eguchi M, Wu Z, Ming W, Iwabuki H, Inukai T, et al. HMGA2 as a potential molecular target in MLL-AF4 positive infant acute lymphoblastic leukemia. Blood. 2014;124(21):2244.

238. Thuault S, Valcourt U, Petersen M, Manfioletti G, Heldin CH, Moustakas A. Transforming growth factor-β employs HMGA2 to elicit epithelial-mesenchymal transition. J Cell Biol. 2006;174(2):175–83.

239. Wend P, Runke S, Wend K, Anchondo B, Yesayan M, Jardon M, et al. WNT10B/β-catenin signalling induces HMGA2 and proliferation in metastatic triple-negative breast cancer. EMBO Mol Med. 2013;5(2):264–79.

240. Gerritsen M, Yi G, Tijchon E, Kuster J, Schuringa JJ, Martens JHA, et al. RUNX1 mutations enhance self-renewal and block granulocytic differentiation in human in vitro models and primary AMLs. Blood Adv. 2019;3(3):320–32.

241. Zou Q, Wu H, Fu F, Yi W, Pei L, Zhou M. RKIP suppresses the proliferation and metastasis of breast cancer cell lines through up-regulation of miR-185 targeting HMGA2. Arch Biochem Biophys. 2016;610:25–32.

242. Lin Y, Liu AY, Fan C, Zheng H, Li Y, Zhang C, et al. MicroRNA-33b inhibits breast cancer metastasis by targeting HMGA2, SALL4 and Twist1. Sci Rep. 2015;5(9995):1–12.

243. Liu S, Patel SH, Ginestier C, Ibarra I, Martin-Trevino R, Bai S, et al. MicroRNA93 regulates proliferation and differentiation of normal and malignant breast stem cells. PLoS Genet. 2012;8(6):1–15.

244. De Martino I, Visone R, Fedele M, Petrocca F, Palmieri D, Hoyos JM, et al. Regulation of microRNA expression by HMGA1 proteins. Oncogene. 2009;28(11):1432–42.

245. Rogalla P, Drechsler K, Frey G, Hennig Y, Helmke B, Bonk U, et al. HMGI-C expression patterns in human tissues: Implications for the genesis of frequent mesenchymal tumors. Am J Pathol. 1996;149(3):775–9.

246. Wu J, Zhang S, Shan J, Hu Z, Liu X, Chen L, et al. Elevated HMGA2 expression is associated with cancer aggressiveness and predicts poor outcome in breast cancer. Cancer Lett. 2016;376(2):284–92.

247. Sarhadi VK, Wikman H, Salmenkivi K, Kuosma E, Sioris T, Salo J, et al. Increased expression of high mobility group A proteins in lung cancer. J Pathol. 2006;209(2):206–12.

248. Wang X, Liu X, Li AY-J, Chen L, Lai L, Lin HH, et al. Overexpression of HMGA2 promotes metastasis and impacts survival of colorectal cancers. Clin Cancer Res. 2011;17(8):2570–80.

249. Hristov AC, Cope L, Delos Reyes M, Singh M, Iacobuzio-Donahue C, Maitra A, et al. HMGA2 protein expression correlates with lymph node metastasis and increased tumor grade in pancreatic ductal adenocarcinoma. Mod Pathol. 2009;22(1):43–9.

250. Marquis M, Beaubois C, Lavallée V-P, Abrahamowicz M, Danieli C, Lemieux S, et al. High expression of HMGA2 independently predicts poor clinical outcomes in acute myeloid leukemia. Blood Cancer J. 2018;8(68):1–12.

251. Langelotz C, Schmid P, Jakob C, Heider U, Wernecke KD, Possinger K, et al. Expression of high-mobility-group-protein HMGI-C mRNA in the peripheral blood is an independent poor prognostic indicator for survival in metastatic breast cancer. Br J Cancer. 2003;88(9):1406–10.

252. Brants JR, Ayoubi TAY, Chada K, Marchal K, Van de Ven WJM, Petit MMR. Differential regulation of the insulin-like growth factor II mRNA-binding protein genes by architectural transcription factor HMGA2. FEBS Lett. 2004;569(1–3):277–83.

253. Cleynen I, Brants JR, Peeters K, Deckers R, Debiec-Rychter M, Sciot R, et al. HMGA2 regulates transcription of the Imp2 gene via an intronic regulatory element in cooperation with Nuclear Factor-κB. Mol Cancer Res. 2007;5(4):363–72.

254. Li Z, Zhang Y, Ramanujan K, Ma Y, Kirsch DG, Glass DJ. Oncogenic NRAS, required for pathogenesis of embryonic rhabdomyosarcoma, relies upon the HMGA2–IGF2BP2 pathway. Cancer Res. 2013;73(10):3041–50.

255. Zhang JY, Chan EKL, Peng XX, Tan EM. A novel cytoplasmic protein with RNA-binding motifs is an autoantigen in human hepatocellular carcinoma. J Exp Med. 1999;189(7):1101–10.

256. Bell JL, Wächter K, Mühleck B, Pazaitis N, Köhn M, Lederer M, et al. Insulin-like Growth Factor 2 mRNA-Binding Proteins (IGF2BPs): Post-transcriptional drivers of cancer progression? Cell Mol Life Sci. 2013;70(15):2657–75.

257. Cao J, Mu Q, Huang H. The roles of Insulin-like Growth Factor 2 mRNA-Binding Protein 2 in cancer and cancer stem cells. Stem Cells Int. 2018;2018:1–15.

258. Nielsen J, Kristensen MA, Willemoës M, Nielsen FC, Christiansen J. Sequential dimerization of human zipcode-binding protein IMP1 on RNA: A cooperative mechanism providing RNP stability. Nucleic Acids Res. 2004;32(14):4368–76.

259. Chao JA, Patskovsky Y, Patel V, Levy M, Almo SC, Singer RH. ZBP1 recognition of β-actin zipcode induces RNA looping. Genes Dev. 2010;24(2):148–58.

260. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010;141(1):129–41.

261. Degrauwe N, Schlumpf TB, Janiszewska M, Martin P, Cauderay A, Provero P, et al. The RNA binding protein IMP2 preserves glioblastoma stem cells by preventing let-7 target gene silencing. Cell Rep. 2016;15(8):1634–47.

262. Li Z, Gilbert JA, Zhang Y, Zhang M, Qiu Q, Shavlakadze T, et al. An HMGA2-IGF2BP2 axis regulates myoblast proliferation and myogenesis. Dev Cell. 2012;23(6):1176–88.

263. Liu FY, Zhou SJ, Deng YL, Zhang ZY, Zhang EL, Wu ZB, et al. MiR-216b is involved in pathogenesis and progression of hepatocellular carcinoma through HBx-miR-216b-IGF2BP2 signaling pathway. Cell Death Dis. 2015;6(3):1–12.

264. Fawzy IO, Hamza MT, Hosny KA, Esmat G, El Tayebi HM, Abdelaziz AI. MiR-1275: A single microRNA that targets the three IGF2-mRNA-binding proteins hindering tumor growth in hepatocellular carcinoma. FEBS Lett. 2015;589(17):2257–65.

265. Liu W, Li Y, Wang B, Dai L, Qian W, Zhang JY. Autoimmune response to IGF2 mRNA-Binding Protein 2 (IMP2/p62) in breast cancer. Scand J Immunol. 2015;81(6):502–7.

266. Barghash A, Helms V, Kessler SM. Overexpression of IGF2 mRNA-Binding Protein 2 (IMP2/p62) as a feature of basal-like breast cancer correlates with short survival. Scand J Immunol. 2015;82(2):142–3.

267. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, Novartis Institutes for BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, et al. Genome-wide association analysis identifies loci for . Science. 2007;316:1331–6.

268. Liu G, Zhu T, Cui Y, Liu J, Liu J, Zhao Q, et al. Correlation between IGF2BP2 gene polymorphism and the risk of breast cancer in Chinese Han women. Biomed Pharmacother. 2015;69(2015):297–300.

269. Janiszewska M, Suvà ML, Riggi N, Houtkooper RH, Auwerx J, Clément-Schatlo V, et al. Imp2 controls oxidative phosphorylation and is crucial for preservin glioblastoma cancer stem cells. Genes Dev. 2012;26(17):1926–44.

270. Farina KL, Hüttelmaier S, Musunuru K, Darnell R, Singer RH. Two ZBP1 KH domains facilitate β-actin mRNA localization, granule formation, and cytoskeletal attachment. J Cell Biol. 2003;160(1):77–87.

271. Nielsen J, Adolph SK, Rajpert-De Meyts E, Lykke-Andersen J, Koch G, Christiansen J, et al. Nuclear transit of human zipcode-binding protein IMP1. Biochem J. 2003;376(2):383–91.

272. Hüttelmaier S, Zenklusen D, Lederer M, Dictenberg J, Lorenz M, Meng XH, et al. Spatial regulation of β-actin translation by Src-dependent phosphorylation of ZBP1. Nature. 2005;438(7067):512–5.

273. Gong C, Li Z, Ramanujan K, Clay I, Zhang Y, Lemire-Brachat S, et al. A long non-coding RNA, lncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. Dev Cell. 2015;34(2):181–91.

274. Dai N, Rapley J, Ange M, Yanik FM, Blower MD, Avruch J. mTOR phosphorylates IMP2 to promote IGF2 mRNA translation by internal ribosomal entry. Genes Dev. 2011;25(11):1159–72.

275. Busch B, Bley N, Müller S, Glaß M, Misiak D, Lederer M, et al. The oncogenic triangle of HMGA2, LIN28B and IGF2BP1 antagonizes tumor-suppressive actions of the let-7 family. Nucleic Acids Res. 2016;44(8):3845–64.

276. Jønson L, Christiansen J, Hansen TVO, Vikeså J, Yamamoto Y, Nielsen FC. IMP3 RNP safe houses prevent miRNA-directed HMGA2 mRNA decay in cancer and development. Cell Rep. 2014;7(2):539–51.
277. Dai N, Ji F, Wright J, Minichiello L, Sadreyev R, Avruch J. IGF2 mRNA binding Protein-2 is a tumor promoter that drives cancer proliferation through its client mRNAs IGF2 and HMGA1. Elife. 2017;6:1–21.
278. Satelli A, Li S. Vimentin in cancer and its potential as a molecular target for cancer therapy. Cell Mol Life Sci. 2011;68(18):3033–46.
279. Tuch BB, Laborde RR, Xu X, Gu J, Chung CB, Monighetti CK, et al. Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. PLoS One. 2010;5(2):1–17.
280. Trinklein ND, Force Aldred S, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. An abundance of bidirectional promoters in the human genome. Genome Res. 2004;14(1):62–6.
281. Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, et al. 5’ CpG island methylation is associated with transcriptional silencing of the tumour suppressor pl6/CDKN2/MTS1 in human cancers. Nat Med. 1995;1(7):686–92.
282. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med. 2003;349(21):2042–54.
283. Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setién F, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res. 2007;67(4):1424–9.
284. Feinberg AP, Vogelstein B. Hypomethylation of ras oncogenes in primary human cancers. Biochem Biophys Res Commun. 1983;111(1):47–54.
285. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009;1(2):239–59.
286. Doi A, Park I, Wen B, Murakami P, Aryee MJ, Herb B, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet. 2009;41(12):1350–3.
287. Chan YA, Aristizabal MJ, Lu PYT, Luo Z, Hamza A, Kobor MS, et al. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-Chip. PLoS Genet. 2014;10(4):1–15.
288. Dai Z, Dai X. Antisense transcription is coupled to nucleosome occupancy in sense promoters. Bioinformatics. 2012;28(21):2719–23.
289. Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet. 2004;36(8):900–5.
290. Axel R. Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. Biochemistry. 1975;14(13):2921–5.
291. Yuan W, Zhou J, Tong J, Zhuo W, Wang L, Li Y, et al. ALBA protein complex reads genic R-loops to maintain genome stability in Arabidopsis. Sci Adv. 2019;5(eaav9040):1–11.
292. Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, et al. A genomic code for nucleosome positioning. Nature. 2006;442(7104):772–8.
293. Daenen F, van Roy F, De Bleser PJ. Low nucleosome occupancy is encoded around functional human transcription factor binding sites. BMC Genomics. 2008;9(332):1–9.
294. Meng L, Person RE, Beaudet AL. Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. Hum Mol Genet. 2012;21(13):3001–12.

295. Crossley MP, Bocek M, Cimprich KA. R-loops as cellular regulators and genomic threats. Mol Cell. 2019;73(3):398–411.

296. Boros-Oláh B, Dobos N, Hornyák L, Szabó Z, Karányi Z, Halmos G, et al. Drugging the R-loop interactome: RNA-DNA hybrid binding proteins as targets for cancer therapy. DNA Repair. 2019;84(102642):1–10.

297. Xu H, Di Antonio M, McKinney S, Mathew V, Ho B, O'Neil NJ, et al. CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nat Commun. 2017;8(14432).

298. Gibbons HR, Shaginurova G, Kim LC, Chapman N, Spurlock CF, Aune TM. Divergent lncRNA GATA3-AS1 regulates GATA3 transcription in T-Helper 2 cells. Front Immunol. 2018;9:2512.

299. Dumelie JG, Jaffrey SR. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. Elife. 2017;6(e28306):1–39.

300. Yan Q, Shields EJ, Bonasio R, Sarma K. Mapping native R-loops genome-wide using a targeted nuclease approach. Cell Rep. 2019;29(5):1369–80.

301. Chen JY, Zhang X, Fu XD, Chen L. R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. Nat Protoc. 2019;14(5):1661–85.

302. Hosono Y, Niknafs YS, Prensner JR, Iyer MK, Dhanasekaran SM, Mehra R, et al. Oncogenic role of THOR, a conserved cancer/testis long noncoding RNA. Cell. 2017;171(7):1559–72.

303. Zhang X, Xue C, Lin J, Ferguson JF, Weiner A, Liu W, et al. Interrogation of nonconserved human adipose lincRNAs identifies a regulatory role of linc-ADAL in adipocyte metabolism. Sci Transl Med. 2018;10(446):1–32.

304. Ennajdaoui H, Howard JM, Sterne-Weiler T, Jahanbani F, Coyne DJ, Uren PJ, et al. IGF2BP3 modulates the interaction of invasion-associated transcripts with RISC. Cell Rep. 2016;15(9):1876–83.

305. Chatterjee S, Fasler M, Büssing I, Großhans H. Target-mediated protection of endogenous microRNAs in C. elegans. Dev Cell. 2011;20(3):388–96.

306. Pitchiaya S, Heinicke LA, Park JI, Cameron EJ, Walter NG. Resolving sub-cellular miRNA trafficking and turnover at single-molecule resolution. Cell Rep. 2017;19(3):630–42.

307. Peng S, Chen LL, Lei XX, Yang L, Lin H, Carmichael GG, et al. Genome-wide studies reveal that Lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. Stem Cells. 2011;29(3):496–504.

308. Polesskaya A, Cuvellier S, Naguibneva I, Duquet A, Moss EG, Harel-Bellan A. Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. Genes Dev. 2007;21(9):1125–38.

309. Wilbert ML, Huelga SC, Kapeli K, Stark TJ, Tiffany Y, Chen SX, et al. LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Mol Cell. 2012;48(2):195–206.

310. Hafner M, Max KEA, Bandaru P, Morozov P, Gerstberger S, Brown M, et al. Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. RNA. 2013;19(5):613–26.

311. Feng C, Neumeister V, Ma W, Xu J, Lu L, Bordeaux J, et al. Lin28 regulates HER2 and promotes malignancy through multiple mechanisms. Cell Cycle. 2012;11(13):2486–94.

312. Nguyen LH, Robinton DA, Seligson M, Wu L, Li L, Rakheja D, et al. Lin28b is sufficient to drive liver cancer and necessary for its maintenance in murine models. Cancer Cell. 2014;26(2):248–261.

313. Liu K, Jiao X-D, Hao J-L, Qin BD, Wu Y, Chen W, et al. MTSS1 inhibits

metastatic potential and induces G2/M phase cell cycle arrest in gastric cancer. Onco Targets Ther. 2019;12:5143–52.

314. Emori M, Tsukahara T, Murase M, Kano M, Murata K, Takahashi A, et al. High expression of CD109 antigen regulates the phenotype of cancer stem-like cells/cancer-initiating cells in the novel epithelioid sarcoma cell line ESX and is related to poor prognosis of soft tissue sarcoma. PLoS One. 2013;8(12):e84187.

315. Kainov Y, Favorskaya I, Delektorskaya V, Chemeris G, Komelkov A, Zhuravskaya A, et al. CRABP1 provides high malignancy of transformed mesenchymal cells and contributes to the pathogenesis of mesenchymal and neuroendocrine tumors. Cell Cycle. 2014;13(10):1530–9.

316. Kretz M, Webster DE, Flockhart RJ, Lee CS, Zehnder A, Lopez-Pajares V, et al. Suppression of progenitor differentiation requires the long noncoding RNA ANCR. Genes Dev. 2012;26(4):338–43.

317. Liu G, Zhao G, Chen X, Hao D, Zhao X, Lv X, et al. The long noncoding RNA Gm15055 represses Hoxa gene expression by recruiting PRC2 to the gene cluster. Nucleic Acids Res. 2016;44(6):2613–27.

318. Chien C-S, Wang M-L, Chu P-Y, Chang Y-L, Liu W-H, Yu C-C, et al. Lin28B/Let-7 regulates expression of Oct4 and Sox2 and reprograms oral squamous cell carcinoma cells to a stem-like state. Cancer Res. 2015;75(12):2553–65.

319. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell. 2013;25(1):69–80.

320. Zhao R, Liu Q, Lou C. MicroRNA-299-3p regulates proliferation, migration and invasion of human ovarian cancer cells by modulating the expression of OCT4. Arch Biochem Biophys. 2018;651:21–7.

321. Xu D-D, Zhou P-J, Wang Y, Zhang Y, Zhang R, Zhang L, et al. miR-150 suppresses the proliferation and tumorigenicity of leukemia stem cells by targeting the nanog signaling pathway. Front Pharmacol. 2016;7(439):1–14.

322. Fabjani G, Tong D, Wolf A, Roka S, Leodolter S, Hoecker P, et al. HMGA2 is associated with invasiveness but not a suitable marker for the detection of circulating tumor cells in breast cancer. Oncol Rep. 2005;14(3):737–41.

323. Zhou X, Benson KF, Ashar HR, Chada K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature. 1995;376(6543):771–774.

324. Mari F, Hermanns P, Giovannucci-Uzielli ML, Galluzzi F, Scott D, Lee B, et al. Refinement of the 12q14 microdeletion syndrome: Primordial dwarfism and developmental delay with or without osteopoikilosis. Eur J Hum Genet. 2009;17(9):1141–7.

325. Lynch SA, Foulds N, Thuresson AC, Collins AL, Annerén G, Hedberg BO, et al. The 12q14 microdeletion syndrome: Six new cases confirming the role of HMGA2 in growth. Eur J Hum Genet. 2011;19(5):534–9.

326. Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JRB, et al. A common variant of HMGA2 is associated with adult and childhood height in the general population. Nat Genet. 2007;39(10):1245–50.

327. Dai N, Zhao L, Wrighting D, Krämer D, Majithia A, Wang Y, et al. IGF2BP2/IMP2 deficient mice resist obesity through enhanced translation of Ucp1 mRNA and other mRNAs encoding mitochondrial proteins. Cell Metab. 2015;21(4):609–21.

328. Foti D, Chiefari E, Fedele M, Iuliano R, Brunetti L, Paonessa F, et al. Lack of the

architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. Nat Med. 2005;11(7):765–73.

329. Copley MR, Babovic S, Benz C, Knapp DJHF, Beer PA, Kent DG, et al. The Lin28b–let-7–Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. Nat Cell Biol. 2013;15(8):916–25.

330. Bridge JA, Liu J, Qualman SJ, Suijkerbuijk R, Wenger G, Zhang J, et al. Genomic gains and losses are similar in genetic and histologic subsets of rhabdomyosarcoma, whereas amplification predominates in embryonal with anaplasia and alveolar subtypes. Genes Chromosom Cancer. 2002;33(3):310–21.

331. Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, et al. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov. 2014;4(2):216–31.

332. Pappo A, Vassal G, Crowley JJ, Bolejack V, Hogendoorn PCW, Chugh R, et al. A phase 2 trial of R1507, a monoclonal antibody to the Insulin-Like Growth Factor-1 Receptor (IGF-1R), in patients with recurrent or refractory rhabdomyosarcoma, osteosarcoma, synovial sarcoma, and other soft tissue sarcomas: Results of a sarcoma alliance. Cancer. 2014;120(16):2448–56.

333. Kim A, Widemann BC, Krailo M, Jayaprakash N, Fox E, Weigel B, et al. Phase 2 trial of sorafenib in children and young adults with refractory solid tumors: A report from the Children’s Oncology Group. Pediatr Blood Cancer. 2015;62(9):1562–6.

La presente tesis ha sido desarrollada siendo su autora, Cristina Oliveira Mateos, beneficiaria del Programa Predoctoral de Formación de Personal Investigador No Doctor del Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (Beca PRE_2013_1_1009).