UNIVERSITY OF NAPLES "FEDERICO II"

**DOCTORATE**
**MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY**

**XXIX CYCLE**

# *HMGA1* PSEUDOGENES AS ONCOGENIC COMPETITIVE ENDOGENOUS RNAs


**TUTOR**
**Prof. ALFREDO FUSCO**
**Dr. FRANCESCO ESPOSITO**

**CANDIDATE**
**Dr. MARCO DE MARTINO**

**COORDINATOR**
**Prof. VITTORIO ENRICO AVVEDIMENTO**


**ACADEMIC YEAR 2015-2016**

# INDEX

# ABSTRACT

The High Mobility Group A family is composed of non-histone chromosomal proteins that orchestrate the assembly of the nucleoprotein complexes involved in gene transcription, replication, and chromatin structure. HMGA1 overexpression is a feature of human cancer, and has a causal role in cell transformation. However, the molecular basis of this upregulation has not been completely elucidated. The identification and validation of numerous *HMGA1*-targeting microRNAs demonstrates that cells are sensitive to even subtle increases in HMGA1 abundance, thus highlighting the importance of microRNA-mediated HMGA1 regulation in cancer.

Pseudogenes have long been considered as non-functional genomic sequences. However, recent evidence suggests that many of them might have some form of biological activity, and the possibility of functionality through a microRNA-mediated pathway. In fact, recent studies show that microRNAs could act as a regulatory language, through which messenger RNAs, transcribed pseudogenes, and long non-coding RNAs crosstalk with each other and form a previously unknown regulatory network. RNA transcripts involved in this network have been termed "competitive endogenous RNAs", since they influence each other's level by competing for the same pool of microRNAs through microRNA response elements on their target transcripts.

Our research group identified two *HMGA1* pseudogene-derived RNA transcripts, *HMGA1P6* and *HMGA1P7*, that competing with *HMGA1* for microRNA binding, lead to the upregulation of *HMGA1* cellular levels, exerting an oncogenic role. We demonstrate that the overexpression of these *HMGA1* pseudogenes increases the levels of *HMGA1* and other cancer-related genes by inhibiting the suppression of their synthesis mediated by microRNAs. Moreover, *HMGA1* pseudogenes were found overexpressed in several human cancer types. Interestingly, preliminary results showed that mice engineered to overexpress *HMGA1P6* or *HMGA1P7* develop malignant B cell lymphomas. Therefore, the *HMGA1*-pseudogene engineered mouse models demonstrate the oncogenic potential of these pseudogenes and indicate that ceRNA-mediated microRNA sequestration may contribute to the development of cancer.

# 1. BACKGROUND

## 1.1 HMGA proteins

The high mobility group A (HMGA) family is constituted by four proteins: HMGA1a, HMGA1b, HMGA1c and HMGA2. Two different genes encode them: *HMGA1* gene produces the HMGA1 proteins through alternative splicing[1,2]. *HMGA1* is situated at chromosomal locus 6p21 in human and in the t-complex locus on chromosome 17 in mice, whereas *HMGA2* is located at chromosomal locus 12q14-15 in humans and at the *pigmy* locus on chromosome 10 in mice, and they are well preserved during the evolution: only rare differences can be found between the human and the murine HMGA sequences.

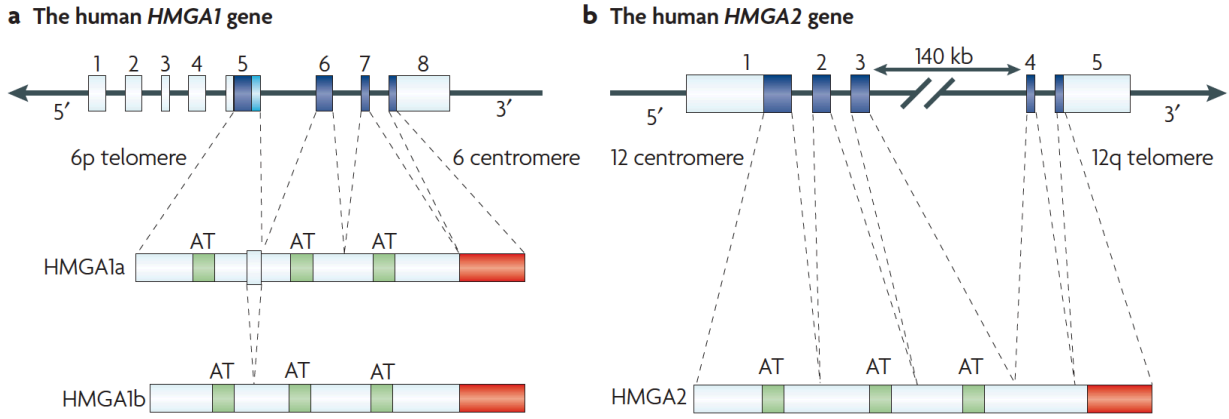



**Figure 1.** Characteristics of the HMGA genes and proteins. a) The human *HMGA1* gene consists of 8 exons, among which only exons 5 to 8 are transcribed and translated. Each exon from 5 to 7 encodes a single AT-hook domain; the 8$^{th}$ encodes the acidic C-terminal tail and contains the 3' Untraslated Region. Between the first and the second AT-hook there are 11 additional amino acids in HMGA1a protein compared to HMGA1b. b) *HMGA2* gene consists of 5 exons coding for only one protein and each of the first three exons codes for ah AT-hook domain. A huge intronic sequence separates the third from the fourth exon. From: Fusco, A. & Fedele, M. Roles of HMGA proteins in cancer. *Nature Reviews Cancer* 7, 899-910 (2007).

HMGA proteins bind the minor groove of DNA in AT-rich sequences. Their DNA-binding domain is situated in the amino-terminal region of the protein and contains three short basic domains, the so-called AT-hooks[3]. HMGA proteins do not have transcriptional activity *per se*; however, by interacting with the transcriptional machinery they alter chromatin structure and thereby regulate, negatively or positively, the transcriptional activity of several genes[4,5] (Table 1).

**Table 1.** Genes regulated by HMGA proteins

**POSITIVE REGULATION**

***Vascular endothelial tissue related***

E-selectin (Lewis *et al.* 1994)
IGFBP-1 (Allander *et al.* 1997)
COX-2 (Ji *et al.* 1998)
SM22q (Chin *et al.* 1998)
INOS (Perrella *et al.* 1999)
CD44 (Foster *et al.* 2000)

***Immune system related***

IL-2 (Reeves *et al.* 1987)
TNF-β (Fashena *et al.* 1992)
INF-β (Thanos *et al.* 1992)
IL-2Rα (John *et al.* 1995)
HLA-II (Abdulkadir *et al.*1995)
MSGA/GROα (Wood *et al.* 1995)
GM-CSF (Himes *et al.* 1996)
IgG heavy chain (Sobasjima *et al.* 1997)
c-fos (Chin *et al.* 1998)
CXCL1 (Nirodi *et al.* 2000)

***Viral genes***

JV virus early and late genes (Leger *et al.* 1995)
HSV-1 IE3 (French *et al.*1996)
HIV-1 LTR (Farnet *et al.* 1997)
HSV-1 EBNA1 (Sears *et al.* 2003)
BV EBNA1 (Sears *et al.* 2003)

***Cell cycle related***

Cyclin A (Tessari *et al.* 2003)
Cyclin E (Fedele *et al.* 2006)
CDC2 (Fedele *et al.* 2006)
CDC6 (Fedele *et al.* 2006)
CDC25A (Fedele *et al.* 2006)

***Others***

Tyrosinase (Sato *et al.* 1994)
PKCγ (Xiao *et al.* 1996)
Rhodopsin (Chau *et al.* 2000)
Neurogranin IRC3 (Xiao *et al.* 2000)
Leptin (Melillo *et al.* 2001)
Mdm2 (Pierantoni *et al.* 2006)
Bcl-2 (Esposito *et al.* 2008)

**NEGATIVE REGULATION**

GP 91-phox (Skalnik *et al.* 1992)
IL-4 (Chuvpilo *et al.* 1993)
IgE (Kim *et al.* 1995)
TCRα (Bagga *et al.* 1997)
b-globin (Chase *et al.* 1999)
a-EnaC (Zentner et al 2001)
BRCA1 (Baldassarre *et al.* 2003)
RAG2 (Battista *et al.* 2005)
Bax (Pierantoni et al 2006)
p21 (Pierantoni *et al.* 2006)
Hand-1 (Martinez Hoyos *et al.* 2008)
CBX-7 (Forzati *et al.* 2012)

## 1.2 HMGA functions in development

The high expression of HMGA proteins during embryogenesis suggests that they have a key role in development. Indeed, the phenotype characterization of mice knockout for each of the *HMGA* genes revealed the involvement of these proteins in different aspects of development. Cardiac hypertrophy and type 2 diabetes were observed in *Hmga1*-null and heterozygous mice[6], suggesting that quantitatively appropriate expression of the HMGA1 proteins are required for cardiomyocitic cell growth and function of the insulin pathway. *Hmga2*-null and heterozygous mice showed a pigmy phenotype with a decreased body size and a drastic reduction of the fat tissue[7], suggesting a critical role of the *Hmga2* gene in the control of body growth and adipocyte proliferation and differentiation. Besides HMGA1 is present in mitotic cells (spermatogonia and primary spermocytes), whereas HMGA2 is highly expressed in meiotic and post-meiotic cells[8,9]. Therefore, even though HMGA1 and HMGA2 may have overlapping functions, they seem to have different role in development.

## 1.3 HMGA expression in normal and neoplastic tissues

In normal adult tissues the HMGA protein expression is absent. In particular, *HMGA2* gene is not expressed in any of the several adult mouse and human tissue tested, whereas *HMGA1* gene is expressed at very low levels in adult murine and human tissues: a higher expression was observed only in testis, skeletal muscle and thymus.

However, in malignant tumours HMGA proteins are expressed at high levels. In particular, HMGA are expressed in all malignant tissues analysed, including pancreas, thyroid, colon, breast, lung, ovary, uterine cervix and body, prostate, gastric carcinomas and in some forms of leukaemia (Table 2). Importantly, overexpression of HMGA1 correlated with malignancy and especially with metastatic ability. Therefore, HMGA1 expression might be a candidate biomarker for cancer diagnosis and prognosis[10].

The induction of the *HMGA* genes in malignant transformation probably occurs through oncofetal transcriptional mechanisms, which have not yet been well characterized. It is known that the elevated expression of HMGA1 in cancer cells requires a complex cooperation between SP1 family members and AP1 factors, induced by the activation of Ras GTPase signaling[11].

## 1.4 HMGA as cellular oncogene

Several studies, both *in vitro* and *in vivo*, clearly established the oncogenic potential of HMGA proteins. Transfection of an antisense construct for the *HMGA2* cDNA into normal thyroid cells (FRTL5), followed by infection with transforming myeloproliferative sarcoma virus or Kirsten murine sarcoma virus (KiMSV), generated cell lines that expressed significant levels of retroviral transforming oncogenes v-mos or v-ras-Ki and removed the dependency on thyroid-stimulating hormones[12]. However, in contrast with the untransfected cells or cells transfected with the sense construct, those containing the antisense construct did not demonstrate the appearance of any malignant phenotypic markers[12].

Also the block of HMGA1 protein expression prevented thyroid cell transformation by KiMSV[13], thus strongly indicating that HMGA1 protein plays a crucial role and HMGA2 an accessory role in thyroid cell transformation. Consistent with these data, the expression of an adenovirus carrying the *HMGA1* gene in an antisense orientation suppressed HMGA1 protein synthesis and induced cell death of two human thyroid anaplastic carcinoma cell lines, but not normal thyroid cells[14].

More direct evidence of the role of HMGA proteins in malignant transformation came from work by Wood *et al.* They showed that increased expression of HMGA1 and HMGA2

leads to neoplastic transformation of Rat1a and CB33 cells, and induced distant metastasis when these cells are injected into athymic mice[15]. Moreover, human breast epithelial cells harbouring a tetracycline-regulated *HMGA1* transgene form both primary and metastatic tumours in nude mice only when the transgenes are actively expressed[16]. The generation of transgenic mice overexpressing either the *HMGA1* or the *HMGA2* gene confirmed their oncogenicity *in vivo*. In fact, both the *HMGA1* and *HMGA2* transgenic mice develop GH/PRL-secreting pituitary adenomas and T/NK lymphomas[17-19].

**Table 2.** Cancers associated with aberrant expression of HMGA proteins

***Overexpression of full-length proteins***
Lewis lung carcinoma (Giancotti *et al.* 1989)
Prostate (Tamimi *et al.* 1993)
Thyroid neoplasias (Chiappetta *et al.* 1995)
Colorectal (Fedele *et al.* 1996)
Squamous carcinoma of uterine cervix (Bandiera *et al.* 1998)
Non-small cell lung carcinoma (Rogalla *et al.* 1998)
Neuroblastomas (Giannini *et al.* 1999)
Burkitt's lymphoma (Wood *et al.* 2000)
Lipomas (Fedele *et al.* 2001)
Pancreatic duct cell carcinoma (Abe *et al.* 2002)
Breast (Baldassarre *et al.* 2003)
Lymphoblastic leukemia (Pierantoni *et al.* 2003)
Ovarian carcinoma (Masciullo *et al.* 2003)
Pituitary tumors (Fedele *et al.* 2006)
Testicular germ cell tumors (Esposito *et al.* 2008)
Oesophageal squamous cell carcinoma (Palumbo *et al.* 2016)

***Chromosomal translocation/AT hook rearrangements***
Myeloid leukemias (Elton *et al.* 1986)
Thyroid neoplasias (Chiappetta *et al.* 1995)
Pulmonary chondroid hamartomas (Kazmierczak *et al.* 1996)
Uterine leiomyomas (Henning *et al.* 1996)
Endometrial polyps (Henning *et al.* 1996)
Breast hamartoma (Dal Cin *et al.* 1997)
Lipomas (Fedele *et al.* 2001)

## 1.5 Role of HMGA proteins in neoplastic transformation

Several mechanisms have been proposed to account for the transforming ability of HMGA proteins and most of them are based on the ability of these proteins to regulate the expression of genes that have a crucial role in the control of cell proliferation and invasion. For example, HMGA proteins increase the activity of several members of AP-1 transcription factors family, which play an important role in the regulation of cell proliferation, tumorigenesis and metastasis[20].

HMGA1 also improves the transcriptional activity of NF-kB causing the expression of inflammatory proteins (iNOS, COX2, E-selectin, IL-2, IL-4 and GM-CSF) and it has been demonstrated that inflammation represents a risk factor for most types of cancer[21]. Moreover HMGA proteins overexpression impairs DNA repair negatively modulating the expression of some genes involved in this process and because they compete with p53 and human MutS homologue proteins for Holliday junction binding, exert a negative influence on the DNA mismatch repair response[22]. In addition, HMGA1 interacts with p53 and interferes with p53-mediated transcription of genes involved in regulation of apoptosis and cell cycle but promotes the transcription of the p53 inhibitor MDM2[23]. This results in a reduction of p53-dependent apoptosis. In particular, HMGA1 antagonizes the p53-mediated transcriptional repression of Bcl-2 gene[24].

Finally, HMGA1 alters the expression of genes involved in epithelial-mesenchymal transition (EMT), which is a common phenomenon in epithelial tumours[25].

Therefore, HMGA overexpression is a feature of human cancer, and has a causal role in cell transformation. However, the molecular basis of this upregulation has not been completely elucidated.

## 1.6 MicroRNA and HMGA microRNA-dependent regulation

MicroRNAs (miRNAs) are endogenous, small non-coding single-stranded RNAs of ~22 nucleotides in length, which function at post-transcriptional level as negative regulators of gene expression.

The analysis of genomic positions of miRNAs shows that most miRNAs genes are situated in intergenic regions, but they are also found within exonic or intronic regions in either sense or antisense orientation. The miRNAs situated within introns are called “mirtrons”[26]. miRNAs genes can be structured as individual ones or organized as clusters representing miRNAs families, which are commonly associated in sequence and function. miRNAs are primarily produced by RNA polymerase II (RNA pol II) from their own promoter or from promoter of the host gene in which they are localized. RNA pol II synthesized large miRNA precursors called primary-miRNAs (pri-miRNAs)[27]. Clustered miRNAs might be transcribed from a single transcription unit as polycistronic primary-miRNA.

miRNAs biogenesis takes place both in the nucleus and in the cytoplasm: firstly, pri-miRNAs are cleaved into the nucleus by RNAse III Drosha, associated to a double stranded RNA-binding protein DGCR8 known as the microprocessor complex, producing ~70 nucleotides precursor miRNA (pre-miRNA) products, which locally fold into stable secondary stem-loop

structures[28]. Then, pre-miRNAs stuctures are recognized by the transporter Exportin 5, which mediates the translocation to the cytoplasm[29]. Here the second cleavage process (dicing) occurs, carried out by the RNAse III enzyme Dicer associated to TRBP (TAR RNA-binding protein) or protein activator of the interferon induced protein kinase (also known as PRKRA), and Argonaute (AGO1-4), which cuts pre-miRNA hairpin producing a transitory miRNA/miRNA* duplex, which is composed of the mature miRNA guide (miRNA), generally selected according to thermodynamic properties, and the complementary passenger strand (miRNA*), usually subjected to degradation. miRNA/miRNA* duplex is then loaded into the miRNA-associated RNA-induced silencing complex (RISC), where the mature miRNA is able to control gene expression at post-transcriptional level by binding through or partially through the so-called miRNA response element (MREs), mainly situated on 3'Untraslated Region (3'UTR), and leading to mRNA degradation or translation inhibition, respectively [30] (Figure 2)**.**

Bionformatic analyses show that miRNAs account for 1-5% of the human genome but they can control at least 30% of protein-coding genes[31,32]. To date, 1800 distinct miRNAs molecules have been identified within the human genome, and lot of these are very well conserved through the evolution[33].

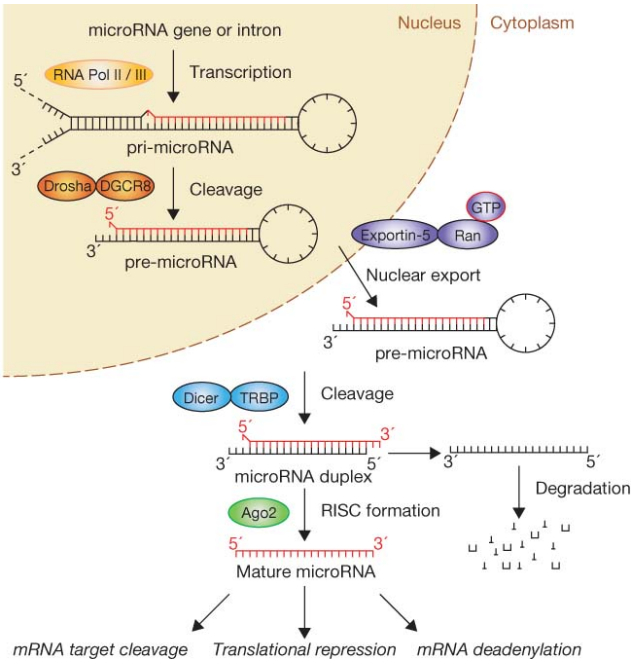



**Figure 2.** The current model for the biogenesis and post-transcriptional suppression of miRNAs. From: Winter, J. *et al.* Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology* 11, 228 - 234 (2009)

Recent studies have shown that HMGA protein levels are regulated by miRNAs. Indeed, the presence of negatively acting regulatory elements within the *HMGA2* 3'UTR was first shown by experiments that found an upregulation in luciferase activity by truncating the *HMGA2*

3'UTR[34]. However, a mechanism accounting for these results, based on the involvement of miRNAs, was described: multiple target sites for let-7, one of the founding members of the miRNA family, were found in the 3'UTR of the *HMGA2* gene, suggesting a role for this miRNA in the negative regulation of HMGA2[35,36]. Indeed, ectopic expression of let-7 reduced *HMGA2* expression by promoting degradation of its mRNA[35]. These findings were similar to those obtained by other groups, who showed that the let-7 family of miRNAs suppresses *HMGA2* in a mouse model system[36], in head and neck cancers[37] and in uterine leiomyomas[38]. Other studies demonstrate the direct targeting of *HMGA1* by miR-15[39], miR-16[39], miR-26a[40], miR-196a[39], miR-214[41], miR-761[41] and let-7, and of *HMGA2* by miR-16[39] and let-7[35]. Taken together, all these reports underline the importance exerted by the *HMGA1* and *HMGA2* miRNA-dependent regulation in cancer progression.

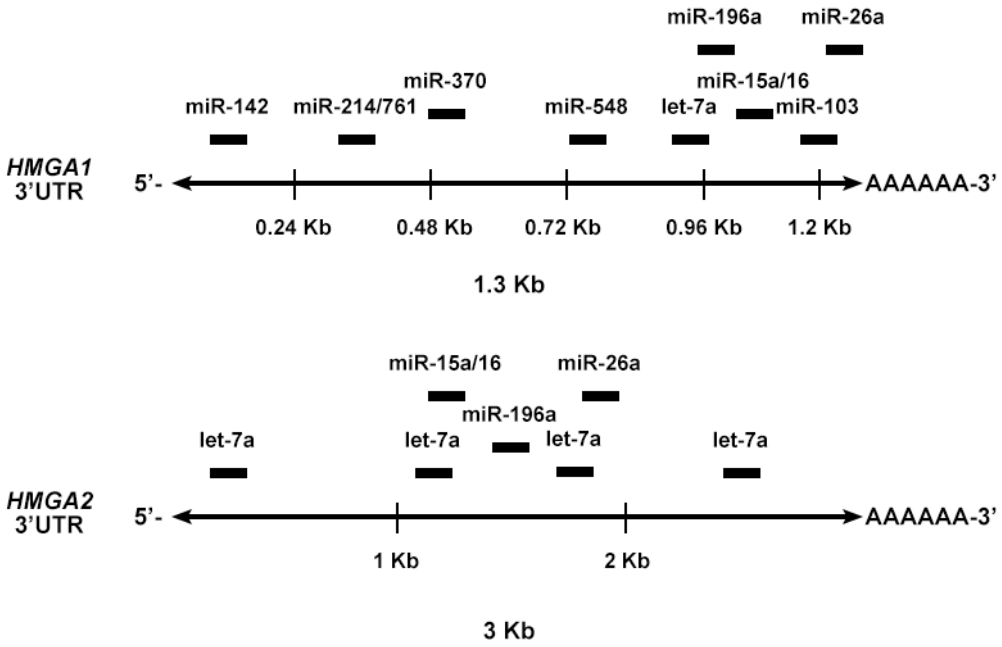



**Figure 3.** Binding sites of *HMGA1/2*-targeting miRNAs. Schematic representation of human *HMGA1* and *HMGA2* 3'UTR, and the relative position of some predicted binding sites for *HMGA*-targeting miRNAs.

## 1.7 Pseudogenes and the RNA languages

Discovered in 1977, pseudogenes were dismissed as "junk DNA"[42]. They were not formally studied; indeed, extensive effort went into developing strategies to avoid their accidental detection[42-44].

However, over the past two decades, multiple classes of regulatory RNAs, both long RNAs (>few hundred nucleotides) and short RNAs (<200 nucleotides) that do not encode proteins,

have been identified. Long noncoding RNAs (lncRNAs) comprise intronic and intergenic ncRNAs and natural antisense transcripts (NATs)[45-47].

Extensive study of the short and long types of ncRNAs has shown that the noncoding part of our genome is far from non-functional[46]. Furthermore, the use of tiling arrays and high-throughput sequencing strategies has shown that more than 90% of the human genome is transcribed, whereas only 1 to 2% encodes proteins[48]. Thus, it may be that there are more biologically functional ncRNAs than protein-encoding ones, which provides a rationale for the annotation of the noncoding transcribed genome.

The attribution of function to specific pseudogenes has raised these underappreciated molecules to the status of a new class of regulatory lncRNAs involved in both physiological and pathological processes[49]. Pseudogenes are structurally similar to genes that encode functional proteins, but pseudogenes contain "defects" that, in most cases, render them unable to encode fully functional proteins[50]. Two categories of pseudogenes have been identified: non-processed and processed pseudogenes. In particular, the non-processed pseudogenes are often found in clustered gene families. These faulty genes have arisen by tandem duplication, representing an example of non-processed pseudogenes. The non-processed pseudogenes contain counterparts of both exons and introns and sometimes also of upstream promoter regions because they originate from a duplication at genomic DNA level. However, even if the non-processed pseudogene has an high homology rate with the functional gene, closer examination will identify several "molecular defects" such as inappropriate termination codons in exons, aberrant splice junctions, and so on. Certain types of subchromosomal region, especially pericentromeric and subtelomeric regions, are highly unstable. They are prone to recombination events that can result in duplicated gene segments (containing both exons and introns) being distributed to other chromosomal locations.

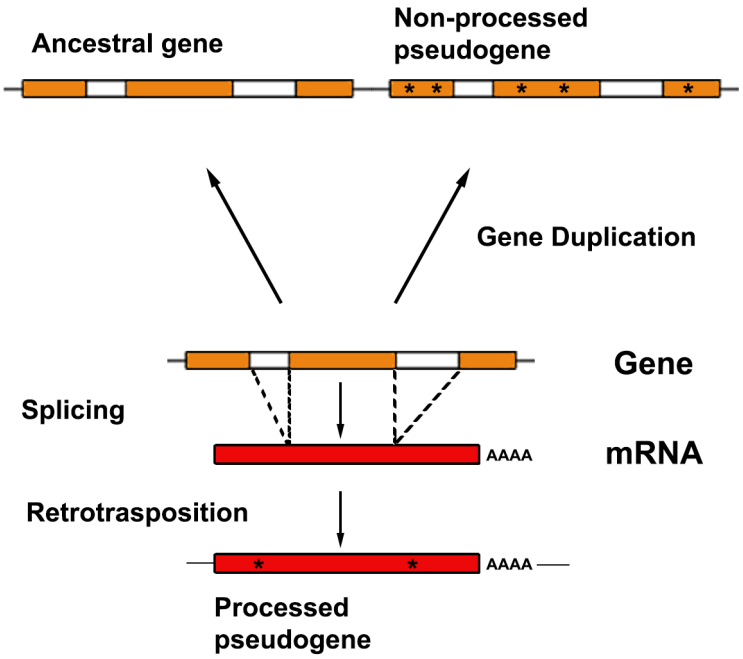



**Figure 4.** Pseudogene classes. (Upper) Non-processed pseudogenes originate from gene duplication and are located on the same chromosome of the parental gene from which they are derived. (Lower) Processed pseudogenes arise by retrotransposition and they are located on a different chromosome than the parental gene.

Processed pseudogenes are defective copies of a gene that contain only exonic sequences and lack an intronic sequence or upstream promoter sequences. They originate from an ancestral retrotransposition: cellular reverse transcriptases can use processed mRNAs to produce cDNA that may then fuse into DNA. Processed pseudogenes are common in interspersed gene families. The human genome is estimated to contain more than 18000 pseudogenes, two-thirds of which are processed. Lower organisms such as *Caenorhabditis elegans* have a comparable number of protein-coding genes as humans[51]. However, the human genome is ~30 times larger than that of *Caenorhabditis elegans*, suggesting that the noncoding portion of the genome is of crucial importance in dictating the greater complexity of higher eukaryotes[52,53]. Several lines of evidence suggest that pseudogenes may have physiological functions. The low frequency of nonsynonymous versus synonymous mutations that characterize pseudogenes indicates that their sequence is prevented from neutral drifting[54]. The active preservation of pseudogene sequence explains why, after millions of years since they arose, examples still exist of gene-pseudogene pairs that show more than 90% sequence identity[55]. It also explains the numerous examples of actively transcribed processed pseudogenes that show high evolutionary conservation among primates[56,57]. Consistent with the notion that they exert biological functions, the expression of pseudogenes is a regulated process. Studies that employed high-throughput techniques and

included pseudogenes in the analysis of differentially expressed RNAs have reported that the global expression profile of pseudogenes differs in different lines and under different conditions[58]. For example, the pseudogene transcriptome can vary during physiological processes, such as neural differentiation[59], as well as in association with patho-physiological conditions. Furthermore, various pseudogenes show a spatio-temporal expression pattern distinct from that of their coding counterparts.

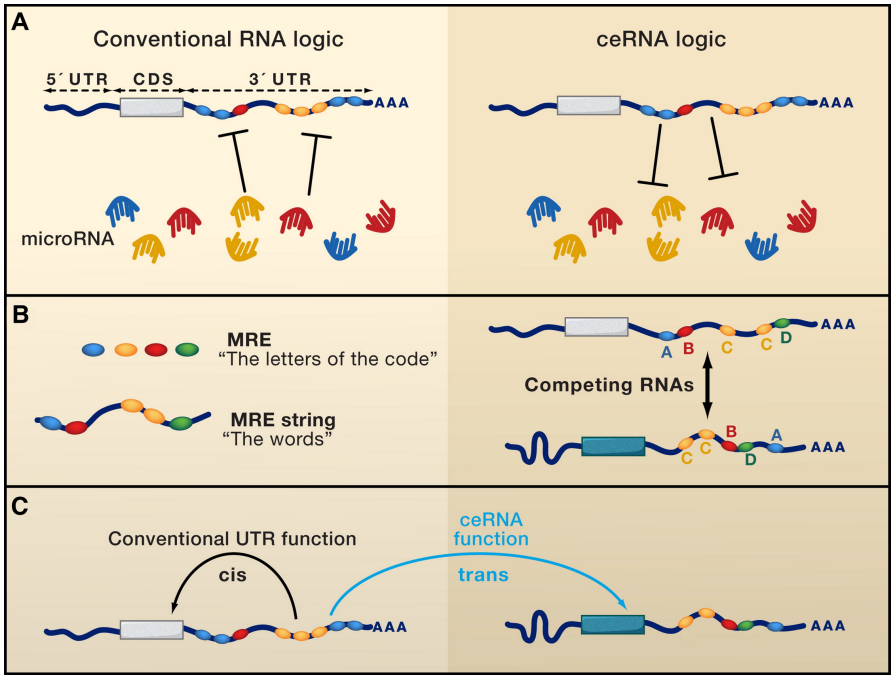



**Figure 5.** ceRNA interaction. (A) The interaction between mRNAs and miRNAs could be reciprocal, (B) producing a crosstalk between mRNA molecules through miRNAs and MREs. (C) In this scenario the 3'UTR region of a mRNA molecule can act not only *in cis*, but also *in trans*, regulating other mRNAs. From: Salmena *et al.* A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 146, 353-358 (2011).

In this scenario, miRNAs can be sequestered by pseudogene RNAs. As above-mentioned, miRNAs can bind to several complementary sequences that mainly lie on the 3'UTR of targets genes to lead to RNA degradation or translation repression[60]. The activity of miRNAs is influenced by the concomitant presence of multiple legitimate targets (either coding or non-coding). Therefore, the targets operate as competitive endogenous RNAs (ceRNAs) and shield each other from miRNA-mediated action by competing for the microRNA binding to the relative MREs. Thus, ceRNA activity confers to RNA molecules the capacity to crosstalk and to regulate *in trans* each other by influencing the availability of shared microRNAs[61]. Furthermore, ceRNA hypothesis gives a role to all transcripts carrying at least one MRE, regardless of their protein-coding ability. Consequently, MREs can be considered as the letters of a ''RNA language'' by which transcripts can actively interact each other controlling their respective expression levels.

Indeed, coding RNAs[62-64], and different classes of ncRNAs, including pseudogenes[62,65,66], have been found to act as ceRNAs.

Furthermore, the ceRNA hypothesis may expand the regulatory function of 3'UTRs. Indeed, 3'UTRs classically act as *cis* regulatory elements, regulating the stability of their own transcripts. However, following ceRNA hypothesis, 3'UTRs may also act in *trans* to modulate gene expression through microRNA binding[67].

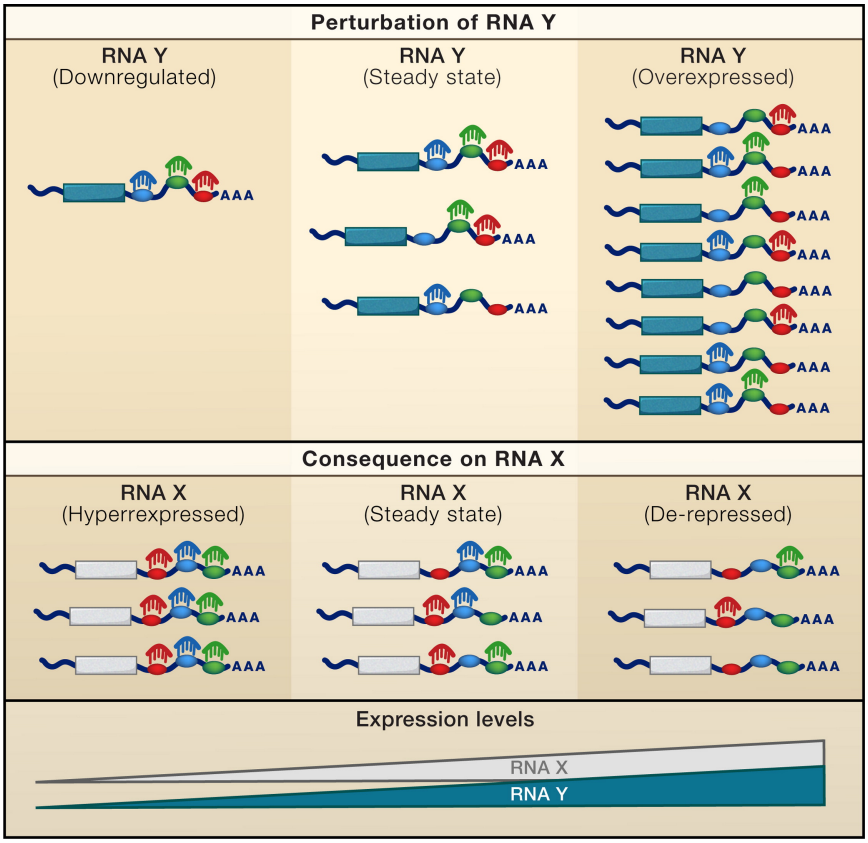



**Figure 6.** Decoy function. Several RNA transcripts (e.g., RNA X and RNA Y) have seed sequences for the same miRNAs. As proposed by ceRNA hypothesis, the overexpression of a MRE-sharing transcript (e.g., by transcriptional activation) results in the increment of MRE concentration in the cell, leading to the derepression of other transcripts that contain the same MRE sequences. From: Salmena *et al.* A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 146, 353-358 (2011).

Pseudogenes are particularly well suited to function as ceRNAs for two reasons. First, many are not translated; therefore, they can take part in microRNA binding without interference from the translation machinery[68]. Second, pseudogenes have high sequence homology with their parental genes; thus, they should play the role of multifunctional decoys that can protect their parental genes from all shared microRNAs. In other words, when a specific mRNA is overexpressed, the repression exerted by its targeting microRNAs would be diluted because the total number of MREs exceeds that of the microRNAs themselves. Thus, altering the expression

levels of an individual ceRNA would have repercussions on other ceRNAs with which it shares MREs.

Some specific features of pseudogenes suggest that they may be involved in the pathogenesis of human cancer. Indeed, Poliseno *et al.* reported that *PTEN pseudogene 1* (*PTENpg1*) is transcribed in human tissues and cancers and it contains MREs for some of the same miRNAs that target its corresponding protein-coding gene, *PTEN*[62]. By sequestering miRNAs that would regulate *PTEN,* the corresponding pseudogene derepresses the protein-coding gene from miRNA regulation. Moreover, the human *BRAF* pseudogene (*BRAFP1*) has been recently found overexpressed in various tumour types, suggesting that it may contribute to cancer progression[69]. Karreth *et al.* demonstrated the ceRNA role of both mouse *Braf-rs1* pseudogene (*Braf-rs1*) and its human orthologous, *BRAFP1*, eliciting the expression of BRAF and the activation of MAPK cascade both *in vitro* and *in vivo*. Indeed, miRNA bioinformatics analysis showed that murine *Braf-rs1* and *B-raf* share 53 miRNAs, as well as human *BRAFP1* and *BRAF* share 40 miRNAs. Thus, the *BRAF* pseudogene, *via* sequestration of common miRNAs, may work as decoy for *BRAF* in mice and humans then upregulating BRAF and promoting MAPK signalling and tumorigenesis. Furthermore, mice overexpressing *Braf-rs1* develop an aggressive tumour similar to the human diffuse large B cell lymphoma. In addition, several transcriptional or genomic aberrations of *BRAFP1* were frequently found in multiple human cancers, including B cell lymphomas. Taken together, pseudogenes are deeply involved in ceRNA hypothesis and give rise to large-scale controlling system across the transcriptome, critically increasing the functional data of human genome and acquiring main roles in physiological and pathological conditions.

## 1.8 HMGA1 Pseudogenes

The analysis of the human genome by bionformatic database revealed the presence of eight processed *HMGA1* pseudogenes.

*HMGA1P1* and *HMGA1P2* pseudogenes are located on Xp21.3 and 4q13.3 chromosome, respectively. They are not conserved during the evolution, but are only found in human genome. There are few mutations that distinguish *HMGA1P1* and *HMGA1P2* from *HMGA1a*. These changes in DNA cause few errors in protein sequence that importantly do not affect their translationability. Indeed, preliminary data show that expressing vectors for *HMGA1P1* and *HMGA1P2* are able to code for proteins. As shown in Figure 7, some HMGA1P1 and HMGA1P2 mutations hit aminoacidic residues that are frequently modified at post-translational level along the HMGA1 protein. Interestingly, these amicoacidic residues are deeply involved in

the regulation of several HMGA1 functions such as chromatin remodelling and protein-protein interactions[70-74]. Therefore, if *HMGA1P1* and *HMGA1P2* pseudogenes coded for proteins, they could represent a sort of competitor proteins for HMGA1 wild-type with different post-translational modifications, altering HMGA1 properties in chromatin remodelling and protein-protein interactions.

*HMGA1P3* pseudogene, only found in human genome, is classified as processed pseudogene and it is located on chromosome 12q24.11. Even though classified as non-coding RNA, it has only four aminoacidic mutations along the protein sequence compared to HMGA1 sequence, without affecting its translationability (Figure 7). Moreover, *HMGA1P3* lacks the C-terminal acidic tail that is a future of HMGA proteins[10]. In fact, it has been revealed that HMGA1a and HMGA1b can be regulated by phosphorylation on three serines situated in the C-terminal tail. Moreover, it is believed that HMGA C-terminal tail may be important in modulating protein-protein interactions and could be involved in enhancing transcription factor activity[75,76]. Finally, expression of truncated *Hmga1b* (*Hmga1/T*) gene, without both the acidic tail and the 3'UTR, significantly enhances growth rate and impairs adipocytic differentiation[77], suggesting that Hmga1/T mutant works in a contrasting manner compared to Hmga1 wild-type. Given this scenario, if *HMGA1P3* coded for a protein, it could represent a truncated form of HMGA1 wild-type with all molecular activities mentioned above.

Interestingly, previous data showed that *HMGA1P1*, *HMGA1P2* and *HMGA1P3* locus can be affected by chromosomal rearrangements in benign human tumours[78]. This study unveils the existence of an interesting pseudogene activation mechanism in tumour, since they could translocate, after a chromosomal rearrangements, under a promoter region or within a functional gene, then coding for new fusion proteins. However, no studies have evaluated *HMGA1P1*, *HMGA1P2* and *HMGA1P3* expression in human normal and malignant tissues where it might have consequences on the function of wild-type HMGA1 protein and then influence cancer progression.

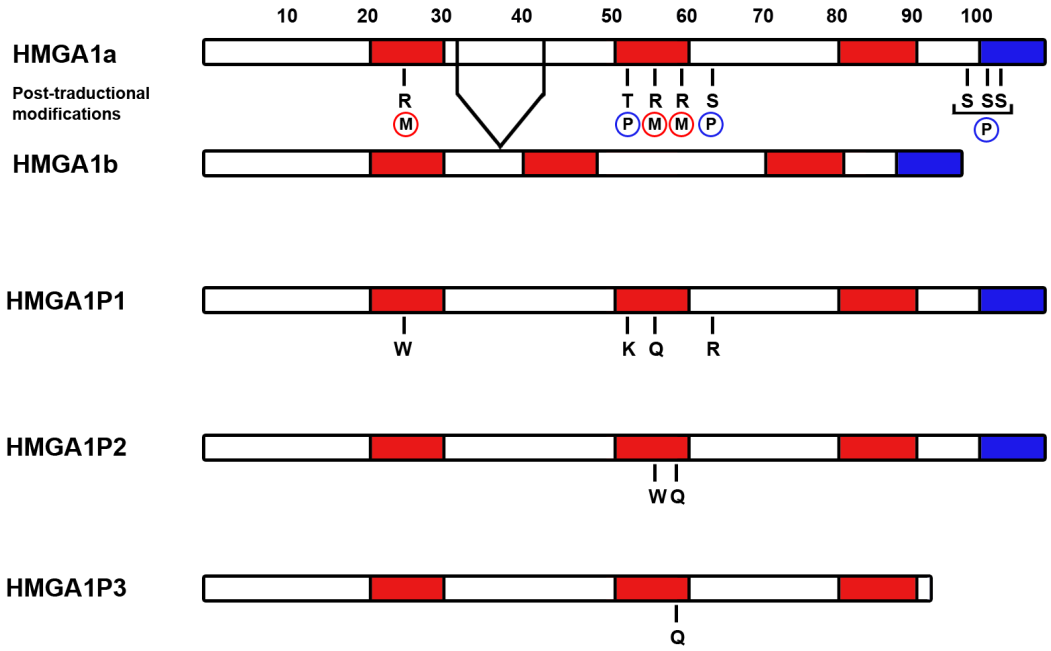



**Figure 7.** Structure of HMGA1Ps and their main mutations compared to HMGA1 proteins. Diagrams illustrating the domain structures of HMGA1a, HMGA1b, HMGA1P1, HMGA1P2 and HMGA1P3. Known post-traslational modifications of human HMGA1a and HMGA1b proteins impaired in HMGA1Ps are highlighted (Phosphorylation in blue, methylation in red). The three At-hooks are in red and the acidic tail in blue.

The non-coding RNA *HMGA1P4* pseudogene is classified as processed pseudogene and is located on the human chromosome 9q34.11. *HMGA1P4* genomic sequence shows low homology with *HMGA1*, and bioinformatics analysis confirms its untraslationability. Another processed pseudogene related to *HMGA1* is *HMGA1P5*, that is located on human chromosome 10q22.2. As *HMGA1P4*, *HMGA1P5* has low homology along *HMGA1* sequence and it may code for a peptide not related to HMGA1 protein. At the moment there are no published studies about these pseudogenes.

The processed pseudogenes *HMGA1P6* and *HMGA1P7* are sited on 13q12.12 and 6q23.2 chromosome, respectively. They are not conserved through the evolution, but they are present only in human genome. These pseudogenes have high sequence homology with *HMGA1* both in the 5' and 3'UTRs and in the coding region (Figure 8). A missense mutation of the start methionine codon avoids *HMGA1P7* mRNA translation whereas *HMGA1P6* bears a mutation in the stop codon, which is postponed several aminoacidic residues downstream, producing a non-traslatable mRNA. In the homology sequences among *HMGA1*, *HMGA1P6* and *HMGA1P7*, we retrieved conserved seed matches for miRNAs that have been predicted (miR-103, miR-142-3p, miR-370 and miR-432) or previously validated (miR-15[39], miR-16[39], miR-26a[40], miR-214[41], miR-548c-3p[79] and miR-761[41]) able to target the *HMGA1* gene. In this context *HMGA1P6* and *HMGA1P7* could act as decoys for *HMGA1*-targeting miRNAs, as competitive endogenous RNAs.

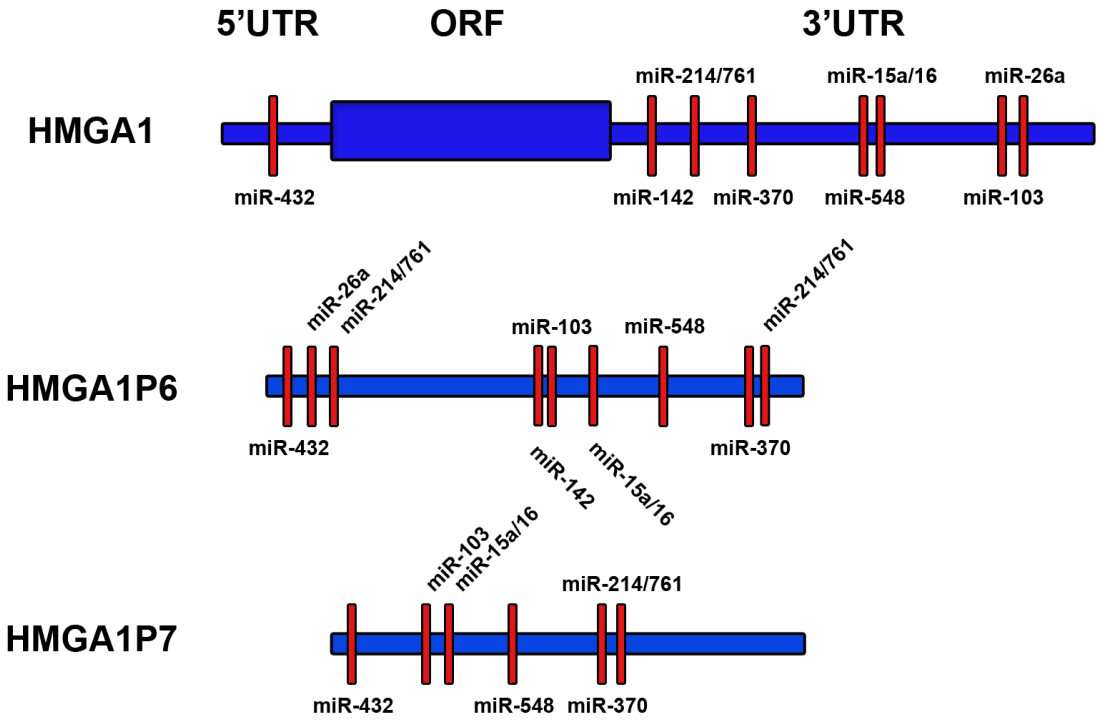



**Figure 8.** *HMGA1*, *HMGA1P6* and *HMGA1P7* share the same MREs. HMGA1 (top), HMGA1P6 (middle) and HMGA1P7 (bottom) mRNAs sequences are reported in blue, whereas the red boxes represent the HMGA1-targeting miRNA seed matches shared by these transcripts.

Finally, *HMGA1-p* is located on chromosome 2p13.2. Its expression is able to induce destabilization of *HMGA1* mRNA. Indeed, it has been demonstrated that the *HMGA1-p* RNA competes with *HMGA1* 3'UTR for a critical RNA stability factor, the alpha C-binding protein (αCP1)[80]. The *HMGA1-p* was found overexpressed in diabetic patients then causing a significant destabilization of *HMGA1* mRNA with consequent loss of INSR expression, which is regulated by HMGA1, and then generating an insulin resistance phenotype[80]. Therefore, this study established a novel mechanistic linkage between *HMGA1-p* pseudogene expression and type 2 diabetes mellitus.

# 2. AIM

Non-coding RNAs, including pseudogenes, have long been considered as non-functional genomic relicts of evolution, but a large body of evidence now suggests they are important in both physiology and disease. Pseudogenes are usually defined as defunct copies of genes that have lost their potential as DNA templates for functional products because they harbour premature or delayed stop codons, deletions/insertions and frameshift mutations that abrogate their translation into functional proteins. Since miRNAs repress target gene expression by binding to complementary sequences in the 3'UTR of target mRNA, pseudogenes can be

targeted by miRNAs that modulate the expression of coding genes. Indeed, several pseudogene transcripts exert regulatory control of their ancestral gene expression levels by competing for the same miRNAs, which is in keeping with the notion that miRNA activity is theoretically affected by the availability of target MREs in the cellular milieu. Given this scenario, the aim of the present study is to investigate the possible oncogenic function of *HMGA1P6* and *HMGA1P7* exerted by a competitive endogenous RNAs mechanism and the consequences of this new mechanism especially in the process of carcinogenesis.
This dissertation is based upon the following publications (attached at end):

Esposito, F., **De Martino, M.**, Petti, M. G., Forzati, F., Tornincasa, M., Federico, A., Arra, C., Pierantoni, G. M. & Fusco, A. *HMGA1* pseudogenes as candidate proto-oncogenic competitive endogenous RNAs. *Oncotarget* **5**, 8341-8354 (2014).

Esposito, F., **De Martino, M.**, Forzati, F. & Fusco, A. *HMGA1*-pseudogene overexpression contributes to cancer progression. *Cell Cycle* **13**, 3636-3639 (2014).

Esposito, F., **De Martino, M.**, D'Angelo, D., Mussnich, P., Raverot, G., Jaffrain-Rea, M. L., Fraggetta, F., Trouillas, J. & Fusco, A. *HMGA1*-pseudogene expression is induced in human pituitary tumors. *Cell Cycle* **14**, 1471-1475 (2015).

**De Martino, M.**, Forzati, F., Arra, C., Fusco, A. & Esposito, F. *HMGA1*-pseudogenes and cancer. *Oncotarget* **7**, 28724-28735 (2016).

**De Martino, M.**, Forzati, F., Marfella, M., Pellecchia, S., Arra, C., Terracciano, L., Fusco, A. & Esposito, F. *HMGA1P7*-pseudogene regulates H19 and Igf2 expression by a competitive endogenous RNA mechanism. *Sci Rep.* **6**, 37622 (2016).

# 3. MATERIALS AND METHODS

## 3.1 Cell culture and transfections

HEK293, MCF7, 8505c, and MEF (from 12.5-day-old embryos) cells were maintained in DMEM supplemented with 10% foetal calf serum (Thermo Fisher Scientific), glutamine and antibiotics. NIH3T3 cells were cultured in DMEM supplemented with 10% calf serum (Thermo Fisher Scientific), glutamine and antibiotics. MycoAlert (Lonza) was regularly used to test that cells were not infected by mycoplasma. Lipofectamine plus (Thermo Fisher Scientific) reagent was used to transfect the cells according to manufacturer's instructions. The transfected cells were selected in a medium containing geneticin (Sigma). Transfection efficiency was tested for each experiment by assessing GFP signal. To inhibit *HMGA1P6*, *HMGA1P7*, *DICER* and *Dicer* expression, small interfering RNAs and corresponding scrambled small interfering RNAs were designed and used as suggested by the manufacturer (RIBOXX).

## 3.2 Human cancer samples

Neoplastic and normal human thyroid tissues were obtained from surgical specimens and immediately frozen in liquid nitrogen. Thyroid tumours were collected at the Service d'Anatomo-Pathologie, Centre Hospitalier Lyon Sud, Pierre Bénite, France. The tumour samples were frozen until required for RNA or protein extraction. TissueScan Ovarian Cancer Tissue qRT-PCR Panel were purchased from Origene (HORT302).

Surgical samples of pituitary tumours were obtained from patients operated in 2 centers: Lyon, France (29 tumours) and at the Neuromed Institute, Pozzilli, Italy (12 tumours).

Normal and neoplastic human breast tissues were obtained from surgical specimens and immediately frozen in liquid nitrogen. 32 breast samples were collected at the Institute of Pathology, University of Basel, Switzerland. The tumour samples were frozen until required for RNA extraction. We declare that informed consent for the scientific use of biological material was obtained from all patients.

## 3.3 RNA extraction and quantitative reverse transcription PCR

Total RNA was extracted from tissues and cell cultures with Trizol (Thermo Fisher Scientific) according to the manufacturer's instructions. For mRNA detection, we reverse transcribed total RNA by using the QuantiTect Reverse Transcription Kit (Qiagen), and then

Real-time PCR (qRT-PCR) was performed by using Power SYBR Green PCR Master Mix (BioRad) and the following primers:

*hHMGA1*-Fw 5'-aaggggcagacccaaaaa-3'

*hHMGA1*-Rev 5'-tccagtcccagaaggaagc-3'

*hHMGA1P6*-Fw 5'-gcagacccacaaaactgga-3'

*hHMGA1P6*-Rev 5'-gagcaaagctgtcccatcc-3'

*hHMGA1P7*-Fw 5'-gctccttctcggctcctc-3'

*hHMGA1P7*-Rev 5'-gcttgggcctcttttatgg-3'

*hG6PD*-Fw 5'-acagagtgagcccttcttcaa-3'

*hG6PD*-Rev 5'-ataggagttgcgggcaaag-3'

*mHmga1*-Fw 5'-ggcagacccaagaaactgg-3'

*mHmga1*-Rev 5'-ggcactgcgagtggtgat-3'

*mG6pd*-Fw 5'-cagcggcaactaaactcaga-3'

*mG6pd*-Rev 5'-ttccctcaggatcccacac-3'

*mHmga2*-Fw 5'-aaggcagcaaaaacaagagc-3'

*mHmga2*-Rev 5'-ttgtggccatttcctaggtc-3'

*mEzh2*-Fw 5'-tggaagcagcggaggata-3'

*mEzh2*-Rev 5'-gtcactggtgactgaacactcc-3'

*mVegf*-Fw 5'-aaaaacgaaagcgcaagaaa-3'

*mVegf*-Rev 5'-tttctccgctctgaacaagg-3'

*mIgf2* Fw 5′-cctccttacccaacttcaggt-3′

*mIgf2* Rv 5′-aagagatgagaagcaccaacatc-3′

*mh19* Fw 5′-atgtcttcatttctccctatagcc-3′

*mh19* Rv 5′-gtcatcctcgccttcagtg-3′

*mCol6a3* Fw 5′-ggaggtgtacaggaagttccac-3′

*mCol6a3* Rev 5′-gactgagccgtcaaagagga-3′

*mMki67* Fw 5′-gctgtcctcaagacaatcatca-3′

*mMki67* Rev 5′-ggcgttatcccaggagact-3′

*mGpx3* Fw 5′-gtgaacggggagaaagagc-3′

*mGpx3* Rev 5′-tgagcccaggagttctgc-3′

*mLeprel1* Fw 5′-tggaccctctttaccgagaa-3′

*mLeprel1* Rev 5′-tgatccaagatggcaatcac-3′

*hActin Fw* 5′-ccaaccgcgagaagatga-3′

*hActin Rv* 5′-ccagaggcgtacagggatag-3′

*hH19* Fw 5′-ttacttcctccacggagtcg-3′

*hH19* Rv 5′-gagctgggtagcaccatttc-3′

*hIGF2* Fw 5′-gctggcagaggagtgtcc-3′

*hIGF2* Rv 5′-gggattcccattggtgtct-3′

The $2^{-\Delta\Delta Ct}$ formula was used to calculate the differential gene expression[81].

## 3.4 Plasmids

For transfection of miRNA oligonucleotides, cells were transfected with 50 nmol/ml of miRNA precursors or with a control no-targeting scrambled oligonucleotides (Thermo Fisher Scientific) using siPORT neoFX Transfection Agent (Thermo Fisher Scientific).

For the *HMGA1P6* expression construct (pCAG-*HMGA1P6*) and the *HMGA1P6* luciferase reporter construct (pGL3-*HMGA1P6*), the entire sequence of *HMGA1P6* gene (ENST00000418454.1) was amplified by using the primers

Fw *HMGA1P6* 5'-tcctctaattgggactccga-3'

Rev *HMGA1P6* 5'-ttactcagatcccaggcaga-3'

The amplified fragment was cloned into pCAG vector and into pGL3-Control firefly luciferase reporter vector (Promega), respectively. For the *HMGA1P7* construct (pCAG-*HMGA1P7*) and the *HMGA1P7* luciferase reporter construct (pGL3-*HMGA1P6*), the entire sequence of the *HMGA1P7* gene (ENST00000406908.1) was amplified by using the primers

Fw *HMGA1P7* 5'-agccagtcgagctggaggtc-3'

Rev *HMGA1P7* 5'-ctgcaatgtgtactcagagc-3'

The amplified fragment was cloned as described for the *HMGA1P6* constructs. For *Igf2* luciferase reporter construct (pGL3-*Igf2*), the miRNA seed sequence containing fragment of *Igf2* gene (ENSMUST00000000033) was amplified by using the primers:

*Igf2* Fw 5′-aatttctagacccaaaatctcacttttccc-3′

*Igf2* Rev 5′-aatttctagagatggcccataggtgtgctc-3′

The amplified fragment was cloned into pGL3-Control luciferase reporter vector (Promega).

All the generated vectors were confirmed by sequencing. The Renilla luciferase vector (pRL-CMV), for transient transfection efficiency, was purchased from Promega. The 3′UTR region of the *HMGA1* gene has been previously described[79].

## 3.5 Protein extraction, western blotting and antibodies

Cells were lysed in lysis buffer containing 1% NP40, 1 mM EDTA, 50 mMTris-HCl (pH 7.5) and 150 mM NaCl, supplemented with complete protease inhibitors mixture (Roche Branford, CT, USA). Total proteins were separated by SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose membranes (Amersham, Rainham, UK) by elettroblotting. Membranes were blocked with 5% non-fat dry milk and incubated with primary antibodies.

The primary antibodies used were anti-EZH2 (AC22) from Cell Signaling; anti-Vinculin (sc-7649), anti-Actin (sc-1615), anti-γ-Tubulin (sc-17787), anti-GAPDH (sc-32233) from Santa Cruz Biotechnology; Anti-VEGF from Abcam; anti-IGF2 (#32592) from Sabbiotech. Antibodies versus HMGA1 and HMGA2 are polyclonal Ab raised against a synthetic peptide located in the NH2-terminal region[82,83].

## 3.6 Dual-luciferase reporter assay

For dual-luciferase reporter assay, $3\times10^5$ HEK293 cells were co-transfected in 6-well plates with the pGL3-*HMGA1P6* or the pGL3-*HMGA1P7* luciferase reporter vectors, together with the Renilla luciferase plasmid and miRNA precursors or a control no-targeting scrambled oligonucleotides (Ambion), using siPORT neoFX Transfection Agent (Ambion). pGL3-*Igf2* luciferase reporter vector was co-transfected in NIH3T3 cells using the same experimental conditions. The pRL-TK control vector expressing Renilla luciferase (Promega) was used for normalization of cell number and transfection efficiency. Luciferase activity was measured 48 hours after transfection using the Dual-Luciferase Reporter.

## 3.7 Growth curve assay

For each experimental point $3\times10^4$ cells were plated in a 60 mm plate. Cells were counted in triplicate with Burker hemocytometer chamber[84].

## 3.8 Cell migration and invasion assay

Transwell motility and invasion assays were done using 8 micron pore, 6.5-mm polycarbonate transwell filters (Corning Costar Corp.). Cells ($3x10^4$) suspended in serum-free medium were seeded on the upper surface of the filters and allowed to migrate toward 10% FBS-containing media in the bottom compartment for 24 hours. The cells remaining on the upper surface were wiped off with a cotton swab and the cells that had migrated to the underside of the transwell filters were fixed, stained with crystal violet. For cell invasion assay $1x10^5$ cells were

placed on Matrigel basement membrane (BD Biosciences) in the upper chamber of the filters in 24-well with 300 µl DMEM/10%FBS in the lower chamber and incubated for 48 hours. Invaded cells on the lower surface of the Matrigel membrane were fixed and stained with crystal violet[85].

## 3.9 Flow cytometric analysis

HEK293 cells were transfected with *HMGA1P6*, *HMGA1P7* and the empty vector, and analysed by flow cytometry after 48 hours of growth under normal culture conditions. Primary MEFs were obtained from 12.5-day-old embryos. The MEFs were minced and used to establish single cell suspensions and then analysed by flow cytometry after 48 hours of growth under normal culture conditions. Briefly, cells were harvested in PBS containing 2 mmol/l EDTA, washed once with PBS, and fixed for 2 hours in cold ethanol (70%). Fixed cells were washed once in PBS and treated with 40 µg/ml RNase A in PBS for 30 minutes. They were then washed once in PBS and stained with 50 µg/ml propidium iodide (Roche). Stained cells were analysed with a fluorescence activated cell sorter (FACS) Calibur (Becton-Dickinson), and the data were analysed using a mod-t cell cycle analysis programme.

For spleen FACS analyses, organs from transgenic and WT mice were removed, pressed through a stainless steel mesh, and suspended in PBS. After one wash in PBS, the cells were resuspended in $NH_4Cl$ and washed twice more in PBS. Aliquots of lymphocytes ($1x10^6$) were placed in 96-well round-bottom plates. For hematopoietic lineage analysis, we used monoclonal antibodies specific for the following: NK-FITC (PK146), CD3 APC-H7 (560176), CD19 PE/cy7 (HIB19). All antibodies were from BioLegend.

## 3.10 Generation and genotyping of transgenic mice

The 3.5 kb *HMGA1P7* of the pCAG-*HMGA1P7* expression plasmid was excised with *Sal*I & *Hind*III restriction endonucleases by cleaving 10 µg of the plasmid. The fragment was purified from SeaKem GTG agarose (avoiding exposure to UV light) using the Qbiogene Geneclean Spin kit, then dialysed 24 hours against 2l microinjection buffer (10 mM Tris.HCl pH 7.2, 0.1 mM EDTA), and diluted to a concentration of 4 ng/µl. The DNA was injected in three sessions into C57BL/6N- derived zygotes. For this purpose, C57BL/6N female mice (bred at PolyGene from parents obtained from Charles River) were superovulated at 28-34 days of age and mated in the PolyGene mouse facility to C57BL/6N breeder males, originally also obtained from Charles River. Injected zygotes were cultivated overnight and transferred into pseudopregnant B6CBAF1 females, also from Charles River. The animals were kept in individually ventilated cages. Injections were performed at the PolyGene labs in Rümlang, Switzerland. The same

methodology was used to generate *HMGA1P6* transgenic mice. Pups were biopsied at weaning and analysed for transgene integration by PCR. All mice were maintained under standardized nonbarrier conditions in the Laboratory Animal Facility of the Istituto dei Tumori di Napoli (Naples, Italy), and all studies were conducted in accordance with Italian regulations for experimentations on animals.

## 3.11 SA-beta-gal assay

$4\times10^4$ MEFs, plated 24 hours before the assay, were washed twice with PBS and immersed in fixation buffer (2% [w/v] formaldehyde, 0.2% [w/v] glutaraldehyde in PBS) for 7 minutes. After 3 additional PBS washes, the cells were allowed to stain overnight in staining solution (40 mM citric acid/sodium phosphate, pH 6.0; 150 mM NaCl; 2.0 mM MgCl2; 1 mg/ml X-gal) at 37°C without $CO_2$ to avoid changes in pH. The next day, the staining solution was replaced with PBS, and the stained and unstained cells were counted by light microscopy (at least 24 fields).

## 3.12 RNA-sequencing

RNA samples were initially checked for quality and quantity using a Bioanalyzer with the total RNA Pico chip (Agilent Technologies, Inc) and a Qubit® with RNA Assay Kit (Thermo Fisher Scientific Inc) respectively.

Spike-In Mix 1 and Spike-In Mix 2, each containing the full complement of 92 polyadenylated transcripts from the ERCC plasmid reference library, were added to samples. SOLiD™ Total RNA-Seq Kit (Life Technologies Corporation) was used to convert RNA transcripts into a cDNA library, starting from low input amounts of poly(A) RNA, for analysis on the 5500 Genetic Analysis System. First of all, mRNA was incubated in a thermal cycler at 95 °C for 10 minutes to fragment the RNA by chemical hydrolysis. The RNA digested was hybridized and ligated with Solid specific adaptors. Two rounds of size selection using Agencourt® AMPure® XP Reagent were performed to increase the percentage of library inserts that were in the desired size range, >150 bp. The purified DNA was amplified and barcoded by 18 PCR cycles to enable sequencing of all the samples in a single multiplexed SOLiD System sequencing run.

The yield and size distribution of the amplified DNA libraries was assessed running the samples on an Agilent® 2100 BioanalyzerTM Instrument with the DNA HS Kit following the manufacturer's instructions and Qubit® dsDNA HS kit®.

Ninety nanograms of each library was pooled together and one E80 emulsion was prepared following SOLiD EZ BeadTM System. About 400 millions of pooled templated beads were

deposited on a 4 lanes of a 6 lanes slide and the sequencing was performed up to a read length of 50 bp, based on 5500 Genetic Analysis System Run sequencer protocol.

## 3.13 Bionformatic analysis

Four samples were analysed: two from WT and two from *HMGA1P7* transgenic MEFs. The comparison performed was WT versus transgenic, two biological replicates for condition.

Sequencing reads in SOLiD "xsq" format were mapped against the reference genome (UCSC GRC38/mm10); reference gene structure was Refseq from the refGene.txt file of the UCSC genome browser FTP site; the mapping so ware was the Whole Transcriptome Analysis module from the Lifescope 2.5.1 Genomic Analysis So ware analysis suite from Applied Biosystems/ThermoFisher Scientific.

A filter file, containing 6415 sequences (sequencing adaptors; barcodes; tRNAs; rRNAs; rRNA fragments; repetitive sequences; ERCC RNA sequences) was used to filter the transcripts for non-significant reads and (2) to quantify the absolute expression using the External RNA Controls Consortium (ERCC) RNA Spike-In Mix.

The genome-mapped reads were then correlated with Refseq genes and the resulting gene-associated read counts were analysed with a Genomnia proprietary procedure based on the Bioconductor library edgeR[86]. The chosen limit for evaluating differential expression was 5 counts per millions in at least half of the examined samples. The normalization procedure used was the standard for edgeR (TMM). Genes were called differentially expressed when the comparison was evaluated with a FDR < 0.05. Absolute gene expression was evaluated from the read counts in RPKM (Reads per kilo base per million mapped reads).

Primary gene annotation was performed using the Bioconductor libraries biomaRt and GOstats, while functional clustering of the genes was performed using the DAVID functional annotation web site (https://david. ncifcrf.gov/).

## 3.14 Histological analysis

For light microscopy, tissues were fixed by immersion in 10% formalin and embedded in paraffin by standard procedures. Sections of 5 μm were stained with hematoxylin and eosin reagents. Frozen sections, 4–8 μm thick, of WT and transgenic tissues were cut in a frozen microtome and allowed to dry for 1h at room temperature, before fixing in acetone for 10 min. The slides were air dried for 2 h at room temperature and then placed in PBS for 5 min before the immunoperoxidase staining procedure, using anti-CD79 (ab79414) from Abcam.

## 3.15 Statistical analysis

Data were analysed using a two-sided unpaired t test (GraphPad Prism, GraphPad Software, Inc.). Values of $P < 0.05$ were considered statistically significant. Regression analysis, correlation coefficients and statistical analysis were generated using GraphPad Prism, GraphPad Software, Inc.

# 4. RESULTS

## 4.1 *HMGA1P6* and *HMGA1P7* act as ceRNAs for *HMGA1*

Firstly, to assess the capability of HMGA1-targeting miRNAs to bind *HMGA1P6* and *HMGA1P7* (*HMGA1Ps*), we transfected miR-15, miR-16, miR-214, miR-761 into MCF7 cells (human breast adenocarcinoma), and evaluated *HMGA1P6*, *HMGA1P7* and *HMGA1* mRNA levels by qRT-PCR. As shown in Figure 9a, the transfection of the *HMGA1*-targeting miRNAs significantly reduced *HMGA1*, *HMGA1P6* and *HMGA1P7* mRNA levels.

To evaluate the direct interaction between *HMGA1*-targeting miRNAs and *HMGA1Ps* mRNAs, we cloned the full-length *HMGA1P6* and *HMGA1P7* mRNAs downstream of the luciferase open reading frame. These reporter vectors were transfected into human embryonic kidney (HEK293) cells together with miRNA precursors and a control scrambled oligonucleotide. As shown in Figure 9b, the luciferase signal was lower in miRNA-transfected cells compared with the scrambled oligonucleotide. These data suggest that the *HMGA1Ps* and *HMGA1* are subjected to the same miRNA-mediated post-trascriptional regulation.

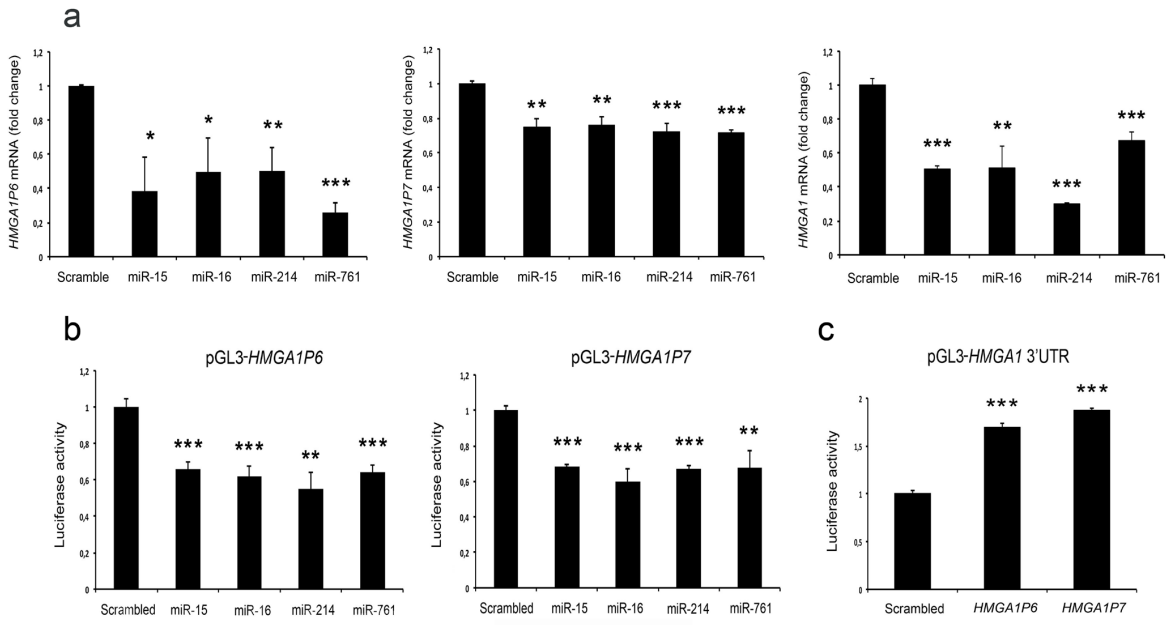



**Figure 9.** *HMGA1P6* and *HMGA1P7* are targeted by *HMGA1*-targeting miRNAs. (A) qRT-PCR analysis of *HMGA1P6* (left), *HMGA1P7* (middle) and *HMGA1* (right) mRNA from MCF7 cells transfected with scrambled-oligonucleotide, miR-15, miR-16, miR-214 and miR-761.(B) *HMGA1Ps* were cloned into the pGL3 control vector. Relative luciferase activity in HEK293 cells transiently transfected with miR-15, miR-16, miR-214, miR-761 and a control scrambled oligonucleotide. (C) The 3'UTR of *HMGA1* was cloned into the pGL3 control vector. Relative luciferase activity in HEK293 cells transiently transfected with the empty vector, *HMGA1P6* and *HMGA1P7*. The results are reported as the mean of values. Error bars represent mean ± SD; n=3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

Subsequently, we transfected a reporter vector carrying *HMGA1* 3'UTR downstream of the luciferase open reading frame to verify the ability of *HMGA1Ps* to act as decoy for *HMGA1*-targeting miRNAs. We transfected this reporter vector into HEK293 cells together with

*HMGA1P6*- and *HMGA1P7*- expressing vectors. As expected, luciferase activity was higher in *HMGA1Ps*-transfected cells compared with control (Figure 9c). Moreover, the transfection of growing amounts of *HMGA1P6*- or *HMGA1P7*- expressing vectors resulted in the overexpression of HMGA1 levels (Figure 10a) due to the reduction of negative regulation exerted by miRNAs. Conversely, *HMGA1Ps* knockdown produced a reduction of *HMGA1* mRNA and protein levels (Figure 10b), reflecting the results obtained by *HMGA1Ps* overexpression. Therefore, both *HMGA1P6* and *HMGA1P7* compete for the endogenous miRNA-binding sites.

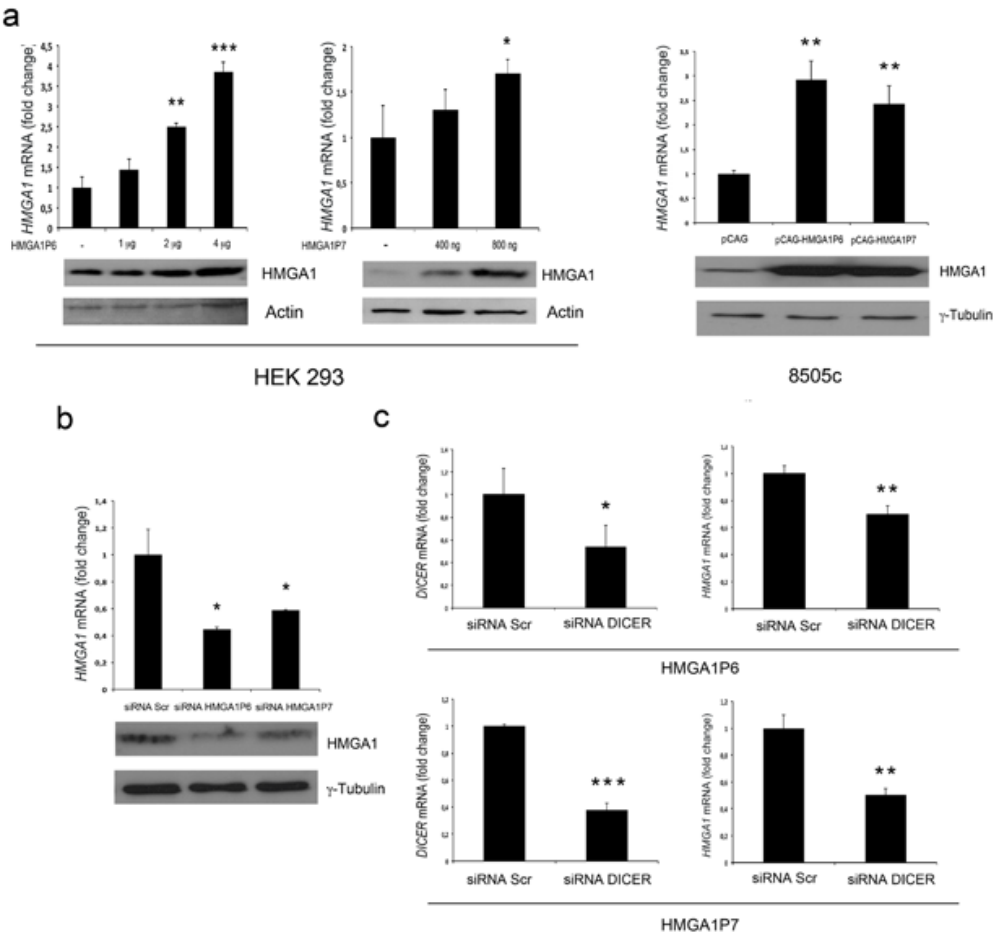



**Figure 10.** HMGA1 is positively regulated by *HMGA1Ps*. (A) (Upper panels) qRT-PCR analysis of *HMGA1* mRNA levels in HEK293 and 8505c cells transfected with the empty vector, *HMGA1P6* and *HMGA1P7*. (Lower panels) Western blot analysis of HMGA1 protein levels from the same samples shown in the upper panels. (B) (Upper panel) qRT-PCR analysis of *HMGA1* mRNA levels in 8505c cells transfected with the scrambled oligonucleotide, siRNA-*HMGA1P6* and siRNA-*HMGA1P7*s. (Lower panel) Western blot analysis of HMGA1 protein levels from the same samples shown in the upper panel. (C) *HMGA1* mRNA levels 24 h after the transfection of *HMGA1P6* and *HMGA1P7* in scrambled oligonucleotide or siRNA-*DICER* 8505c transfected cells. The results are reported as the mean of values; Error bars represent means ± SD; n =3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

Furthermore, the upregulation of *HMGA1* induced by overexpression of *HMGA1Ps* was decreased in DICER-knockdown cells (Figure 10c). In fact, silencing of DICER, the enzyme that catalyses the last step of miRNA maturation, results in a reduction of mature miRNAs compared with control cells. These results support the notion that *HMGA1Ps* need mature miRNAs to regulate HMGA1 levels, acting as ceRNAs.

## 4.2 *HMGA1P6* and *HMGA1P7* exert oncogenic activity

*HMGA1Ps* can be transcribed but not translated. However, they can act as miRNA decoys, derepressing HMGA1 mRNA and protein levels (Figure 10a). To evaluate the functional effects of *HMGA1P6* and *HMGA1P7* overexpression, we analysed their role in cellular proliferation, apoptosis, migration and invasion.

As shown in Figure 11a and 11b, HEK293 cells and 8505c cells (derived from human anaplastic thyroid carcinoma) transfected with *HMGA1P6*- or *HMGA1P7*- expressing vectors grew faster than the empty-vector transfected cells. Cell cycle data of *HMGA1P6* and *HMGA1P7* overexpressing cells showed an increased number of cells in the S phase and a reduced number of cells in G1 compared with control cells (Figure 11c). These results were confirmed also by *HMGA1Ps* silencing. In fact, the silenced cells for both *HMGA1P6* and *HMGA1P7* grew slower than the control (Figure 11d).

To analyse the role of *HMGA1* pseudogenes in apoptotic cell death, we incubated HEK293 cells with doxorubicin in the presence or absence of the *HMGA1Ps*. As shown in Figure 11e, *HMGA1P6* and *HMGA1P7* overexpressing cells showed a reduced apoptosis rate induced by doxorubicin. These results are in agreement with the finding that HMGA1 overexpression induces resistance to apoptosis[10].

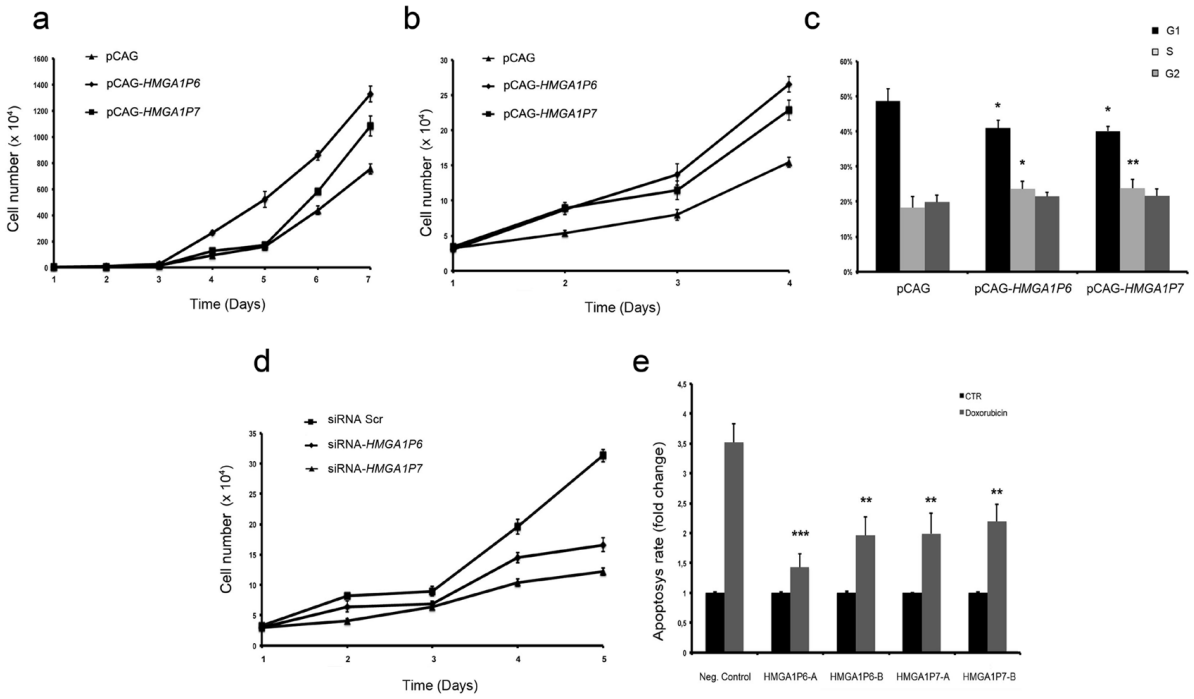



**Figure 11.** *HMGA1P6* and *HMGA1P7* expression increases cell proliferation and reduces apoptosis. (A) and (B) HEK293 and 8505c cell proliferation in *HMGA1P6*- and *HMGA1P7*-transfected cells. (C) HEK293 cells were transfected with the control, *HMGA1P6* or *HMGA1P7* vectors. The DNA content of the transfected HEK293 cells was analysed by flow cytometry after propidium iodine staining. (D) 8505c cell proliferation in siRNA-*HMGA1P6*- and siRNA-*HMGA1P7*-transfected cells. (E) *HMGA1P6*- and *HMGA1P7*-transfected cells were treated with doxorubicin, and apoptosis was assessed by FACS. The results are reported as the mean of values; Error bars represent means ± SD; n =3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

Furthermore, we performed cell migration and invasion assays in *HMGA1Ps* transfected cells. As expected, cell migration was significantly higher in two cell lines overexpressing *HMGA1Ps* than in control cells (Figure 12a). Suitably, the invasion matrigel assay reported a higher invasion capability in *HMGAPs* overexpressing cells compared to control (Figure 12b). These results indicate that *HMGA1Ps* are strongly involved in cell proliferation, motility and invasion.

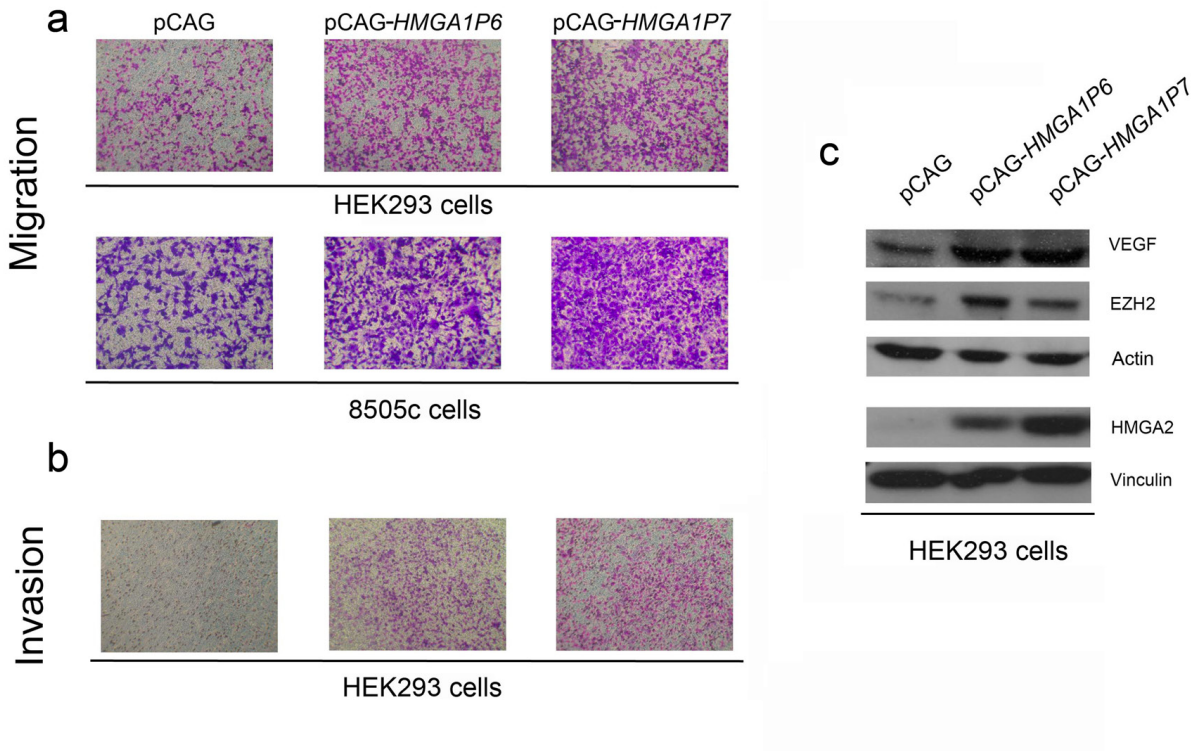



**Figure 12.** The expression of the *HMGA1Ps* affects cell migration and invasion. (A) Cell migration assays of HEK293 and 8505c cells transfected with *HMGA1P6* or *HMGA1P7* or with a control vector. One representative experiment is reported. (B) Cell invasion assays of HEK293 cells transfected with *HMGA1Ps* or with the backbone vector. One representative experiment is reported. (C) Extracts from HEK293 transfected with *HMGA1P6* or *HMGA1P7* or with a control vector were analysed by Western blotting.

Interestingly, although *HMGA1Ps* exert a miRNA decoy activity on *HMGA1* oncogene, bioinformatic analysis revealed that they contain sequences that can be targeted by miRNAs that target *High Mobility Group A2* (*HMGA2*), *Vascular Endothelial Growth Factor* (*VEGF*) and *Enhancer of Zeste Homolog 2* (*EZH2*), all of which are known to be involved in carcinogenesis. Thus, we found that *HMGA1Ps* overexpression increased the level of the proteins coded for by these genes (Figure 12c). Taken together these data underline the role of *HMGA1* pseudogenes in cancer progression by acting as decoys for cancer-related genes other than *HMGA1*.

## 4.3 Correlation of HMGA1 and the overexpression of *HMGA1Ps* in human cancer

To analyse whether *HMGA1Ps* act as decoys in the regulation of HMGA1 protein levels also in human cancer, we evaluated HMGA1 and *HMGA1Ps* expression in a panel of differentiated and undifferentiated thyroid carcinomas by Western-Blotting and qRT-PCR. As shown in Figure 13a, papillary thyroid carcinomas (PTC), which are well differentiated and poorly aggressive, expressed low levels of *HMGA1P6* and *HMGA1P7*, as well as low levels of

HMGA1 protein (Figure 13b). Conversely, anaplastic thyroid carcinoma (ATC), which is one of the most aggressive human neoplasia, expressed very high *HMGA1Ps* levels that, moreover, correlated with HMGA1 protein levels (Figure 13a and 13b). Similar data were obtained in human ovarian carcinomas (Figure 13c and 13d). The direct correlation between *HMGA1* and *HMGA1P6* expression (r=0.6553, P<0.0001) and between *HMGA1* and *HMGA1P7* expression (r=0.7001, P<0.0001) indicates that these genes are co-regulated.

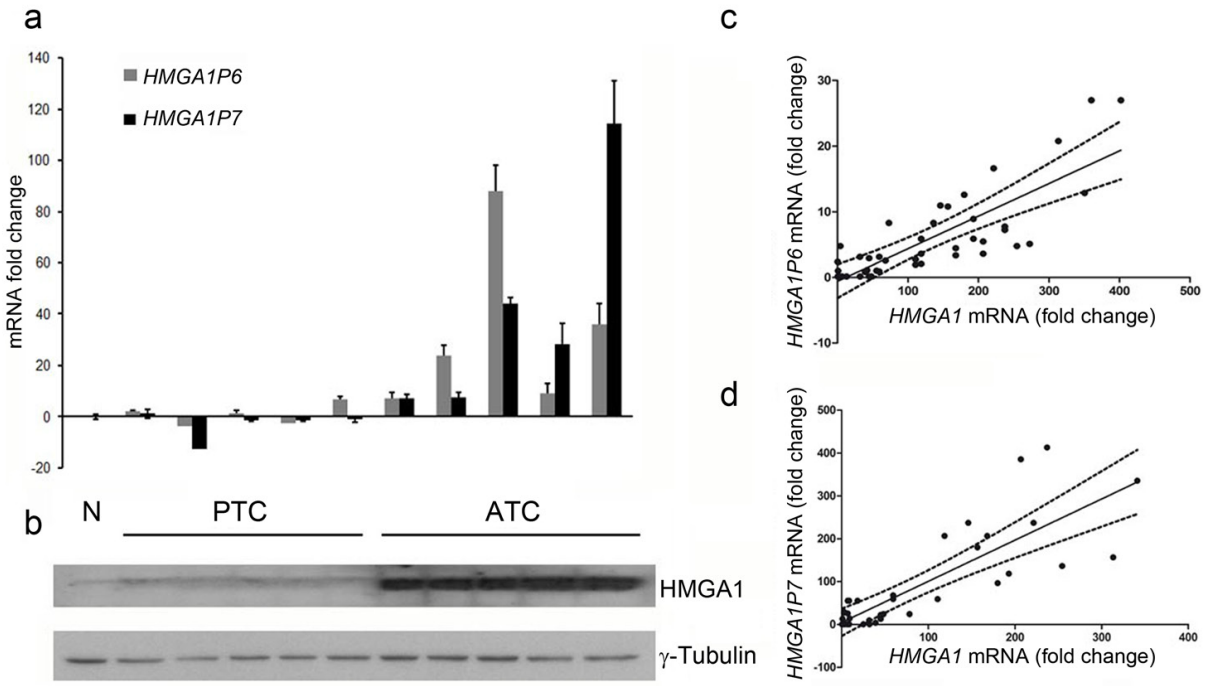



**Figure 13.** HMGA1 protein expression positively correlates with the expression of the *HMGA1Ps* in ATC. (A) *HMGA1P6* and *HMGA1P7* qRT-PCR analysis in normal thyroid tissue (NT), papillary thyroid carcinoma (PTC) and anaplastic thyroid carcinoma (ATC). The results are reported as the mean of expression values. The error bars represent mean ± SD; n = 3. (B) Western blot analysis of HMGA1 protein expression in the same samples as in A. (C) and (D) Ovarian sample expression values derived from commercial sources were combined for correlation analysis. Linear regressions of *HMGA1 vs. HMGA1P6* (C) and *HMGA1 vs. HMGA1P7* (D) are shown.

Furthermore, we analysed *HMGA1Ps* expression in a panel of 41 human pituitary tumours (PT), including 14 growth hormone tumours (GH) with acromegaly and 27 non-functioning pituitary adenomas (NFPA) or gonadotroph FSH-LH tumours. In fact, several evidences showed that HMGA proteins act as drivers of human PT[10,87-89], and that HMGA expression levels are associated to several miRNAs able to target both *HMGA1* and *HMGA2* mRNAs in PT[39,79]. As shown in Figure 14a and 14b, *HMGA1P6* and *HMGA1P7* were differently expressed with regard to the immunohistochemical type. Furthermore, we analysed the HMGA1 in the same samples (Figure 14c). Intriguingly, a significant linear correlation was found between *HMGA1* and *HMGA1P6* expression (r=0.5952, P<0.0001), suggesting that these genes are co-regulated (Figure 14d). On the contrary, there was no correlation between *HMGA1* and *HMGA1P7* in the whole tumours panel. However, further analysis of *HMGA1* and *HMGA1Ps* co-regulation

revealed some differences according to tumour type. Indeed, in the subset of GH tumours, HMGA1 strongly correlated with both *HMGA1P6* (r=0.8593, P<0.0001) and *HMGA1P7* (r=0.6772, p<0.0001) expression (Figure 14e). Moreover, other evidences show a direct co-regulation of *HMGA1* and *HMGA1Ps* levels also in other human carcinomas such as squamous laryngeal carcinomas and endometrial carcinomas. Taken together, these data strongly support the hypothesis that *HMGA1Ps* act as ceRNAs in several human cancer types, represent a novel potential mechanism by which *HMGA1* expression levels are upregulated.

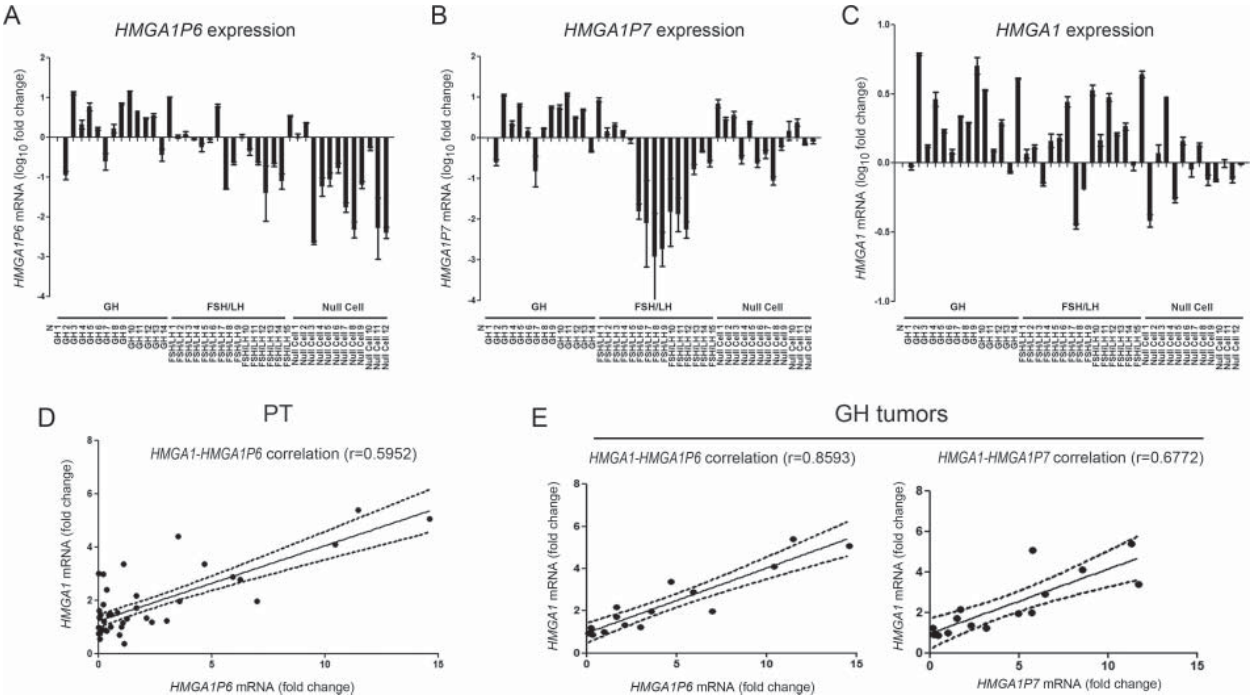



**Figure 14.** *HMGA1* and *HMGA1Ps* expression positively correlates in pituitary tumours. (A) *HMGA1P6*, (B) *HMGA1P7* and (C) *HMGA1* qRT-PCR analysis in normal pituitary gland (N), GH, FSH/LH and null cell tumours. (D, E) The obtained values were then combined for correlation analysis. (D) Linear regression of *HMGA1 vs. HMGA1P6* in the whole series of PT. (E) Linear regressions of *HMGA1 vs*. *HMGA1P6* (left panel) and *HMGA1* (right panel) *vs.HMGA1P7* in GH tumours.

## 4.4 *HMGA1P7* transgenic mouse embryonic fibroblasts grow faster and senesce later

To investigate the role of *HMGA1Ps* in a more physiological context, we generated transgenic mice overexpressing *HMGA1P7*. Firstly, we verified the expression of *HMGA1P7* in transgenic mice using RNA derived from several organs such as liver, spleen, lung and mouse embryonic fibroblasts (MEFs). As expected, *HMGA1P7* levels were high in the tissues of transgenic mice (TG) and not revealed in wild-type (WT) counterpart (Figure 15a). Then, to verify the *HMGA1P7* ceRNA mechanism also *in vivo*, we analyse *hmga1* levels. Consistently,

*hmga1* transcript and protein levels were upregulated in *HMGA1P7*-transgenic MEFs in comparison to WT control (Figure 15b). Remarkably, *HMGA1P7*-transgenic MEFs also expressed higher levels of the Hmga2, Ezh2 and Vegf proteins that are involved in carcinogenesis and are coded for by genes that share miRNAs with *HMGA1P7* (Figure 15c).

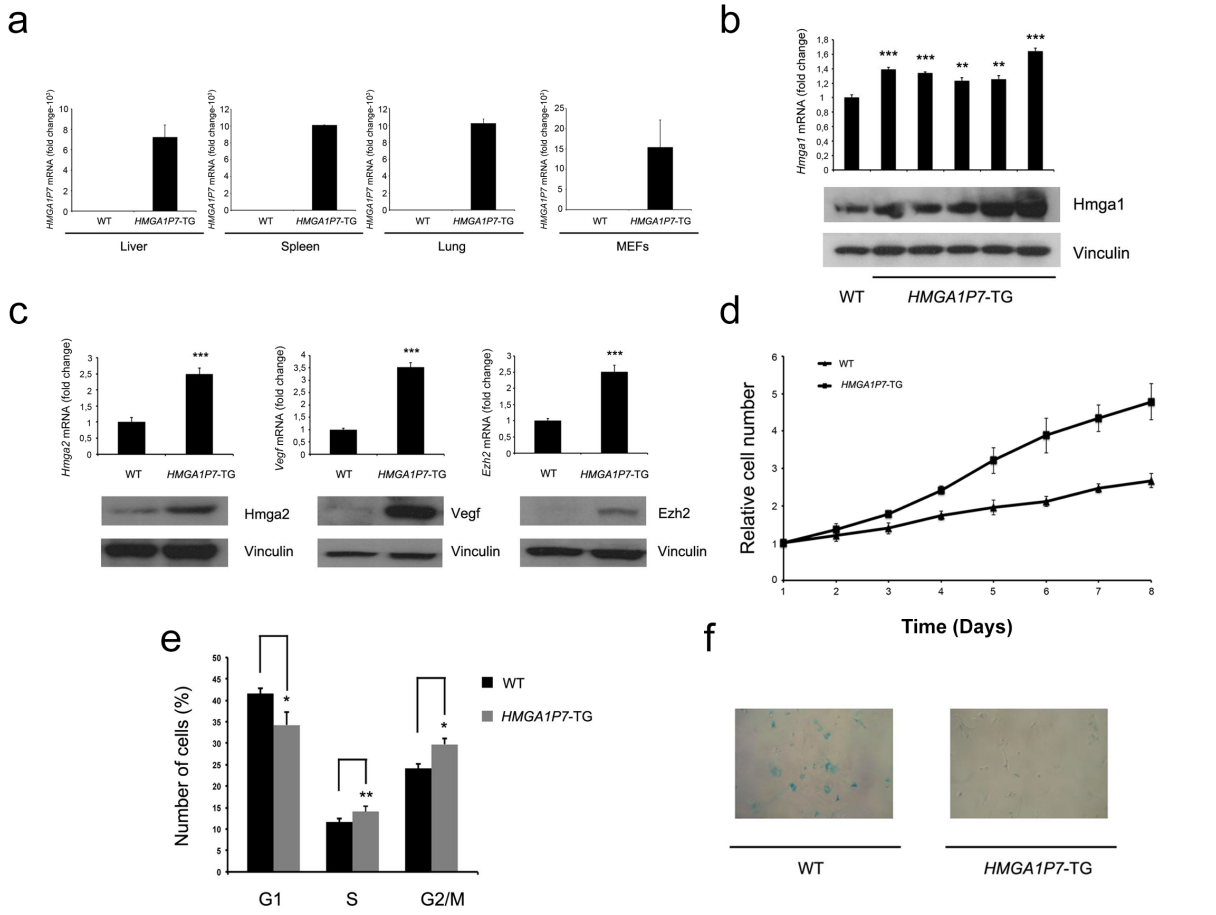



**Figure 15.** *HMGA1P7* overexpressing MEFs show a higher growth rate, and lower susceptibility to senescence. (A) qRT-PCR analysis of total RNA from livers, spleens, lungs, and MEFs of WT and *HMGA1P7* transgenic mice. (B) (upper panel) qRT- PCR analysis of *HMGA1* mRNA levels in WT and *HMGA1P7* transgenic MEFs. (Lower panel) Western blot analysis of HMGA1 protein expression in the same samples. (C) qRT-PCR and Western blot analysis of genes that share common miRNAs with *HMGA1*. (D) MEFs were prepared from WT and *HMGA1P7* overexpressing embryos at 12.5 dpc. At culture passage 3, they were plated and counted daily for 8 days. (F) Propidium iodide flow cytometry of asynchronous growing WT and *HMGA1P7* overexpressing MEFs. (G) Light microscopy of representative WT and *HMGA1P7* overexpressing MEFs stained for β-galactosidase activity at culture passages 6. The results are reported as the mean of values with *error bars* indicating SD (mean ± SD); n =3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

In agreement with the previous results, the growth rate of *HMGA1P7* transgenic MEFs was significantly higher than the wild-type counterpart (Figure 15d). Indeed, we evaluated whether the higher growth rate of *HMGA1P7*-MEFs was triggered by an altered cell cycle. So, we examined asynchronously growing MEFs by flow cytometry. The number of *HMGA1P7*-MEFs

was lower in G1 and higher in the S phase of the cell cycle compared with WT MEFs (Figure 15e).

Then, we assessed the susceptibility of MEFs to senescence by evaluating senescence-associated β-gal (SA-β-gal) activity. In particular, SA-β-gal activity was detected in WT MEFs but not in *HMGA1P7*-transgenic ones (Figure 15f). Importantly, analogous data were obtained for *HMGA1P6* transgenic MEFs (data not shown). Therefore, these results suggest that *HMGA1Ps* overexpression decreased susceptibility to cellular senescence.

## 4.5 RNA-seq on *HMGA1P7* transgenic MEFs

Since *HMGA1* pseudogenes contain several seed sequences for miRNAs their overexpression should derepress the expression of some other genes. In particular, we focused on the identification of novel *HMGA1P7* ceRNA interactors differentially expressed in *HMGA1P7* transgenic MEFs with respect to wild-type ones. To identify the genes controlled by *HMGA1P7* pseudogene, we examined the whole transcriptome of WT and *HMGA1P7* transgenic MEFs by RNA-seq analyses. To this aim, the entire population of RNA transcripts extracted from WT and *HMGA1P7*-MEFs were sequenced. The genome-wide RNA expression profiles studies reveal that about one hundred fifty transcripts (32 upregulated and 116 downregulated) were controlled by *HMGA1P7* expression with a significant fold-change variation (FDR adjusted p-value of 0.05). To validate the results obtained by RNA-seq, we analysed the expression of some upregulated genes such as *Collagen Type VI Alpha 3* (*Col6a3*), *Marker of Proliferation Ki-67* (*Mki67*), *H19*, *Insulin-like growing factor 2 (Igf2)* and downregulated genes such as *Glutathione Peroxidase 3* (*Gpx3*), *Leprecan-Like 1* (*Leprel1*) by qRT-PCR. As shown in Figure 16, the quantitative qRT-PCR analyses confirmed the data obtained from the RNA-seq analyses. Interestingly, these genes have been related to several human cancers (colon, gastric, liver, breast and haematological cancers), and are considered possible therapeutic targets[90-95].

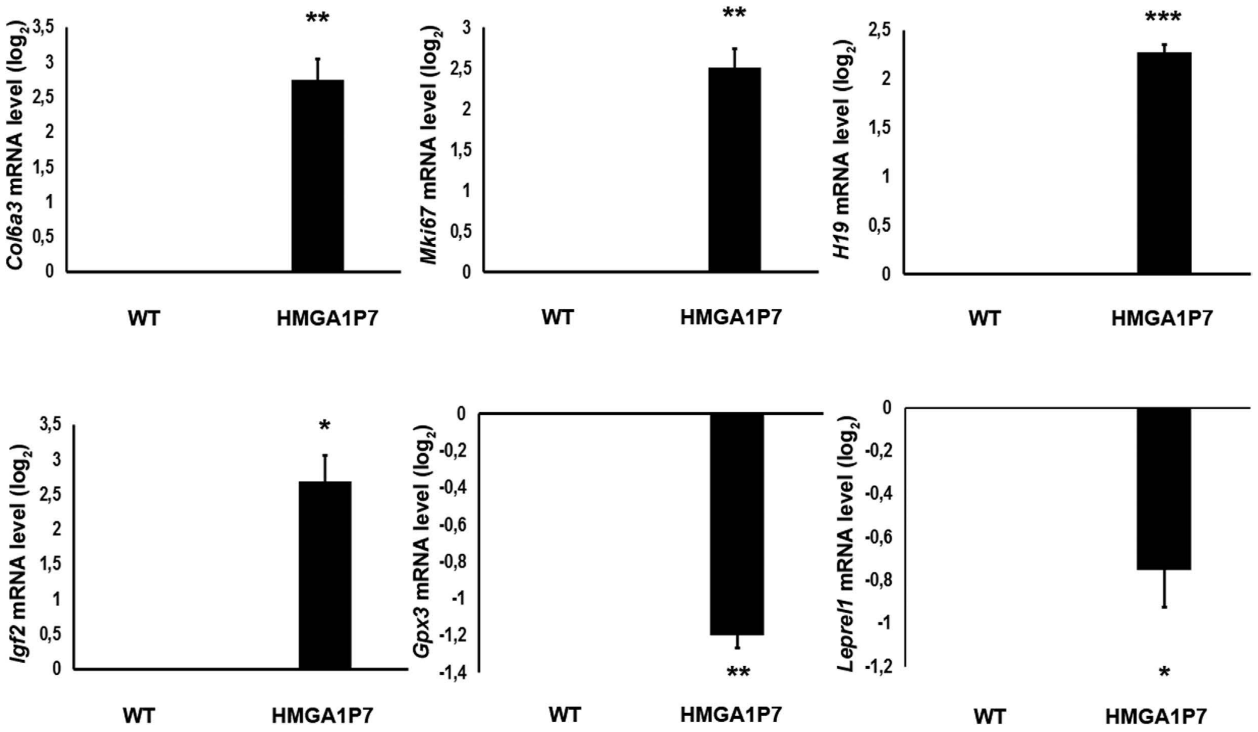



**Figure 16.** Validation of RNA-seq analyses on *HMGA1P7* MEFs. qRT-PCR analysis of selected deregulated genes from RNA-seq performed on WT and *HMGA1P7* transgenic MEFs. The results are reported as the mean of values. The error bars represent mean ± SD; *P < 0.05 **P < 0.01 ***P < 0.001 (t test).

Among the differentially expressed mRNAs found in *HMGA1P7* overexpressing MEFs, we focused our attention on *H19* and *Igf2* since they, other than to be involved in carcinogenesis, showed the highest fold change among the upregulated genes, and are also targeted by several miRNAs that are able to bind to the *HMGA1P7* mRNA. *H19* and *IGF2* genes are closely linked, showing highly similar patterns of gene expression, but they are reciprocally imprinted (Figure 17). In fact, *H19* is expressed solely from the maternally inherited chromosome, whereas *IGF2* expression is from the paternal chromosome. In particular, the noncoding *H19* has a critical role in genomic imprinting during cell growth and development[96]. The loss of imprinting results in misexpression of *H19* and was detected in many tumours including hepatocellular[97], bladder[98], gastric[99,100] and colon[101] cancer. *IGF2* codes for a mitogenic growth factor that is active in early development and has a critical role in embryonic and foetal growth[102]. Increased expression of *IGF2* is a common feature of both paediatric and adult malignancies[102], and mounting evidence implicates *IGF2* as a major factor contributing to oncogenesis[102-104].

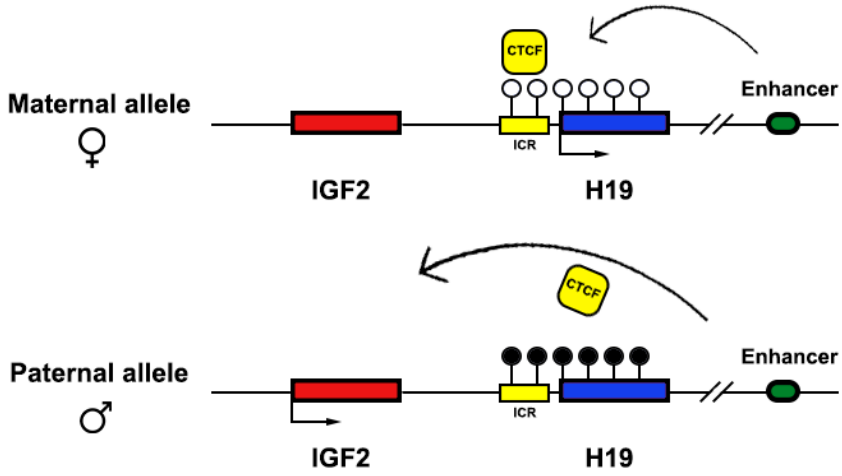



**Figure 17.** Regulation of the paternal and maternal expression in the *IGF2–H19* imprinted locus. The white lollipops represent unmethylated CpG islands and the black lollipops represent methylated CpG islands. The red box represents *IGF2* gene whereas the blue box represents *H19* gene. The yellow and green circles represent the CTCF insulator protein and a downstream enhancer, respectively.

Therefore, we confirmed the upregulation of both *Igf2* and *H19* in several *HMGA1P7* transgenic mice derived organs, such as heart, spleen and kidney (Figure 18a, 18b and 18c). Moreover, as expected from previous results, qRT-PCR and Western blot data showed an upregulation of both Igf2 and *H19* levels in *HMGA1P7* overexpressing NIH3T3 cell line (Figure 18d), supporting the hypothesis that *HMGA1P7* should act as ceRNA for *H19* and *Igf2*.

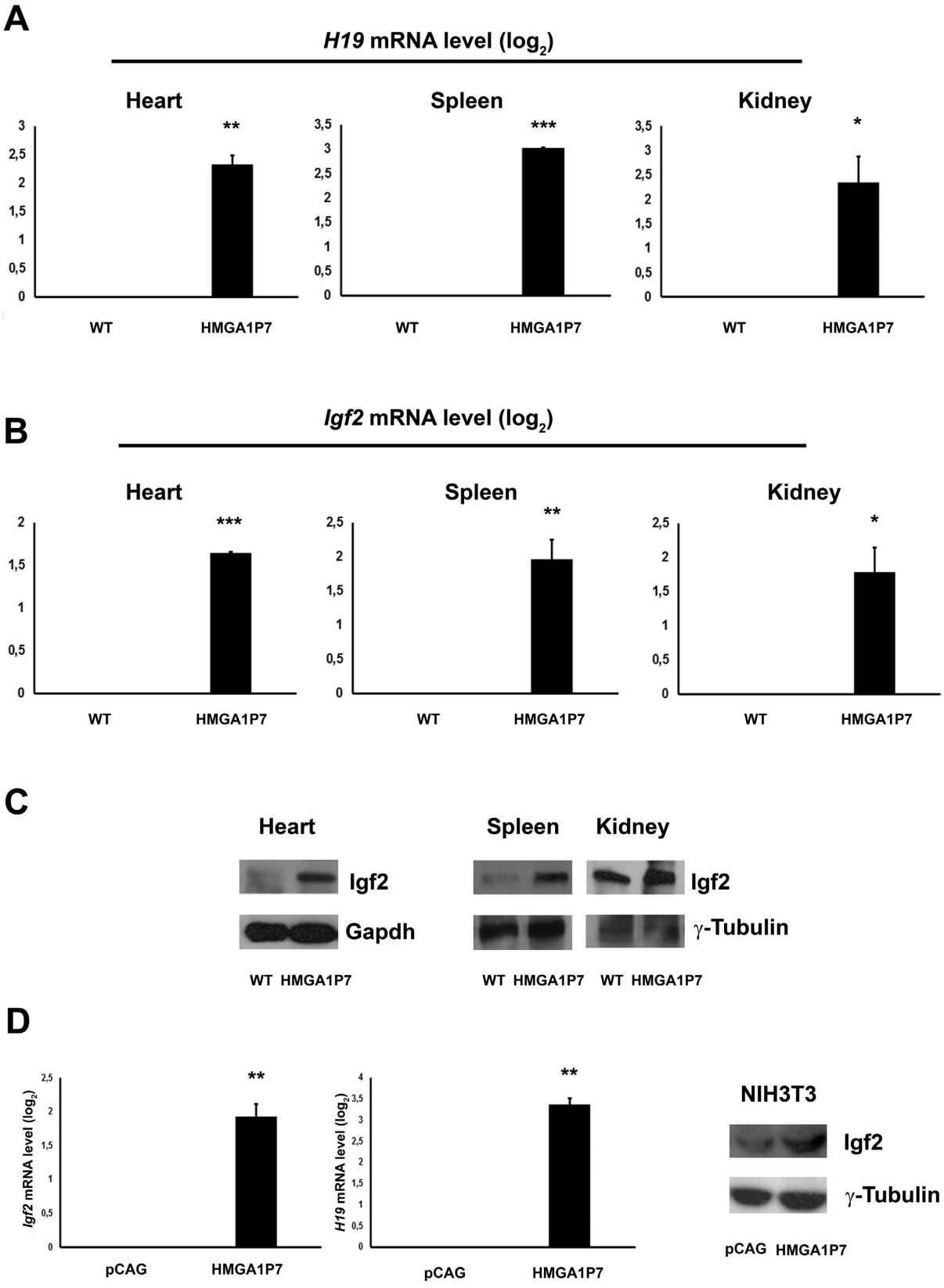



**Figure 18.** *H19* and *Igf2* are positively regulated by *HMGA1P7*. (A) qRT-PCR analysis of *H19* from hearts, spleens and kidneys of WT and *HMGA1P7* transgenic mice. (B) qRT-PCR analysis of *Igf2* from hearts, spleens and kidneys of WT and *HMGA1P7* transgenic mice. (C) Western blot analysis of Igf2 from heart, spleen and kidney of WT and *HMGA1P7* transgenic mice. (D) Left Panel, qRT-PCR analysis of *H19* and *Igf2* from control and *HMGA1P7* overexpressing NIH3T3 cells. Right Panel, Western blot analysis of Igf2 from control and *HMGA1P7* overexpressing NIH3T3 cells. The results are reported as the mean of values. The error bars represent mean ± SD; *P < 0.05 **P < 0.01 ***P < 0.001 (t test).

## 4.6 *HMGA1P7* is a sponge for *H19* and *Igf2* targeting miRNAs

To verify whether the effect of *HMGA1P7* pseudogene on *H19* and *Igf2* expression is dependent on sharing targeting-miRNAs, we assessed the capability of *HMGA1P7*-targeting miRNAs to bind to *H19* and *Igf2* mRNAs. To this aim, we transfected the above-mentioned *HMGA1P7*-targeting miRNAs (miR-15, miR-16, miR-214, miR-761) into NIH3T3 cells, analysing *H19* and *Igf2* mRNA levels. As reported in Figure 19a, after the transfection of *HMGA1P7*-targeting miRNAs we found a significant reduction of *H19* and *Igf2* levels. To establish whether the *HMGA1P7*-targeting miRNAs directly interacted with *Igf2* transcript, we cloned the *Igf2* 3'UTR downstream of the luciferase open reading frame. This reporter vector was transfected into NIH3T3 cell line together with miRNA precursors and a control non-targeting scrambled oligonucleotides. The luciferase signal was considerably lower after transfection with miR-15, miR-16, miR-214, miR-761 in comparison with the cells transfected with the scrambled oligonucleotide (Figure 19b).

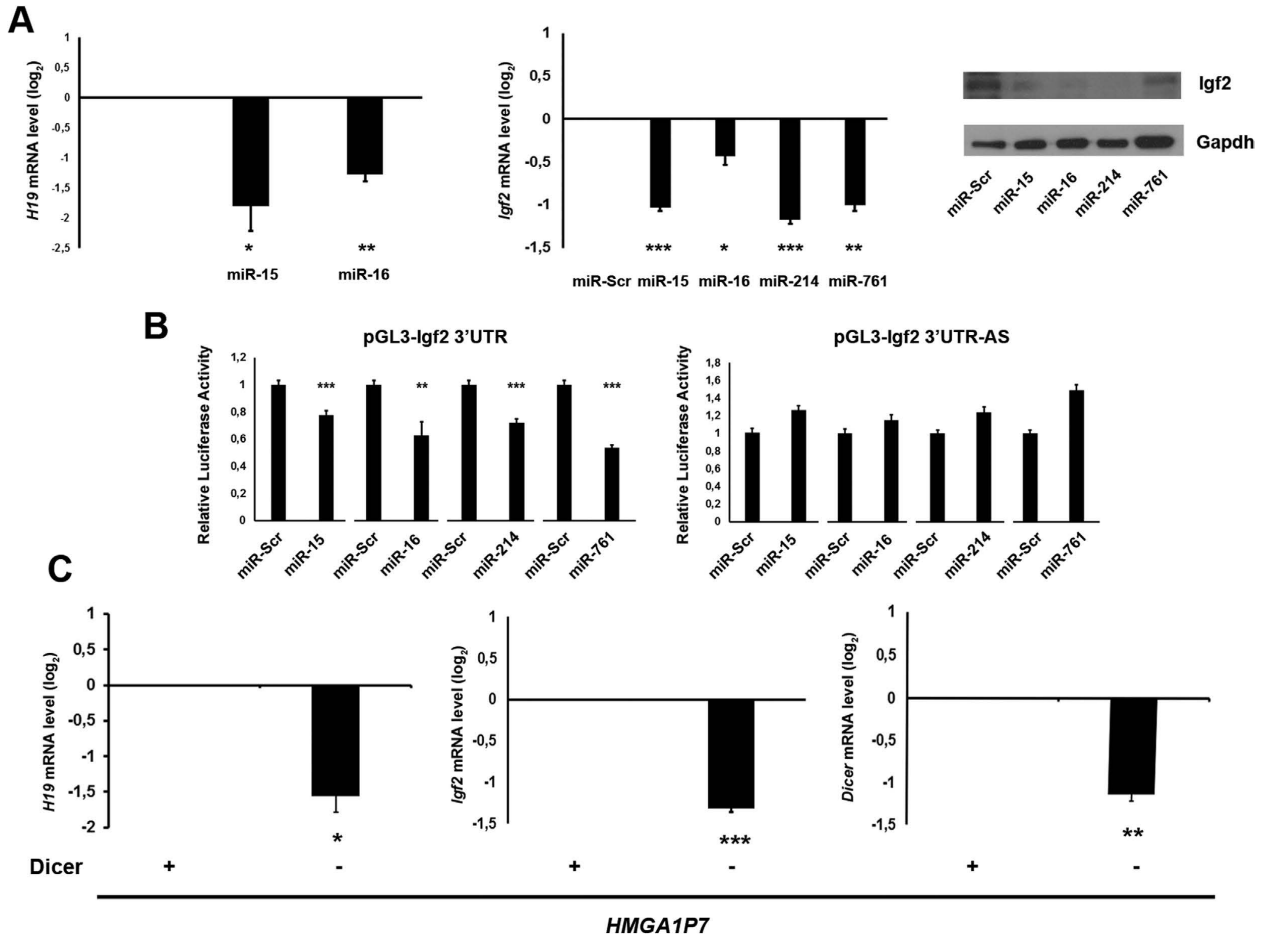



**Figure 19.** *H19* and *Igf2* are targeted by *HMGA1P7*-targeting miRNAs. (A) qRT-PCR analysis of *H19* (left Panel) and *Igf2* (middle Panel) mRNA from the NIH3T3 cells transfected with scrambled-oligonucleotide, miR-15, miR-16, miR-214 and miR-761. Right Panel, Western blot analysis of Igf2 in the same samples as in the middle panel. (B) *Igf2* was cloned into the pGL3 control vector. Relative luciferase activity in HEK293 cells transiently transfected with miR-15, miR-16, miR-214, miR-761 and a control scrambled oligonucleotide. (C) *H19* and *Igf2* mRNA levels 24h after the transfection of *HMGA1P7* in scrambled oligonucleotide or siRNA-Dicer NIH3T3 transfected cells.

Finally, the overexpression of *H19* and *Igf2* exerted by *HMGA1P7* upregulation was depleted in Dicer-knockdown cells, suggesting that *HMGA1P7*, *H19* and *Igf2* follow the same miRNA-mediated post-transcriptional regulation (Figure 19c). These data are consistent with the hypothesis that *HMGA1P7* requires mature miRNAs to regulate *H19* and *Igf2* levels.

## 4.7 *HMGA1P7, H19* and *IGF2* expression positively correlates in human breast cancer

Then, we investigated whether *HMGA1P7* works as ceRNA through, or partially through *H19*, *IGF2* and *HMGA1* in breast cancer human cells. As expected, we found the upregulation of *H19*, *IGF2* and *HMGA1* following *HMGA1P7* overexpression in MCF7 cells (Figure 20a). Moreover, MCF7-*HMGA1P7* cells grew faster than the control transfected cells as consequence of *HMGA1P7* ceRNA pathway activation (Figure 20b).

To confirm the ceRNA role of *HMGA1P7* and the regulation on *H19* and *IGF2* expression levels also in human cancer, we evaluated the expression these genes in a panel of breast carcinoma samples by qRT-PCR, since *H19* and *IGF2* have been reported to be overexpressed in this type of tumor[105,106]. As reported in Figure 20c, *HMGA1P7* was overexpressed in most of carcinoma samples as well as *H19* and *IGF2*. Moreover, the direct correlation between *HMGA1P7* and *H19* expression (Spearman $r = 0{,}8656$; $p < 0{,}001$) and between *HMGA1P7* and *IGF2* expression (Spearman $r = 0{,}7958$; $p < 0{,}001$) underlines that these genes are co-regulated (Figure 20d). Altogether, these results strongly support the idea that *HMGA1P7* could act as ceRNAs in human breast cancer and represent a novel potential mechanism accounting for *H19* and *IGF2* upregulation in these tumours.

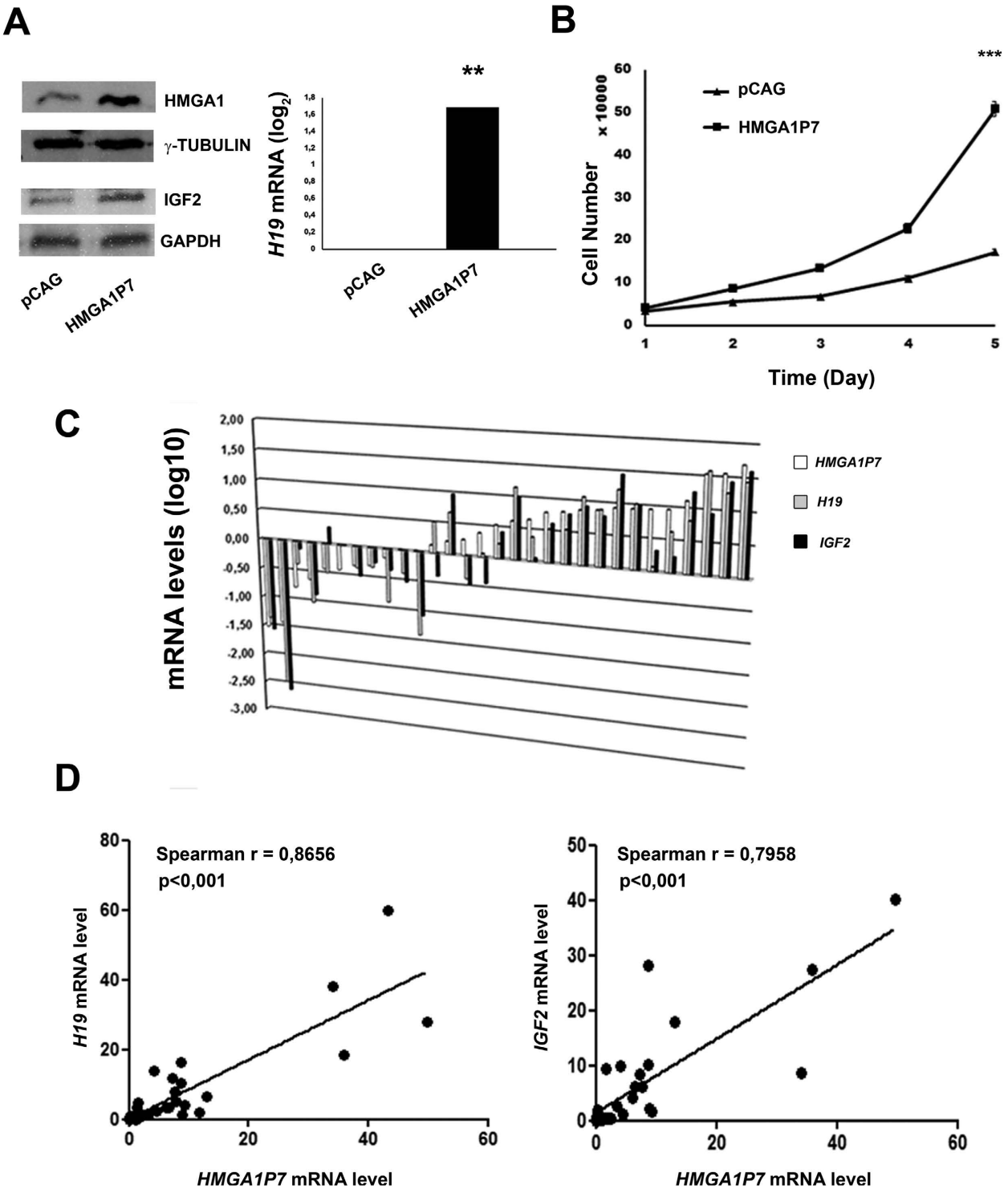



**Figure 20.** *H19* and *IGF2* expression positively correlates with *HMGA1P7* in breast cancer. (A) Left Panel, Western blot analysis of HMGA1 and IGF2 protein levels in control and *HMGA1P7*-overexpressing MCF7 cells. Right Panel, qRT-PCR analysis of *H19* expression of control and *HMGA1P7*-overexpressing MCF7 cells. (B) MCF7 cell proliferation of control and *HMGA1P7*-overexpressing cells (C) qRT-PCR analysis in tumour and normal breast tissues. The fold change indicates the relative change in expression levels between tumour samples and normal samples, assuming that the value of normal sample is equal to 1. (D) Correlation analysis of *HMGA1P7 vs. H19* and *HMGA1P7 vs. IGF2* are shown. The Spearman's rank correlation coefficient is shown.

## 4.8 *HMGA1Ps* act as oncogenes also *in vivo*

Remarkably, preliminary results showed that about 50% of *HMGA1Ps* mice killed at 20 months of age displayed a lymphoid pathology. In fact, the *HMGA1Ps* transgenic mice spleens were characterized by splenomegaly and by a lymphoid hyperplasia (Figure 21a). In particular, by histological analysis, we found the expansion and merging of adjacent white pulp areas with loss of normal structures and germinal centers. In addition, a higher magnification showed a monotonous lymphoid population composed of fairly large rounded cells with scant cytoplasm and round nuclei with finely dispersed chromatin and inconspicuous nucleoli (Figure 21b). Immunohistochemical analysis showed that the examined samples were all positive to CD79, a marker of B-cell population. Moreover, organ-infiltrated cells were detectable in the liver, kidneys and lungs (Figure 21c). Furthermore, FACScan analysis of lymphocytes isolated from pathological spleens using the CD19, CD3, and NK anti-mouse antibodies revealed a predominantly $CD19^+$ population (Figure 21d), confirming the neoplastic expansion of B-cells. Although further analyses are required, these preliminary results underlined the oncogenic role of *HMGA1Ps* also *in vivo*, indicating that ceRNA-mediated microRNA sequestration may contribute to the development of cancer.

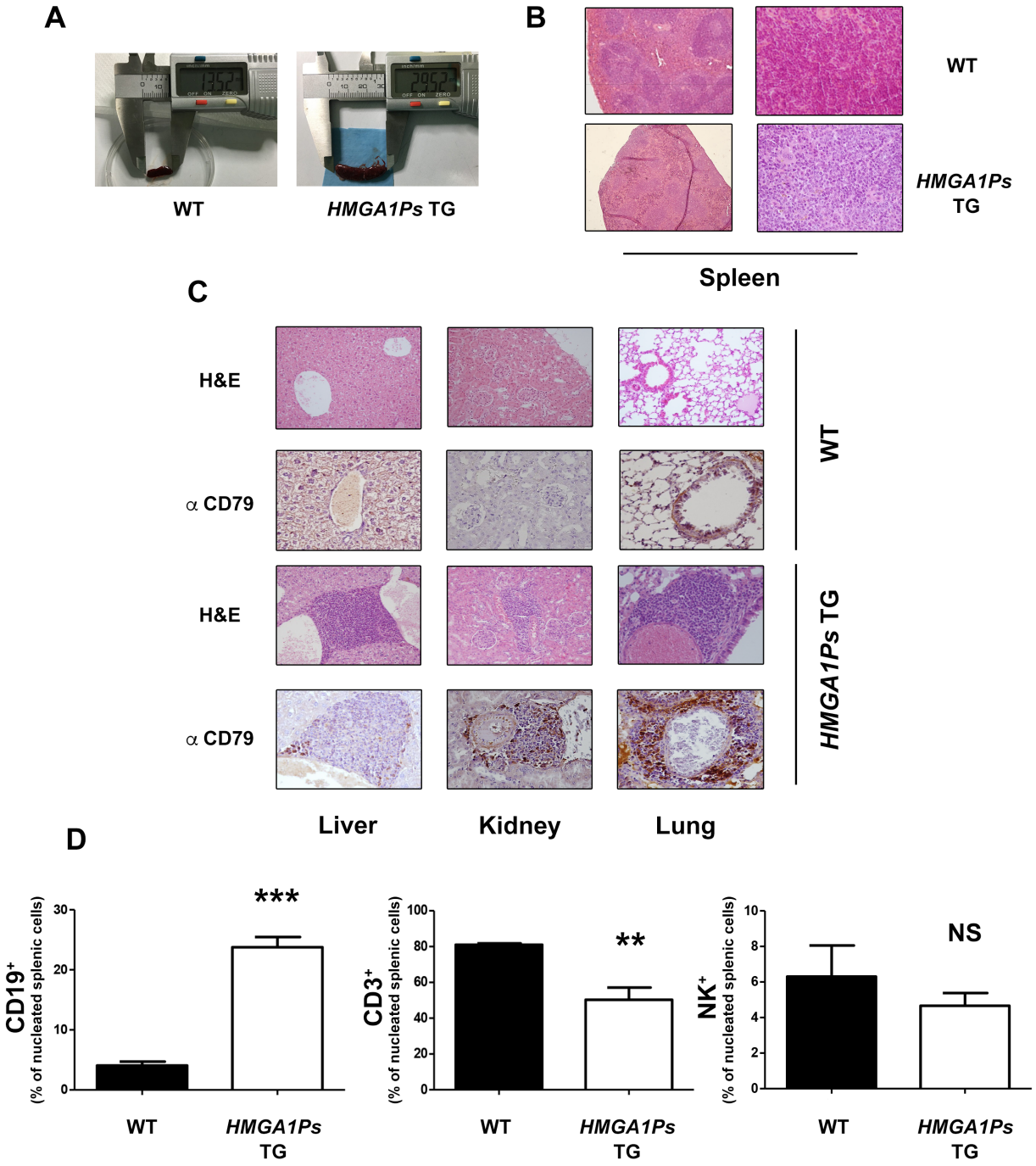



**Figure 21.** *HMGA1Ps* expression *in vivo* results in lymphoid malignancy. (A) Size of WT and *HMGA1Ps* transgenic mouse spleens. (B) Histological analysis of WT and *HMGA1Ps* transgenic mouse spleens. Haematoxylin and Eosin staining. (C) Histological and immunohistochemical analyses of WT and *HMGA1Ps* transgenic mouse livers, kidneys and lungs. (D) FACScan analysis of splenic cells isolated from WT and *HMGA1Ps* transgenic mice using CD19, CD3, and NK1.1 anti-mouse antibodies.

# 5. DISCUSSION

HMGA protein overexpression plays a critical role in carcinogenesis[10]. Even though recent results evidence that HMGA expression is deeply regulated by the action of several miRNAs[39-41,79], their upregulation in human cancer has not been completely elucidated. Recent findings ascribe an important role in the regulation of gene expression to pseudogenes. Indeed, the mammalian genome contains about 20.000 pseudogenes[107], more than those present in other organisms. The biological meaning of pseudogenes was completely unknown until few years ago, but very recent studies have reported their pivotal role in controlling gene expression mainly functioning as sponge for miRNAs, since the presence of the same miRNA responsive elements in gene and in the pseudogene regions could inhibit the miRNA negative regulation of their protein-coding target gene. Accordingly, recent results demonstrate that some pseudogenes are able to modulate the expression of oncogenes or tumour suppressor genes suggesting a role of pseudogenes in carcinogenesis[108,109].

Then, looking for the presence of *HMGA1* pseudogenes in the human genome by bioinformatic analysis, we have identified eight *HMGA1* pseudogenes. From the examination of *HMGA1* pseudogene sequences it emerges that they can regulate *HMGA1* expression and function. In particular, three *HMGA1* pseudogenes (*HMGA1P1*, *HMGA1P2* and *HMGA1P3*), whose translationability seems not be altered by mutations, could represent a sort of competitor proteins for HMGA1 wild-type with different post-translational modifications, deregulating HMGA1 properties in chromatin remodelling and protein-protein interactions. Furthermore, Chiefari *et al.* reported the function of another *HMGA1* pseudogene (*HMGA1-p*), which is able to compete with *HMGA1* 3'UTR for a critical RNA stability factor, αCP1[80]. Interestingly, *HMGA1-p* was found upregulated in diabetic patients then causing a significant destabilization of *HMGA1* mRNA with consequent loss of *INSR* expression, which is regulated by HMGA1, and then generating the insulin resistance phenotype[80]. However, we focused on the role of two pseudogenes, *HMGA1P6* and *HMGA1P7*, which produced a non-translational mRNAs and can act on the stability of *HMGA1* transcript by shielding it from miRNAs able to target this gene.

Therefore, we investigated whether *HMGA1* pseudogenes influence HMGA1 protein levels, and, consequently play a key role in cancer progression. Firstly, we demonstrated that the overexpression of *HMGA1Ps* was able to increase both HMGA1 mRNA and protein levels by protecting *HMGA1* transcripts from the negative regulation of several *HMGA1* targeting miRNAs, namely, miR-15, miR-16, miR-214, miR-761[39-41,79]. Consistent with these data, our functional analyses showed that *HMGA1Ps* overexpressing cell lines grew faster, exhibiting a lower number of G1-phase cells and a higher S-phase cells compared with the backbone vector-

transfected ones. Therefore, *HMGA1P6* and *HMGA1P7* influences cell cycle progression through their ability to increase HMGA1 protein levels, which is involved in the regulation of the G1-S transition phase of the cell cycle[79]. Moreover, *HMGA1P6* and *HMGA1P7* overexpression increased cell migration, and invasiveness, and decreased the apoptotic rate.

These results encouraged us to verify a possible *HMGA1Ps* overexpression in human cancer. Interestingly, *HMGA1P6* and *HMGA1P7* were abundantly expressed in anaplastic thyroid carcinoma, which are very aggressive and express very high HMGA1 protein levels. Conversely, *HMGA1P6* and *HMGA1P7* expression was low in papillary thyroid carcinoma, which are well differentiated and poorly aggressive, and express moderate HMGA1 protein levels. Similar results were obtained analysing *HMGA1Ps* expression levels in human ovarian carcinomas. Moreover, we found a direct correlation between *HMGA1* and *HMGA1Ps* in a series of human pituitary tumours, in particular in the somatotroph ones. Recent data confirmed the overexpression of *HMGA1Ps* also in endometrioid endometrial carcinomas (EEC). In particular, *HMGA1P6* and *HMGA1P7* expression was significant associated with tumor staging increasing and positively correlated with that of *HMGA1* levels. Therefore, the evaluation of *HMGA1Ps* levels should be used as EEC diagnostic and prognostic marker. However, in oesophageal squamous cell carcinoma there was not a deregulation of *HMGA1* pseudogenes. It is likely that the cellular context, the abundance of miRNAs and other ceRNA interactors should influence *HMGA1* pseudogenes levels, promoting or inhibiting their oncogenic activity.

Intriguingly, *HMGA1* pseudogenes seem to affect cancer progression also by binding to the same miRNAs that can target other genes involved in carcinogenesis process. Indeed, we found in *HMGA1P6*, *HMGA1P7* genes and also in *HMGA1* UTR regions several sequences that are potential target sites for few miRNAs able to target some cancer-related genes such as *HMGA2*, *EZH2*, *VEGF*. These genes were effectively found upregulated in *HMGA1Ps* overexpressing cells with respect to the control ones. This has important consequences. In fact, it means that *HMGA1*-pseudogene expression can exert the upregulation of HMGA2 and EZH2 protein levels then deeply contributing to cancer progression. Such a mechanism is likely to occur in anaplastic thyroid carcinoma where the overexpression of *EZH2* is a molecular feature[85].

The generation of transgenic mice overexpressing *HMGA1P6* or *HMGA1P7* has supported their oncogenic role. Indeed, MEFs deriving from *HMGA1P6* and *HMGA1P7* overexpressing mice showed a higher growth rate and lower susceptibility to senescence with respect to the WT counterpart.

Then, to identify other additional mRNAs deregulated in *HMGA1P7* transgenic MEFs, we performed an RNA-seq analysis. We found that the expression of several genes were influenced

by *HMGA1P7* including also genes involved in cancer progression such as *Col6a3*, *Mki67*, *H19*, *Igf2, Gpx3* and *Leprel1*[90-95,105,106]. Indeed, oncomine analyses and tissue-microarray immunohistochemistry showed overexpression of COL6A3 in colorectal carcinomas that was significantly and directly correlated with Dukes stage, T stage, stage, recurrence and smoking status and then with a poor prognosis[30]. The MKi-67 protein (also known as Ki67) is a cellular marker for proliferation. Ki-67 protein is expressed during all active phases of the cell cycle (G1, S, G2, and mitosis), but is absent from resting cells (G0)[92]. *GPX3* gene codes for the Glutathione peroxidase 3, also known as plasma glutathione peroxidase (GPx-P), the variations in activity of GPX1, GPX2, and GPX3 isoforms may be associated with the development of cancers, for example, prostate cancer or even colon cancer[110]. Leprecan-like 1 is a potential tumour suppressor gene since it has been demonstrated to be downregulated in the hepatocarcinoma tissues and its overexpression inhibits cancer cell proliferation and colony formation through regulation of the cell cycle by downregulation of cyclins[111].

Deregulation of *H19* noncoding gene was found in many tumours such as hepatocellular and bladder cancer[97,98]. Finally, *IGF2* overexpression is widely reported in paediatric and adult tumors[102], and several studies involve *IGF2* as a key factor leading to cancerogenesis[102-104].

Among the most deregulated genes, we selected and studied *H19* non-coding gene and *Igf2* that share several miRNAs with *HMGA1P7*. Therefore, we demonstrate that *HMGA1P7* overexpression increases *H19* and *Igf2* levels inhibiting their mRNA suppression by miRNAs that target *HMGA1P7* gene, namely, miR-15, miR-16, miR-214, and miR-761. Noteworthy, we showed that expression of *HMGA1P7* significantly correlates with *H19* and *IGF2* levels in human breast cancer, suggesting the upregulation of *HMGA1P7* may increase *H19* and *IGF2* expression by a ceRNA mechanism then contributing to cancer progression.

Furthermore, the oncogenic role of *HMGA1Ps* is also supported by preliminary data that report the development of malignant haematological neoplasias, diagnosed as B-cell lymphomas, in *HMGA1Ps* transgenic mice. In fact, pathological mice showed splenomegaly with the loss of the normal structures. Moreover, histological analysis revealed a splenic monotonous population. We determined the immunophenotype by flow citometry. The cells expressing CD19 were increased in spleens, suggesting an expansion of B-cell population. Accordingly, by immunohistochemistry tumour cells stained positive for CD79. These results seem to confirm the oncogenic role of *HMGA1* pseudogenes also *in vivo*. In fact, besides the lines of evidence obtained by *in vitro* experiments, formal proof of the causal link existing between pseudogenes and the pathogenesis of human cancer has to be confirmed by genetically engineered mouse models.

Recently, Karreth *et al.* reported the consequences of *Braf-rs1* overexpression *in vivo*[69]. The effects of the induction of *Braf-rs1* expression were dramatic: the mice were characterized by splenomegaly and enlarged lymph nodes, all symptoms that an in-depth flow cytometric analysis revealed to be the result of an aggressive form of diffuse large B-cell lymphoma. *Braf-rs1*-induced lymphomas were further characterized. Indeed, histological analyses proved that, consistent with the ceRNA hypothesis, *Braf-rs1*-driven tumours are characterized by increased Braf and pErk levels[69]. Indeed, these genetically engineered mouse models represented a proof of principle that aberrantly expressed pseudogenes are necessary and sufficient to cause cancer by working as ceRNAs for their oncogenic parental genes.

However, although *HMGA1Ps* transgenic mice develop B-cell malignancy, further investigations are required to unveil the molecular mechanisms underlying the development of this disease. Preliminary data showed that *HMGA1Ps* upregulation does not exert Hmga1 overexpression in splenic cells. Thus, we investigated other *HMGA1Ps* ceRNA interactors involved in cancer progression. Interestingly, we found the upregulation of *Ezh2* in pathological spleens consistently with previous our data demonstrating the interaction between *HMGA1Ps* and *Ezh2* through a ceRNA mechanism in *in vitro* experiments. Interestingly, it has been previously shown that engineered mouse models for *EZH2* develop B-cell lymphoma[112,113], suggesting that EZH2 upregulation might should drive the *HMGA1Ps* transgenic mice malignancy[114,115]. However, further analyses are needed to confirm this new ceRNA pathway.

Then, the data reported here confirm the oncogenic role of the *HMGA1* pseudogenes that is exerted by the increased expression through a ceRNA mechanism of *HMGA1* and other cancer-related genes. Future studies are, however, needed to characterize other genes regulated by the *HMGA1* pseudogenes and thereby better define the mechanisms by which they can contribute to cancer progression.

# REFERENCES


1. Johnson, K. R., Lehn, D. A. & Reeves, R. Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y. *Mol. Cell Biol.* **9**, 2114–2123 (1989).
2. Nagpal, S. *et al.* Retinoid-dependent recruitment of a histone H1 displacement activity by retinoic acid receptor. *J. Biol. Chem.* **6**, 22563–22568 (1999).
3. Reeves, R. & Nissen, M. S. The AT DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.* **265**, 8573–8582 (1990).
4. Thanos, D. & Maniatis, T. The high mobility group protein HMG I(Y) is required for NF-κB-dependent virus induction of the human IFN-β gene. *Cell* **27**, 777–789 (1992).
5. Thanos, D., Du, W. & Maniatis, T. The high mobility group protein HMG I(Y) is an essential structural component of a virus-inducible enhancer complex. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 73–81 (1993).
6. Foti, D. *et al.* Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. *Nat Med.* **11**, 765-773 (2005).
7. Zhou, X., Benson, K. S., Ashar, H. R. & Chada, K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* **376**, 771-774 (1995).
8. Chieffi, P. *et al*. HMGA1 and HMGA2 protein expression in mouse spermatogenesis. *Oncogene* **21**, 3544-3550 (2002).
9. Di Agostino, S. *et al.* Phosphorilation of high-mobility group protein A2 by Nek2 kinase during the first meiotic division in mouse spermatocytes. *Mol Biol Cell* **15**, 1224-1232 (2004).
10. Fusco, A. & Fedele, M. Roles of HMGA proteins in cancer. *Nat Rev Cancer* **7**, 899-910 (2007).
11. Cleynen, I. *et al.* Transcriptional control of the human high mobility group A1 gene: basic and oncogenic Ras-regulated expression. *Cancer Res.* **67**, 4620-4629 (2007).
12. Berlingieri, M. T. *et al.* Inhibition of HMGI-C protein synthesis suppresses retrovirally induced neoplastic transformation of rat thyroid cells. *Mol. Cell Biol.* **15**, 1545–1553 (1995).
13. Berlingieri, M. T., Pierantoni, G. M., Giancotti, V., Santoro, M. & Fusco, A. Thyroid cell transformation requires the expression of the HMGA1 proteins. *Oncogene* **21**, 2971–2980 (2002).

14. Scala, S., Portella, G., Fedele, M., Chiappetta, G. & Fusco, A. Adenovirus-mediated suppression of HMGI(Y) protein synthesis as potential therapy of human malignant neoplasias. *Proc. Natl Acad. Sci. USA*. **97**, 4256–4261 (2000).
15. Wood, L. J., Maher, J. F., Bunton, T. E. & Resar, L. M. The oncogenic properties of the HMG-I gene family. *Cancer Res.* **60**, 4256–4261 (2000).
16. Reeves, R., Edberg, D. D. & Li, Y. Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. *Mol. Cell Biol.* **21**, 575–594 (2001).
17. Baldassarre, G. *et al.* Onset of natural killer cell lymphomas in transgenic mice carrying a truncated HMGI-C gene by the chronic stimulation of the IL-2 and IL-15 pathway. *Proc. Natl Acad. Sci. USA*. **98**, 7970–7975 (2001).
18. Fedele, M. *et al.* Overexpression of the HMGA2 gene in transgenic mice leads to the onset of pituitary adenomas. *Oncogene* **21,** 3190-3198 (2002).
19. Fedele, M. *et al.* Transgenic mice overexpressing the wild-type form of the HMGA1 gene develop mixed growth hormone/prolactin cell pituitary adenomas and natural killer lymphomas. *Oncogene* **24**, 3427-3435 (2005).
20. Angel, P. & Karin, M. The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. *Biochim Biophys Acta* **1072**, 129-157 (1991).
21. Mantovani, F. *et al.* NF-κB mediated transcriptional activation is enhanced by the architectural factor HMGI-C. *Nucleic Acids Res.* **26**, 1433–1439 (1998).
22. Subramanian, D. & Griffith, J. D. Interactions between p53, hMSH2-hMSH6 and HMGAI(Y) on Holliday junctions and bulged bases. *Nucleic Acids Res.* **30**, 2427-2434 (2002).
23. Pierantoni, GM. *et al.* High mobility group A1 (HMGA1) proteins interact with p53 and inhibit its apoptotic activity. *Cell Death Differ.* **13**, 1554-1563 (2006).
24. Esposito, F. *et al.* High-mobility group A1 proteins regulate p53-mediated transcription of Bcl-2 gene. *Cancer Res.* **70**, 5379-5388 (2010).
25. Pegoraro, S. *et al.* HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. *Oncotarget* **4**, 1293-1308 (2013).
26. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83-86 (2007).
27. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23,** 4051-4060 (2004).

28. Lee, Y. *et al*. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425,** 415-419 (2003).
29. Yi, R., Qin, Y., Macara, I. G. & Cullen B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17,** 3011-3016 (2003).
30. Vasudevan, S., Tong, Y. & Steitz J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318**, 1931-1934 (2007).
31. Rajewsky, N. L(ou)sy miRNA targets? *Nat. Struct. Mol. Biol.* **13**, 754-755 (2006).
32. Berezikov, E. *et al.* Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**, 21-24 (2005).
33. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res.* **32**, 109-111 (2004).
34. Borrmann, L., Wilkening, S. & Bullerdieck J. The expression of HMGA genes is regulated by their 3'UTR. *Oncogene* **20**, 4537-4541 (2001).
35. Lee, Y. S. & Dutta, A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.* **21**, 1025-1030 (2007).
36. Mayr, C., Hemann, M. T. & Bartel, D. P. Disrupting the pairing between *let-7* and *Hmga2* enhances oncogenic transformation. *Science* **315**, 1576-1579 (2007).
37. Hebert, C., Norris, K., Scheper, M. A., Nikitakis, N. & Sauk, J. J. High Mobility Group A2 is a target for miRNA-98 in head and neck squamous cell carcinoma. *Molecular Cancer* **6**, 5-16 (2007).
38. Wang, T. *et al.* A micro-RNA signature associated with race, tumor size, and target gene activity in human uterine leiomyomas. *Genes Chromosomes Cancer* **46**, 336-347 (2007).
39. Palmieri, D. *et al.* Downregulation of HMGA-targeting microRNAs has a critical role in human pituitary tumorigenesis. *Oncogene* **31**, 3857-3865 (2012).
40. Li, W., Yuan, Y., Huang, L., Qiao, M. & Zhang, Y. Metformin alters the expression profiles of microRNAs in human pancreatic cancer cells. *Diabetes Res Clin Pract*. **96**, 187– 195 (2012).
41. Sehic, A., Risnes, S., Khuu, C., Khan, Q. E. & Osmundsen, H. Effects of in vivo transfection with anti-miR-214 on gene expression in murine molar tooth germ. *Physiol Genomics* **43**, 488–498 **(**2011).
42. Muro, E. M., Mah, N. & Andrade-Navarro, M. A. Functional evidence of post-transcriptional regulation by pseudogenes. *Biochimie* **93**, 1916-1921 (2011).
43. Williams, T. K., Yeo, C. J. & Brody, J. Does this band make sense? Limits to expression based cancer studies. *Cancer Lett.* **271**, 81-84 (2008).

44. Brent, M. R. Steady progress and recent breakthroughs in the accuracy of automated genome annotation. *Nat. Rev. Genet.* **9**, 62–73 (2008).
45. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* **10**, 637-643 (2009).
46. Gutschner, T. & Diederichs, S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* **9**, 703-719 (2012).
47. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12**, 861–874 (2011).
48. Carninci, P., Yasuda, J. & Hayashizaki, Y. Multifaceted mammalian transcriptome. *Curr. Opin. Cell Biol.* **20**, 274-280 (2008).
49. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904-914 (2011).
50. D'Errico, I., Gadaleta, G. & Saccone, C. Pseudogenes in metazoa: Origin and features. *Brief. Funct. Genomics Proteomics* **3**, 157–167 (2004).
51. Baltimore, D. Our Genome Unveiled. *Nature* **409**, 814-816 (2001).
52. Costa, F. F. Non-coding RNAs, epigenetics and complexity. *Gene* **410**, 9-17 (2008).
53. Mattick, J. S. The genetic signature of noncoding RNAs. *PLoS Genet.* **5**, e1000459 (2009).
54. Pink, R. C. *et al.* Pseudogenes: Pseudo-functional or key regulators in health and disease? *RNA* **17**, 792-798 (2011).
55. Balakirev, E. S. & Ayala, F. J. Pseudogenes: Are they "junk" or functional DNA? *Annu. Rev. Genet.* **37**, 123-151 (2003).
56. Khachane, A. N. & Harrison, P. M. Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics* **10**, 435 (2009).
57. Chang, L. Y. & Slightom, J. L. Isolation and nucleotide sequence analysis of the beta-type globin pseudogene from human, gorilla and chimpanzee. *J. Mol. Biol.* **180**, 767-784 (1984).
58. Kalyana-Sundaram, S. *et al.* Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* **149**, 1622-1634 (2012).
59. Lin, M. *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS ONE* **6**, e23356 (2011).
60. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* **12**, 99-110 (2011).

61. Marques, A. C., Tan, J. & Ponting, C. P. Wrangling for microRNAs provokes much crosstalk. *Genome Biol.* **12**, 132 (2011).
62. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033-1038 (2010).
63. Lee, D. Y. *et al.* A 3 -untranslated region (3'UTR) induces organ adhesion by regulating miR-199a* functions. *PLoS ONE* **4**, e4527 (2009).
64. Lee, D. Y. *et al.* Expression of versican 3 -untranslated region modulates endogenous microRNA functions. *PLoS ONE* **5**, e13599 (2010).
65. Franco-Zorrilla, J. M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**, 1033-1037 (2007).
66. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358-369 (2011).
67. Rastinejad, F. *et al.* Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation. *Cell* **72**, 903-917 (1993).
68. Gu, S., Jin, L., Zhang, F., Sarnow, P. & Kay, M. A. Biological basis for restriction of microRNA targets to the 3 untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.* **16**, 144-150 (2009).
69. Karreth, F. A. *et al.* The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell* **161**, 319-332 (2015).
70. Zhang, Q. & Wang, Y. High mobility group proteins and their post-translational modifications. *Biochim Biophys Acta* **1784**, 1159-1166 (2008).
71. Sgarra, R. *et al.* During apoptosis of tumor cells HMGA1a protein undergoes methylation: Identification of the modification site by mass spectrometry. *Biochemistry* **42**, 3575-3585 (2003).
72. Nissen, M. S., Langan, T. A. & Reeves, R. Phosphorylation by cdc2 kinase modulates DNA binding activity of high mobility group I nonhistone chromatin protein. *J Biol Chem.* **266**, 19945-19952 (1991).
73. Sgarra, R. *et al*. The AT- hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6. *J Biol Chem.* **281**, 3764-3772 (2006).
74. Reeves, R. Molecular biology of HMGA proteins: Hubs of nuclear function. *Gene* **277**, 63-81 (2001).

75. Currie, R. A. Functional interaction between the DNA binding subunit trimerization domain of NF-Y and the high mobility group protein HMG-I(Y). *J Biol Chem.* **272**, 30880-30888 (1997).
76. Sgarra, R. *et al*. Macroscopic differences in HMGA oncoproteins post-translational modifications: C-terminal phosphorylation of HMGA2 affects its DNA binding properties. *J Proteome Res.* **8**, 2978-2989 (2009).
77. Pierantoni, G. M. *et al*. A truncated HMGA1 gene induces proliferation of the 3T3-L1 preadipocytic cells: a model of human lipomas. *Carcinogenesis* **24**, 1861-1869 (2003).
78. Rogalla, P., Blank, C., Helbig, R., Wosniok, W. & Bullerdiek, J. Significant correlation between the breakpoints of rare clonal aberrations in benign solid tumors and the assignment of HMGIY retropseudogenes. *Cancer Genet Cytogenet.* **130**, 51-56 (2001).
79. D'Angelo, D. *et al*. Altered microRNA expression profile in human pituitary GH adenomas: Down-regulation of miRNA targeting HMGA1, HMGA2, and E2F1. *J Clin Endocrinol Metab.* **97**, E1128-1138 (2012).
80. Chiefari, E. *et al*. Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. *Nat Commun.* **1**, 40 (2010).
81. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
82. Esposito, F. *et al.* Aurora B expression in post-puberal testicular germ cell tumours. *J Cell Physiol*. **221**, 435–439 (2009).
83. Esposito, F. *et al.* The high-mobility group A1-estrogen receptor β nuclear interaction is impaired in human testicular seminomas. *J Cell Physiol*. **227**, 3749–3755 (2012).
84. Esposito, F. *et al.* Down-regulation of the miR-25 and miR-30d contributes to the development of anaplastic thyroid carcinoma targeting the polycomb protein EZH2. *J Clin Endocrinol Metab.* **97**, 710-718 (2012).
85. Borbone, E. *et al*. Enhancer of zeste homolog 2 overexpression has a role in the development of anaplastic thyroid carcinomas. *J Clin Endocrinol Metab.* **96**, 1029–1038 (2011).
86. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288–4297 (2012).

87. Pierantoni, G. M. *et al.* High-mobility group A2 gene expression is frequently induced in non-functioning pituitary adenomas (NFPAs), even in the absence of chromosome 12 polysomy. *Endocr Relat Cancer* **12**, 867-874 (2005).
88. Fedele, M., Pierantoni, G. M., Visone, R. & Fusco, A. Critical role of the HMGA2 gene in pituitary adenomas. *Cell Cycle* **5**, 2045-2048 (2006).
89. Fedele, M. *et al.* HMGA2 induces pituitary tumorigenesis by enhancing E2F1 activity. *Cancer Cell* **9**, 459-471 (2006).
90. Qiao, J. *et al.* Stroma derived COL6A3 is a potential prognosis marker of colorectal carcinoma revealed by quantitative proteomics. *Oncotarget* **6**, 29929–29946 (2015).
91. Xie, X., Liu, X., Zhang, Q. & Yu, J. Overexpression of collagen VI α3 in gastric cancer. *Oncol Lett*. **7**, 1537–1543 (2014).
92. Li, L. T., Jiang, G., Chen, Q. & Zheng, J. N. Ki67 is a promising molecular target in the diagnosis of cancer. *Mol Med Rep*. **11**, 1566–1572 (2015).
93. Zhou, J. D. *et al.* GPX3 hypermethylation serves as an independent prognostic biomarker in non-M3 acute myeloid leukemia. *Am J Cancer Res*. **5**, 2047–2055 (2015).
94. Qi, X. *et al.* Clinical significance and therapeutic value of glutathione peroxidase 3 (GPx3) in hepatocellular carcinoma. *Oncotarget* **5**, 11103–11120 (2014).
95. Shah, R. *et al.* e prolyl 3-hydroxylases P3H2 and P3H3 are novel targets for epigenetic silencing in breast cancer. *Br J Cancer* **100**, 1687–1696 (2009).
96. Gabory, A., Jammes, H. & Dandolo, L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays* **32**, 473–480 (2010).
97. Imad, J. *et al.* The H19 non-coding RNA is essential for human tumor growth. *PLoS ONE* **2**, e845 (2007).
98. Liu, C. *et al.* H19-derived miR-675 contributes to bladder cancer cell proliferation by regulating p53 activation. *Tumour Biol*. **37**, 263–270 (2015).
99. Zhou, X., Yin, C., Dang, Y., Ye, F. & Zhang, G. Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. *Sci Rep*. **5**, 11516 (2015).
100. Chen, J. S., Zhang, X. Q., Lv, J. M., Li, Y. & Liu, T. X. X. H19 serves as a diagnostic biomarker and up-regulation of H19 expression contributes to poor prognosis in patients with gastric cancer. *Neoplasma* **63**, 223–230 (2016).
101. Liang, W. C. *et al.* The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. *Oncotarget* **6**, 22513–22525 (2015).

102. Foulstone, E. *et al.* Insulin-like growth factor ligands, receptors, and binding proteins in cancer. *J Pathol*. **205**, 145–153 (2005).
103. Nielsen, H. M. *et al.* Copy number variations alter methylation and parallel IGF2 overexpression in adrenal tumors. *Endocr Relat Cancer* **22**, 953–967 (2015).
104. Brouwer-Visser, J. & Huang, G. S. IGF2 signaling and regulation in cancer. *Cytokine Growth Factor Rev*. **26**, 371–377 (2015).
105. Shetty, P. J. *et al.* Regulation of IGF2 transcript and protein expression by altered methylation in breast cancer. *J Cancer Res Clin Oncol*. **137**, 339–345 (2011).
106. Sun, H. *et al.* H19 lncRNA mediates 17β-estradiol-induced cell proliferation in MCF-7 breast cancer cells. *Oncol Rep*. **33**, 3045–3052 (2015).
107. Hirotsune, S. *et al.* An expressed pseudogene regulates the messenger RNA stability of its homologous coding gene. *Nature* **423**, 91-96 (2003).
108. Poliseno L. Pseudogenes: Newly Discovered Players in Human Cancer. *Sci Signal.* **5**, (2012).
109. Poliseno, L., Marranci, A. & Pandolfi P. P. Pseudogenes in Human Cancer. *Front Med.* **2**, (2015).
110. Zmorzyński, S., Świderska-Kołacz, G., Koczkodaj, D. & Filip, A. A. Significance of polymorphisms and expression of enzyme- encoding genes related to glutathione in hematopoietic cancers and solid tumors. *Biomed Res Int*. **2015**, 853573 (2015).
111. Wang, J. *et al.* LEPREL1 expression in human hepatocellular carcinoma and its suppressor role on cell proliferation. *Gastroenterol Res Pract*. **2013**, 109759 (2013).
112. Béguelin, W. *et al.* EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* **23**, 677-692 (2013).
113. Berg, T. *et al.* A transgenic mouse model demonstrating the oncogenic role of mutations in the polycomb-group gene EZH2 in lymphomagenesis. *Blood* **123**, 3914-3924 (2014).
114. Caganova, M. *et al.* Germinal center dysregulation by histone methyltransferase EZH2 promotes lymphomagenesis. *J Clin Invest.* **123**, 5009-5022 (2013).
115. Tian, X., Pelton, A., Shahsafaei, A. & Dorfman, D. M. Differential expression of enhancer of zeste homolog 2 (EZH2) protein in small cell and aggressive B-cell non-Hodgkin lymphomas and differential regulation of EZH2 expression by p-ERK1/2 and MYC in aggressive B-cell lymphomas. *Mod Pathol.* **29**, 1050-1057 (2016).

# LIST OF PUBBLICATIONS

1. Esposito, F., **De Martino, M.**, Petti, M. G., Forzati, F., Tornincasa, M., Federico, A., Arra, C., Pierantoni, G. M. & Fusco, A. HMGA1 pseudogenes as candidate proto-oncogenic competitive endogenous RNAs. *Oncotarget* **5**, 8341-8354 (2014).

2. Esposito, F., **De Martino, M.**, Forzati, F. & Fusco, A. HMGA1-pseudogene overexpression contributes to cancer progression. *Cell Cycle* **13**, 3636-3639 (2014).

3. Esposito, F., **De Martino, M.**, D'Angelo, D., Mussnich, P., Raverot, G., Jaffrain-Rea, M. L., Fraggetta, F., Trouillas, J. & Fusco, A. HMGA1-pseudogene expression is induced in human pituitary tumors. *Cell Cycle* **14**, 1471-1475 (2015).

4. Leone, V., Langella, C., Esposito, F., **De Martino, M.**, Decaussin-Petrucci, M., Chiappetta, G., Bianco, A. & Fusco, A. miR-130b-3p upregulation contributes to the development of thyroid adenomas targeting CCDC6 Gene. *Eur Thyroid J.* **4**, 213-221 (2015).

5. **De Martino, M.**, Forzati, F., Arra, C., Fusco, A. & Esposito, F. HMGA1-pseudogenes and cancer. *Oncotarget* **7**, 28724-28735 (2016).

6. Palumbo, A., Da Costa, N. M., Esposito, F., **De Martino, M.**, D'Angelo, D., de Sousa, V. P., Martins, I., Nasciutti, L. E., Fusco, A. & Ribeiro Pinto, L. F. HMGA2 overexpression plays a critical role in the progression of esophageal squamous carcinoma. *Oncotarget* **7**, 25872-25884 (2016).

7. Palumbo, A., Da Costa, N. M., **De Martino, M.**, Sepe, R., Pellecchia, S., de Sousa, V. P., Nicolau Neto, P., Kruel, C. D., Bergman, A., Nasciutti, L. E., Fusco, A. & Pinto, L. F. UBE2C is overexpressed in ESCC tissues and its abrogation attenuates the malignant phenotype of ESCC cell lines. *Oncotarget* **7**, 65876-65887 (2016).

8. **De Martino, M.**, Forzati, F., Marfella, M., Pellecchia, S., Arra, C., Terracciano, L., Fusco, A. & Esposito, F. HMGA1P7-pseudogene regulates H19 and Igf2 expression by a competitive endogenous RNA mechanism. *Sci Rep.* **6**, 37622 (2016).

9. Forzati, F., **De Martino, M.**, Esposito, F., Sepe, R., Pellecchia, S., Malapelle, U., Pellino, G., Arra, C. & Fusco, A. miR-155 is positively regulated by CBX7 in mouse embryonic fibroblasts and colon carcinomas, and targets the KRAS oncogene. *BMC Cancer* **17**, 170 (2017).

# *HMGA1* pseudogenes as candidate proto-oncogenic competitive endogenous RNAs

**Francesco Esposito[1], Marco De Martino[1], Maria Grazia Petti[1], Floriana Forzati[1], Mara Tornincasa[1], Antonella Federico[1], Claudio Arra[2], Giovanna Maria Pierantoni[1], Alfredo Fusco[1],**

[1] Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Scuola di Medicina e Chirurgia di Napoli, Università degli Studi di Napoli "Federico II", Naples, Italy.

[2] Istituto Nazionale dei Tumori, Fondazione Pascale, Naples, Italy.

**Correspondence to**
Dr. Alfredo Fusco, **e-mail:** alfusco@unina.it

***Key words:*** *HMGA1P6, HMGA1P7, HMGA1*, ceRNA, miRNA

***Received:*** May 06, 2014 ***Accepted:*** July 08, 2014 ***Published:*** July 15, 2014

## ABSTRACT

**The High Mobility Group A (HMGA) are nuclear proteins that participate in the organization of nucleoprotein complexes involved in chromatin structure, replication and gene transcription. HMGA overexpression is a feature of human cancer and plays a causal role in cell transformation. Since non-coding RNAs and pseudogenes are now recognized to be important in physiology and disease, we investigated *HMGA1* pseudogenes in cancer settings using bioinformatics analysis. Here we report the identification and characterization of two *HMGA1* non-coding pseudogenes, *HMGA1P6* and *HMGA1P7*. We show that their overexpression increases the levels of HMGA1 and other cancer-related proteins by inhibiting the suppression of their synthesis mediated by microRNAs. Consistently, embryonic fibroblasts from *HMGA1P7*-overexpressing transgenic mice displayed a higher growth rate and reduced susceptibility to senescence. Moreover, *HMGA1P6* and *HMGA1P7* were overexpressed in human anaplastic thyroid carcinomas, which are highly aggressive, but not in differentiated papillary carcinomas, which are less aggressive. Lastly, the expression of the *HMGA1* pseudogenes was significantly correlated with HMGA1 protein levels thereby implicating *HMGA1P* overexpression in cancer progression. In conclusion, *HMGA1P6* and *HMGA1P7* are potential proto-oncogenic competitive endogenous RNAs.**

## INTRODUCTION

The High-Mobility Group A (HMGA) family consists of three proteins: HMGA1a, HMGA1b, and HMGA2 [1]. HMGA proteins do not have transcriptional activity *per se*; however, by interacting with the transcription machinery, they alter the chromatin structure and thereby regulate the transcriptional activity of various genes [2, 3]. The levels of HMGA proteins are low or absent in normal cells and adult tissues [4]. In contrast, their constitutive expression is remarkably high in neoplastically transformed cells and in embryonic cells [5-7]. Their overexpression is associated with a highly malignant phenotype and correlates with the presence of metastasis and reduced survival [8, 9]. Several studies implicate the expression of the HMGA genes in the process of carcinogenesis [10-18]. However, although HMGA overexpression is known to play a critical role in malignant cell transformation, the mechanisms regulating HMGA protein levels remain largely obscure.

Non-coding RNAs, including pseudogenes, have long been viewed as non-functional genomic relicts of evolution, but a large body of evidence now suggests they are important in both physiology and disease. Pseudogenes are usually defined as defunct copies of genes that have lost their potential as DNA templates for functional products [19-26] because they harbour premature or delayed stop codons, deletions/insertions and frameshift mutations that abrogate their translation into functional proteins. There are two types of pseudogenes: processed

pseudogenes, which have been retrotransposed back into a genome *via* an RNA intermediate; and nonprocessed pseudogenes, which are the genomic remnants of duplicated genes or residues of dead genes. Processed pseudogenes contain no introns, and share 5' and 3' untranslated region (UTR) sequences with their ancestral genes [27]. Since miRNAs repress target gene expression by binding to complementary sequences in the 3' UTR of target mRNA, pseudogenes can be targeted by miRNAs that modulate the expression of coding genes. Indeed, several pseudogene transcripts exert regulatory control of their ancestral gene expression levels by competing for the same miRNAs [28], which is in keeping with the notion that miRNA activity is theoretically affected by the availability of target microRNA response elements (MRE) in the cellular milieu [28-30]. Given this scenario, we studied the possible functional relationship between the mRNAs produced by the *HMGA1* oncogene and its pseudogenes (*HMGA1Ps*), and the consequences of this interaction especially in the process of carcinogenesis in which HMGA1 overexpression plays a critical role.

## RESULTS

### *HMGA1P6* and *HMGA1P7* are targeted by *HMGA1*-targeting miRNAs

We first identified seven *HMGA1Ps* by bioinformatics analysis. Of these, we focused on the *HMGA1P6* and *HMGA1P7* processed pseudogenes located at 13q12.12 and 6q23.2, respectively, because of their very high sequence homology with *HMGA1* in the coding region and in the 5' and 3' UTRs (Figure 1A). A missense mutation of the initiator methionine codon prevents translation of *HMGA1P7* whereas *HMGA1P6* is non-protein coding since it carries a mutation in the stop codon and so generates a non-translatable mRNA. Within the high homology regions, we found perfectly conserved seed matches for miRNAs that have been predicted (miR-103, miR-142-3p, miR-370, and miR-432) or already demonstrated (miR-15 [31], miR-16 [31], miR-26a [32], miR-214 [33], miR-548c-3p [34] and miR-761 [33]) to target the *HMGA1* gene (Figure 1B and 1C).

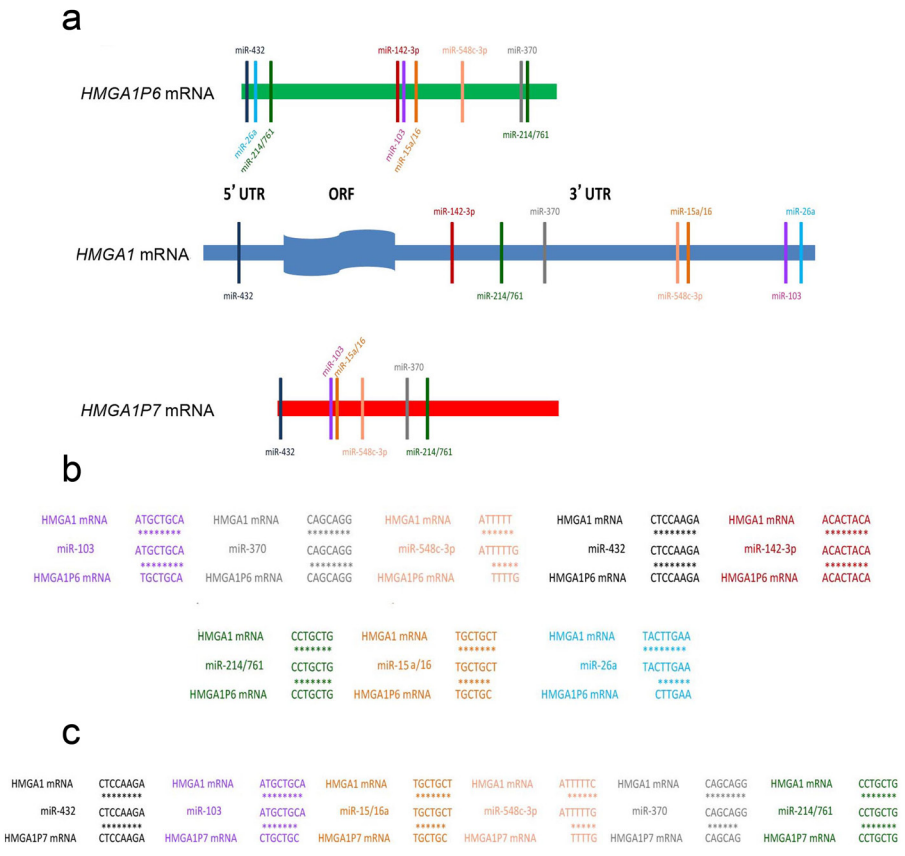



**Figure 1: *HMGA1P6* and *HMGA1P7* show the same seed sequences of *HMGA1*-targeting miRNAs. (A)** *HMGA1* (middle), *HMGA1P6* (top) and *HMGA1P7* (bottom) UTRs contain highly conserved regions. HMGA1-targeting miRNA seed matches within the high homology region are conserved between *HMGA1* and *HMGA1Ps*. **(B)** and **(C)** binding of HMGA1-targeting miRNAs to *HMGA1P6* (B) and *HMGA1P7* (C).

To evaluate the ability of these miRNAs to target *HMGA1P6* and *HMGA1P7*, we transfected miR-15, miR-16, miR-214 and miR-761 into MCF7 cells (human breast adenocarcinoma), and measured *HMGA1P6*, *HMGA1P7*, and *HMGA1* mRNA levels using Real-time PCR and PCR primer sets (see Methods) that discriminate the three mRNA transcripts. As shown in Figure 2A, the transfection of the *HMGA1*-targeting miRNAs led to a significant reduction of *HMGA1*, *HMGA1P6* and *HMGA1P7* mRNA levels.

To determine whether the *HMGA1*-targeting miRNAs directly interacted with the *HMGA1P* mRNAs, we inserted the full-length *HMGA1P6* and *HMGA1P7* mRNAs downstream of the luciferase open reading frame. These reporter vectors were transfected into human embryonic kidney (HEK293) cells together with miRNA precursors and a control non-targeting scrambled oligonucleotide. The luciferase activity was much lower after miR-15, miR-16, miR-214 and miR-761 transfection compared with the scrambled oligonucleotide (Figure 2B). These results indicate that the *HMGA1Ps* and *HMGA1* undergo the same miRNA-mediated post-transcriptional regulation.

## *HMGA1P6* and *HMGA1P7* act as decoys for *HMGA1*-targeting miRNAs

Subsequently, we examined the ability of *HMGA1P6* and *HMGA1P7* to function as a decoy for *HMGA1*-targeting miRNAs using a vector carrying the 3' UTR of the *HMGA1* mRNA downstream of the luciferase open reading frame. This reporter vector was transfected into HEK293 cells together with *HMGA1P6*- or *HMGA1P7*-expressing vectors. As expected, luciferase activity was much higher in *HMGA1Ps*-transfected cells than in the control vector (Figure 2C). Moreover, overexpression of different amounts of *HMGA1P6* or *HMGA1P7* drastically and dose-dependently reduced the effects exerted by miRNA on the levels of both the *HMGA1* transcript and protein (Figure 3A). Conversely, *HMGA1P6* and *HMGA1P7* knockdown resulted in decreased *HMGA1* mRNA and protein levels (Figure 3B) thereby mirroring the results obtained with *HMGA1P6* and *HMGA1P7* overexpression (Figure 3A). Therefore, the *HMGA1Ps* compete for the endogenous miRNA-binding sites.

The upregulation of HMGA1 induced by overexpression of *HMGA1P6* and *HMGA1P7* was blunted

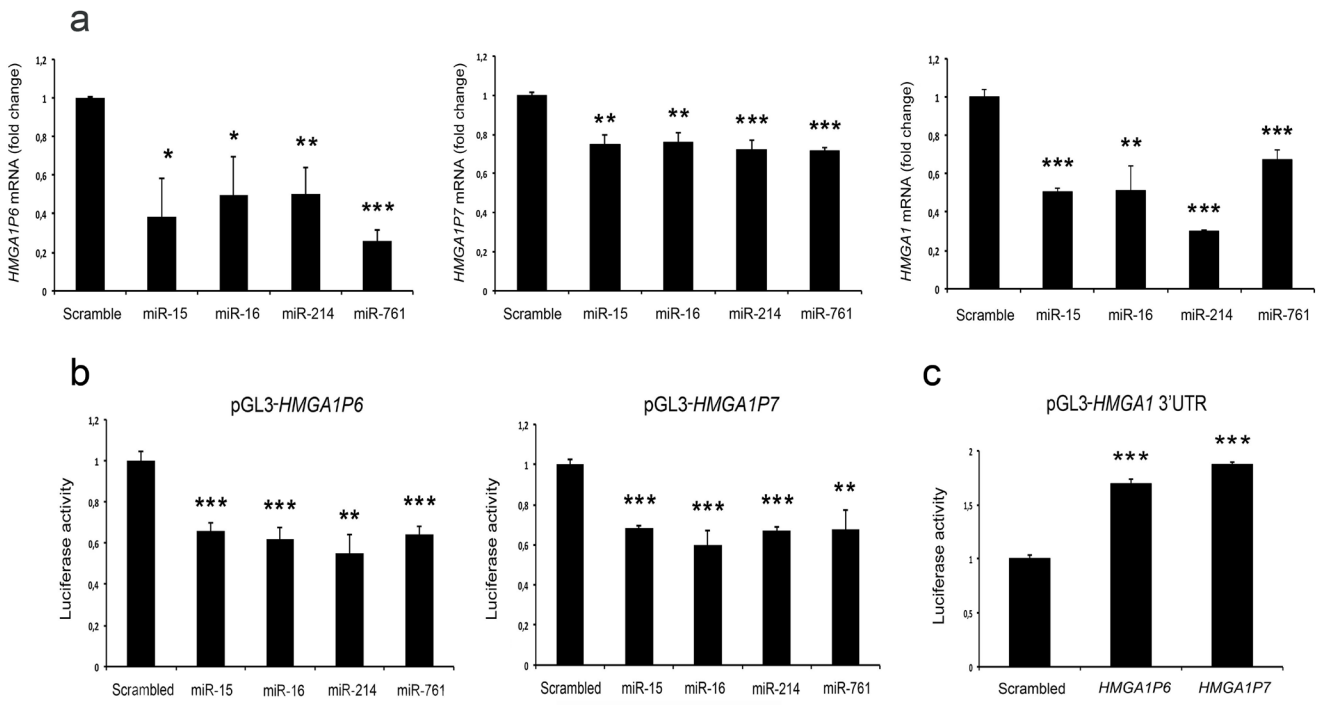



**Figure 2: *HMGA1P6* and *HMGA1P7* are targeted by *HMGA1*-targeting miRNAs. (A)** qRT-PCR analysis of *HMGA1P6* (left), *HMGA1P7* (middle) and *HMGA1* (right) mRNA from the MCF7 cells transfected with scrambled-oligonucleotide, miR-15, miR-16, miR-214 and miR-761.**(B)** *HMGA1Ps* were cloned into the pGL3 control vector. Relative luciferase activity in HEK293 cells transiently transfected with miR-15, miR-16, miR-214, miR-761 and a control scrambled oligonucleotide. **(C)** The 3'UTR of *HMGA1* was cloned into the pGL3 control vector. Relative luciferase activity in HEK293 cells transiently transfected with the empty vector, *HMGA1P6* and *HMGA1P7*. The results are reported as the mean of values. Error bars represent mean ± SD; n=3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

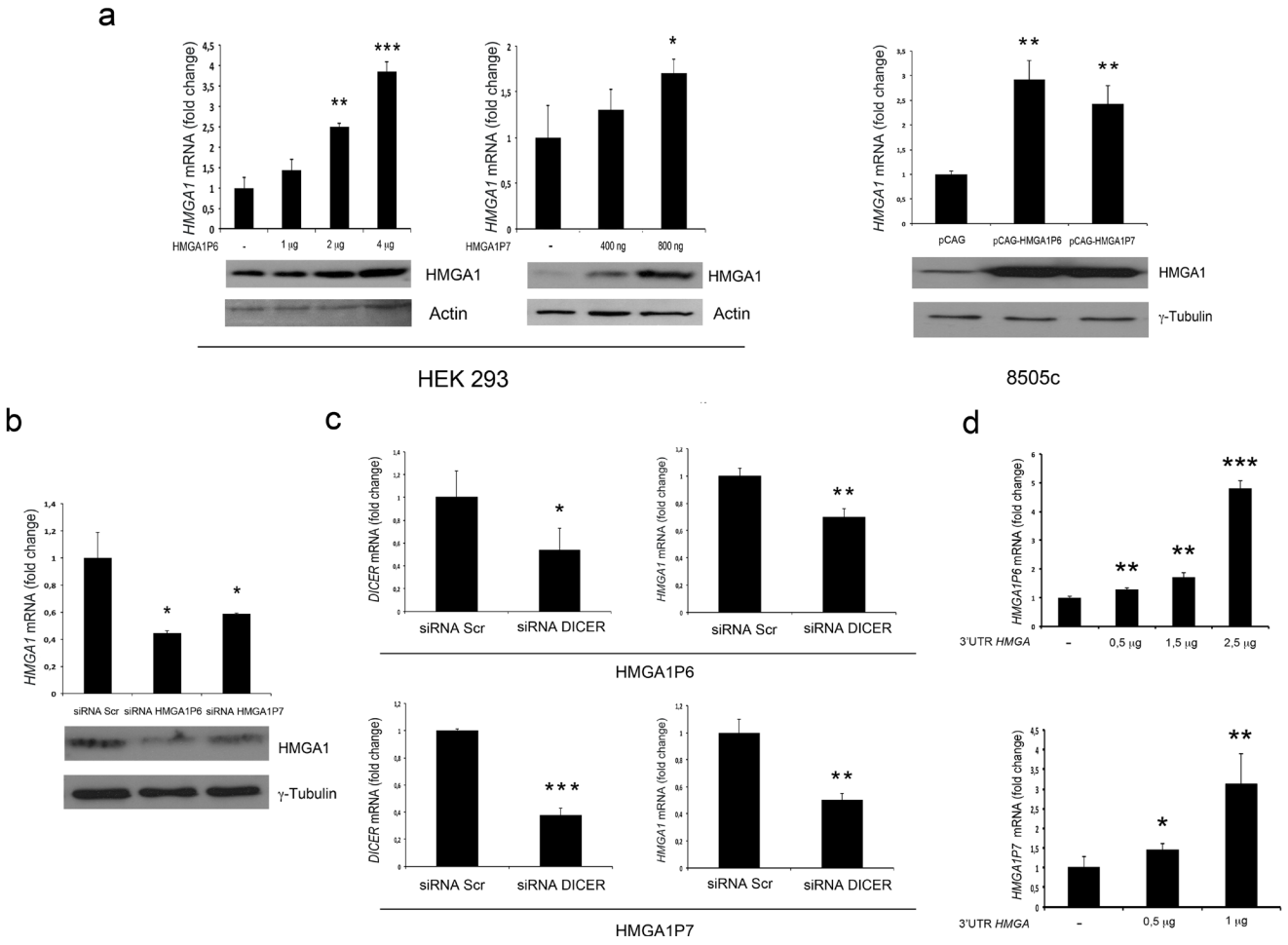



**Figure 3: HMGA1 is positively regulated by *HMGA1Ps*.** **(A)** (upper panels) qRT-PCR analysis of *HMGA1* mRNA levels in HEK293 and 8505c cells transfected with the empty vector, *HMGA1P6* and *HMGA1P7*. (Lower panels) Western blot analysis of HMGA1 protein levels from the same samples shown in the upper panels. **(B)** (Upper panel) qRT-PCR analysis of *HMGA1* mRNA levels in 8505c cells transfected with the scrambled oligonucleotide, siRNA-*HMGA1P6* and siRNA-*HMGA1P7*s. (Lower panel) Western blot analysis of HMGA1 protein levels from the same samples shown in the upper panel. **(C)** *HMGA1* mRNA levels 24 h after the transfection of *HMGA1P6* and *HMGA1P7* in scrambled oligonucleotide or siRNA-*DICER* 8505c transfected cells. **(D)** *HMGA1P6* and *HMGA1P7* mRNA levels after the transfection of the 3'UTR of the *HMGA1* plasmids in MCF7 cells. The results are reported as the mean of values; Error bars represent means ± SD; n =3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

in DICER-silenced cells (Figure 3C). In fact, silencing of DICER, the enzyme that catalyses the last step of miRNA maturation, leads to reduced levels of mature miRNAs compared to control cells. These results support the notion that *HMGA1P6* and *HMGA1P7* require mature miRNAs to regulate HMGA1 levels. Finally, as expected from our observation that *HMGA1Ps* increase HMGA1 levels, we found that the *HMGA1* 3' UTR upregulates *HMGA1P* levels (Figure 3D).

## *HMGA1P6* and *HMGA1P7* exert oncogenic activity

The *HMGA1* pseudogenes can be transcribed but they cannot code for protein. However, the above-reported results suggest that they derepress *HMGA1* transcript and protein levels (Figure 3A). To evaluate the functional consequences of *HMGA1P6* and *HMGA1P7* overexpression, we investigated their role in cellular proliferation, apoptosis, migration and invasion in cells expressing HMGA1.

As shown in Figure 4A and B, HEK293 cells and 8505c cells (derived from a human anaplastic thyroid carcinoma) transfected with *HMGA1P6*- or *HMGA1P7*-expressing vectors grew significantly faster than the empty vector-transfected cells. Cell cycle analysis of the cells overexpressing *HMGA1P6* and *HMGA1P7* revealed an increased number of cells in the S phase and a reduced number of cells in G1 compared with control cells (Figure 4C). This was not unexpected given the increased

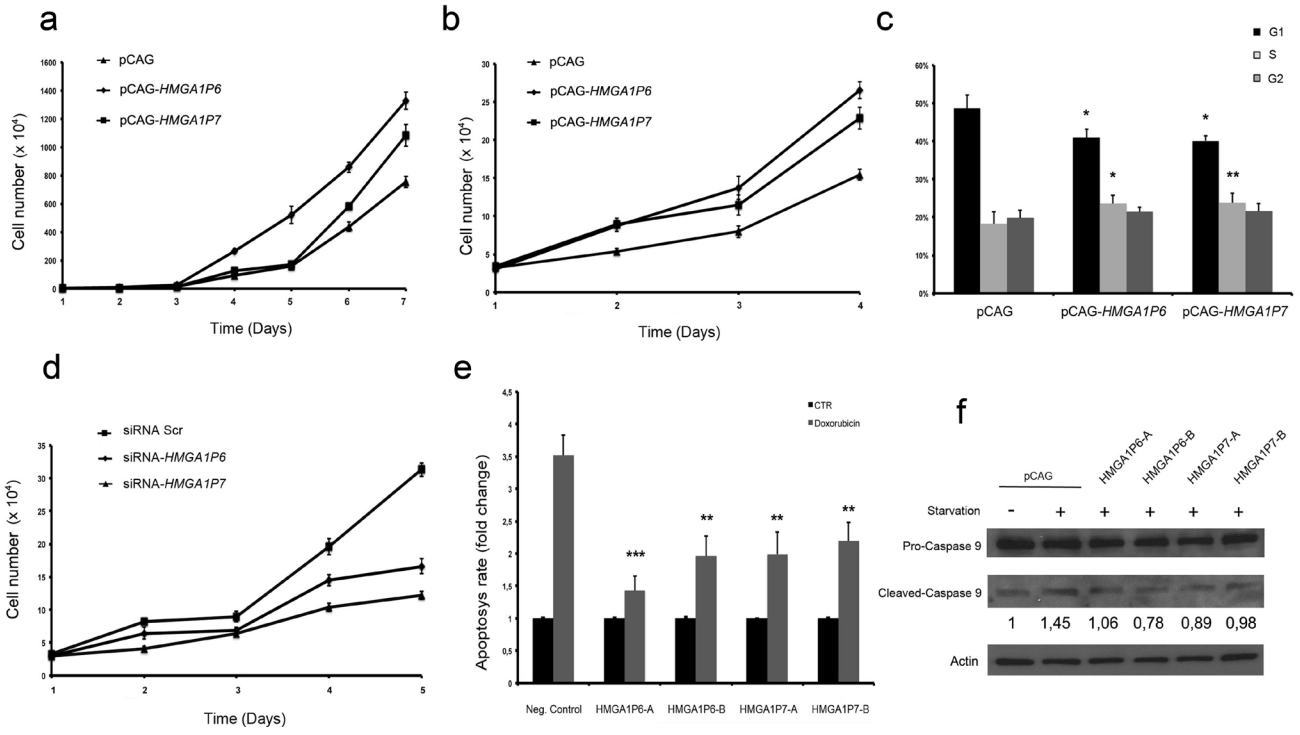



**Figure 4: *HMGA1P6* and *HMGA1P7* expression increases cell proliferation and reduces apoptosis. (A)** and **(B)** HEK293 and 8505c cell proliferation in *HMGA1P6*- and *HMGA1P7*-transfected cells. **(C)** HEK293 cells were transfected with the control, *HMGA1P6* or *HMGA1P7* vectors. The DNA content of the transfected HEK293 cells was analyzed by flow cytometry after propidium iodine staining. **(D)** 8505c cell proliferation in siRNA-*HMGA1P6*- and siRNA-*HMGA1P7*-transfected cells. **(E)** *HMGA1P6*- and *HMGA1P7*-transfected cells were treated with doxorubicin, and apoptosis was assessed by FACS. **(F)** HEK293 cells were starved, and apoptosis was assessed by Western blot analysis of Caspase 9 cleavage. The results are reported as the mean of values; Error bars represent means ± SD; n =3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

HMGA1 levels induced by *HMGA1P6* and *HMGA1P7* expression. Moreover, in 8505c cells knocked down for the *HMGA1Ps,* we found that 8505c-siRNA-*HMGA1P6* and 8505c-siRNA-*HMGA1P7* cells grew at a significantly slower rate than the 8505c-siRNA negative control (Figure 4D). Interestingly, cell cycle analysis of the 8505C-siRNA-*HMGA1P6* and 8505C-siRNA-*HMGA1P7* cells revealed an increase in the number of cells in the sub-G1 phase, which corresponds to apoptotic cells, compared with control cells (data not shown). This result is in agreement with the finding that HMGA silencing induces apoptosis in cancer cells [12].

To probe further the role of *HMGA1* pseudogenes in apoptotic cell death, we incubated HEK293 cells with doxorubicin in the presence or absence of the *HMGA1Ps.* As shown in Figure 4E, *HMGA1P6* and *HMGA1P7* overexpression significantly reduced the programmed cell death induced by doxorubicin. The same result was obtained with HEK293 cells in which apoptosis was induced by serum-starvation. Indeed, the overexpression of the *HMGA1Ps* counteracted caspase 9 cleavage (Figure 4F).

Since HMGA1 promotes cell migration and invasion [8] we carried out cell migration and invasion assays in cells transfected with the *HMGA1Ps*. As expected, cell migration was significantly higher in HEK293 and 8505c cells overexpressing *HMGA1P6* or *HMGA1P7* than in control cells (Figure 5A). Moreover, 8505c-siRNA-*HMGA1P6* and 8505c-siRNA-*HMGA1P7* cells migrated more slowly than the 8505c-siRNA negative control (Figure 5B). Accordingly, the invasion matrigel assay revealed invasion activity in HEK293 cells transfected with *HMGA1P6* or *HMGA1P6* (Figure 5C). Similar results were obtained with the *HMGA1Ps*-8505c cells (data not shown). These results indicate that cell proliferation, motility and invasion is driven by regulation *HMGA1Ps*-mediated of *HMGA1*.

Bioinformatic analysis revealed that *HMGA1P6* and *HMGA1P7* contain sequences that can be targeted by miRNAs that target *High Mobility Group A2 (HMGA2)*, *Vascular Endothelial Growth Factor (VEGF)* and *Enhancer of Zeste Homolog 2 (EZH2)*, all of which are known to be involved in carcinogenesis [34-36]. Accordingly, we found that *HMGA1P6* or *HMGA1P7*

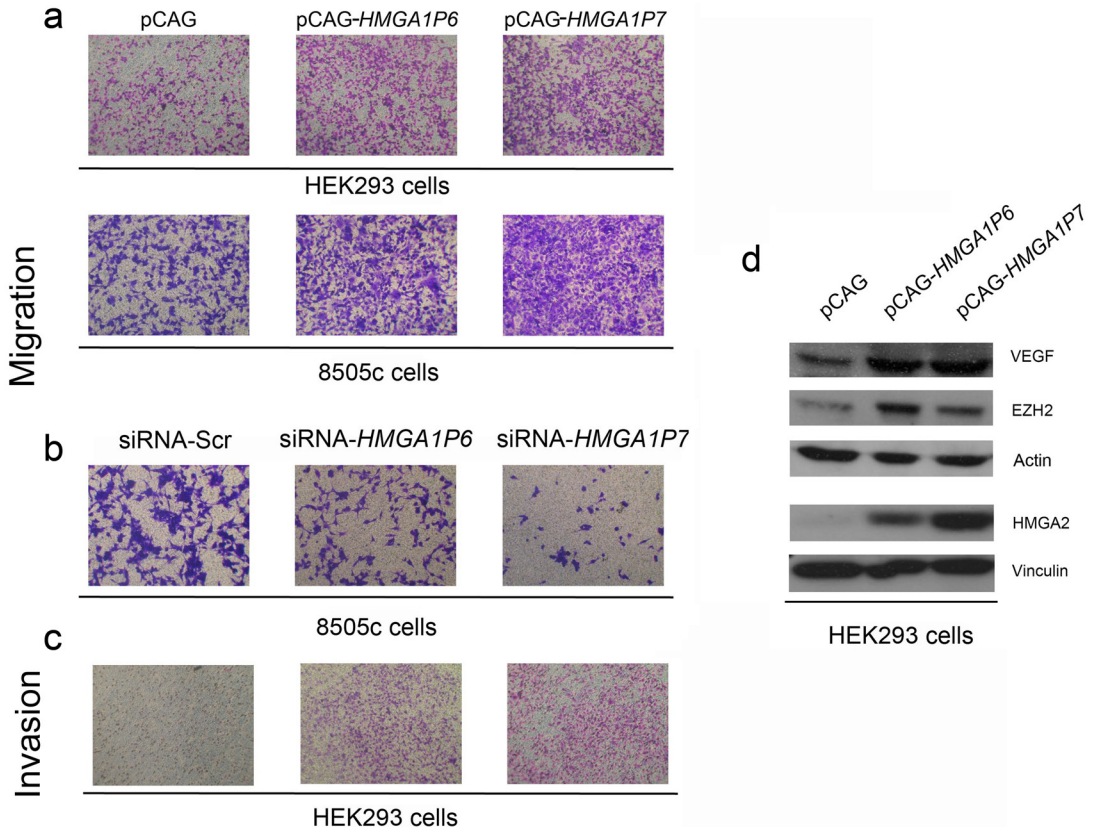



**Figure 5: The expression of the *HMGA1Ps* affects cell migration and invasion.** **(A)** Cell migration assays of HEK293 and 8505c cells transfected with *HMGA1P6* or *HMGA1P7* or with a control vector. One representative experiment is reported. **(B)** Cell migration assays of 8505c cells transfected with siRNA-*HMGA1P6* or siRNA-*HMGA1P7* or with a empty vector. One representative experiment is reported. **(C)** Cell invasion assays of HEK293 cells transfected with *HMGA1Ps* or with the backbone vector. One representative experiment is reported. **(D)** Extracts from HEK293 transfected with *HMGA1P6* or *HMGA1P7* or with a control vector were analyzed by Western blotting.

overexpression increased the level of the proteins coded for by these genes (Figure 5D). Consequently, it appears that *HMGA1P6* and *HMGA1P7* expression may contribute to cancer progression by acting as decoys for cancer-related genes other than *HMGA1*.

## Correlation of HMGA1 and the overexpression of the *HMGA1Ps* in human cancer

To verify whether the two *HMGA1Ps* function as decoys in the regulation of HMGA1 protein levels also in human cancer, we analyzed the expression of HMGA1 and of the *HMGA1Ps* in a panel of differentiated and undifferentiated thyroid carcinomas by Western Blotting and Real-time PCR. As shown in Figure 6, papillary (PTC) thyroid carcinomas, which are well differentiated and poorly aggressive, expressed low levels of *HMGA1P6* and *HMGA1P7* (Figure 6A). Conversely, anaplastic thyroid carcinoma (ATC), which is one of the most aggressive human tumours, expressed very high *HMGA1P* levels that, moreover, correlated with HMGA1 protein levels (Figure 6B). Accordingly, HMGA1 expression, which was undetectable in normal thyroid tissue, was much higher in ATC than in PTC. Similar results were obtained in human ovarian carcinomas (see Figure 6C and 6D). The direct correlation between *HMGA1* and *HMGA1P6* expression ($r=0.6553$, $P<0.0001$) and between *HMGA1* and *HMGA1P7* expression ($r=0.7001$, $P<0.0001$) suggests that these genes are co-regulated (Figure 6C and 6D). Taken together, these results indicate that *HMGA1P6, HMGA1P7* and *HMGA1* expression is correlated with cancer aggressiveness.

## *HMGA1P7* overexpressing mouse embryonic fibroblasts grow faster and senesce later

To establish the role of the *HMGA1Ps in vivo*, we generated transgenic mice overexpressing *HMGA1P7*. The expression of *HMGA1P7* in transgenic mice was verified by RT-PCR using RNAs extracted from liver, spleen, lung

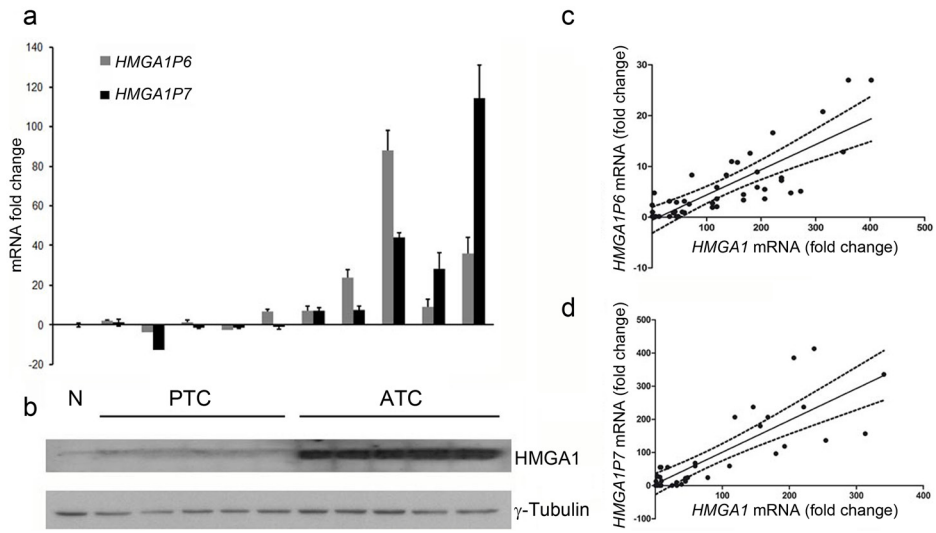



**Figure 6: HMGA1 protein expression positively correlates with the expression of the *HMGA1Ps* in ATC. (A)** *HMGA1P6* and *HMGA1P7* qRT-PCR analysis in normal thyroid tissue (NT), papillary thyroid carcinoma (PTC) and anaplastic thyroid carcinoma (ATC). The results are reported as the mean of expression values. The error bars represent mean ± SD; n = 3. **(B)** Western blot analysis of HMGA1 protein expression in the same samples as in A. **(C)** and **(D)** Ovarian sample expression values derived from commercial sources were combined for correlation analysis. Linear regressions of *HMGA1* versus *HMGA1P6* (C) and *HMGA1* versus *HMGA1P7* (D) are shown.

and mouse embryonic fibroblasts (MEFs). *HMGA1P7* mRNA levels were high in the tissues of *HMGA1P7* transgenic mice and absent in the WT counterpart (Figure 7A). Consistently, *HMGA1* transcript and protein levels were significantly higher in *HMGA1P7*-MEFs than in the WT control (Figure 7B). Notably, *HMGA1P7*-MEFs also expressed increased levels of the Hmga2, Ezh2 and Vegf proteins, which are coded for by genes that share miRNAs with *HMGA1* (Figure 7C), and increased levels of the HMGA1-regulated genes *Ccna*, *Ccnb*, *Ccnd2* and *E2f*-1 (Figure 7D), which play a critical role in cell cycle regulation [8].

As expected from the high HMGA1 expression in the *HMGA1P7*-MEFs, the growth rate of these MEFs was significantly higher than that of the WT controls (Figure 7E). To determine whether the higher growth rate of *HMGA1P7*-MEFs was caused by altered progression through the cell cycle, we examined asynchronously growing MEFs by flow cytometry. The number of *HMGA1P7*-MEFs was lower in G1 and higher in the S phase of the cell cycle compared with WT MEFs (Figure 7F).

We, next, examined the susceptibility of MEFs to senescence by measuring senescence-associated β-gal (SA-β-gal) activity. At culture passage 6, SA-β-gal activity was present in WT MEFs, but not in the *HMGA1P7* transgenic counterparts (Figure 7G). These findings indicate that *HMGA1P7* overexpression reduced susceptibility to cellular senescence.

### Identification of the genes modulated by *HMGA1P7* expression

To identify the genes regulated by *HMGA1P7* expression, we analyzed the expression profile of WT and *HMGA1P7* transgenic MEFs in microarray analyses. To this aim, RNAs extracted from WT and *HMGA1P7*-MEFs were hybridized to the Affymetrix GeneChip Mouse Gene 2.0 ST oligonucleotide arrays. Seventy transcripts that had a significant fold change variation ($p < 0.05$) were examined as candidate genes involved in *HMGA1P7* tumour-promoting activity. Interestingly, we found five upregulated cancer-related genes (*Epha3, Hjurp, Kif26, S1pr3* and *Pde3B*) that shared miRNAs with the *HMGA1P7* transcript. These genes are involved in various human cancers (glioblastoma, breast and hematological cancers), and are candidate therapeutic targets [37-41]. Real-time PCR experiments confirmed upregulation of all these genes in *HMGA1P7*-MEFs (Supplementary Figure 1). These results support the concept that *HMGA1P7* modulates the expression of several cancer-related genes by acting as a ceRNA.

## DISCUSSION

The HMGA proteins play a critical role in carcinogenesis. Recently, several miRNAs have been demonstrated to target these genes [31-34], and their dysregulation may contribute to HMGA1 protein

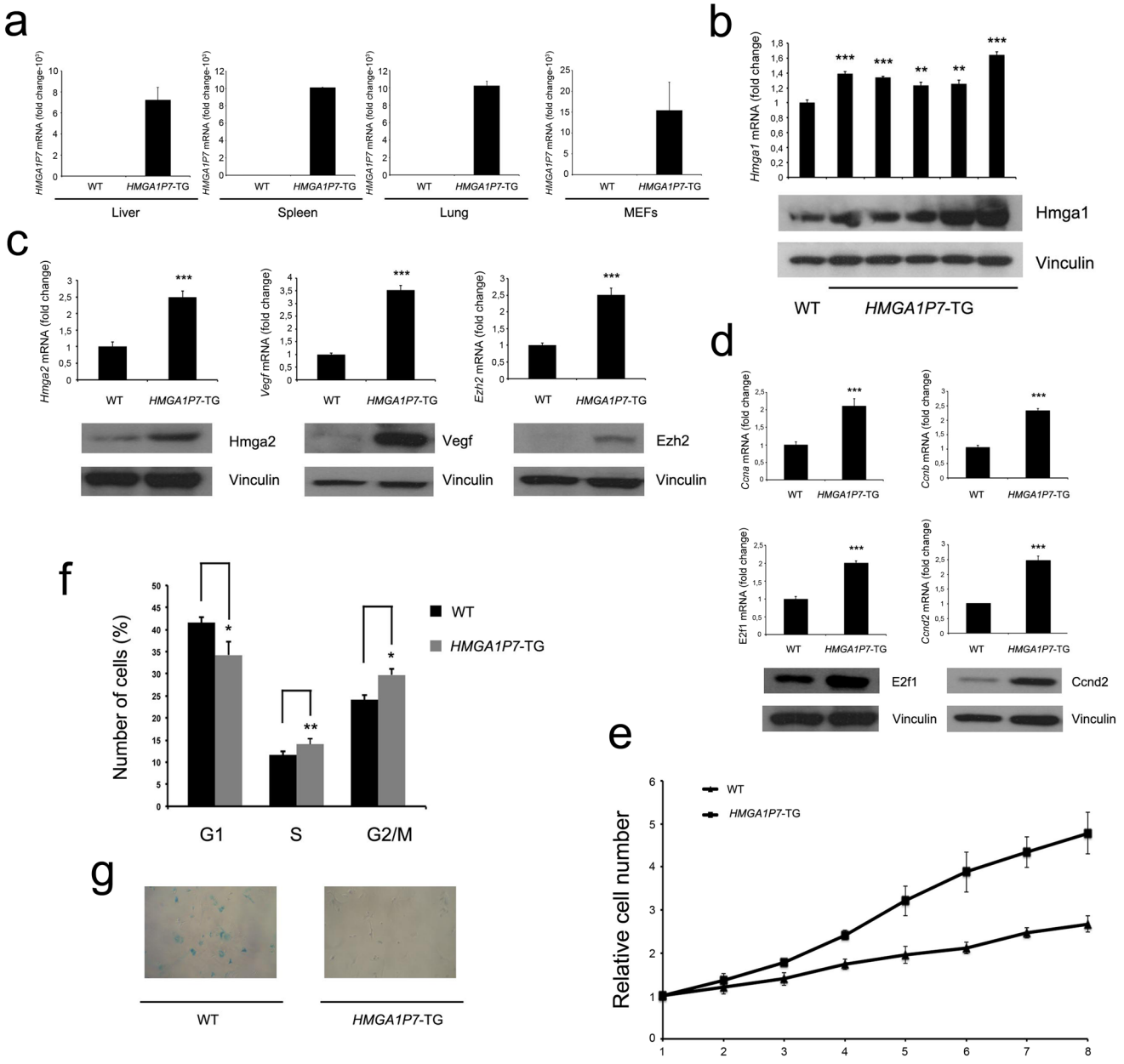



**Figure 7: *HMGA1P7* overexpressing MEFs show a higher growth rate, and lower susceptibility to senescence.** **(A)** qRT-PCR analysis of total RNA from livers, spleens, lungs, and MEFs of WT and *HMGA1P7* transgenic mice. **(B)** (upper panel) qRT-PCR analysis of *HMGA1* mRNA levels in WT and *HMGA1P7* transgenic MEFs. (Lower panel) Western blot analysis of HMGA1 protein expression in the same samples. **(C)** and **(D)** qRT-PCR and Western blot analysis of genes that share common miRNAs with *HMGA1* (Left panel) and *HMGA1*-regulated genes (Right panel). **(E)** MEFs were prepared from WT and *HMGA1P7* overexpressing embryos at 12.5 dpc. At culture passage 3, they were plated and counted daily for 8 days. **(F)** Propidium iodide flow cytometry of asynchronous growing WT and *HMGA1P7* overexpressing MEFs. **(G)** Light microscopy of representative WT and *HMGA1P7* overexpressing MEFs stained for β-galactosidase activity at culture passages 6. The results are reported as the mean of values with *error bars* indicating SD (mean ± SD); n =3. *, $P < 0.05$ **, $P < 0.01$ ***, $P < 0.001$ (*t* test).

overexpression in human neoplasias [31, 32, 34]. Moreover, an important role in the regulation of protein synthesis has recently been ascribed to pseudogenes [27, 28]: the presence of the same miRNA-targeted seed sequences in the *HMGA1* and in the *HMGA1Ps* UTR regions could block the access of miRNAs to their protein-coding target genes. Finally, it has been outlined a novel gene-expression pathway in which an *HMGA*

protein-coding gene, *Hmga2*, operates largely independently of its protein-coding function to promote cancer progression as a competing endogenous RNA [42].

We asked whether *HMGA1* pseudogenes affect HMGA1 protein levels, and, consequently, whether they play a critical role in cancer progression. We focused on *HMGA1P*6 and *HMGA1P*7, which have conserved seed matches for miRNAs targeting the *HMGA1* gene in the high homology regions. We demonstrate that overexpression of these *HMGA1* pseudogenes increases HMGA1 protein levels, and inhibits the suppression of HMGA1 protein synthesis by miRNAs that target the *HMGA1* gene, namely, miR-15, miR-16, miR-214, and miR-761 [31-34]. Consistent with these results, our functional studies demonstrate that *HMGA1P*6 and *HMGA1P*7 overexpression increases the cell growth rate by decreasing the number of G1-phase cells and increasing the number of S-phase cells, compared with the backbone vector-transfected cells. Therefore, *HMGA1P*6 and *HMGA1P*7 affect cell cycle progression, as expected, given their ability to increase the protein levels of HMGA1, which is involved in the regulation of the G1-S transition phase of the cell cycle [34]. Moreover, *HMGA1P*6 and *HMGA1P*7 overexpression increased cell migration and invasiveness, and decreased the apoptotic rate.

These results prompted us to verify whether *HMGA1P*6 and *HMGA1P*7 overexpression is involved also in human carcinogenesis. Interestingly, *HMGA1P*6 and *HMGA1P*7 were abundantly expressed in ATC, which are very aggressive and express very high HMGA1 protein levels [43]. Conversely, *HMGA1P*6 and *HMGA1P*7 expression was low in PTC, which are well differentiated and poorly aggressive, and express moderate HMGA1 protein levels. We obtained similar results in human ovarian carcinomas suggesting that *HMGA1Ps* can regulate HMGA1 protein levels also *in vivo*.

Interestingly, *HMGA1P6* and *HMGA1P7* seem to affect cancer progression also by binding to the same miRNAs that target proteins involved in cancer progression. Indeed, the overexpression of the *HMGA1Ps* increased also the levels of HMGA2, VEGF and EZH2 that are coded for by genes targeted by *HMGA1*-targeting miRNAs. Notably EZH2, which is involved in carcinogenesis, is overexpressed in ATC but not in PTC [44].

Data obtained with transgenic mice overexpressing *HMGA1P7* and with the relative MEFs support the concept that *HMGA1P7* plays an oncogenic role. Indeed, MEFs derived from transgenic mice overexpressing *HMGA1P7* show a higher growth rate, and lower susceptibility to senescence with respect to the WT counterpart. Moreover, flow cytometry showed an increase of cells in S phase as expected given the ability of HMGA1 to increase the E2F1 activity [45].

In contrast to a report that ectopic overexpression of *HMGA1* reduces the lifespan of IMR90 cells [46], *HMGA1P7*-MEFs that have more abundant HMGA1 protein levels, senesce later with respect to WT MEFs. In agreement with our findings, we obtained the opposite result in *Hmga1*-null MEFs [47]. It is likely that the cellular context influences the effect exerted by HMGA proteins on cell growth. Moreover, the different experimental approach, one *in vivo* and one *in vitro*, may account for these contradictory results. In fact, discrepancies between *in vitro* and transfection approaches were reported in a study of the p53 pathway [48]. The behaviour of the *Hmga1*-null and *HMGA1P7*-MEFs described here supports the oncogenic role of HMGA overexpression, which is a feature of malignant neoplasias. In conclusion, our finding that HMGA1P7-overexpressing MEFs grow faster and senesce later than their WT counterpart sustains our model in which *HMGA1Ps* act as ceRNAs that regulate *HMGA1* and other genes by competing for shared miRNAs thus contributing to cancer progression.

## MATERIALS AND METHODS

### Cell culture and transfections

HEK293, MCF7, 8505c, and MEF (from 12.5-day-old embryos) cells were maintained in DMEM supplemented with 10% foetal calf serum (GIBCO; Invitrogen), glutamine and antibiotics. Cells were regularly tested with MycoAlert (Lonza) to ascertain that cells were not infected with mycoplasma. Cells were transfected using Lipofectamine plus reagent (Invitrogen) according to the manufacturer's instructions. The transfected cells were selected in a medium containing geneticin (Sigma). For each transfection, several geneticin-resistant mass cell populations were isolated and expanded for further analysis. Transfection efficiency was verified for each experiment by evaluating GFP expression. To inhibit *HMGA1P6* and *HMGA1P7* expression, small interfering RNAs and corresponding scramble small interfering RNAs were designed and used as suggested by the manufacturer (RIBOXX).

### Human thyroid and ovary tissue samples

Neoplastic and normal human thyroid tissues were obtained from surgical specimens and immediately frozen in liquid nitrogen. Thyroid tumours were collected at the Service d'Anatomo-Pathologie, Centre Hospitalier Lyon Sud, Pierre Bénite, France. The tumour samples were frozen until required for RNA or protein extraction. We declare that informed consent for the scientific use of biological material was obtained from all patients.

TissueScan Ovarian Cancer Tissue Real-time PCR Panel were purchased from Origene (HORT302).

## RNA extraction and quantitative reverse transcription PCR

Total RNA was extracted from tissues and cell cultures with Trizol (Gibco) according to the manufacturer's instructions. For mRNA detection, we reverse transcribed total RNA from cell lines by using the QuantiTect Reverse Transcription Kit (Qiagen), and then Real-time PCR was performed by using Power SYBR Green PCR Master Mix (Applied Biosystems) and the following primers:

*HMGA1*-Fw 5'-aaggggcagacccaaaaa-3'
*HMGA1*-Rev 5'-tccagtcccagaaggaagc-3'
*HMGA1P6*-Fw 5'-gcagacccacaaaactgga-3'
*HMGA1P6*-Rev 5'-gagcaaagctgtcccatcc-3'
*HMGA1P7*-Fw 5'-gctccttctcggctcctc-3'
*HMGA1P7*-Rev 5'-gcttgggcctcttttatgg-3'
G6PD-Fw 5'-acagagtgagcccttcttcaa-3'
G6PD-Rev 5'-ataggagttgcgggcaaag-3'
*Hmga1*-Fw 5'-ggcagacccaagaaactgg-3'
*Hmga1*-Rev 5'-ggcactgcgagtggtgat-3'
*Ccna*-Fw 5'-cttggctgcaccaacagtaa-3'
*Ccna*-Rev 5'-caaactcagttctcccaaaaaca-3'
*Ccnb*-Fw 5'-gcgctgaaaattcttgacaac-3'
*Ccnb*-Rev 5'-ttcttagccaggtgctgcat-3'
*G6pd*-Fw 5'-cagcggcaactaaactcaga-3'
*G6pd*-Rev 5'-ttccctcaggatcccacac-3'
*Epha3*-Fw 5'-tggctccttggacagtttct-3'
*Epha3*-Rev 5'-ttcccacaagctccatgact-3'
*Hjurp*-Fw 5'-gagaactggccatcttgcag-3'
*Hjurp*-Rev 5'-aaggtgtttccgggcact-3'
*Kif26b*-Fw 5'-aagaggcaggctctcaagc-3'
*Kif26b*-Rev 5'-gcagagaaagcaagggtcctt-3'
*S1pr3*-Fw 5'-agatgcgccttgcagaac-3'
*S1pr3*-Rev 5'-agagtggtggtgggttcct-3'
*Pde3B*-Fw 5'-ccttgtatttcccgagaacagat-3'
*Pde3B*-Rev 5'-ggtaatgaggtttacaccactgc-3'
*Hmga2*-Fw 5'-aaggcagcaaaaacaagagc-3'
*Hmga2*-Rev 5'-ttgtggccatttcctaggtc-3'
*Ezh2*-Fw 5'-tggaagcagcggaggata-3'
*Ezh2*-Rev 5'-gtcactggtgactgaacactcc-3'
*Vegf*-Fw 5'-aaaaacgaaagcgcaagaaa-3'
*Vegf*-Rev 5'-tttctccgctctgaacaagg-3'

The $2^{-\Delta\Delta Ct}$ formula was used to calculate the differential gene expression.

## Plasmids

For transfection of miRNA oligonucleotides, cells were transfected with 50 nmol/ml of miRNA precursors or with a control no-targeting scrambled oligonucleotides (Ambion, Austin, TX) using siPORT neoFX Transfection Agent (Ambion). For the *HMGA1P6* expression construct (pCAG-*HMGA1P6*) and the *HMGA1P6* luciferase reporter construct (pGL3-*HMGA1P6*), the entire sequence of *HMGA1P6* gene (ENST00000418454.1) was amplified by using the primers Fw *HMGA1P6* 5'-tcctctaattgggactccga-3' and Rev *HMGA1P6* 5'-ttactcagatcccaggcaga-3'. The amplified fragment was cloned into pCAG vector kindly given by Dr. S. Soddu, and into pGL3-Control firefly luciferase reporter vector (Promega), respectively. For the *HMGA1P7* construct (pCAG-*HMGA1P7*) and the *HMGA1P7* luciferase reporter construct (pGL3-*HMGA1P6*), the entire sequence of the *HMGA1P7* gene (ENST00000406908.1) was amplified by using the primers Fw *HMGA1P7* 5'-agccagtcgagctggaggtc-3' and Rev *HMGA1P7* 5'-ctgcaatgtgtactcagagc-3'. The amplified fragment was cloned as described for the *HMGA1P6* constructs. All the generated vectors were confirmed by sequencing. The Renilla luciferase vector (pRL-CMV), for transient transfection efficiency, was purchased from Promega. The 3′ UTR region of the *HMGA1* gene has been previously described [34].

## Protein extraction, western blotting and antibodies

Protein extraction and Western blotting were performed as previously described [49]. The primary antibodies used were anti-EZH2 (AC22) and anti-Cyclin D2 (2924) from Cell Signaling; anti-Actin (sc-1615), anti-Vinculin (sc-7649), anti-γ-Tubulin (sc-17787), and anti-E2f1 (sc-193) from Santa Cruz Biotechnology; anti-VEGF (ab46154) from Abcam. Antibodies versus the HMGA1 and HMGA2 proteins are described elsewhere [50, 51]. Blots were visualized by using the Western blotting detection reagents (GE Healthcare).

## Cell migration and invasion assay

Cell migration and invasion experiments were performed as previously described [44].

## Dual-luciferase reporter assay

For dual-luciferase reporter assay, $3 \times 10^5$ HEK293 cells were co-transfected in 6-well plates with the pGL3-*HMGA1P6* or the pGL3-*HMGA1P7* luciferase reporter vectors, together with the Renilla luciferase plasmid and miRNA precursors or a control no-targeting scrambled oligonucleotides (Ambion), using siPORT neoFX Transfection Agent (Ambion). The pRL-TK control vector expressing Renilla luciferase (Promega) was used for normalization of cell number and transfection efficiency. Luciferase activity was measured 48 hours after transfection using the Dual-Luciferase Reporter

Assay System (Promega) with a Lumat LB 9507 apparatus (Berthold Technologies).

### Flow cytometric analysis

HEK293 cells were transfected with *HMGA1P6*, *HMGA1P7* and the empty vector, and analysed by flow cytometry after 48 hours of growth under normal culture conditions. Primary MEFs were obtained from 12.5-day-old embryos. The MEFs were minced and used to establish single cell suspensions and then analysed by flow cytometry after 48 hours of growth under normal culture conditions. Briefly, cells were harvested in PBS containing 2 mmol/l EDTA, washed once with PBS, and fixed for 2 hours in cold ethanol (70%). Fixed cells were washed once in PBS and treated with 40 μg/ml RNase A in PBS for 30 minutes. They were then washed once in PBS and stained with 50 μg/ml propidium iodide (Roche). Stained cells were analysed with a fluorescence activated cell sorter (FACS) Calibur (Becton-Dickinson), and the data were analysed using a mod-fit cell cycle analysis programme.

### Generation and genotyping of mutant mice

The 3.5 kb *HMGA1P7* of the pCAG-*HMGA1P7* expression plasmid was excised with *Sal*I & *Hind*III restriction endonucleases by cleaving 10 μg of the plasmid. The fragment was purified from SeaKem GTG agarose (avoiding exposure to UV light) using the Qbiogene Geneclean Spin kit, then dialysed 24 h against 2 l microinjection buffer (10 mM Tris.HCl pH 7.2, 0.1 mM EDTA), and diluted to a concentration of 4 ng/μl. The DNA was injected in three sessions into C57BL/6N-derived zygotes. For this purpose, C57BL/6N female mice (bred at PolyGene from parents obtained from Charles River) were superovulated at 28-34 days of age and mated in the PolyGene mouse facility to C57BL/6N breeder males, originally also obtained from Charles River. Injected zygotes were cultivated overnight and transferred into pseudopregnant B6CBAF1 females, also from Charles River. The animals were kept in individually ventilated cages. Injections were performed at the PolyGene labs in Rümlang, Switzerland. Pups were biopsied at weaning and analysed for transgene integration by PCR, using the PCR primer combination: Fw 5'-ggcatgtcccactctatt-3'; Rev 5'-caattcctgcaatgtgtactc-3'. All mice were maintained under standardized nonbarrier conditions in the Laboratory Animal Facility of the Istituto dei Tumori di Napoli (Naples, Italy), and all studies were conducted in accordance with Italian regulations for experimentations on animals.

### SA-β-gal assay

$4 \times 10^4$ MEFs, plated 24 hours before the assay, were washed twice with PBS and immersed in fixation buffer (2% [w/v] formaldehyde, 0.2% [w/v] glutaraldehyde in PBS) for 7 minutes. After 3 additional PBS washes, the cells were allowed to stain overnight in staining solution (40 mM citric acid/sodium phosphate, pH 6.0; 150 mM NaCl; 2.0 mM MgCl2; 1 mg/ml X-gal) at 37°C without CO2 to avoid changes in pH. The next day, the staining solution was replaced with PBS, and the stained and unstained cells were counted by light microscopy (at least 24 fields).

### Microarray analyses

RNAs extracted from *HMGA1P7* transgenic and WT MEFs (two biological replicates for each sample) were hybridized to the Affymetrix GeneChip Mouse Gene 2.0 ST oligonucleotide arrays. Hybridization, washing, staining, scanning, and data analysis were performed by the Affymetrix Microarray Unit at the IFOM-IEO campus, Milan, Italy, according to the manufacturer's instructions. Data were analyzed using Partek Genomics Suite version 6.6. Transcripts showing a significant fold change variation ($p<0.05$) were examined.

### Statistical analysis

Data were analyzed using a two-sided unpaired Student's t test (GraphPad Prism, GraphPad Software, Inc.). Values of $P<0.05$ were considered statistically significant. The mean +/− s.d. of three or more independent experiments is reported. Regression analyses and correlation coefficients were generated using GraphPad Prism, GraphPad Software, Inc.

## ACKNOWLEDGMENTS

We thank Mario Berardone for the art-work and Jean Ann Gilder (Scientific Communication srl) for substantive text editing.

## GRANT AND FUND SUPPORT

This work was supported by grants from Associazione Italiana per la Ricerca sul Cancro (AIRC-IG 5346), the Project "Invecchiamento" of the National Research Council (CNR).

## AUTHOR CONTRIBUTIONS

F.E. and A.F. conceived and designed the study. F.E., with help from M.D.M., M.G.P., F.F., M.T., A.F., G.M.P. and A.F., performed all experiments, and F.E., M.D.M and A.F. analyzed most data. F.E. and A.F. wrote the manuscript with contributions from all authors.

## AUTHOR INFORMATION

The authors declare no competing financial interests.

## REFERENCES


1. Johnson KR, Lehn DA and Reeves R. Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y. Mol Cell Biol. 1989; 9:2114–2123. doi:10.1128/MCB.9.5.2114.
2. Grosschedl R, Giese K and Pagel J. HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet. 1994; 10:94–100. doi:10.1016/0168-9525(94)90232-1.
3. Thanos D and Maniatis T. The high mobility group protein HMG I(Y) is required for NF-κB-dependent virus induction of the human IFN-β gene. Cell. 1992; 71:777–789. doi:10.1016/0092-8674(92)90554-P.
4. Zhou X, Benson KF, Ashar HR and Chada K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature. 1995; 376:771–774. doi:10.1038/376771a0.
5. Chiappetta G, Avantaggiato V, Visconti R, Fedele M, Battista S, Trapasso F, Merciai BM, Fidanza V, Giancotti V, Santoro M, Simeone and Fusco A. High level expression of the HMGA1 gene during embryonic development. Oncogene. 1996; 13:2439–2446.
6. Pierantoni GM, Agosti V, Fedele M, Bond H, Caliendo I, Chiappetta G, Lo Coco F, Pane F, Turco MC, Morrone G, Venuta S and Fusco A. High-mobility group A1 proteins are overexpressed in human leukaemias. Biochem J. 2003; 372:145–150.
7. Chieffi P, Battista S, Barchi M, Di Agostino S, Pierantoni GM, Fedele M, Chiariotti L, Tramontano D and Fusco A. HMGA1 and HMGA2 protein expression in mouse spermatogenesis. Oncogene. 2002; 21:3644–3650.
8. Fusco A and Fedele M. Roles of the HMGA proteins in cancer. Nat Rev Cancer. 2007; 7:899–910. doi:10.1038/nrc2271.
9. Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del Sal G and Manfioletti G. HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. Oncotarget. 2013; 4:1293–1308.
10. Berlingieri MT, Manfioletti G, Santoro M, Bandiera A, Visconti R, Giancotti V and Fusco A. Inhibition of HMGI-C protein synthesis suppresses retrovirally induced neoplastic transformation of rat thyroid cells. Mol Cell Biol. 1995; 15:1545–1553.
11. Berlingieri MT, Pierantoni GM, Giancotti V, Santoro M and Fusco A. Thyroid cell transformation requires the expression of the HMG1 proteins. Oncogene. 2002; 21:2971–2980. doi:10.1038/sj/onc/1205368.
12. Scala S, Portella G, Fedele M, Chiappetta G and Fusco A. Adenovirus mediated suppression of the HMGI(Y) protein synthesis as a potential therapy of human malignant neoplasias. Proc Natl Acad Sci USA. 2000; 97:4256–4261. doi:10.1073/pnas.070029997.
13. Wood LJ, Maher JF, Bunton TE and Resar LM. The oncogenic properties of the HMG-I gene family. Cancer Res. 2000; 60:4256–4261.
14. Reeves R, Edberg DD and Li Y. Architectural transcription factor HMGI (Y) promotes tumor progression and mesenchymal transition of human epithelial cells. Mol Cell Biol. 2001; 21:575–594. doi:10.1128/MCB.21.2.575-594.2001.
15. Xu Y, Sumter TF, Bhattacharya R, Tesfaye A, Fuchs EJ, Wood LJ, Huso DL and Resar LM. The HMG-I oncogene causes highly penetrant, aggressive lymphoid malignancy in transgenic mice and is overexpressed in human leukemia. Cancer Res. 2004; 64:3371–3375. doi:10.1158/0008-5472.CAN-04-0044.
16. Hillion J, Dhara S, Sumter TF, Mukherjee M, Di Cello F, Belton A, Turkson J, Jaganathan S, Cheng L, Ye Z, Jove R, Aplan P and Lin YW. The high-mobility group A1a/signal transducer and activator of transcription-3 axis: an achilles heel for hematopoietic malignancies?. Cancer Res. 2008; 68:10121–10127. doi:10.1158/0008-5472.CAN-08-2121.
17. Tesfaye A, Di Cello F, Hillion J, Ronnett BM, Elbahloul O, Ashfaq F, Dhara S, Prochownik E, Tworkoski K, Reeves R, Elbahlouh O, Kowalski J and Bhattacharya R. The high-mobility group A1 gene up-regulates cyclooxygenase 2 expression in uterine tumorigenesis. Cancer Res. 2007; 67:3998–4004. doi:10.1158/0008-5472.CAN-05-1684.
18. Puca F, Colamaio M, Federico A, Gemei M, Tosti N, Bastos AU, Del Vecchio L, Pece S, Battista S and Fusco A. HMGA1 silencing restores normal stem cell characteristics in colon cancer stem cells by increasing p53 levels. Oncotarget. 2014; 5:3234–3245.
19. Vanin EF. Processed pseudogenes: Characteristics and evolution. Annu Rev Genet. 1985; 19:253–272. doi:10.1146/annurev.ge.19.120185.001345.
20. Mighell AJ, Smith NR, Robinson PA and Markham AF. Vertebrate pseudogenes. FEBS Lett. 2000; 468:109–114. doi:10.1016/S0014-5793(00)01199-6.
21. Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T and Gerstein M. Molecular fossils in the human genome: Identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res. 2002; 12:272–280. doi:10.1101/gr.207102.
22. Balakirev ES and Ayala FJ. Pseudogenes: Are they "junk" or functional DNA?. Annu Rev Genet. 2003; 37:123–151. doi:10.1146/annurev.genet.37.040103.103949.
23. Zhang Z, Harrison PM, Liu Y and Gerstein M. Millions of years of evolution preserved: A comprehensive catalog of the processed pseudogenes in the human genome. Genome Res. 2003; 13:2541–2558. doi:10.1101/gr.1429003.

24. Zhang Z and Gerstein M. Large-scale analysis of pseudogenes in the human genome. Curr Opin Genet Dev. 2004; 14:328–335. doi:10.1016/j.gde.2004.06.003.

25. Zheng D, Zhang Z, Harrison PM, Karro J, Carriero N and Gerstein M. Integrated pseudogene annotation for human chromosome 22: Evidence for transcription. J Mol Biol. 2005; 349:27–45. doi:10.1016/j.jmb.2005.02.072.

26. D'Errico I, Gadaleta G and Saccone C. Pseudogenes in metazoa: origin and features. Brief Funct Genomic Proteomic. 2004; 3:157–167. doi:10.1093/bfgp/3.2.157.

27. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ and Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010; 465:1033–1038. doi:10.1038/nature09144.

28. Arvey A, Larsson E, Sander C, Leslie CS and Marks DS. Target mRNA abundance dilutes microRNA and siRNA activity. Mol Syst Biol. 2010; 6:363. doi:10.1038/msb.2010.24.

29. Salmena L, Poliseno L, Tay Y, Kats L and Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?. Cell. 2011; 146:353–358. doi: 10.1016/j.cell.2011.07.014.

30. Tay Y, Katz L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Poliseno L, Provero P, Di Cunto F, Lieberman J, Rigoutsos I and Pandolfi PP. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell. 2011; 147:344–357. doi: 10.1016/j.cell.2011.09.029.

31. Palmieri D, D'Angelo D, Valentino T, De Martino I, Ferraro A, Wierinckx A, Fedele M, Trouillas J and Fusco A. Downregulation of HMGA-targeting microRNAs has a critical role in human pituitary tumorigenesis. Oncogene. 2012; 31:3857–3865. doi: 10.1038/onc.2011.557.

32. Li W, Yuan Y, Huang L, Qiao M and Zhang Y. Metformin alters the expression profiles of microRNAs in human pancreatic cancer cells. Diabetes Res Clin Pract. 2012; 96:187–195. doi: 10.1016/j.diabres.2011.12.028.

33. Sehic A, Risnes S, Khuu C, Khan QE and Osmundsen H. Effects of in vivo transfection with anti-miR-214 on gene expression in murine molar tooth germ. Physiol Genomics. 2011; 43:488–498. doi:10.1152/physiolgenomics.00248.

34. D'Angelo D, Palmieri D, Mussnich P, Roche M, Wierinckx A, Raverot G, Fedele M, Croce CM, Trouillas J and Fusco A. Altered MicroRNA Expression Profile in Human Pituitary GH Adenomas: Down-Regulation of miRNA Targeting HMGA1, HMGA2, and E2F1. J Clin Endocrinol Metab. 2012; 97:E1128–38. doi:10.1210/jc.2011-3482.

35. Hua Z, Lv Q, Ye W, Wong CKA, Cai G, Gu D, Ji Y, Zhao C, Wang J, Yang BB and Zhang Y. MiRNA-directed regulation of VEGF and other angiogenic factors under hypoxia. PLoS One. 2006; 1:e116. doi:10.1371/journal.pone.0000116.

36. Huang SD, Yuan Y, Zhuang CW, Li BL, Gong DJ, Wang SG, Zeng ZY and Cheng HZ. MicroRNA-98 and microRNA-214 post-transcriptionally regulate enhancer of zeste homolog 2 and inhibit migration and invasion in human esophageal squamous cell carcinoma. Mol Cancer. 2012; 11:51. doi:10.1186/1476-4598-11-51.

37. Day BW, Stringer BW, Al-Ejeh F, Ting MJ, Wilson J, Ensbey KS, Jamieson PR, Bruce ZC, Lim YC, Offenhauser C, Charmsaz S, Cooper LT and Ellacott JK. EphA3 maintains tumorigenicity and is a therapeutic target in glioblastoma multiforme. Cancer Cell. 2013; 23:238–248. doi:10.1016/j.ccr.2013.01.007.

38. Valente V, Serafim RB, de Oliveira LC, Adorni FS, Torrieri R, Tirapelli DP, Espreafico EM, Oba-Shinjo SM, Marie SK, Paçò-Larson ML and Carlotti CG Jr. Modulation of HJURP (Holliday Junction-Recognizing Protein) levels is correlated with glioblastoma cells survival. PLoS One. 2013; 8:e62200. doi:10.1371/journal.pone.0062200.

39. Wang Q, Zhao ZB, Wang G, Hui Z, Wang MH, Pan JF and Zheng H. High expression of KIF26B in breast cancer associates with poor prognosis. PLoS One. 2013; 8:e61640. doi: 10.1371/journal.pone.0061640.

40. Moon E, Lee R, Near R, Weintraub L, Wolda S and Lerner A. Inhibition of PDE3B augments PDE4 inhibitor-induced apoptosis in a subset of patients with chronic lymphocytic leukemia. Clin Cancer Res. 2002; 8:589–595.

41. Hsu A, Zhang W, Lee JF, An j, Ekambaram P, Liu J, Honn KV, Klinge CM and Lee MJ. Sphingosine-1-phosphate receptor-3 signaling up-regulates epidermal growth factor receptor and enhances epidermal growth factor receptor-mediated carcinogenic activities in cultured lung adenocarcinoma cells. Int J Oncol. 2012; 40:1619–1626. doi:10.3892/ijo.2012.1379.

42. Kumar MS, Armenteros-Monterroso E, East P, Chakravorty P, Matthews N and Winslow MM. HMGA2 functions as a competing endogenous RNA to promote lung cancer progression. Nature. 2014; 505:212–217. doi:10.1038/nature12785.

43. Chiappetta G, Tallini G, De Biasio MC, Manfioletti G, Martinez-Tello FJ, Pentimalli F, de Nigris F, Mastro A, Botti G, Fedele M, Berger N, Santoro M, Giancotti V and Fusco A. Detection of high mobility group I HMGI(Y) protein in the diagnosis of thyroid tumors: HMGI(Y) expression represents a potential diagnostic indicator of carcinoma. Cancer Res. 1998; 58:4193–4198.

44. Borbone E, Troncone G, Ferraro A, Jasencakova Z, Stojic L, Esposito F, Horning N, Fusco A and Orlando V. Enhancer of zeste homolog 2 overexpression has a role in the development of anaplastic thyroid carcinomas. J Clin Endocrinol Metab. 2011; 96:1029–1038. doi:10.1210/jc.2010-1784.

45. Palmieri D, Valentino T, De Martino I, Esposito F, Cappabianca P, Wierinckx A, Vitiello M, Lombardi G,

Colao A, Trouillas J, Pierantoni GM, Fusco A and Fedele M. PIT1 upregulation by HMGA proteins has a role in pituitary tumorigenesis. Endocr Relat Cancer. 2012; 19:123–135. doi: 10.1530/ERC-11-0135.

46. Narita M, Narita M, Krizhanovsky V, Nuñez S, Chicas A, Hearn SA, Myers MP and Lowe SW. A novel role for high-mobility group a proteins in cellular senescence and heterochromatin formation. Cell. 2006; 126:503–514. doi:10.1016/j.cell.2006.05.052.
47. Federico A, Forzati F, Esposito F, Arra C, Palma G, Barbieri A, Palmieri D, Fedele M, Pierantoni GM, De Martino I and Fusco A. Hmga1/Hmga2 double knockout mice display a "superpygmy" phenotype. Biol Open. 2014; doi: 10.1242/bio.20146759.
48. Toledo F and Wahl GM. Regulating the p53 pathway: *in vitro* hypotheses, *in vivo veritas*. Nat Rev Cancer. 2006; 6:909–923. doi:10.1038/nrc2012.
49. Esposito F, Boscia F, Gigantino V, Tornincasa M, Fusco A, Franco R and Chieffi P. The high-mobility group A1-estrogen receptor β nuclear interaction is impaired in human testicular seminomas. J Cell Physiol. 2012; 227:3749–3755. doi:10.1002/jcp.24087.
50. Esposito F, Tornincasa M, Chieffi P, De Martino I, Pierantoni GM and Fusco A. High-mobility group A1 proteins regulate p53-mediated transcription of Bcl-2 gene. Cancer Res. 2010; 70:5379–5388. doi:10.1158/0008-5472.CAN-09-4199.
51. Piscuoglio S, Zlobec I, Pallante P, Sepe R, Esposito F, Zimmermann A, Diamantis I, Terracciano L, Fusco A and Karamitopoulou E. HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. Histopathology. 2012; 60:397–404. doi: 10.1111/j.1365-2559.2011.04121.x.

Cell Cycle
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/kccy20

# HMGA1-pseudogene overexpression contributes to cancer progression

Francesco Esposito[a], Marco De Martino[a], Floriana Forzati[a] & Alfredo Fusco[ab]

[a] Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Scuola di Medicina e Chirurgia di Napoli; Università degli Studi di Napoli "Federico II,"; Naples, Italy

[b] Instituto Nacional de Câncer (INCA); Rio de Janeiro, Brazil

Accepted author version posted online: 30 Oct 2014.Published online: 31 Dec 2014.

Cell Cycle 13:23, 3636–3639; December 1, 2014; © 2014 Taylor & Francis Group, LLC

# HMGA1-pseudogene overexpression contributes to cancer progression

Francesco Esposito[1], Marco De Martino[1], Floriana Forzati[1], and Alfredo Fusco[1,2,*]

[1]Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Scuola di Medicina e Chirurgia di Napoli; Università degli Studi di Napoli "Federico II,"; Naples, Italy; [2]Instituto Nacional de Câncer (INCA); Rio de Janeiro, Brazil



**Two pseudogenes for *HMGA1*, whose overexpression has a critical role in cancer progression, have been identified. They act as decoy for miRNAs that are able to target the *HMGA1* gene then enhancing cell proliferation and migration. Moreover, these pseudogenes contain sequences that are potential target sites for cancer-related miRNAs. Interestingly, *HMGA1* pseudogenes are highly expressed in human anaplastic thyroid carcinomas, that is one of the most aggressive tumor in mankind, but almost undetectable in well differentiated thyroid carcinomas.**

The human genome has been sequenced, and no more than 2% codes for proteins building our bodies.[1] Therefore, a key question is to identify which is the meaning of the other 98% of the genome. Many evidences show that the mammalian genome is able to generate numerous previously undiscovered transcripts called "non coding RNA" (ncRNA).[1] However, the role of these ncRNAs remains largely unknown. The ncRNAs include different classes: the "short interfering RNAs" (siRNAs), the "Natural Antisense Transcripts" (NATs), the "microRNAs" (miRNAs), long non coding RNAs (lncRNAs), and pseudogenes.[2]

siRNAs are double-stranded small interfering RNAs of ~21 base pairs in length that serve as effector molecules of sequence-specific gene silencing. They are highly conserved across species.[3]

NATs are RNAs that are at least in part complementary to other endogenous RNAs. They might be transcribed *in cis* from opposing DNA strands at the same genomic locus (*cis*-NATs) or *in trans* at separate loci (*trans*-NATs).[4] NATs biological role is based on gene knockdown induced by base-pairing of sense and anti-sense strands. To date, in human have been identified cis-encoded exon-overlapping sense-antisense (SA) in a number of 7356 NATs.[4]

Among *trans*-NATs the most studied are miRNAs. They are small RNAs (18-21 nt) that can inhibit mRNA expression by binding to 3′ Untranslated Region (3′UTR) with perfect or imperfect match, suppressing mRNA translation or affecting RNA stability.[5] Krek et al. reported that vertebrate *microRNAs* target, usually, almost 200 transcripts each one.[6,7]

Long non coding RNAs (lncRNAs) are non-protein coding transcripts longer than 200 nucleotides.[8] This rather arbitrary limit distinguishes lncRNAs from small regulatory RNAs such as miRNAs, siRNAs, Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), and other short RNAs.[8]

Pseudogenes are dysfunctional relatives of genes that have lost their protein-coding skill or are otherwise no longer expressed in the cell.[9-11] They are characterized by a mixture of homology to a known gene and nonfunctionality: every pseudogene has a DNA sequence that is similar to some functional genes, but they are unable to produce functional final protein products.[9-11]

There are 3 major families of pseudogenes: processed (or retrotransposed), non-processed (or duplicated), and disabled pseudogenes. Processed pseudogenes originate from a segment of mRNA transcript of a gene that is spontaneously reverse transcribed back into DNA and inserted into chromosomal DNA. Once these pseudogenes are inserted back into the genome, they generally include a poly-A tail, and their introns are

*Correspondence to: Alfredo Fusco; Email: alfusco@unina.it

Submitted: 10/02/2014

Revised: 10/06/2014

Accepted: 10/06/2014

http://dx.doi.org/10.4161/15384101.2014.974440

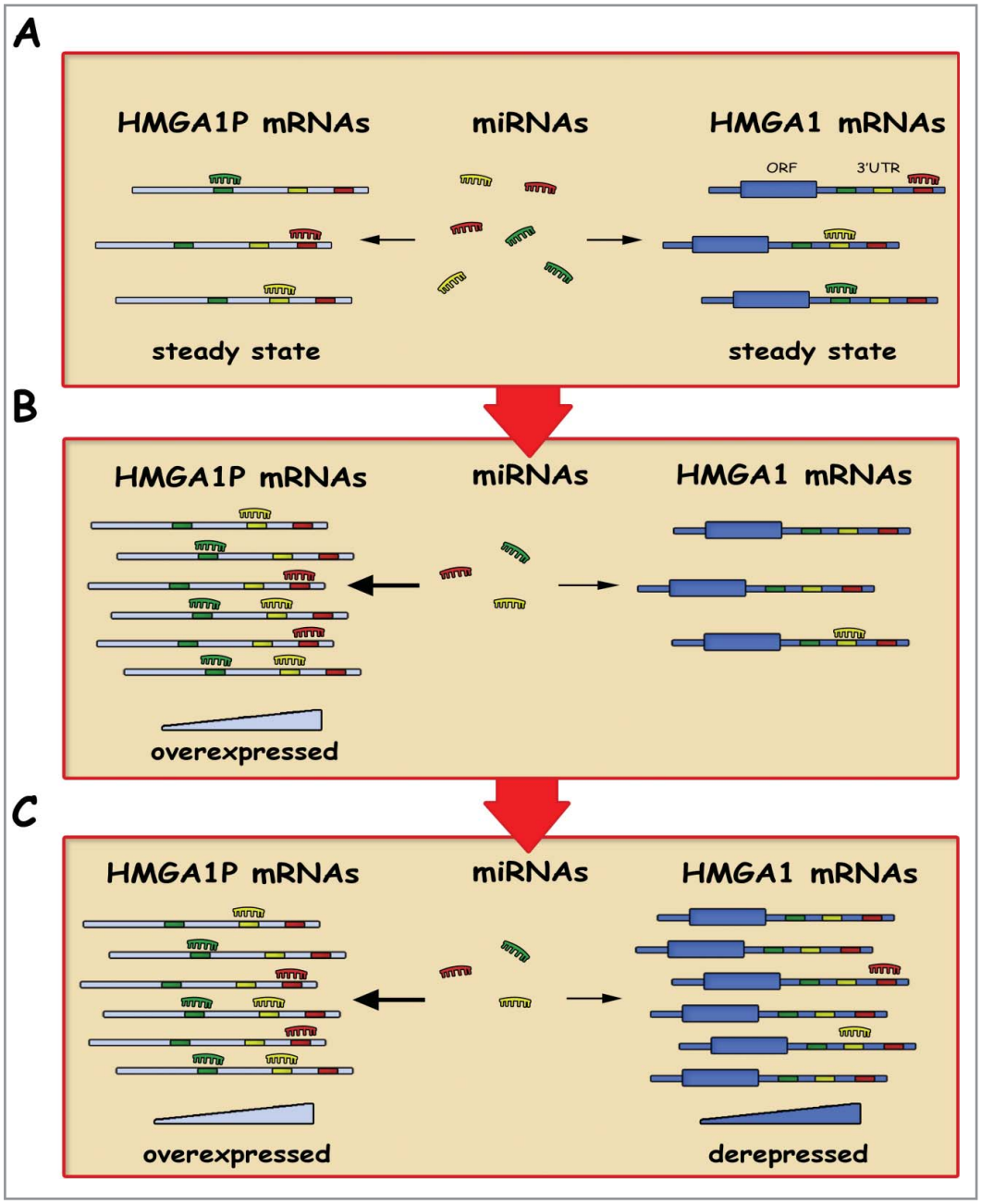



**Figure 1.** Decoy function of *HMGA1Ps*. Several RNA transcripts have seed sequences for the same miRNAs (**A**). As proposed by Esposito et al., overexpression of *HMGA1Ps* mRNAs increases cellular concentrations of particular seed sequences (**B**), resulting in the derepression of *HMGA1* transcripts that contain the same seed sequences (**C**).

frequently spliced out.[12] Non-processed pseudogenes arise as a consequence of a gene duplication event and then acquire mutations making them nonfunctional. Duplicated pseudogenes usually include all the same characteristics of genes from they originate, as well as an intact exon-intron structure and promoter sequences.[12] Disabled genes (or unitary pseudogenes) present different mutations that stop a gene from being productively transcribed or translated.[13]

Recently, Esposito et al. have identified 2 processed pseudogenes, *HMGA1P6* and *HMGA1P7*, belonging to the *HMGA1* gene family. They are placed at 13q12.12 and 6q23.2, respectively, and have been reported to have a critical role in the process of carcinogenesis.[14] The *High Mobility Group A1* (*HMGA1*) gene codes for the HMGA1 proteins, HMGA1a and HMGA1b.[15] These proteins are non-histone chromatinic proteins that bind to DNA and organize chromatin architecture, interacting with several transcription factors, and then regulate the expression of several genes, positively or negatively.[15] Their role in carcinogenesis is widely accepted. Indeed, these proteins are expressed at high levels for the duration of embryogenesis and in malignancies where their expression levels point out a poor prognosis of the cancer patients, whereas their expression is brought down in adult normal tissues.[16-18] Moreover, the *knock-down* of *HMGA1* expression prevents thyroid cells transformation and leads cancer cells of diverse tissue origin to apoptosis.[19-21] On the contrary, their overexpression *in vitro* induces mouse and rat fibroblast transformation,[22] and transgenic mice overexpressing *hmga1* develop several neoplasias including pituitary adenomas, Natural Killer (NK)/T-cell lymphomas,[24] lipomas, cervix and body adenocarcinomas.[15]

*HMGA1* pseudogenes, *HMGA1P6* and *HMGA 1P7*, show just few mismatches all over the coding sequence and the 5' and 3' UTRs of *HMGA1*. They have conserved seed matches for miRNAs that have been previously confirmed to target the *HMGA1* gene. Subsequently, the authors show that these pseudogenes equally operate as decoys for *HMGA1*-targeting miRNAs (**Fig. 1**). Indeed, *HMGA1P6* and *HMGA1P7* overexpression increases HMGA1 protein levels whereas their silencing results in decreased *HMGA1* mRNA and protein levels.[14] Consistently

with the *HMGA1P6* and *HMGA1P7* decoy function, cells overexpressing them show an enhanced migration, invasiveness, and a faster proliferation ability.[14] Opposite results are obtained when these pseudogenes are silenced with also an increase in apoptotic cells following a reduced *HMGA1* protein levels, as already observed when *HMGA1* is knocked down in thyroid cells.[21]

The generation of transgenic mice overexpressing *HMGA1P6* or *HMGA1P7* has confirmed their oncogenic activity. Indeed, mouse embryonic fibroblasts (MEFs) deriving from *HMGA1P6* or *HMGA1P7*[14] overexpressing mice grow faster and senesce later than their wild-type counterparts.

However, we retain that the key point of this study is the finding of a role of these pseudogenes in human carcinogenesis. In fact, anaplastic thyroid carcinoma (ATC), that represents one of the most aggressive tumors in the mankind, evidenced a exceptionally high expression of the *HMGA1* pseudogenes essentially correlating with the HMGA1 protein levels.[14] Conversely, their expression is almost undetectable in papillary and follicular thyroid carcinomas, that are well differentiated and much less aggressive. Analogous results were obtained when *HMGA1Ps* expression was analyzed in human ovarian[14] and larynx carcinomas, and pituitary adenomas (manuscript in preparation).

Remarkably, the presence in the *HMGA1P6*, *HMGA1P7*, and also *HMGA1* UTR regions of sequences that are potential target sites for cancer-related miRNAs targeting genes such as *High Mobility Group A2 (HMGA2), Enhancer of Zeste Homolog 2 (EZH2), Vascular Endothelial Growth Factor (VEGF)*, and *Ephrin Type-A Receptor 3 (Epha3)*, that are effectively upregulated in *HMGA1P6* and *HMGA1P7*[14] overexpressing cells and MEFs with respect to the control cells. This has important consequences. In fact, it means that high *HMGA1* gene or its pseudogene expression allows an increase in HMGA2 and EZH2 protein levels then contributing to cancer progression.[14] Such a mechanism is likely to occur in ATC where the overexpression of *EZH2* has been detected in ATC but not in the undifferentiated thyroid carcinomas.[25,26]

Interestingly, Kumar et al. have recently demonstrated that the *HMGA2* 3′ UTR contains 7 conserved seed sequences for let-7, which has been previously demonstrated to constrain lung cancer development. They identified 6 ceRNA targets that are regulated by *hmga2* in a let7-dependent manner: Transforming growth factor β receptor III (tgfbr3), Angiopoietin-related protein 2 (Angptl2), *Fibronectin Type III Domain Containing Protein 3 (Fndc3)*, *Ski-like protein (Skil)*

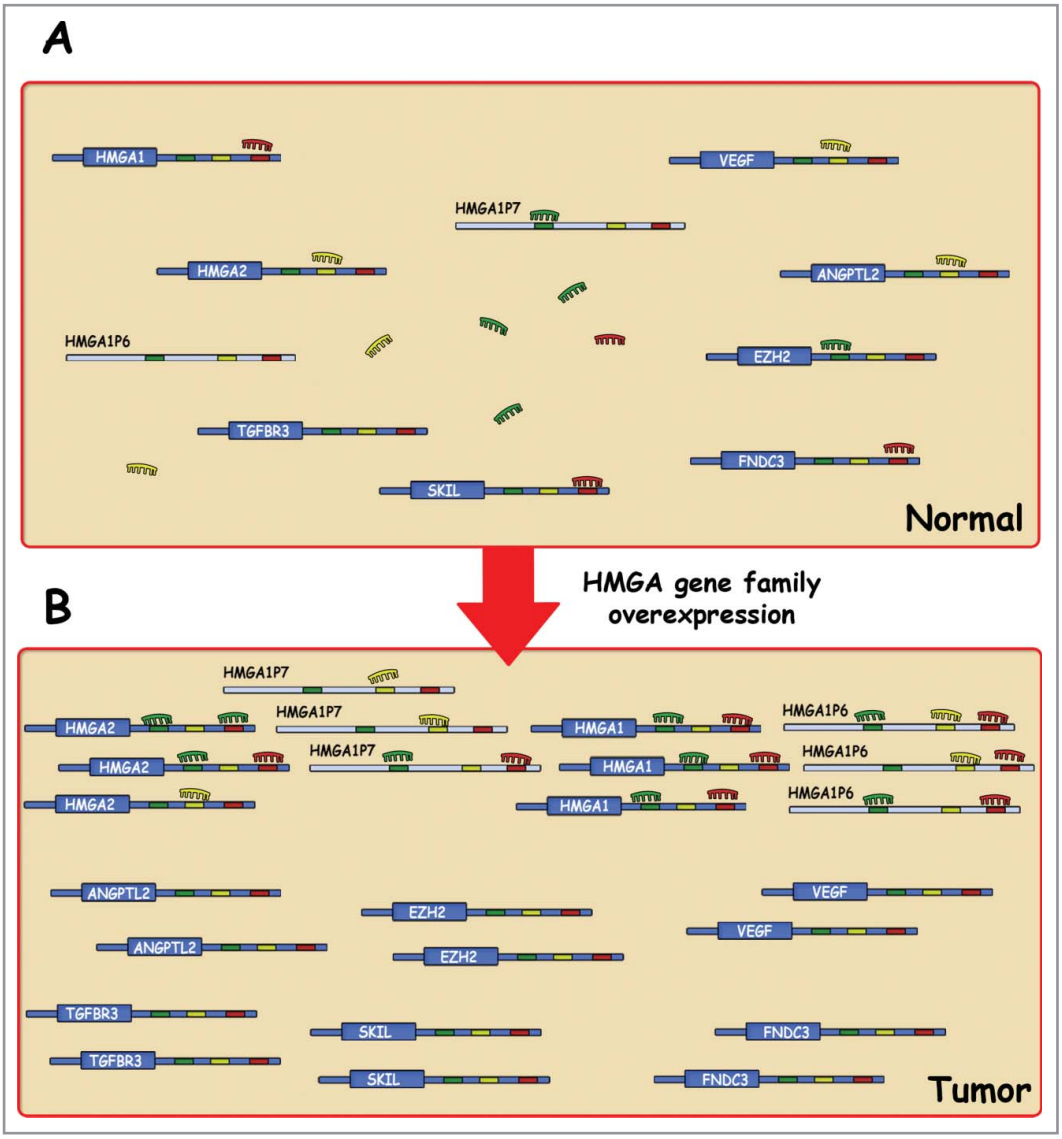



**Figure 2.** A ceRNA model for the *HMGA* gene family. *HMGA* RNA transcripts have seed sequences for the same miRNAs shared with other transcripts (**A**). As proposed by Esposito and Kumar, in malignancies, the HMGA gene family overexpression increases cellular concentrations of particular seed sequences, resulting in the derepression of several cancer-related gene transcripts that contain the same seed sequences (**B**).

and *Hmga1*.[27] Then, on the basis of these results also *HMGA1*, through its 3'UTR, may function as decoy for *HMGA2* and other cancer-related gene expression. Therefore, a synergism between members of HMGA protein family might be envisaged based not only on common functions but also on the ability of the *HMGA* mRNAs to act as decoy for miRNAs able to target themselves.

In conclusion, the results published by Esposito et al. indicate that also the expression of the *HMGA1* pseudogenes contribute to carcinogenesis and, together with the paper by Kumar et al., reveal another mechanism by which *HMGA* gene family has a critical role in cancer progression based on the ability to regulate gene expression also as non-coding RNAs (**Fig. 2**). Therefore, these new reports make even more important the *HMGA* gene family in cancer diagnosis and prognosis, and as potential target for an innovative cancer therapy.



### References


1. Furuno M, Pang KC, Ninomiya N, Fukuda S, Frith MC, Bult C, Kai C, Kawai J, Carninci P, Hayashizaki Y, et al. Clusters of Internally Primed Transcripts Reveal Novel Long Noncoding RNAs. PLoS Genet 2006; 2:e37; PMID:16683026; http://dx.doi.org/10.1371/journal.pgen.0020037
2. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell 2009; 136:629-41; PMID:19239885; http://dx.doi.org/10.1016/j.cell.2009.02.006
3. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. Argonaute2, a link between genetic and biochemical analyses of RNAi. Science 2001; 293:1146-50; PMID:11498593; http://dx.doi.org/10.1126/science.1064023
4. Zhang Y, Li J, Kong L, Gao G, Liu QR, Wei L. NATsDB: Natural Antisense Transcripts DataBase. Nucl Acids Res 2007; 35:D156-61; PMID:17082204; http://dx.doi.org/10.1093/nar/gkl782
5. Zeng Y, Yi R, Cullen BR. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A 2003; 100:9779-84; PMID:12902540; http://dx.doi.org/10.1073/pnas.1630797100
6. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. PLoS Biol 2004; 2:e363; PMID:15502875; http://dx.doi.org/10.1371/journal.pbio.0020363
7. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, et al. Combinatorial microRNA target predictions. Nat Genet 2005; 37:495-500; PMID:15806104; http://dx.doi.org/10.1038/ng1536
8. Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. Gene regulation by the act of long non-coding RNA transcription. BMC Biol 2013; 11:59; PMID:23721193; http://dx.doi.org/10.1186/1741-7007-11-59
9. Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. FEBS Lett 2000; 468:109-14; PMID:10692568; http://dx.doi.org/10.1016/S0014-5793(00)01199-6
10. Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T, Gerstein M. Molecular fossils in the human genome: Identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res 2002; 12:272-80; PMID:11827946; http://dx.doi.org/10.1101/gr.207102
11. Balakirev ES, Ayala FJ. Pseudogenes: Are they "junk" or functional DNA? Annu Rev Genet 2003; 37:123-51; PMID:14616058; http://dx.doi.org/10.1146/annurev.genet.37.040103.103949
12. 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; 465:1033-38; PMID:20577206; http://dx.doi.org/10.1038/nature09144
13. Zhang ZD, Frankish A, Hunt T, Harrow J, Gerstein M. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. Genome Biol 2010; 11:R26; PMID:20210993; http://dx.doi.org/10.1186/gb-2010-11-3-r26
14. Esposito F, De Martino M, Petti MG, Forzati F, Tornincasa M, Federico A, Arra C, Pierantoni GM, Fusco A. HMGA1 pseudogenes as candidate proto-oncogenic competitive endogenous RNAs. Oncotarget 2014; 5:8344-54; PMID:25268743.
15. Fusco A, Fedele M. Roles of the HMGA proteins in cancer. Nat Rev Cancer 2007; 7:899-910; PMID:18004397; http://dx.doi.org/10.1038/nrc2271
16. Chiappetta G, Avantaggiato V, Visconti R, Fedele M, Battista S, Trapasso F, Merciai BM, Fidanza V, Giancotti V, Santoro M, et al. High level expression of the HMGA1 gene during embryonic development. Oncogene 1996; 13:2439-46; PMID:8957086
17. Piscuoglio S, Zlobec I, Pallante P, Sepe R, Esposito F, Zimmermann A, Diamantis I, Terracciano L, Fusco A, Karamitopoulou E. HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. Histopathology 2012; 60:397-404; PMID:22276603; http://dx.doi.org/10.1111/j.1365-2559.2011.04121.x
18. Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del Sal G, Manfioletti G. HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. Oncotarget 2013; 4:1293-1308; PMID:23945276
19. Berlingieri MT, Manfioletti G, Santoro M, Bandiera A, Visconti R, Giancotti V, Fusco A. Inhibition of HMGI-C protein synthesis suppresses retrovirally induced neoplastic transformation of rat thyroid cells. Mol Cell Biol 1995; 15:1545-53; PMID:7862147
20. Berlingieri MT, Pierantoni GM, Giancotti V, Santoro M, Fusco A. Thyroid cell transformation requires the expression of the HMG1 proteins. Oncogene 2002; 21:2971-80; PMID:12082527; http://dx.doi.org/10.1038/sj.onc.1205368
21. Scala S, Portella G, Fedele M, Chiappetta G, Fusco A. Adenovirus mediated suppression of the HMGI(Y) protein synthesis as a potential therapy of human malignant neoplasias. Proc Natl Acad Sci USA. 2000; 97:4256-61; http://dx.doi.org/10.1073/pnas.070029997
22. Vallone D, Battista S, Pierantoni GM, Fedele M, Casalino L, Santoro M, Viglietto G, Fusco A, Verde P. Neoplastic transformation of rat thyroid cells requires the junB and fra-1 gene induction which is dependent on the HMGI-C gene product. EMBO J 1997; 16:5310-21; PMID:9311991; http://dx.doi.org/10.1093/emboj/16.17.5310
23. Palmieri D, Valentino T, De Martino I, Esposito F, Cappabianca P, Wierinckx A, Vitiello M, Lombardi G, Colao A, Trouillas J, et al. PIT1 upregulation by HMGA proteins has a role in pituitary tumorigenesis. Endocr Relat Cancer 2012; 19:123-35; PMID:22199144; http://dx.doi.org/10.1530/ERC-11-0135
24. Fedele M, Pentimalli F, Baldassarre G, Battista S, Klein-Szanto AJ, Kenyon L, Visone R, De Martino I, Ciarmiello A, Arra C, et al. Transgenic mice overexpressing the wild-type form of the HMGA1 gene develop mixed growth hormone/prolactin cell pituitary adenomas and natural killer cell lymphomas. Oncogene 2005; 24:3427-35; PMID:15735694; http://dx.doi.org/10.1038/sj.onc.1208501
25. Borbone E, Troncone G, Ferraro A, Jasencakova Z, Stojic L, Esposito F, Horning N, Fusco A, Orlando V. Enhancer of zeste homolog 2 overexpression has a role in the development of anaplastic thyroid carcinomas. J Clin Endocrinol Metab 2011; 96:1029-38; PMID:21289264; http://dx.doi.org/10.1210/jc.2010-1784
26. Esposito F, Tornincasa M, Pallante P, Federico A, Borbone E, Pierantoni GM, Fusco A. Down-regulation of the miR-25 and miR-30d contributes to the development of anaplastic thyroid carcinoma targeting the polycomb protein EZH2. J Clin Endocrinol Metab 2012; 97:E710-8; PMID:22399519; http://dx.doi.org/10.1210/jc.2011-3068
27. Kumar MS, Armenteros-Monterroso E, East P, Chakravorty P, Matthews N, Winslow MM. HMGA2 functions as a competing endogenous RNA to promote lung cancer progression. Nature 2014; 505:212-17; PMID:24305048; http://dx.doi.org/10.1038/nature12785



Cell Cycle

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/kccy20

# HMGA1-pseudogene expression is induced in human pituitary tumors

Francesco Esposito[a], Marco De Martino[a], Daniela D'Angelo[a], Paula Mussnich[a], Gerald Raverot[bc], Marie-Lise Jaffrain-Rea[de], Filippo Fraggetta[f], Jacqueline Trouillas[bg] & Alfredo Fusco[ah]

[a] Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Medicina Molecolare e Biotecnologie Mediche; Scuola di Medicina e Chirurgia di Napoli; Università degli Studi di Napoli "Federico II", Naples, Italy

[b] U1028; CNRS UMR 5292; Lyon Neuroscience Research Center; Neuro-oncology and Neuroinflammation Team; Lyon, France

[c] Department of endocrinology; Groupement Hospitalier Est; Hospices Civils de Lyon, Lyon, France

[d] Department of Clinical Applied Sciences and Biotechnology; University of L'Aquila; L'Aquila, Italy

[e] Neuromed; IRCCS, Pozzilli, Italy

[f] Department of Pathology; Ospedale Cannizzaro; Catania, Italy

[g] Centre of pathology; Groupement Hospitalier Est; Hospices Civils de Lyon; Lyon, France

[h] Instituto Nacional de Câncer - INCA; Rio de Janeiro, Brazil

Accepted author version posted online: 20 Apr 2015.

To cite this article: Francesco Esposito, Marco De Martino, Daniela D'Angelo, Paula Mussnich, Gerald Raverot, Marie-Lise Jaffrain-Rea, Filippo Fraggetta, Jacqueline Trouillas & Alfredo Fusco (2015) HMGA1-pseudogene expression is induced in human pituitary tumors, Cell Cycle, 14:9, 1471-1475

To link to this article: http://dx.doi.org/10.1080/15384101.2015.1021520

Cell Cycle 14:9, 1471–1475; May 1, 2015; © 2015 Taylor & Francis Group, LLC

# *HMGA1*-pseudogene expression is induced in human pituitary tumors

Francesco Esposito[1], Marco De Martino[1], Daniela D'Angelo[1], Paula Mussnich[1], Gerald Raverot[2,3], Marie-Lise Jaffrain-Rea[4,5], Filippo Fraggetta[6], Jacqueline Trouillas[2,7], and Alfredo Fusco[1,8,*]

[1]Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Medicina Molecolare e Biotecnologie Mediche; Scuola di Medicina e Chirurgia di Napoli; Università degli Studi di Napoli "Federico II", Naples, Italy; [2]U1028; CNRS UMR 5292; Lyon Neuroscience Research Center; Neuro-oncology and Neuroinflammation Team; Lyon, France; [3]Department of endocrinology; Groupement Hospitalier Est; Hospices Civils de Lyon, Lyon, France; [4]Department of Clinical Applied Sciences and Biotechnology; University of L'Aquila; L'Aquila, Italy; [5]Neuromed; IRCCS, Pozzilli, Italy; [6]Department of Pathology; Ospedale Cannizzaro; Catania, Italy; [7]Centre of pathology; Groupement Hospitalier Est; Hospices Civils de Lyon; Lyon, France; [8]Instituto Nacional de Câncer - INCA; Rio de Janeiro, Brazil

**Keywords:** HMGA1P6, HMGA1P7, HMGA1, ceRNA, miRNA



Numerous studies have established that High Mobility Group A (HMGA) proteins play a pivotal role on the onset of human pituitary tumors. They are overexpressed in pituitary tumors, and, consistently, transgenic mice overexpressing either the *Hmga1* or the *Hmga2* gene develop pituitary tumors. In contrast with HMGA2, HMGA1 overexpression is not related to any rearrangement or amplification of the *HMGA1* locus in these tumors. We have recently identified 2 *HMGA1* pseudogenes, *HMGA1P6* and *HMGA1P7*, acting as competitive endogenous RNA decoys for *HMGA1* and other cancer related genes. Here, we show that *HMGA1* pseudogene expression significantly correlates with *HMGA1* mRNA levels in growth hormone and nonfunctioning pituitary adenomas likely inhibiting the repression of *HMGA1* through microRNAs action. According to our functional studies, these *HMGA1* pseudogenes enhance the proliferation and migration of the mouse pituitary tumor cell line, at least in part, through their upregulation. Our results point out that the overexpression of *HMGA1P6* and *HMGA1P7* could contribute to increase *HMGA1* levels in human pituitary tumors, and then to pituitary tumorigenesis.

## Introduction

The High Mobility Group A (HMGA) proteins are non-histone chromatinic proteins involved in transcriptional regulation of gene expression.[1] The HMGA protein family consists of 3 proteins: HMGA1a, HMGA1b, and HMGA2 encoded by 2 different genes, with the HMGA1 proteins being products of the same gene generated through alternative splicing.[1] HMGA overexpression is a feature of malignant neoplasias and its causal role in cell transformation and cancer progression is supported by many studies.[1-3]

We have already reported several evidences that HMGA proteins act as drivers of human pituitary tumors (PT),[4-8] with both the HMGA proteins overexpressed. However, in these tumors, only the overexpression of HMGA2 is associated to gene rearrangement and amplification following trisomy of chromosome 12.[4] Consistently, it is well known that transgenic mice overexpressing either *hmga1* or *hmga2* develop PT,[5-7] and that the HMGA overexpression is associated to the downregulation of several miRNAs able to target both *HMGA1* and *HMGA2* mRNAs in PT (miR-15, miR-16, miR-23b, miR-26a, miR-34b, miR-130b, miR-196a2, miR-326, miR-432, miR-548c-3p, miR-570, miR-603, and Let-7a).[9-11]

We have recently identified 2 *HMGA1* non-coding pseudogenes, *HMGA1P6* and *HMGA1P7*, having conserved seed matches for miRNAs targeting the *HMGA1* gene. The overexpression of *HMGA1* pseudogene (*HMGA1Ps*) increases HMGA1 protein levels, working as competitive endogenous RNA (ceRNA), thereby inhibiting the suppression of HMGA1 protein synthesis by miRNAs.[12,13] Since the *HMGA1Ps* untranslated regions (UTRs) contain also seed sequences for miRNAs able to target HMGA2, their overexpression leads also to increased HMGA2 protein levels. *HMGA1Ps* also show oncogenic activity by inhibiting apoptosis and increasing cell proliferation and migration.[12,13]

The aim of this study has been to investigate the expression and the role of *HMGA1Ps* in PT.

### HMGA1 *and* HMGA1Ps *expression positively correlates in pituitary tumors*

Briefly, we analyzed the expression of *HMGA1Ps* by qRT-PCR in a panel of 41 human PT, including 14 growth hormone (GH) tumors with acromegaly and 27 nonfunctioning pituitary adenomas (NFPA) or gonadotroph FSH-LH tumors detected by immunohistochemistry (IHC). As shown in **Fig. 1** (Panels A and

*Correspondence to: Alfredo Fusco; Email: alfusco@unina.it
Submitted: 12/09/2014; Revised: 02/15/2015; Accepted: 02/15/2015
http://dx.doi.org/10.1080/15384101.2015.1021520

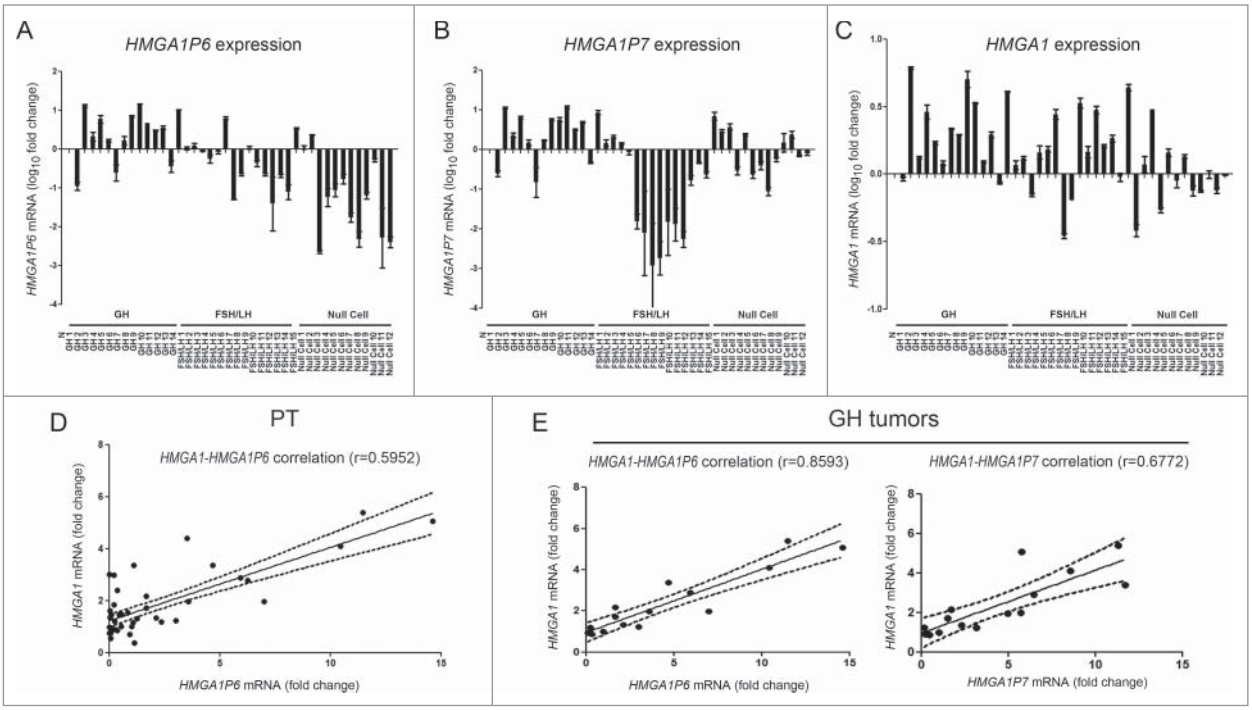



**Figure 1.** *HMGA1* and *HMGA1Ps* expression positively correlates in pituitary tumors. (**A**) *HMGA1P6*, (**B**) *HMGA1P7* and (**C**) *HMGA1* qRT-PCR analysis in normal pituitary gland (N), GH, FSH/LH and null cell tumors. **(D, E)** The obtained values were then combined for correlation analysis. (**D**) Linear regression of *HMGA1* versus *HMGA1P6* in the whole series of PT. (**E**) Linear regressions of *HMGA1* vs. *HMGA1P6* (left panel) and *HMGA1* (right panel) versus *HMGA1P7* in GH tumors.

B), *HMGA1P6* and *HMGA1P7* were differently expressed with regard to the IHC type, when compared with the normal pituitary gland. To verify whether the 2 *HMGA1Ps* may also function as ceRNAs in the regulation of *HMGA1* mRNA levels in PT, we analyzed *HMGA1* expression in these same samples (**Fig. 1C**).

As indicated in **Figure 1**, Panel D, a significant linear correlation was found between *HMGA1* and *HMGA1P6* expression ($r = 0.5952$, $P < 0.0001$), suggesting that these genes are co-regulated, while there is no correlation between *HMGA1* and *HMGA1P7* expression in the whole tumor series. However, further analysis of *HMGA1* and *HMGA1Ps* co-regulation disclosed some differences according to tumor type. Indeed, in the subset of GH tumors, *HMGA1* strongly correlated with *HMGA1P6* expression ($r = 0.8593$, $P < 0 .0001$), and also with *HMGA1P7* expression ($r = 0.6772$, $P < 0.0001$) (**Fig. 1E**). Taken together, these data strongly support the hypothesis that *HMGA1Ps* could act as ceRNAs in PT and represent a novel potential mechanism of *HMGA1* upregulation in these tumors, in particular in GH tumors with acromegaly.

***HMGA1Ps* expression increases AtT20 cell proliferation and migration acting as ceRNAs for HMGA1**

Subsequently, we examined the ability of *HMGA1P6* and *HMGA1P7* to function as a decoy for *HMGA1*-targeting miRNAs in the pituitary cell line AtT20. **Figure 2**, Panel A, confirms successful transfection of both genes. As shown in **Figure 2**, Panel B, overexpression of either *HMGA1P6* or *HMGA1P7* was associated with HMGA1 protein overexpression. As expected from our previously results,[12] overexpression of *HMGA1P6* or *HMGA1P7* reduced the effects exerted by miRNA on the HMGA1 levels (**Fig. 2B**). Then, to evaluate the functional consequences of *HMGA1P6* and *HMGA1P7* overexpression in PT, we investigated their role in cellular proliferation, and migration in AtT20 cells. To this aim, AtT20 cells were transfected with the *HMGA1P6, HMGA1P7*, and the control vector, and counted each day for 9 d. **Figure 2C**, shows that the growth rate of AtT20 following the transfection of the *HMGA1Ps* was higher compared with the cells transfected with the control vector. Finally, we carried out cell migration assays in the same *HMGA1Ps* overexpressing cells especially considering that HMGA1 promotes cell migration.[1] As expected from the increased HMGA1 protein levels in the pseudogene-transfected cells,[12] cell migration was significantly higher in AtT20 cells overexpressing *HMGA1P6* or *HMGA1P7* than in control cells (**Fig. 2D**).

## Discussion

The critical role of High Mobility Group A proteins on the onset of human PT has been widely accepted. In fact, the overexpression of both *HMGA* genes in PT has been reported by numerous studies,[4-8] and consistently the development of GH-PRL tumors is a feature of transgenic mice overexpressing either the *Hmga1* or the *Hmga2* gene.[5, 6]

The results reported here indicate that *HMGA1Ps*, which act as miRNA sponges for *HMGA1* genes, also contribute to

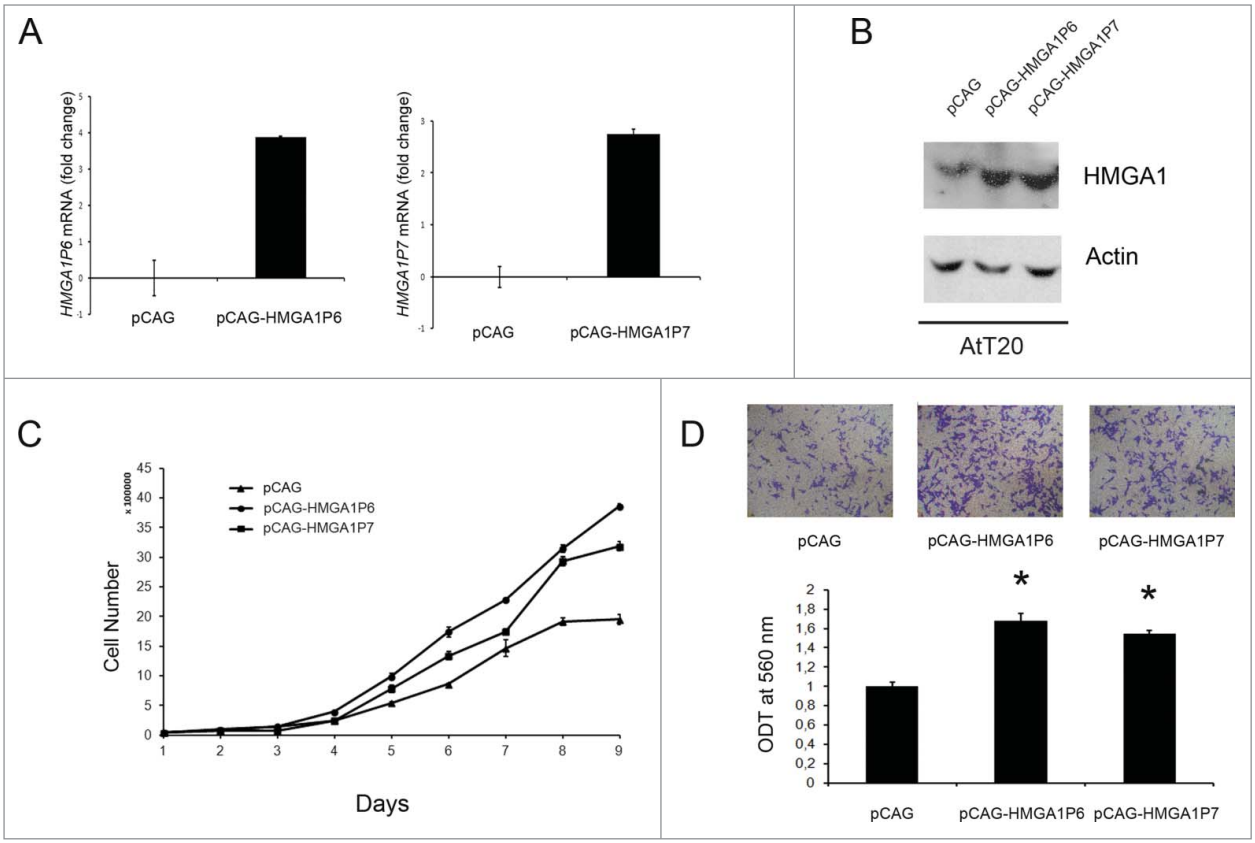



**Figure 2.** *HMGA1Ps* expression increases AtT20 cell proliferation and migration acting as ceRNAs for HMGA1. (**A**) qRT-PCR analysis of *HMGAP6* and *HMGA1P7* mRNA levels in AtT20 cells transfected with the empty, *HMGA1P6* or *HMGA1P7* vectors. (**B**) Western blot analysis of HMGA1 protein levels from the same samples shown in A. (**C**) AtT20 cell proliferation in *HMGA1P6*- and *HMGA1P7*-transfected cells. (**D**) Cell migration assays of AtT20 cells transfected with *HMGA1P6, HMGA1P7* or with a control vector (Upper panel). Migrated cells were quantified and expressed as mean $\pm$ SD *, $P < 0.05$ (*t* test) (Lower panel).

pituitary tumorigenesis by enhancing pituitary cell proliferation and migration.

Since no rearrangement or amplification of the *HMGA1* locus have been detected in PT, where HMGA1 is overexpressed,[4-8] *HMGA1Ps* overexpression contributes to high HMGA1 protein levels detected in PT together with tumor downregulation of miRNAs targeting *HMGA1* (*miR-15, miR-16, miR-23b, miR-26a, miR-34b, miR-130b, miR-196a2, miR-326, miR-432, miR-548c-3p, miR-570, miR-603, and Let-7a*).[9-11]

Indeed, we found a direct correlation between *HMGA1* and *HMGA1Ps* expression in a series of human PT, in particular in the somatotroph ones. Then, functional assays revealed that *HMGA1P6* and *HMGA1P7* increase cell proliferation and migration in pituitary cell line AtT20. This is in agreement with effects of HMGA1, which accelerates the G1-S transition by increasing E2F1 activity, and enhances cell migration in pituitary cell lines.

Noteworthy, *HMGA1P6, HMGA1P7*, as well as *HMGA1* 3' UTR are potential ceRNAs for other cancer-related genes such as *High Mobility Group A2 (HMGA2)* and *Vascular Endothelial Growth Factor* (*VEGF)* which may further contribute to pituitary tumorigenesis.

In conclusion, the results reported here clearly evidence that *HMGA1* pseudogene overexpression contributes to pituitary tumor behavior, thereby disclosing an additional mechanism accounting for the high expression of *HMGA1* (and likely *HMGA2*) in PT. Consequently, these results further support the perspective of an innovative molecular therapy of PT by restoring the expression of miRNAs able to target the *HMGA* genes and/or blocking that of the *HMGA1* pseudogenes.

## Methods

### Cell culture and transfections

AtT20 cells were maintained in DMEM supplemented with 10% foetal calf serum (Euroclone; Milan, Italy), glutamine and antibiotics. Cells were repeatedly tested with MycoAlert (Lonza; Slough, UK) to ensure that cells were not infected with mycoplasma. Cells were transfected using Lipofectamine plus reagent (Life Technologies Italia; Monza, Italy) according to the manufacturer's instructions. The transfected cells were selected in a medium containing geneticin (Sigma; St. Louis, USA). For each transfection, several geneticin-resistant mass cell populations were isolated and expanded for additional analysis. Transfection efficiency was established for each experiment by evaluating GFP

expression. *HMGA1P6* and *HMGA1P7* expression vectors have been previously described.[12]

### Tissue collection and RNA isolation

Surgical samples of PT were obtained from patients operated in 2 centers: Lyon, France (29 tumors - n°1 to n°29) and at the Neuromed Institute, Pozzilli, Italy (12 tumors - n°30 to n°41). Among these, 14 presented with acromegaly and 27 with clinically non-functioning tumors, respectively. According to diagnostic immunohistochemistry with pituitary hormones, immunostaining for GH was confirmed in tumors from acromegalic patients whereas the large majority of clinically non-functioning tumors showed some degree of immunopositivity for FSH and/or LH and the few immunonegative samples were assimilated to gonadotroph tumors. For each tumor, fragments were fixed in the Bouin-Holland fluid or formol and embedded in paraffin for pathological diagnosis, including IHC. Other fragments were immediately frozen and stored at −80°C. We declare that informed consent for the scientific use of biological material was obtained from all patients.

### RNA extraction and quantitative reverse transcription PCR

Total RNA was extracted from tissues with Trizol (Life Technologies Italia; Monza, Italy) according to the manufacturer's instructions. For mRNA detection, total RNA was reverse transcribed by using the QuantiTect Reverse Transcription Kit (Qiagen; Valencia, USA), and then Real-time PCR was performed by using Power SYBR Green PCR Master Mix (Applied Biosystems-Life Technologies Italia; Monza, Italy) and the following primers:

*HMGA1*-Fw 5′-aaggggcagacccaaaaa-3′
*HMGA1*-Rev 5′-tccagtcccagaaggaagc-3′
*HMGA1P6*-Fw 5′-gcagacccacaaaactgga-3′
*HMGA1P6*-Rev 5′-gagcaaagctgtcccatcc-3′
*HMGA1P7*-Fw 5′-gctccttctcggctcctc-3′
*HMGA1P7*-Rev 5′-gcttgggcctcttttatgg-3′
G6PD-Fw 5′-acagagtgagcccttcttcaa-3′
G6PD-Rev 5′-ataggagttgcgggcaaag-3′

The $2^{-\Delta\Delta Ct}$ formula was used to calculate the differential gene expression.

### Protein extraction, western blotting and antibodies

Protein extraction and Western blotting were performed as before described.[14] Antibodies against the HMGA1 proteins are described elsewhere.[15] Blots were visualized by using the Western blotting detection reagents (Thermo Scientific, Waltham, USA).

### Cell migration assay

Cell migration experiments were performed as previously described.[16]

### Statistical analysis

Data were analyzed using a 2-sided unpaired Student's t test (GraphPad Prism, GraphPad Software, Inc.). Values of $P < 0.05$ were considered statistically significant. Regression analyses and correlation coefficients were generated using GraphPad Prism, GraphPad Software, Inc.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.


### Acknowledgments

We thank Mario Berardone for the artwork. We are grateful to E. Jouhanneau, V. Esposito and F. Giangaspero for tumors collection.

### Funding

This work was supported by grants from Associazione Italiana per la Ricerca sul Cancro-AIRC (IG 11477), P.O.R. Campania FSE 2007-2013 - Progetto CREMe - CUP B25B09000050007, "Progetto di Interesse strategico Invecchiamento (PNR-CNR Aging Program) PNR-CNR 2012–2014", Progetto PON01-02782 "Nuove strategie nanotecnologiche per la messa a punto di farmaci e presidi diagnostici diretti verso cellule cancerose circolanti", CNR Epigenomics Flagship Project "EPIGEN".


### References


1. Fusco A, Fedele M. Roles of the HMGA proteins in cancer. Nat Rev Cancer 2007; 7: 899-910; PMID:18004397; http://dx.doi.org/10.1038/nrc2271
2. Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del Sal G, Manfioletti G. HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. Oncotarget 2013; 4: 1293-308; PMID:23945276
3. Puca F, Colamaio M, Federico A, Gemei M, Tosti N, Bastos AU, Del Vecchio L, Pece S, Battista S, Fusco A. HMGA1 silencing restores normal stem cell characteristics in colon cancer stem cells by increasing p53 levels. Oncotarget 2014; 5: 3234-45; PMID:24833610
4. Pierantoni GM, Finelli P, Valtorta E, Giardino D, Rodeschini O, Esposito F, Losa M, Fusco A, Larizza L. High-mobility group A2 gene expression is frequently induced in non-functioning pituitary adenomas (NFPAs), even in the absence of chromosome 12 polysomy. Endocr Relat Cancer 2005; 12: 867-74; PMID:16322327; http://dx.doi.org/10.1677/erc.1.01049
5. Fedele M, Pierantoni GM, Visone R, Fusco A. Critical role of the HMGA2 gene in pituitary adenomas. Cell Cycle 2006; 5: 2045-8; PMID:16969098; http://dx.doi.org/10.4161/cc.5.18.3211
6. Fedele M, Visone R, De Martino I, Troncone G, Palmieri D, Battista S, Ciarmiello A, Pallante P, Arra C, Melillo RM, et al. HMGA2 induces pituitary tumorigenesis by enhancing E2F1 activity. Cancer Cell 2006; 9: 459-71; PMID:16766265; http://dx.doi.org/10.1016/j.ccr.2006.04.024
7. Fedele M, Fusco A. Role of the high mobility group A proteins in the regulation of pituitary cell cycle. J Mol Endocrinol 2010; 44: 309-18; PMID:20219853; http://dx.doi.org/10.1677/JME-09-0178
8. Palmieri D, Valentino T, De Martino I, Esposito F, Cappabianca P, Wierinckx A, Vitiello M, Lombardi G, Colao A, Trouillas J, et al. PIT1 upregulation by HMGA proteins has a role in pituitary tumorigenesis. Endocr Relat Cancer 2012; 19: 123-35; PMID:22199144; http://dx.doi.org/10.1530/ERC-11-0135
9. Palmieri D, D'Angelo D, Valentino T, De Martino I, Ferraro A, Wierinckx A, Fedele M, Trouillas J, Fusco A. Downregulation of HMGA-targeting microRNAs has a critical role in human pituitary tumorigenesis. Oncogene 2012; 31: 3857-65; PMID:22139073; http://dx.doi.org/10.1038/onc.2011.557
10. D'Angelo D, Palmieri D, Mussnich P, Roche M, Wierinckx A, Raverot G, Fedele M, Croce CM, Trouillas J, Fusco A. Altered microRNA expression profile in human pituitary GH adenomas: down-regulation of miRNA targeting HMGA1, HMGA2, and E2F1. J Clin Endocrinol Metab 2012; 97: E1128-38; PMID:22564666; http://dx.doi.org/10.1210/jc.2011-3482
11. Leone V, Langella C, D'Angelo D, Mussnich P, Wierinckx A, Terracciano L, Raverot G, Lachuer J, Rotondi S, Jaffrain-Rea ML, et al. Mir-23b and miR-130b expression is downregulated in pituitary adenomas. Mol Cell Endocrinol 2014; 390: 1-7; PMID:24681352; http://dx.doi.org/10.1016/j.mce.2014.03.002

12. Esposito F, De Martino M, Petti MG, Forzati F, Tornincasa M, Federico A, Arra C, Pierantoni GM, Fusco A. HMGA1 pseudogenes as candidate proto-oncogenic competitive endogenous RNAs. Oncotarget 2014; 5: 8341-54; PMID:25268743
13. Esposito F, De Martino M, Forzati F, Fusco A. HMGA1-pseudogene overexpression contributes to cancer progression. Cell Cycle 2014; 13: 3636-9; PMID:25483074; http://dx.doi.org/10.4161/15384101.2014.974440
14. Martinez Hoyos J, Ferraro A, Sacchetti S, Keller S, De Martino I, Borbone E, Pallante P, Fedele M, Montanaro D, Esposito F, et al. HAND1 gene expression is negatively regulated by the High Mobility Group A1 proteins and is drastically reduced in human thyroid carcinomas. Oncogene 2009; 28: 876-85; PMID:19060921; http://dx.doi.org/10.1038/onc.2008.438
15. Piscuoglio S, Zlobec I, Pallante P, Sepe R, Esposito F, Zimmermann A, Diamantis I, Terracciano L, Fusco A, Karamitopoulou E. HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. Histopathology 2012; 60: 397-404; PMID:22276603; http://dx.doi.org/10.1111/j.1365-2559.2011.04121.x
16. Borbone E, Troncone G, Ferraro A, Jasencakova Z, Stojic L, Esposito F, Horning N, Fusco A, Orlando V. Enhancer of zeste homolog 2 overexpression has a role in the development of anaplastic thyroid carcinomas. J Clin Endocrinol Metab 2011; 96: 1029-38; PMID:21289264; http://dx.doi.org/10.1210/jc.2010-1784

# *HMGA1*-pseudogenes and cancer

**Marco De Martino[1], Floriana Forzati[1], Claudio Arra[2], Alfredo Fusco[1] and Francesco Esposito[1]**

[1] Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Scuola di Medicina e Chirurgia di Napoli, Università degli Studi di Napoli "Federico II", via Pansini 5, Naples, Italy

[2] Istituto Nazionale dei Tumori, Fondazione Pascale, via Mariano Semmola, Naples, Italy

***Correspondence to**: Francesco Esposito, **email**: francesco.esposito2@unina.it*

*Alfredo Fusco, **email**: alfusco@unina.it*

***Keywords***: Pseudogenes, HMGA, cancer, ceRNA

***Received**: November 02, 2015* ***Accepted**: February 05, 2016* ***Published**: February 16, 2016*

## ABSTRACT

**Pseudogenes are DNA sequences with high homology to the corresponding functional gene, but, because of the accumulation of various mutations, they have lost their initial functions to code for proteins. Consequently, pseudogenes have been considered until few years ago dysfunctional relatives of the corresponding ancestral genes, and then useless in the course of genome evolution. However, several studies have recently established that pseudogenes are owners of key biological functions. Indeed, some pseudogenes control the expression of functional genes by competitively binding to the miRNAs, some of them generate small interference RNAs to negatively modulate the expression of functional genes, and some of them even encode functional mutated proteins. Here, we concentrate our attention on the pseudogenes of the *HMGA1* gene, that codes for the HMGA1a and HMGA1b proteins having a critical role in development and cancer progression. In this review, we analyze the family of *HMGA1* pseudogenes through three aspects: classification, characterization, and their possible function and involvement in cancer.**

## INTRODUCTION

The term "pseudogene" comes from the Greek word "pseudo" meaning false. Pseudogenes are also identified as "genomic fossils"[1]. They are outdated DNA sequences that lack protein coding ability because of the presence of frame shift mutations and early or delayed stop codons, even though they look like functional genes [2]. They are considered nonfunctional relatives of ancestral functional genes that might have lost their function during evolution [3]. Pseudogenes have been found in plants [4], bacteria [5], yeast [6], insects [7], nematodes [8] and mammals [9]. Based on their origins, pseudogenes have been classified into: (i) Processed pseudogenes - produced by mRNA retrotransposition [10]; (ii) Duplicated pseudogenes (called unprocessed pseudogenes) - originated from the duplication of functional genes that than become nonfunctional because of acquired mutations [10]; (iii) Unitary or Disabled pseudogenes - originated by mutations in the functional protein coding sequences [2].

Nowadays, pseudogene origin, evolution and function are only partially understood. The first paper about the biological role of a pseudogene was published about 16 years ago [11]. In fact, Korneev et al. reported that the neuronal nitric oxide synthase pseudogene worked as natural antisense in regulating neuronal nitric oxide synthase gene expression. However, recent studies have shown more functional roles for pseudogenes, associating them to long non-coding RNAs (lncRNAs) family [10, 12-14]. In fact, Poliseno et al. reported that *PTEN* pseudogene 1 (*PTENpg1*) is transcribed in human tissues and cancers and harbor microRNA (miRNA) response elements (MREs) for some of the same miRNAs that target its corresponding protein-coding gene, *PTEN* [12]. By sequestering miRNAs that would modulate *PTEN*, the corresponding pseudogene derepresses the protein-coding genes from miRNA regulation [12]. In this way transcripts could talk each other by competing for the same miRNAs, acting as competitive endogenous RNAs (ceRNAs) [15, 16]. In addition, Johnsson et al. characterized two *PTENpg1* antisense RNA isoforms, alpha and beta. The alpha isoform acts on the *PTEN* promoter inhibiting *PTEN* transcription by epigenetic mechanisms. On the contrary, the beta isoform directly interacts with *PTENpg1* RNA,

which influences PTEN protein by changing *PTENpg1* stability and miRNA decoy activity [17]. Therefore, the overexpression of PTENpg1 sustains PTEN expression acquiring oncosuppressive functions [12].

Moreover, the human BRAF pseudogene (*BRAFP1*) has been recently found overexpressed in various tumor types, suggesting that it may contribute to cancer development. Karret et al. demonstrated the ceRNA role of both mouse *Braf-rs1* pseudogene (*Braf-rs1*)and its human ortholog, *BRAFP1*, eliciting the expression of BRAF and the activation of MAPK cascade both *in vitro* and *in vivo*. Indeed, miRNA bioinformatic analysis showed that murine *Braf-rs1* and *B-Raf* share 53 miRNAs, equally human *BRAFP1* and *BRAF* share 40 miRNAs. Thus, the *BRAF* pseudogene, *via* sequestration of common miRNAs, may work as a decoy for *BRAF* in mice and humans then upregulating *BRAF* and promoting MAPK signaling and tumorigenesis. Furthermore, mice overexpressing *Braf-rs1* develop an aggressive tumor similar to the human diffuse large B cell lymphoma. In addition, several transcriptional or genomic aberrations of *BRAFP1* were frequently found in multiple human cancers, including B cell lymphomas [18]. Taken together, pseudogenes are deeply involved in ceRNA hypothesis and give rise to large-scale controlling system across the transcriptome, critically increasing the functional data of human genome and acquiring main roles in physiological and pathological conditions [19].

Besides pseudogene-derived small RNAs have been demonstrated to have a role in chromatin repression [20]. Latest evidences show contribution of pseudogenes in regulating development and disease by encoding peptides or proteins [21-23]. Interestingly, Kandouz et al. detected the expression of *Cx43* pseudogene (*psiCx43*) in several cancer cell lines demonstrating its translationability in a protein of 43 kDa. Moreover, the psiCx43 protein overexpression was able to induce translational inhibition of Cx43 acting as a posttranscriptional regulator of Cx43, whose expression in cancer slows growth and renders the cells more sensitive to cytotoxic chemotherapeutics [23]. Finally, it has been reported that pseudogenes produce small interfering RNAs (siRNAs) in African Trypanosoma brucei and suppress several kinds of functional protein-coding genes through RNA interference pathway [24].

The family of High-Mobility Group A (HMGA) is composed of four proteins, HMGA1a, HMGA1b, HMGA1c, encoded by *HMGA1* gene at the end of alternative splicing, and HMGA2, encoded by the *HMGA2* gene [25]. *HMGA1* is located on chromosome 6p21 in humans and in the t-complex locus on mouse chromosome 17, whereas *HMGA2* is found on chromosome 12q13-15 in humans and at the *pigmy* locus on chromosome 10 in mice [26]. *HMGA1* and *HMGA2* genes are well conserved through the species, in fact, only few differences have been found between the human and the murine *HMGA* sequence

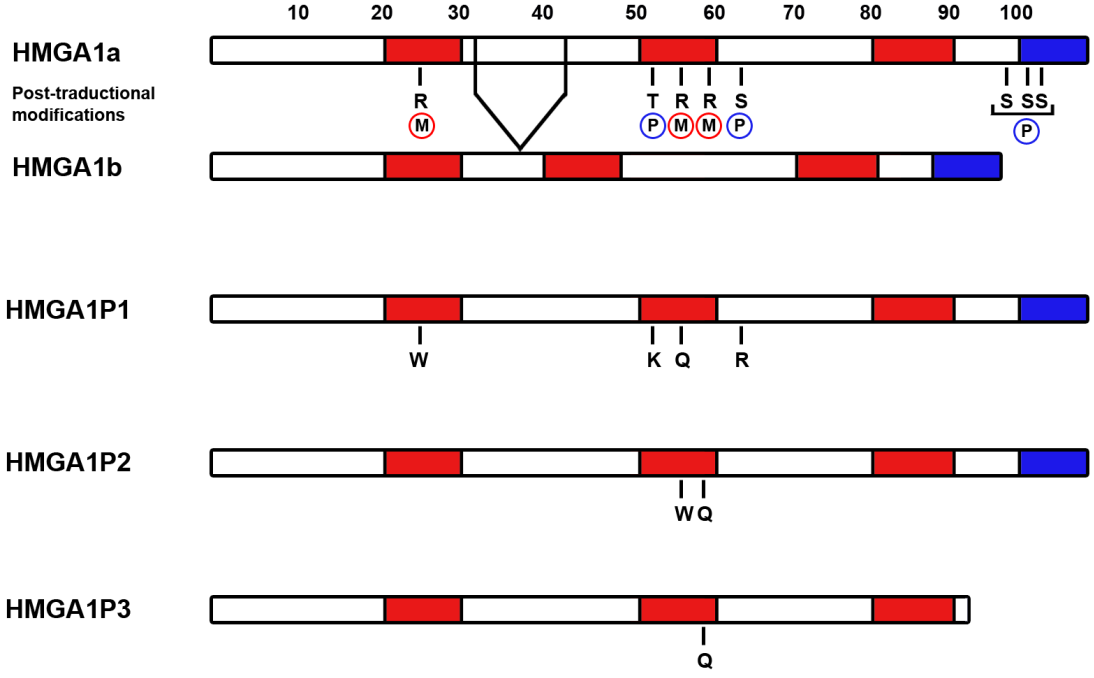



**Figure 1: Structure of HMGA1Ps and their main mutations with respect to HMGA1 proteins.** Diagrams illustrating the domain structures of HMGA1a, HMGA1b, HMGA1P1, HMGA1P2 and HMGA1P3. Known post-traductional modifications of human HMGA1a and HMGA1b proteins impaired in HMGA1Ps are highlighted (Phosphorylation in blue, methylation in red). The three AT-hooks are in red and the acidic tail in blue.

**Table 1: *HMGA1* Pseudogenes family**

| Gene | Location | Function | Main mutations | Reference |
|---|---|---|---|---|
| *HMGA1P1* | Xp21.3 | Competitor protein for HMGA1 with different post-translational modifications.* | arg25, thr53, arg57, ser64 | 91-95,99 |
| *HMGA1P2* | 4q13.3 | Competitor protein for HMGA1 with different post-translational modifications.* | arg57, arg59 | 94,99 |
| *HMGA1P3* | 12q24.11 | Truncated form of HMGA1 with all molecular activities mentioned above.* | arg59, c-terminal tail deletion | 94,96,97,99 |
| *HMGA1P4* | 9q34.11 | - | - | - |
| *HMGA1P5* | 10q22.2 | HMGA1 non-homologous peptide* | - | - |
| *HMGA1P6* | 13q12.12 | Sustains the overexpression of several cancer-related genes by ceRNA mechanism | Stop codon | 100-102 |
| *HMGA1P7* | 6q23.2 | Sustains the overexpression of several cancer-related genes by ceRNA mechanism. | Start codon | 100-102 |
| *HMGA1-p* | 2p13.2 | Competes with *HMGA1* 3' UTR for the binding to αCP1 RNA stability factor. | Few mutations | 106 |

*The function of these pseudogenes still needs to be validated.

[26]. They are non-histone chromosomal proteins, also identified as "architectural transcriptional factors" since they do not show a direct transcriptional activity, but modify the chromatin structure thanks to their DNA-binding domains, called "AT hooks", by which they bind the DNA minor groove at AT-rich nucleotide sequences modulating the gene transcription [27-29]. HMGA proteins are expressed at low levels in adult tissues, but their expression is copious during embryogenesis [26], suggesting their important role in development. Indeed, the phenotypic study of *Hmga1* knock out mice showed that this protein has a critical role in different aspects of development [26, 30]. In particular, cardiac hypertrophy and type 2 diabetes were reported in *Hmga1*-null and heterozygous mice meaning that a correct quantity of HMGA1 protein is necessary for cardiomyocytic cell growth and regulation of the insulin pathway [26, 31-33]. In fact, the downregulation of HMGA1 protein leads to a reduced insulin receptor (INSR) expression in patients with insulin resistance and type 2 diabetes. The recovery of HMGA1 levels improved INSR gene transcription, restoring both expression of insulin receptor protein on cell surface and insulin-binding capacity [31].

It is worth noting that HMGA1 has been found abundantly expressed in all human neoplastic tissues analyzed, including, prostate [34-36], colon [37-39], breast [40-42], gastric [43-45], lung [46-48], testis [49-51], pancreas [52-54], ovary [55-57], thyroid carcinomas [58-60] and also in some forms of leukemia [61-63]. Importantly, HMGA1 expression level has been correlated with an advanced stage, occurrence of distant metastases and reduced survival in colorectal carcinomas [64-67]. To further sustain the HMGA1 function in cancer development, its expression levels have been associated with histologic grade of breast and ovarian carcinomas, where HMGA1 expression gradually enhances from no expression in normal breast tissue, to modest expression in hyperplastic lesions to overexpression in ductal carcinomas [67, 68], and augments from faintly expressed in ovarian carcinomas with low invasive potential to extremely expressed in invasive carcinomas [55, 69]. Importantly, HMGA overexpression plays a causal role in cell transformation. Indeed, their upregulation is able to transform rat fibroblasts [70] and human epithelial breast cells [71] and the block of HMGA1 protein expression prevented thyroid cell transformation induced by Kirsten Murine Sarcoma Virus, and it induced cell death into human thyroid anaplastic cell lines [72, 73]. Furthermore, the silencing of HMGA1 expression in colon cancer stem cells restores normal stem cell characteristics, reducing sphere-forming efficiency and recovering the asymmetric division pattern [39]. Finally, HMGA1 transgenic mice develop several benign or malignant neoplasias, such as GH/PRL-secreting pituitary adenomas, T-cell acute lymphoblastic leukemia and T/NK lymphomas [26, 37, 74].

The molecular processes involved in cell transformation induced by the *HMGA* genes are based on their capacity to positively or negatively control the expression of genes and miRNAs, small noncoding RNAs engaged in gene regulation [75, 76] and implicated in the regulation of cellular proliferation, invasion and apoptosis [77-79].

The upregulation of the *HMGA* genes in cancer

may occur through oncofetal transcriptional mechanisms, which have not been elucidated yet. It is known that the high expression of HMGA1 in cancer cells needs a close cooperation between SP1 family elements and AP1 proteins, stimulated by the activation of Ras GTPase cascade [80]. Furthermore, recent studies have demonstrated the miRNAs HMGA proteins regulation by binding its 3' untranslated region (UTR), provoking mRNA degradation or inhibition of its translation [81, 82]. In particular, several studies reported a strong HMGA regulation by miRNAs in pituitary adenomas (mir-15, mir-16, miR-34b, mir-214, miR-326, miR-432, miR-548c-3p, miR-570, miR-603 and mir-761) [83-85], in thyroid carcinomas (let-7) [86], in breast cancer (mir-26a, miR-33b) [87, 88]. Moreover, the loss of *HMGA2* 3'UTR, commonly found in benign tumors of mesenchymal origin, abolishes the inhibition of HMGA2 expression by several miRNAs [89, 90], leading to HMGA2 protein overexpression that accounts for neoplastic transformation.

## *HMGA1* PSEUDOGENES

The analysis of the human genome by bioinformatic database revealed the presence of eight processed *HMGA1* pseudogenes (*HMGA1Ps*): *HMGA1-p, HMGA1P1*, *HMGA1P2*, *HMGA1P3*, *HMGA1P4*, *HMGA1P5*, *HMGA1P6*, and *HMGA1P7* (Table 1).

## *HMGA1P1* AND *HMGA1P2*

*HMGA1P1* and *HMGA1P2* pseudogenes, classified as processed pseudogenes, are located on Xp21.3 and 4q13.3 chromosome, respectively. They are not conserved during the evolution, but are only found in human genome. There are few mutations that distinguish *HMGA1P1* and *HMGA1P2* from *HMGA1a*. These changes in DNA cause few errors in protein sequence that importantly do not affect their translationability. Indeed, our preliminary data show that expressing vectors for *HMGA1P1* and *HMGA1P2* are able to code for proteins detectable by western blotting analysis. In fact, lysates from HMGA1-null cells transfected with the both vectors were positive to HMGA1 antibodies, which recognize the N-terminal aminoacids shared by HMGA1P1, HMGA1P2 and HMGA1. As shown in Figure 1, some HMGA1P1 and HMGA1P2 mutations hit aminoacidic residues that are frequently modified at post-translational level along the HMGA1 protein. Indeed, *HMGA1P1* is mutated at position 25 where a tryptophan residue substitutes an arginine, within the first AT-hook of HMGA1, which has been shown to be a major site of modification in tumor cells [91]. In fact, Sgarra et al. demonstrated that the arginine residue 25 is strictly related to the execution of programmed cell death in tumor cell lines [92]. *HMGA1P1* is also mutated at threonine residue 53 that is substituted

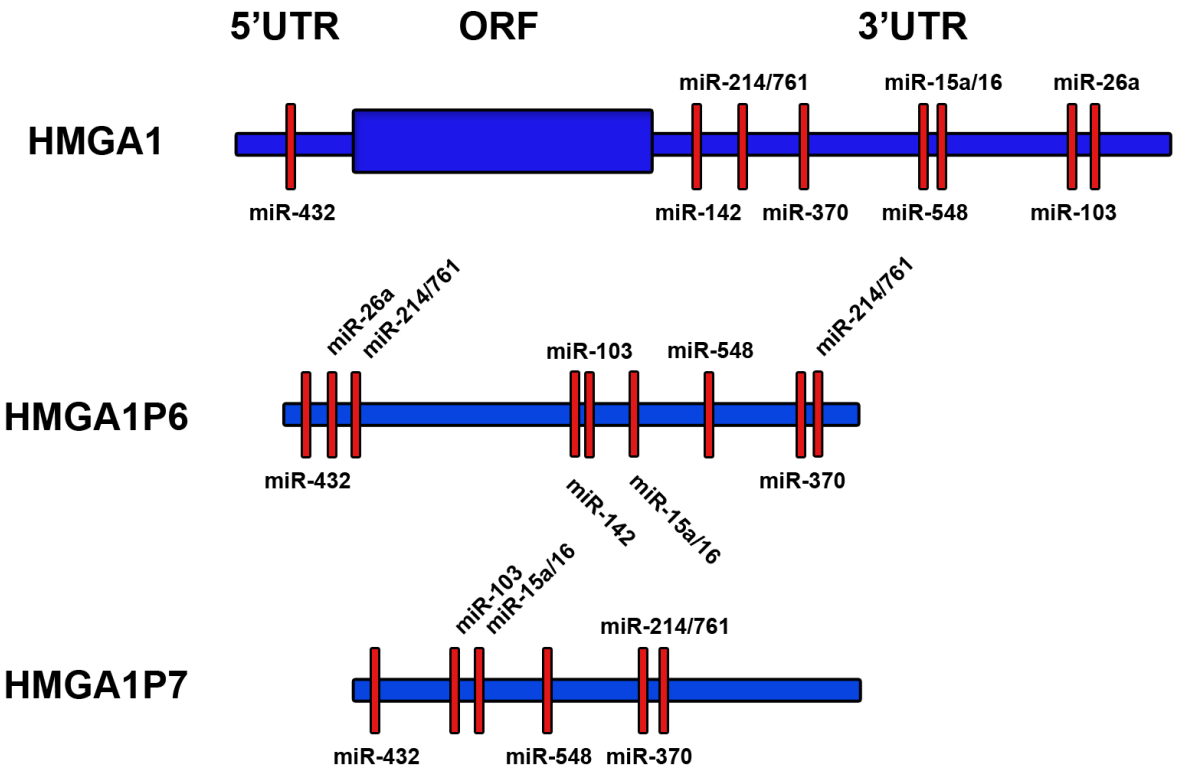



**Figure 2: *HMGA1P6* and *HMGA1P7* mRNA sequence shares HMGA1-targeting miRNAs.** *HMGA1* (top), *HMGA1P6* (middle) and *HMGA1P7* (bottom) mRNA sequences are shown in blue. HMGA1-targeting miRNA seed matches (red boxes) within the high homology regions are shared among *HMGA1*, *HMGA1P6* and *HMGA1P7*.

with a lysine residue. Interestingly, threonine 53 was previously known as the main site of phosphorylation by cdc2 kinase during the cell cycle [93]. Compared with unphosphorylated protein, stoichiometric phosphorylation of recombinant human HMGA1 by cdc2 kinase strongly decreases the binding to DNA. Moreover, the HMGA1 protein arginine residue at position 57, along the second AT-hook, is replaced by a glutamine in HMGA1P1. It has been reported that PRMT6 methylates HMGA1 at the level of arginine 57, which is involved in the affinity for DNA binding and also in protein-protein interaction, thus implying an important role for arginine methylation in modulating HMGA functions [94]. Finally, HMGA1P1 brings a mutation at serine 64 where it shows an arginine residue. This is a Protein Kinase C (PKC) phosphorylation site [95]. Phosphorylation of HMGA1 by PKC resulted in a reduction of DNA-binding affinity as compared with that caused by the phosphorylation with cdc2 kinase, which phosphorylates threonine 53 [91, 95]. Therefore, HMGA1 could be additively phosphorylated by cdc2 kinase and PKC, and the resulting doubly phosphorylated protein exhibits a strong reduction in binding affinity [91, 95].

HMGA1P2 is mutated at arginine residues 57 and 59 where it shows a tryptophan and a glutamine residue, respectively (Figure 1). As well as arginine 57 methylation, it has been reported that HMGA1 arginine 59 is methylated by PRMT6 modulating its ability to bind to DNA and also the protein-protein affinity [94]. Therefore, if *HMGA1P1* and *HMGA1P2* pseudogenes coded for proteins, they could represent a sort of competitor proteins for HMGA1 wild-type with different post-translational modifications, altering HMGA1 properties in chromatin remodeling and protein-protein interactions.

## *HMGA1P3*

*HMGA1P3* pseudogene, only found in human genome, is classified as processed pseudogene and it is located on chromosome 12q24.11. Even though classified as non-coding RNA, it has only four aminoacidic mutations along the protein sequence compared to HMGA1 sequence, without affecting its translationability

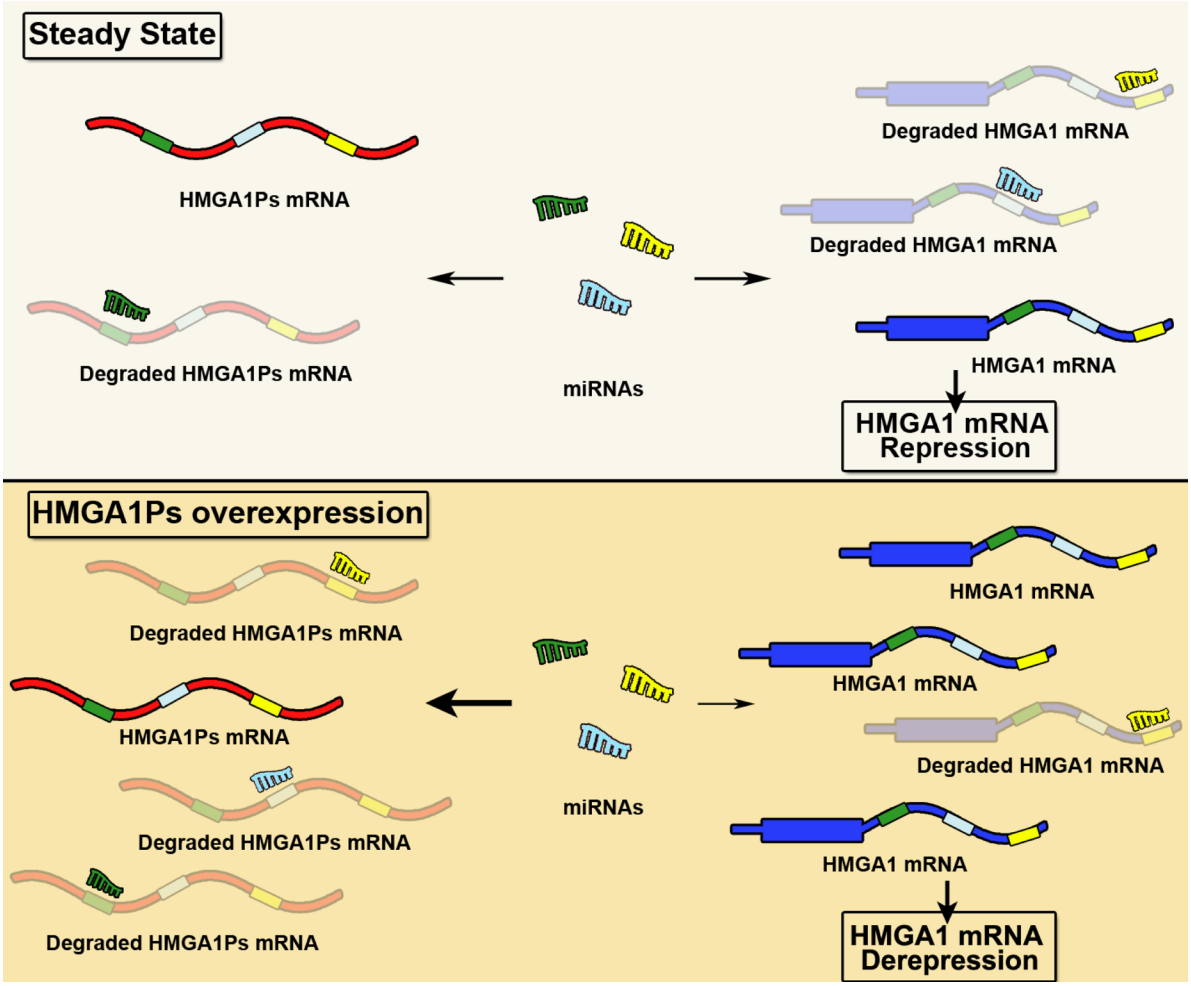



**Figure 3: miRNA decoy function of *HMGA1Ps*.** In the steady state, equilibrium exists between the miRNAs and their targets *HMGA1* and *HMGA1Ps*. By contrast, the overexpression of *HMGA1P6* and *HMGA1P7* results in fewer miRNAs free to bind to *HMGA1*, and thus HMGA1 levels increase.

(Figure 1). Interestingly, arginine 59 is replaced with a glutamine residue as mentioned above for HMGA1P2 [95]. Moreover, *HMGA1P3* pseudogene lacks the C-terminal acidic tail that is a feature of HMGA proteins [26]. In fact, it has been revealed that HMGA1a and HMGA1b are phosphorylated by Casein kinase II (CK2) on three serines situated in the C-terminal tail (S98, S101 and S102) [96, 97]. Moreover, it is believed that the HMGA C-terminal tail may be important in modulating protein-protein interactions [97] and could be involved in enhancing transcription factor activity, but the role of these phosphorylations has been not completely uncovered yet [97]. Finally, expression of a truncated *Hmga1b* gene, without both the acidic tail and the 3' UTR, significantly enhances growth rate and impairs adipocytic differentiation, also suggesting that the *Hmga1/T* mutant works in a contrasting manner [98]. Indeed, transgenic mice overexpressing the HMGA1 wild-type protein showed a reduction of the fat tissue in contrast with the obese phenotype of the *Hmga1/T,* mice even though there are no data that could explain why the wild-type and the truncated form of Hmga1 operate in opposite ways [98].

Given this scenario, if *HMGA1P3* pseudogene coded for protein, it could represent a truncated form of HMGA1 wild-type with all molecular activities mentioned above.

Interestingly, previous studies showed that *HMGA1P1, HMGA1P2* and *HMGA1P3* can be affected by chromosomal rearrangements in benign human tumors [99]. In particular, significantly higher frequency of chromosomal breaks within the chromosomal bands containing these pseudogenes were observed in uterine leiomyomas, lipomas, pleomorphic adenomas, and pulmonary chondroid hamartomas [99]. This study unveils the existence of an interesting pseudogene activation mechanism in tumor, since they could translocate, after chromosomal rearrangements, under a promoter region or within a functional gene, then coding for new fusion proteins.

However, no studies have evaluated *HMGA1P1, HMGA1P2* and *HMGA1P3* expression in human normal and malignant tissues where their possible deregulated

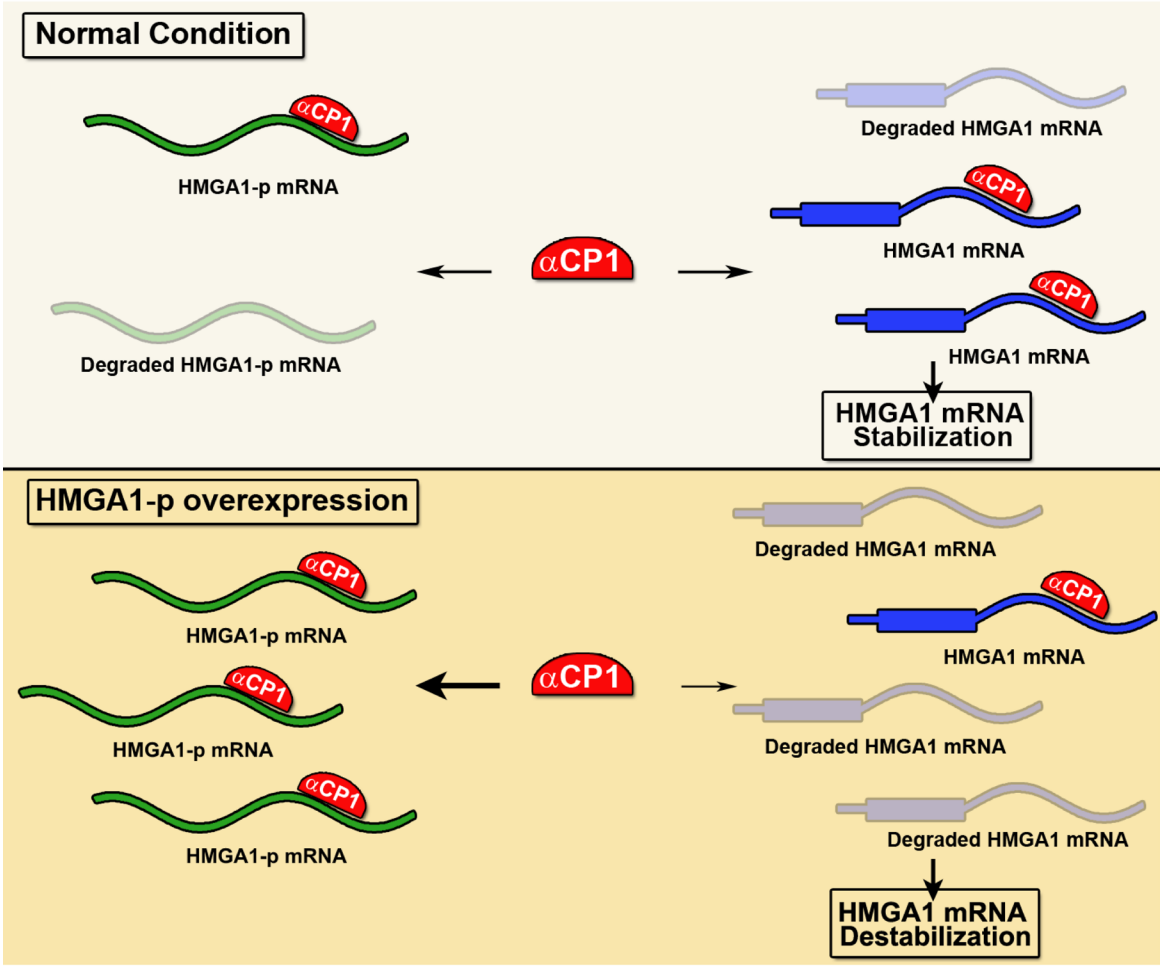



**Figure 4: *HMGA1-p* function model.** In normal condition, the RNA-binding protein αCP1 stabilizes *HMGA1* mRNA by binding to its 3' UTR. In diabetes, the *HMGA1-p* overexpressed transcript competes with *HMGA1* mRNA for the binding to αCP1, increasing the degradation of *HMGA1* mRNA.

expression might have consequences on the function of the wild type HMGA1 protein and then influence cancer progression.

## *HMGA1P4* AND *HMGA1P5*

The non-coding RNA *HMGA1P4* pseudogene is classified as processed pseudogene and is located on the human chromosome 9q34.11. Differently from the above mentioned pseudogenes, *HMGA1P4* genomic sequence shows low homology with *HMGA1*. Moreover, further bioinformatics analysis confirms its untranslationability. Therefore, it could not be classified either as ceRNA or as peptide related to HMGA1.

Another processed pseudogene related to *HMGA1* is *HMGA1P5*. It is present only in humans and located on the chromosome 10q22.2. As *HMGA1P4, HMGA1P5* has low homology along *HMGA1* sequence and it may code for a peptide not related to HMGA1 protein. At the moment there are no published studies about these pseudogenes.

## *HMGA1P6* AND *HMGA1P7*

The processed pseudogenes *HMGA1P6* and *HMGA1P7* are sited on 13q12.12 and 6q23.2 chromosome, respectively. They are not conserved through the evolution, but are present only in human genome [100-102]. These pseudogenes have high sequence homology with *HMGA1* both in the 5' and 3' UTRs and in the coding region (Figure 2). A missense mutation of the start methionine codon avoids *HMGA1P7* mRNA translation whereas *HMGA1P6* bears a mutation in the stop codon, which is postponed several aminoacidic residues downstream, producing a non-translatable mRNA [100-102]. In the homology sequences, among *HMGA1*, *HMGA1P6* and *HMGA1P7*, we retrieved conserved seed matches for miRNAs that have been predicted (miR-103, miR-142-3p, miR-370, and miR-432) or previously validated (miR-15 [83], miR-16 [83], miR-26a [103], miR-214 [104], miR-548c-3p [84] and miR-761 [104]) able to target the *HMGA1* gene (Figure 2).

It has been reported that both *HMGA1P6* and *HMGA1P7* act as decoys for HMGA1-targeting miRNAs. In fact, their overexpression enhances HMGA1 protein levels whereas their knocking down results in the reduction of HMGA1 mRNA and protein amounts (Figure 3) [100-102].

Consistently, these *HMGA1Ps* have also oncogenic action by preventing apoptosis and enhancing cell proliferation and migration [100, 101]. Indeed, overexpression of *HMGA1P6* or *HMGA1P7* increases the growth rate and migration of different cell lines, contributing to tumor development [100-102]. Moreover, the generation of *HMGA1P6* or *HMGA1P7* transgenic mice confirms their oncogenic activity. In fact, mouse embryonic fibroblasts (MEFs) obtained from *HMGA1P6* or *HMGA1P7* [100] transgenic mice grow more rapidly and senesce later than their wild-type counterparts. Remarkably, in *HMGA1P6* and *HMGA1P7* overexpressing cells and MEFs we detected the upregulation of several cancer-related genes such as *High Mobility Group A2* (*HMGA2*), *Enhancer of Zeste Homolog 2* (*EZH2*), *Vascular Endothelial Growth Factor* (*VEGF*), and *Ephrin Type-A Receptor 3* (*Epha3*), with respect to the control cells [100]. This happens because of shared miRNAs targeting *HMGA1P6*, *HMGA1P7*, *HMGA1* and other cancer related genes. Therefore, high *HMGA1* gene or its pseudogene expression allows to increase other oncogene protein levels then contributing to cancer progression. Finally, a direct correlation among *HMGA1*, *HMGA1P6* and *HMGA1P7* expression in a group of human thyroid and ovary tumors has been shown [97-99]. Indeed, papillary thyroid carcinomas (PTC), which are fine differentiated and weakly aggressive, express low levels of *HMGA1P6, HMGA1P7* and *HMGA1*. On the contrary, anaplastic thyroid carcinomas (ATC), which are one of the most malignant human cancers, express very high *HMGA1Ps* levels that, moreover, correlated with HMGA1 protein levels [100]. Similar results were obtained in human ovarian carcinomas [100] and in endometrial carcinomas, where the *HMGA1P6* and *HMGA1P7* expression correlates with the malignancy rate. Interestingly, *HMGA1P6* and *HMGA1P7* were also overexpressed in human pituitary adenomas where the HMGA proteins play a critical role in their development [105]. In particular, *HMGA1P6* and *HMGA1P7* expression significantly correlates with *HMGA1* mRNA in somatotropic and nonfunctioning pituitary adenomas. Moreover, functional studies show that the enforced expression *HMGA1P6* and *HMGA1P7* enhances the proliferation of a pituitary adenoma cell line. Therefore, *HMGA1P6* and *HMGA1P7* overexpression contributes to keep high HMGA1 protein levels enhancing, then, its oncogenic ability.

## *HMGA1-p*

*HMGA1-p* is located on chromosome 2p13.2. Its expression is able to induce destabilization of *HMGA1* mRNA [83]. Indeed, it has been demonstrated that the *HMGA1-p* RNA competes with *HMGA1* 3' UTR for a critical RNA stability factor, the alpha C-binding protein (αCP1) [106]. The *HMGA1-p* was found overexpressed in diabetic patients then causing a significant destabilization of *HMGA1* mRNA with consequent loss of *INSR* expression, which is regulated by HMGA1, then generating the insulin resistance phenotype (Figure 4). Moreover, targeted knockdown of *HMGA1-p* mRNA results in an increase of *HMGA1* mRNA stability and expression levels, with a parallel correction in cell-surface INSR expression and insulin binding capacity [106]. Therefore, this study established a novel mechanistic

linkage between *HMGA1-p* pseudogene expression and type 2 diabetes mellitus.

## CONCLUSIONS AND PERSPECTIVES

The mammalian genome contains an high number of pseudogenes (about 20,000 in humans) [107, 108] more than those present in other organisms. The biological meaning of pseudogenes was completely obscure until few years ago, whereas recent studies have shown their critical role in regulating gene transcription mainly functioning as decoy for miRNAs, and also evidenced a role of pseudogenes in carcinogenesis [109-116]. Interestingly, we have identified, by bioinformatic search, eight pseudogenes for the *HMGA1* gene whose expression is a feature of human malignancies with a key function in promoting cancer progression. From the analysis of the *HMGA1P* sequences it comes up that they could be able to regulate HMGA1 expression and function. Indeed, *HMGA1P6* and *HMGA1P7* act on the stability of *HMGA1* mRNA or by protecting them from miRNAs able to target this gene, whereas *HMGA1-p* competes with *HMGA1* 3' UTR for a critical RNA stability factor, the αCP1. Conversely, HMGA1P1, HMGA1P2 and HMGA1P3 could represent a sort of competitor proteins for HMGA1 wild-type with different post-translational modifications, altering HMGA1 properties in chromatin remodeling and protein-protein interactions. So far, the role of *HMGA1-p* in type 2 diabetes and *HMGA1P6* and *HMGA1P7* in the progression of some human neoplasias appears well documented, but further analysis of their expression in embryonic and adult tissues, and in human carcinomas is required to be deeper investigated. Recently, it has been observed that transgenic mice overexpressing either *HMGA1P6* and *HMGA1P7*, develop lymphomas, infiltrating different organs likely working as ceRNAs for their oncogenic related genes.

Therefore, *HMGA1Ps* represent an epigenetic event, as well as miRNAs, able to regulate HMGA1 activity, and then play a critical role in all the processes such as cancer progression, development, metabolism and many other function in which HMGA1 is involved. The involvement of HMGA1 in all these important cellular processes likely accounts for the need of its fine regulation by using different molecular approaches. Interestingly, recent studies unveil a correlation between *HMGA1P6* and *HMGA1P7* and some clinico-pathological features, opening the perspective of using the evaluation of *HMGA1Ps* expression as diagnostic and prognostic marker, and maybe also in tumor classification. Therefore, the studies summarized here rehabilitate the *HMGA1Ps* from "junk" to a multifunctional pseudogene family that needs to be extensively studied.

## ACKNOWLEDGMENTS

This work was supported by grants from Associazione Italiana per la Ricerca sul Cancro-AIRC (IG 11477), "Progetto di Interesse strategico Invecchiamento (PNR-CNR Aging Program) PNR-CNR 2012-2014" and CNR Epigenomics Flagship Project "EPIGEN".

## REFERENCES

1. Lafontaine I, Dujon B. Origin and fate of pseudogenes in Hemiascomycetes: a comparative analysis. BMC Genomics. 2010;11:260.
2. Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. FEBS Letters. 2000; 468(2-3):109-14.
3. Balakirev ES, Ayala FJ. Pseudogenes: are they "junk" or functional DNA? Annu Rev Genet. 2003;37:123-51.
4. Loguercio LL, Wilkins TA. Structural analysis of a hmg-coA-reductase pseudogene: Insights into evolutionary processes affecting the hmgr gene family in allotetraploid cotton (Gossypium hirsutum L.). Curr Genet. 1998;34(4):241-9.
5. Ochman H, Davalos L. The nature and dynamics of bacterial genomes. Science. 2006;311(5768):1730-3.
6. Harrison P, Kumar A, Lan N, Echols N, Snyder M, Gerstein M. A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. J Mol Biol. 2002;316(3):409-19.
7. Ramos-Onsins S, Aguadé M. Molecular evolution of the Cecropin multigene family in Drosophila: Functional genes *vs*. pseudogenes. Genetics. 1998;150(1):157-71.
8. Harrison PM, Kumar A, Lang N, Snyder M, Gerstein M. A question of size: the eukaryotic proteome and the problems in defining it. Nucleic Acids Res. 2002;30(5):1083-90.
9. Zhang Z, Gerstein M. Large-scale analysis of pseudogenes in the human genome. Current Opinion in Genetics and Development. 2004;14: 328-35.
10. Li W, Yang W, Wang XJ. Pseudogenes: Pseudo or Real Functional Elements? J Genet Genomics. 2013;40(4):171-7.
11. Korneev SA, Park JH, O'Shea M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. J Neurosci. 1999;19(18):7711-20.
12. 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.
13. Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DRF. Pseudogenes: pseudo-functional or key regulators in health and disease? RNA. 2011;17(5):792-8.
14. Peng H, Ishida M, Li L, Saito A, Kamiya A, Iacob R, Dima S, Alexandrescu ST, Grigorie R. Pseudogene INTS6P1 regulates its cognate gene INTS6 through

competitive binding of miR-17-5p in hepatocellular carcinoma. Oncotarget. 2015;6(8):56666-77. doi 10.18632/oncotarget.3290.

15. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353-8.
16. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014;505(7483):344-52.
17. Johnsson P, Ackley A, Vidarsdottir L, Lui W-O, Corcoran M, Grandér D, Morris K V. A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. Nat Struct Mol Biol. 2013;20(4):440-6.
18. Karreth FA, Reschke M, Ruocco A, Ng C, Chapuy B, Léopold V, Sjoberg M, Keane TM, Verma A, Ala U, Tay Y, Wu D, Seitzer N, et al. The BRAF Pseudogene Functions as a Competitive Endogenous RNA and Induces Lymphoma *In Vivo*. Cell. 2015;161(2):319-32.
19. Karreth FA, Ala U, Provero P, Pandolfi PP. Pseudogenes as competitive endogenous RNAs: target prediction and validation. Methods Mol Biol. 2014;1167:199-212.
20. Guo X, Lin M, Rockowitz S, Lachman HM, Zheng D. Characterization of human pseudogene-derived non-coding RNAs for functional potential. PLoS One. 2014;9(4):e93972.
21. Marques AC, Tan J, Lee S, Kong L, Heger A, Ponting CP. Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. Genome Biol. 2012;13(11):R102.
22. Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. Nat Rev Neurosci. 2002;3(7):517-30.
23. Kandouz M, Bier A, Carystinos GD, Alaoui-Jamali MA, Batist G. Connexin43 pseudogene is expressed in tumor cells and inhibits growth. Oncogene. 2004;23(27):4763-70.
24. Wen Y-Z, Zheng L-L, Liao J-Y, Wang M-H, Wei Y, Guo X-M, Qu L-H, Ayala FJ, Lun Z-R. Pseudogene-derived small interference RNAs regulate gene expression in African Trypanosoma brucei. Proc Natl Acad Sci U S A. 2011;108(20):8345-50.
25. Johnson KR, Lehn DA, Reeves R. Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y. Mol Cell Biol. 1989;9(5):2114-23.
26. Fusco A, Fedele M. Roles of HMGA proteins in cancer. Nat Rev Cancer. 2007;7(12):899-910.
27. Grosschedl R, Giese K, Pagel J. HMG domain proteins: Architectural elements in the assembly of nucleoprotein structures. Trends in Genetics. 1994;10(3):94-100.
28. Thanos D, Maniatis T. The high mobility group protein HMG I(Y) is required for NF-kappa B-dependent virus induction of the human IFN-beta gene. Cell. 1992;71(5):777-89.
29. Sepe R, Formisano U, Federico A, Forzati F, Bastos AU, Angelo DD, Cacciola NA, Fusco A, Pallante P. CBX7 and HMGA1b proteins act in opposite way on the regulation of the SPP1 gene expression. Oncotarget. 2015;6(5):2680-92. doi: 10.18632/oncotarget.2777.
30. Federico A, Forzati F, Esposito F, Arra C, Palma G, Barbieri A, Palmieri D, Fedele M, Pierantoni GM, De Martino I, Fusco A. Hmga1/Hmga2 double knock-out mice display a "superpygmy" phenotype. Biol Open. 2014;3(5):372-8.
31. Foti D, Chiefari E, Fedele M, Iuliano R, Brunetti L, Paonessa F, Manfioletti G, Barbetti F, Brunetti A, Croce CM, Fusco A, Brunetti A. Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. Nat Med. 2005;11(7):765-73.
32. Chiefari E, Tanyolaç S, Paonessa F, Pullinger CR, Capula C, Iiritano S, Mazza T, Forlin M, Fusco A, Durlach V, Durlach A, Malloy MJ, Kane JP, et al. Functional variants of the HMGA1 gene and type 2 diabetes mellitus. JAMA. 2011;305(9):903-12.
33. Brunetti A, Chiefari E, Foti D. Recent advances in the molecular genetics of type 2 diabetes mellitus. World J Diabetes. 2014;5(2):128-40.
34. Tamimi Y, Van Der Poel HG, Denyn MM, Umbas R, Karthaus HF, Debruyne FM, Schalken JA. Increased expression of high mobility group protein I(Y) in high grade prostatic cancer determined by *in situ* hybridization. Cancer Res. 1993;53:5512-6
35. Tamimi Y, Van Der Poel HG, Karthaus HF, Debruyne FM, Schalken JA. A retrospective study of high mobility group protein I(Y) as progression marker for prostate cancer determined by *in situ* hybridization. Br J Cancer. 1996;74(4): 573-8.
36. Wei JJ, Wu X, Peng Y, Shi G, Olca B, Yang X, Daniels G, Osman I, Ouyang J, Hernando E, Pellicer A, Rhim JS, Melamed J, et al. Regulation of HMGA1 expression by MicroRNA-296 affects prostate cancer growth and invasion. Clin Cancer Res. 2011;17(6):1297-305.
37. Belton A, Gabrovsky A, Bae YK, Reeves R, Iacobuzio-Donahue C, Huso DL, Resar LM. HMGA1 induces intestinal polyposis in transgenic mice and drives tumor progression and stem cell properties in colon cancer cells. PLoS One. 2012;7(1):e30034.
38. Bush BM, Brock AT, Deng JA, Nelson RA, Sumter TF. The Wnt/β-catenin/T-cell factor 4 pathway up-regulates high-mobility group A1 expression in colon cancer. Cell Biochem Funct. 2013;31(3):228-36.
39. Puca F, Colamaio M, Federico A, Gemei M, Bastos AU, Vecchio L Del, Pece S, Fusco A. HMGA1 silencing restores normal stem cell characteristics in colon cancer stem cells by increasing p53 levels Oncotarget. 2014;5(10):3234-45. doi: 10.18632/oncotarget.1914.
40. Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del Sal G, Manfioletti G. HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. Oncotarget. 2013;4(8):1293-

308. doi: 10.18632/oncotarget.1136.

41. Sepe R, Piscuoglio S, Quintavalle C, Perrina V, Quagliata L, Formisano U, Terracciano LM, Fusco A, Pallante P. HMGA1 overexpression is associated with a particular subset of human breast carcinomas. J Clin Pathol. 2015. pii: jclinpath-2015-202907.

42. Maurizio E, Wiśniewski JR, Ciani Y, Amato A, Arnoldo L, Penzo C, Pegoraro S, Giancotti V, Zambelli A, Piazza S, Manfioletti G, Sgarra R. Translating Proteomic Into Functional Data: An High Mobility Group A1 (HMGA1) Proteomic Signature Has Prognostic Value in Breast Cancer. Mol Cell Proteomics. 2016 ;15(1):109-23

43. Nam ES, Kim DH, Cho SJ, Chae SW, Kim HY, Kim SM, Han JJ, Shin HS, Park YE. Expression of HMGI(Y) associated with malignant phenotype of human gastric tissue. Histopathology. 2003 May;42(5):466-71.

44. Akaboshi S, Watanabe S, Hino Y, Sekita Y, Xi Y, Araki K, Yamamura K, Oshima M, Ito T, Baba H, Nakao M. HMGA1 is induced by Wnt/beta-catenin pathway and maintains cell proliferation in gastric cancer. Am J Pathol. 2009;175(4):1675-85.

45. Jun KH, Jung JH, Choi HJ, Shin EY, Chin HM. HMGA1/HMGA2 protein expression and prognostic implications in gastric cancer. Int J Surg. 2015;24(Pt A):39-44.

46. Hillion J, Wood LJ, Mukherjee M, Bhattacharya R, Di Cello F, Kowalski J, Elbahloul O, Segal J, Poirier J, Rudin CM, Dhara S, Belton A, Joseph B, Zucker S, Resar LMS. Upregulation of MMP-2 by HMGA1 promotes transformation in undifferentiated, large-cell lung cancer. Mol Cancer Res. 2009;7(11):1803-12.

47. Zhang Z, Wang Q, Chen F, Liu J. Elevated expression of HMGA1 correlates with the malignant status and prognosis of non-small cell lung cancer. Tumour Biol. 2015;36(2):1213-9.

48. Pallante P, Sepe R, Puca F, Fusco A. High mobility group a proteins as tumor markers. Front Med (Lausanne). 2015;2:15.

49. Franco R, Esposito F, Fedele M, Liguori G, Pierantoni GM, Botti G, Tramontano D, Fusco A, Chieffi P. Detection of high-mobility group proteins A1 and A2 represents a valid diagnostic marker in post-pubertal testicular germ cell tumours. J Pathol. 2008;214(1):58-64.

50. Esposito F, Boscia F, Gigantino V, Tornincasa M, Fusco A, Franco R, Chieffi P. The high-mobility group A1-estrogen receptor β nuclear interaction is impaired in human testicular seminomas. J Cell Physiol. 2012;227(12):3749-55.

51. Chieffi P. Recent advances in molecular and cell biology of testicular germ-cell tumors. Int Rev Cell Mol Biol. 2014;312:79-100.

52. Trapasso F, Sarti M, Cesari R, Yendamuri S, Dumon KR, Aqeilan RI, Pentimalli F, Infante L, Alder H, Abe N, Watanabe T, Viglietto G, Croce CM, Fusco A. Therapy of human pancreatic carcinoma based on suppression of HMGA1 protein synthesis in preclinical models. Cancer Gene Ther. 2004;11(9):633-41.

53. Piscuoglio S, Zlobec I, Pallante P, Sepe R, Esposito F, Zimmermann A, Diamantis I, Terracciano L, Fusco A, Karamitopoulou E. HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. Histopathology. 2012;60(3):397-404.

54. Hillion J, Smail SS, Di Cello F, Belton A, Shah SN, Huso T, Schuldenfrei A, Nelson DM, Cope L, Campbell N, Karikari C, Aderinto A, Maitra A, Huso DL, Resar LM. The HMGA1-COX-2 axis: a key molecular pathway and potential target in pancreatic adenocarcinoma. Pancreatology. 2012;12(4):372-9.

55. Masciullo V, Baldassarre G, Pentimalli F, Berlingieri MT, Boccia A, Chiappetta G, Palazzo J, Manfioletti G, Giancotti V, Viglietto G, Scambia G, Fusco A. HMGA1 protein over-expression is a frequent feature of epithelial ovarian carcinomas. Carcinogenesis. 2003;24(7):1191-8.

56. Treff NR, Dement GA, Adair JE, Britt RL, Nie R, Shima JE, Taylor WE, Reeves R. Human KIT ligand promoter is positively regulated by HMGA1 in breast and ovarian cancer cells. Oncogene. 2004;23(52):8557-62.

57. Liu Y, Wang Y, Zhang Y, Fu J, Zhang G. Knockdown of HMGA1 expression by short/small hairpin RNA inhibits growth of ovarian carcinoma cells. Biotechnol Appl Biochem. 2012;59(1):1-5.

58. Czyz W, Balcerczak E, Jakubiak M, Pasieka Z, Kuzdak K, Mirowski M. HMGI(Y) gene expression as a potential marker of thyroid follicular carcinoma. Langenbecks Arch Surg. 2004;389(3):193-7.

59. Frasca F, Rustighi A, Malaguarnera R, Altamura S, Vigneri P, Del Sal G, Giancotti V, Pezzino V, Vigneri R, Manfioletti G. HMGA1 inhibits the function of p53 family members in thyroid cancer cells. Cancer Res. 2006;66(6):2980-9.

60. Martinez Hoyos J, Ferraro A, Sacchetti S, Keller S, De Martino I, Borbone E, Pallante P, Fedele M, Montanaro D, Esposito F, Cserjesi P, Chiariotti L, Troncone G, Fusco A. HAND1 gene expression is negatively regulated by the High Mobility Group A1 proteins and is drastically reduced in human thyroid carcinomas. Oncogene. 2009;28(6):876-85.

61. Pierantoni GM, Agosti V, Fedele M, Bond H, Caliendo I, Chiappetta G, Lo Coco F, Pane F, Turco MC, Morrone G, Venuta S, Fusco A. High-mobility group A1 proteins are overexpressed in human leukaemias. Biochem J. 2003;372(Pt 1):145-50.

62. Roy S, Di Cello F, Kowalski J, Hristov AC, Tsai HL, Bhojwani D, Meyer JA, Carroll WL, Belton A, Resar LM. HMGA1 overexpression correlates with relapse in childhood B-lineage acute lymphoblastic leukemia. Leuk Lymphoma. 2013;54(11):2565-7.

63. Mahajan L, Pandit H, Madan T, Gautam P, Yadav AK, Warke H, Sundaram CS, Sirdeshmukh R, Sarma PU,

Kishore U, Surolia A. Human surfactant protein D alters oxidative stress and HMGA1 expression to induce p53 apoptotic pathway in eosinophil leukemic cell line. PLoS One. 2013;8(12):e85046.

64. Abe N, Watanabe T, Sugiyama M, Uchimura H, Chiappetta G, Fusco A, Atomi Y. Determination of high mobility group I(Y) expression level in colorectal neoplasias: a potential diagnostic marker. Cancer Res. 1999 59:1169-74.

65. Chiappetta G, Manfioletti G, Pentimalli F, Abe N, Di Bonito M, Vento MT, Giuliano A, Fedele M, Viglietto G, Santoro M, Watanabe T, Giancotti V, Fusco A. High mobility group HMGI(Y) protein expression in human colorectal hyperplastic and neoplastic diseases. Int J Cancer. 2001;91(2):147-51.

66. Huang ML, Chen CC, Chang LC. Gene expressions of HMGI-C and HMGI(Y) are associated with stage and metastasis in colorectal cancer. Int J Colorectal Dis. 2009;24(11):1281-6.

67. Rogalla P, Drechsler K, Kazmierczak B, Rippe V, Bonk U, Bullerdiek J. Expression of HMGI-C, a member of the high mobility group protein family, in a subset of breast cancers: relationship to histologic grade. Mol Carcinog. 1997;19(3):153-6.

68. Chiappetta G, Botti G, Monaco M, Pasquinelli R, Pentimalli F, Di Bonito M, D'Aiuto G, Fedele M, Iuliano R, Palmieri EA, Pierantoni GM, Giancotti V, Fusco A. HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression. Clin Cancer Res. 2004;10(22):7637-44.

69. Peters DG, Kudla DM, Deloia JA, Chu TJ, Fairfull L, Edwards RP, Ferrell RE. Comparative gene expression analysis of ovarian carcinoma and normal ovarian epithelium by serial analysis of gene expression. Cancer Epidemiol Biomarkers Prev. 2005;14(7):1717-23.

70. Wood LJ, Maher JF, Bunton TE, Resar LMS. The oncogenic properties of the HMG-I gene family. Cancer Res. 2000;60(15):4256-61.

71. Reeves R, Edberg DD, Li Y. Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. Mol Cell Biol. 2001;21(2):575-94.

72. Scala S, Portella G, Fedele M, Chiappetta G, Fusco a. Adenovirus-mediated suppression of HMGI(Y) protein synthesis as potential therapy of human malignant neoplasias. Proc Natl Acad Sci U S A. 2000;97(8):4256-61.

73. Berlingieri MT, Pierantoni GM, Giancotti V, Santoro M, Fusco A. Thyroid cell transformation requires the expression of the HMGA1 proteins. Oncogene. 2002;21(19):2971-80.

74. Xu Y, Sumter TF, Bhattacharya R, Tesfaye A, Fuchs EJ, Wood LJ, Huso DL, Resar LMS. The HMG-I oncogene causes highly penetrant, aggressive lymphoid malignancy in transgenic mice and is overexpressed in human leukemia. Cancer Res. 2004;64(10):3371-5.

75. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific MicroRNAs from mouse. Curr Biol. 2002;12(9):735-9.

76. Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. Nat Rev Cancer. 2015;15(6):321-33.

77. Carleton M, Cleary MA, Linsley PS. MicroRNAs and cell cycle regulation. Cell Cycle. 2007;6(17):2127-32.

78. Ma L, Weinberg R a. MicroRNAs in malignant progression. Cell Cycle. 2008;7(5):570-2.

79. Su Z, Yang Z, Xu Y, Chen Y, Yu Q. MicroRNAs in apoptosis, autophagy and necroptosis. Oncotarget. 2015;6(11): 8474-90. doi: 10.18632/oncotarget.3523.

80. Cleynen I, Huysmans C, Sasazuki T, Shirasawa S, Van De Ven W, Peeters K. Transcriptional control of the human high mobility group A1 gene: Basal and oncogenic Ras-regulated expression. Cancer Res. 2007;67(10):4620-9.

81. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 2004; 116(2):281-97.

82. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6(11):857-66.

83. Palmieri D, D'Angelo D, Valentino T, De Martino I, Ferraro A, Wierinckx A, Fedele M, Trouillas J, Fusco A. Downregulation of HMGA-targeting microRNAs has a critical role in human pituitary tumorigenesis. Oncogene. 2012; 31(34):3857-65.

84. D'Angelo D, Palmieri D, Mussnich P, Roche M, Wierinckx A, Raverot G, Fedele M, Croce CM, Trouillas J, Fusco A. Altered microRNA expression profile in human pituitary GH adenomas: Down-regulation of miRNA targeting HMGA1, HMGA2, and E2F1. J Clin Endocrinol Metab. 2012;97(7):E1128-38.

85. D'Angelo D, Esposito F, Fusco A. Epigenetic Mechanisms Leading to Overexpression of HMGA Proteins in Human Pituitary Adenomas. Front Med. 2015;2:39.

86. Pallante P, Battista S, Pierantoni GM, Fusco A. Deregulation of microRNA expression in thyroid neoplasias. Nat Rev Endocrinol. 2014;10(2):88-101.

87. Zhao X, Yuan Q, Mu D, Sun D, Bo Q, Pan G, Li G, Cui T. MicroRNA-26a inhibits proliferation by targeting high mobility group AT-hook 1 in breast cancer. 2015;8(1):368-73.

88. Lin Y, Liu AY, Fan C, Zheng H, Li Y, Zhang C, Wu S, Yu D, Huang Z, Liu F, Luo Q, Yang CJ, Ouyang G. MicroRNA-33b Inhibits Breast Cancer Metastasis by Targeting HMGA2, SALL4 and Twist1. Sci Rep. 2015;5:9995.

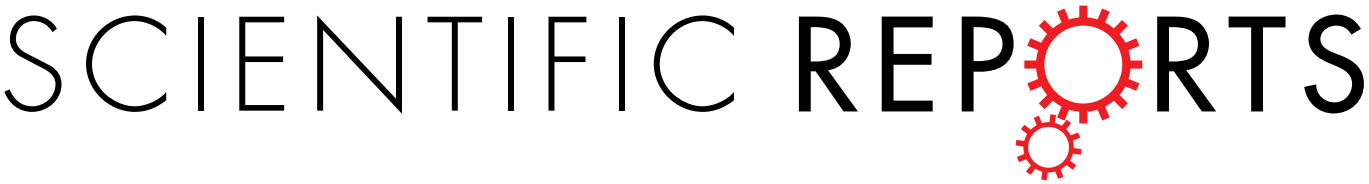
89. Hebert C, Norris K, Scheper MA, Nikitakis N, Sauk JJ. High mobility group A2 is a target for miRNA-98 in head and neck squamous cell carcinoma. Mol Cancer. 2007;6:5.

90. Yong SL, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev. 2007;21(9):1025-30.

91. Zhang Q, Wang Y. High mobility group proteins and their post-translational modifications. Biochim Biophys Acta.

2008;1784(9):1159-66.

92. Sgarra R, Diana F, Bellarosa C, Dekleva V, Rustighi A, Toller M, Manfioletti G, Giancotti V. During apoptosis of tumor cells HMGA1a protein undergoes methylation: Identification of the modification site by mass spectrometry. Biochemistry. 2003;42(12):3575-85.
93. Nissen MS, Langan TA, Reevest R. Phosphorylation by cdc2 kinase modulates DNA binding activity of high mobility group I nonhistone chromatin protein. J Biol Chem. 1991;266(30):19945-52.
94. Sgarra R, Lee J, Tessari MA, Altamura S, Spolaore B, Giancotti V, Bedford MT, Manfioletti G. The AT-hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6. J Biol Chem. 2006;281(7):3764-72.
95. Reeves R. Molecular biology of HMGA proteins: Hubs of nuclear function. Gene. 2001;277(1-2):63-81.
96. Currie RA. Functional interaction between the DNA binding subunit trimerization domain of NF-Y and the high mobility group protein HMG-I(Y). J Biol Chem. 1997;272(49):30880-8.
97. Sgarra R, Maurizio E, Zammitti S, Lo Sardo A, Giancotti V, Manfioletti G. Macroscopic differences in HMGA oncoproteins post-translational modifications: C-terminal phosphorylation of HMGA2 affects its DNA binding properties. J Proteome Res. 2009;8(6):2978-89.
98. Pierantoni GM, Battista S, Pentimalli F, Fedele M, Visone R, Federico A, Santoro M, Viglietto G, Fusco A. A truncated HMGA1 gene induces proliferation of the 3T3-L1 preadipocytic cells: a model of human lipomas. Carcinogenesis. 2003;24(12):1861-9.
99. Rogalla P, Blank C, Helbig R, Wosniok W, Bullerdiek J. Significant correlation between the breakpoints of rare clonal aberrations in benign solid tumors and the assignment of HMGIY retropseudogenes. Cancer Genet Cytogenet. 2001;130(1):51-6.
100. Esposito F, Martino M De, Petti MG, Forzati F, Tornincasa M, Federico A, Arra C, Pierantoni GM, Fusco A. HMGA1 pseudogenes as candidate proto-oncogenic competitive endogenous RNAs. Oncotarget. 2014;5(18):8341-54. doi: 10.18632/oncotarget.2202.
101. Esposito F, De Martino M, Forzati F, Fusco A. HMGA1-pseudogene overexpression contributes to cancer progression. Cell Cycle. 2014;13(23):3636-9.
102. Esposito F, De Martino M, D'Angelo D, Mussnich P, Raverot G, Jaffrain-Rea M-L, Fraggetta F, Trouillas J, Fusco A. HMGA1 -pseudogene expression is induced in human pituitary tumors. Cell Cycle. 2015;14(9):1471-5.
103. Li W, Yuan Y, Huang L, Qiao M, Zhang Y. Metformin alters the expression profiles of microRNAs in human pancreatic cancer cells. Diabetes Res Clin Pract. 2012;96(2):187-95.
104. Sehic A, Risnes S, Khuu C, Khan Q-E-S, Osmundsen H. Effects of *in vivo* transfection with anti-miR-214 on gene expression in murine molar tooth germ. Physiol Genomics. 2011;43(9):488-98.
105. Fedele M, Pierantoni GM, Visone R, Fusco A. Critical role of the HMGA2 gene in pituitary adenomas. Cell Cycle. 2006;5(18):2045-8.
106. Chiefari E, Iiritano S, Paonessa F, Le Pera I, Arcidiacono B, Filocamo M, Foti D, Liebhaber S a, Brunetti A. Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. Nat Commun. 2010;1(4):40.
107. Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, Wynshaw-Boris A, Yoshiki A. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature. 2003;423(6935):91-6.
108. Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu YM, Cao X, Asangani I a., Kothari V, Prensner JR, Lonigro RJ, Iyer MK, Barrette T, Shanmugam A, Dhanasekaran SM, Palanisamy N, Chinnaiyan AM. Expressed pseudogenes in the transcriptional landscape of human cancers. Cell. 2012;149(7):1622-34.
109. Sorge J, Gross E, West C, Beutler E. High level transcription of the glucocerebrosidase pseudogene in normal subjects and patients with Gaucher disease. J Clin Invest. 1990;86(4):1137-41.
110. Renaudie F, Yachou AK, Grandchamp B, Jones R, Beaumont C. A second ferritin L subunit is encoded by an intronless gene in the mouse. Mamm Genome. 1992;2(3):143-9.
111. Chakrabarti R, McCracken JB, Chakrabarti D, Souba WW. Detection of a functional promoter/enhancer in an intron-less human gene encoding a glutamine synthetase-like enzyme. Gene. 1995;153(2):163-9.
112. Thiele H, Berger M, Skalweit A, Thiele BJ. Expression of the gene and processed pseudogenes encoding the human and rabbit translationally controlled turnout protein (TCTP). Eur J Biochem. 2000;267(17):5473-81.
113. Suo G, Han J, Wang X, Zhang J, Zhao Y, Zhao Y, Dai J. Oct4 pseudogenes are transcribed in cancers. Biochem Biophys Res Commun. 2005;337(4):1047-51.
114. Zhang J, Wang X, Li M, Han J, Chen B, Wang B, Dai J. NANOGP8 is a retrogene expressed in cancers. FEBS J. 2006;273(8):1723-30.
115. Poliseno L. Pseudogenes: Newly Discovered Players in Human Cancer. Sci Signal. 2012;5(242):re5-re5.
116. Poliseno L, Marranci A, Pandolfi PP. Pseudogenes in Human Cancer. Front Med. 2015;2:68.



OPEN

Received: 21 April 2016
Accepted: 01 November 2016
Published: 22 November 2016

# *HMGA1P7*-pseudogene regulates *H19* and *Igf2* expression by a competitive endogenous RNA mechanism

**Marco De Martino[1], Floriana Forzati[1], Marianna Marfella[1], Simona Pellecchia[1], Claudio Arra[2], Luigi Terracciano[3], Alfredo Fusco[1] & Francesco Esposito[1]**

**Recent studies have revealed that pseudogene transcripts can function as competing endogenous RNAs, and thereby can also contribute to cancer when dysregulated. We have recently identified two pseudogenes, *HMGA1P6* and *HMGA1P7* for the *HMGA1* gene whose overexpression has a critical role in cancer progression. These pseudogenes work as competitive endogenous RNA decoys for *HMGA1* and other cancer related genes suggesting their role in carcinogenesis. Looking for new *HMGA1* pseudogene ceRNAs, we performed RNA sequencing technology on mouse embryonic fibroblasts deriving from transgenic mice overexpressing *HMGA1P7*. Here, we report that *HMGA1P7* mRNA sustains the *H19* and *Igf2* overexpression by acting as miRNA decoy. Lastly, the expression of *HMGA1P7* was significantly correlated with *H19* and *IGF2* levels in human breast cancer thereby suggesting a role for *HMGA1P7* deregulation in this neoplasia.**

Pseudogenes are a subclass of long non coding RNA (lncRNA) sharing high sequence identity with protein-coding parental counterparts. As stated by the GENCODE pseudogene annotations (v.17), there are almost 15,000 human pseudogenes[1]. They frequently possess features, such as premature stop codons, deletions/insertions, or frame shift mutations, that impede them to produce functional proteins. There are three groups of pseudogenes: processed, duplicated, and unitary[1,2]. Processed pseudogenes do not have introns and are thought to arise from reverse transcription of mRNA followed by reinsertion into the genome[1,2]. Duplicated pseudogenes contain introns and sometimes even upstream regulatory elements since they are produced by gene duplication. For each pseudogene belonging to these two classes there is an associated protein-coding gene that is highly similar in sequence[1,2]. The last type of pseudogenes are the unitary ones, which take place when protein-coding genes accumulate mutations and lose their coding potential[1,2]. Consequently, unitary pseudogenes do not have parental genes. From the time of their discovery in 1977, pseudogenes have commonly been thought as "biologically inconsequential" and non-functional[3]. However, recent studies have unveiled different mechanisms by which pseudogenes control gene expression such as the generation of siRNAs[4,5], competition for RNA-binding proteins or the translation apparatus[6–8], and engagement of proteins by pseudogene antisense RNAs to corresponding sites in the parental gene to modify chromatin transcription and remodeling[9,10]. The latest function identified for pseudogenes is post-transcriptional regulation of mRNA levels by competing for microRNAs (miRNAs). Indeed, processed pseudogenes maintain 5′ and 3′ untranslated region (UTR) sequences of their parental genes[11]. Given that miRNAs inhibit target gene expression by binding to the 3′ UTR, pseudogenes can be targeted by miRNAs that modify the expression of coding genes. Definitely, pseudogene transcripts exert regulatory control of their parental gene expression levels by competing for the same miRNAs[12].

We have recently characterized two processed pseudogenes, *HMGA1P6* and *HMGA1P7*, for the *HMGA1* gene that codes for the HMGA1a and HMGA1b proteins highly overexpressed in most of the human malignancies[13,14].

[1]Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Scuola di Medicina e Chirurgia di Napoli, Università degli Studi di Napoli "Federico II", via Pansini 5, 80131 Naples, Italy. [2]Istituto Nazionale dei Tumori, Fondazione Pascale, via Mariano Semmola, 80131 Naples, Italy. [3]Institute of Pathology, Molecular Pathology Division, University of Basel, Schonbeinstrasse 40, 4003 Basel, Switzerland. Correspondence and requests for materials should be addressed to A.F. (email: alfusco@unina.it) or F.E. (email: francesco.esposito2@unina.it)

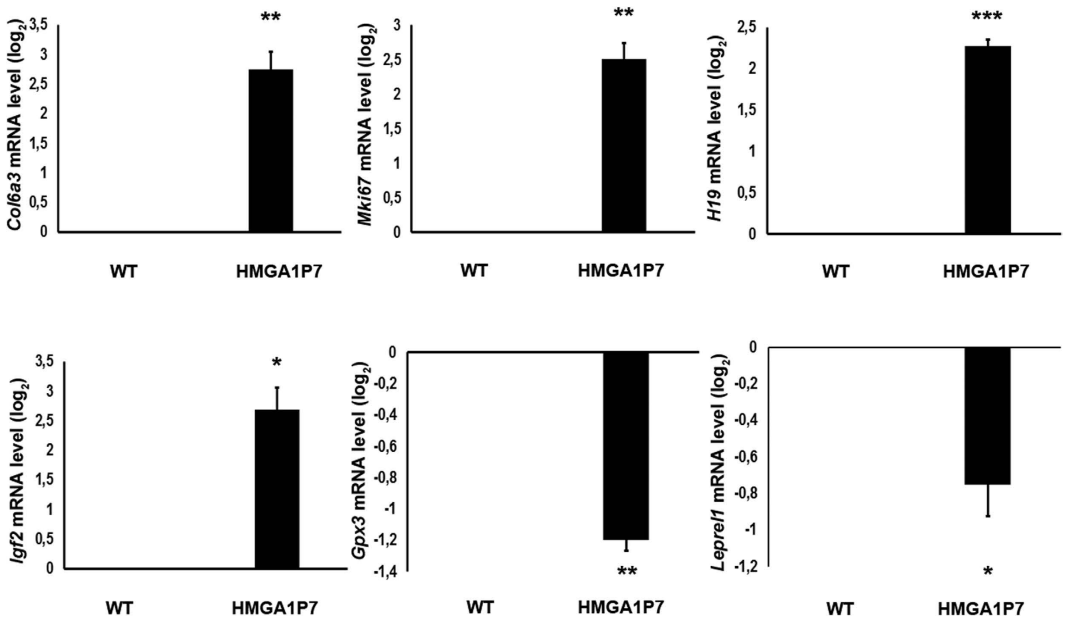



**Figure 1. Validation of RNA-seq analyses on *HMGA1P7* MEFs.** qRT-PCR analysis of selected deregulated genes from RNA-seq performed on WT and *HMGA1P7* transgenic MEFs. The results are reported as the mean of values. The error bars represent mean ± SE; *P < 0.05 **P < 0.01 ***P < 0.001 (t test).

Moreover, it has been previously demonstrated an association between HMGA1 overexpression and a poor patient survival[13], and that their overexpression is even required for cell transformation[15,16], and is able to induce benign and malignant neoplasias in mice[13]. *HMGA1P6* and *HMGA1P7* pseudogenes, present only in human genome, have preserved seed matches for miRNAs targeting the *HMGA1* oncogene. *HMGA1* pseudogenes (*HMGA1Ps*) overexpression, working as competitive endogenous RNA (ceRNA), increases HMGA1 protein levels by blocking the suppression of HMGA1 protein synthesis exerted by miRNAs[17–20]. *HMGA1Ps* have also oncogenic activity by suppressing apoptosis and promoting cell proliferation and migration[17–20]. Moreover, we have previously show that *HMGA1Ps* are overexpressed in anaplastic thyroid carcinomas but not in the differentiated ones, indicating a critical role of them in cancer progression[17]. Since the *HMGA1Ps* contain several seed sequences for miRNAs, their overexpression derepresses the expression of different cancer-related genes, as already demonstrated for *HMGA2*, *VEGF*, *EZH2*[17–20]. Therefore, the aim of this study has been to find novel ceRNA interactors differentially expressed in *HMGAP7* transgenic mouse embryonic fibroblasts (MEFs) with respect to the wild-type (WT) ones, which do not express HMGA1 pseudogenes, using a RNA sequencing (RNA-seq) approach. By this analysis, we found a set of mRNAs up- or down-regulated in *HMGA1P7* overexpressing MEFs in comparison with WT cells. Among them, we focused our attention on two of the most overexpressed and *HMGA1P7* miRNA-sharing genes: *H19* and *insulin-like growing factor 2* (*Igf2*).

*H19* and *IGF2* genes are closely linked, showing highly similar patterns of gene expression, but they are reciprocally imprinted. In fact, *H19* is expressed solely from the maternally inherited chromosome, whereas *IGF2* expression is from the paternal chromosome. In particular, the noncoding *H19* has a critical role in genomic imprinting during cell growth and development[21]. The loss of imprinting results in misexpression of *H19* and was detected in many tumors including hepatocellular[22], bladder[23], gastric[24,25] and colon[26] cancer. *IGF2* codes for a mitogenic growth factor that is active in early development and has a critical role in embryonic and fetal growth[27]. Increased expression of *IGF2* is a common feature of both pediatric and adult malignancies[27], and mounting evidence implicates *IGF2* as a major factor contributing to oncogenesis[27–29]. Here, we report that *HMGA1P7* mRNA induces the *H19* and *Igf2* overexpression by acting as miRNA decoy.

## Results

**RNA-seq on *HMGA1P7* transgenic MEFs.** To identify the genes regulated by *HMGA1P7* expression, we analyzed the whole transcriptome of WT and *HMGA1P7* transgenic MEFs by RNA-seq analyses. To this aim, the entire population of RNA transcripts extracted from WT and *HMGA1P7*-MEFs were sequenced. The genome-wide RNA expression profiles studies reveal that about one hundred fifty transcripts (32 upregulated and 116 downregulated) were regulated by *HMGA1P7* expression with a significant fold-change variation (FDR adjusted p-value of 0.05). To validate the results obtained by RNA-seq, we analyzed the expression of some upregulated genes such as *Collagen Type VI Alpha 3* (*Col6a3*), *Marker of Proliferation Ki-67* (*Mki67*), *H19*, *Igf2* and downregulated genes such as *Glutathione Peroxidase 3* (*Gpx3*), *Leprecan-Like 1* (*Leprel1*) by Real-time PCR (qRT-PCR). As shown in Fig. 1, the quantitative qRT-PCR analyses confirmed the data obtained from the RNA-seq analyses. Interestingly, these genes have been related to several human cancers (colon, gastric, liver, breast and hematological cancers), and are considered possible therapeutic targets[30–35].

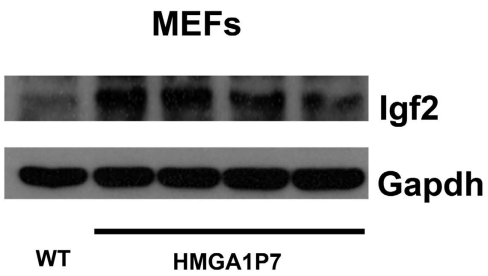



**Figure 2. Igf2 is upregulated in *HMGA1P7* MEFs** . Western blot analysis of Igf2 from WT and *HMGA1P7* transgenic MEFs.

Among the differentially expressed mRNAs found in MEFs overexpressing *HMGA1P7*, we focused our attention on *H19* and *Igf2* since they, other than to be involved in carcinogenesis, showed the highest fold change among the upregulated genes, and are also targeted by several miRNAs that are able to bind to the *HMGA1P7* mRNA. Western blot analysis for Igf2 confirmed the qRT-PCR data (Fig. 2). Moreover, qRT-PCR and Western blot analysis showed that *H19* and *Igf2* were also upregulated in heart, spleen and kidney from *HMGA1P7* adult transgenic mice (Fig. 3A,B and C). As expected from previous results, qRT-PCR shows upregulation of *H19* and *Igf2* following *HMGA1P7* pseudogene overexpression in NIH3T3 cells (Fig. 3D). Western blot confirms *Igf2* upregulation also at protein level in tissues from *HMGA1P7* and in the *HMGA1P7*-transfected NIH3T3 cells (Fig. 3C and D).

Taken together, these results strongly support the hypothesis that *HMGA1P7* could act as ceRNA for *H19* and *Igf2*.

***HMGA1P7* act as decoy for *H19* and *Igf2* targeting miRNAs.** To test whether the effect of the *HMGA1P7* pseudogene on *H19* and *Igf2* expression is dependent on sharing targeting-miRNAs, we evaluated the ability of *HMGA1P7*-targeting miRNAs[17] to bind to *H19* and *Igf2*. To this aim, we transfected miR-15, miR-16, miR-214 and miR-761 (already reported to target *HMGA1P7*)[17] into NIH3T3 cells, and analyzed *H19* and *Igf2* mRNA levels by qRT-PCR. As presented in Fig. 4A, the transfection of the *HMGA1P7*-targeting miRNAs yield a significant reduction of *H19* and *Igf2* mRNA levels. Western blot confirms Igf2 downregulation also at protein level following the transfection of the *HMGA1P7*-targeting miRNAs (Fig. 4A). To define whether the *HMGA1P7*-targeting miRNAs straightly interacted with *Igf2* mRNA, we cloned the *Igf2* 3′ UTR downstream of the luciferase open reading frame. This reporter vector was transfected into NIH3T3 cells together with miRNA precursors and a control non-targeting scrambled oligonucleotide. The luciferase signal was considerably lower after transfection with miR-15, miR-16, miR-214 and miR-761 in comparison with the cells transfected with the scrambled oligonucleotide (Fig. 4B). The overexpression of *H19* and *Igf2* induced by upregulation of *HMGA1P7* was depleted in Dicer-knockdown cells (Fig. 4C) then supporting the hypothesis that *HMGA1P7, H19* and *Igf2* follow the same miRNA-mediated post-transcriptional regulation. In fact, silencing of Dicer, the enzyme that leads miRNA maturation process, results in reduced levels of mature miRNAs compared to control. Moreover, to verify whether *H19* and *Igf2* can act as ceRNA each-other, we transfected siRNA-*Igf2* into NIH3T3 cell line in combination or not with Anti miR-16, which is able to block miR-16 repression on *HMGA1P7*, *Igf2* and *H19*, and a siRNA-control, then evaluating the *H19* mRNA levels. As proposed by our model, siRNA-*Igf2* transfection induces a significant *H19* downregulation, that is reverted by the transfection with the Anti miR-16 oligonucleotide, suggesting that both *H19* and *Igf2* transcripts can talk each-other through miRNAs mediation (Fig. 4D). These data are consistent with the hypothesis that *HMGA1P7* requires mature miRNAs to regulate *H19* and *Igf2* levels.

***HMGA1P7*, *H19* and *IGF2* expression positively correlates in human breast cancer.** Then, we investigated whether *HMGA1P7* functions as ceRNA through, or partially through *H19,* IGF2 and HMGA1 in breast cancer human cells. As expected, we found upregulation of *H19,* IGF2 and HMGA1 following *HMGA1P7* overexpression in MCF7 cells (human breast adenocarcinoma cell line) (Fig. 5A). Moreover, MCF7-*HMGA1P7* cells grow faster than the control transfected cells as consequence of *HMGA1P7* ceRNA pathway activation (Fig. 5B).

To confirm whether *HMGA1P7* works as miRNA sponge for the regulation of *H19* and *IGF2* expression levels also in human cancer, we evaluated the expression of *H19, IGF2* and *HMGA1P7* in a panel of breast carcinoma samples by qRT-PCR, since *H19* and *IGF2* have been reported to be overexpressed in this type of tumor[36,37]. As shown in Fig. 5C, *HMGA1P7* was overexpressed in most of the carcinoma samples as well as *H19* and *IGF2.* Moreover, the direct correlation between *HMGA1P7* and *H19* expression (Spearman $r = 0{,}8656$; $p < 0{,}001$) and between *HMGA1P7* and *IGF2* expression (Spearman $r = 0{,}7958$; $p < 0{,}001$) underlines that these genes are co-regulated (Fig. 5D). Altogether, these results strongly support the idea that *HMGA1P7* could act as ceRNAs in human breast cancer and represent a novel potential mechanism accounting for *H19* and *IGF2* upregulation in these tumors.

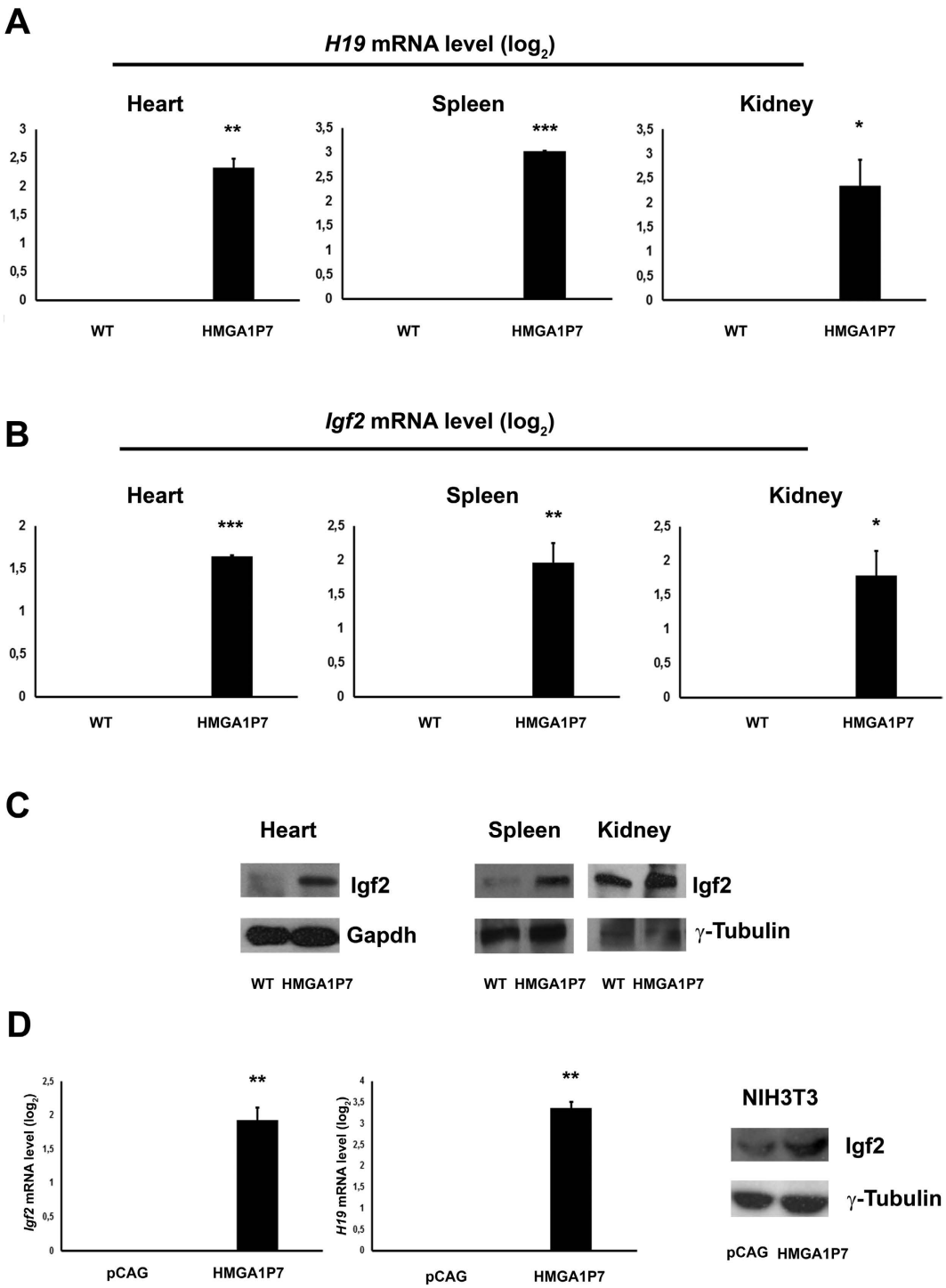



**Figure 3.** ***H19* and *Igf2* are positively regulated by *HMGA1P7*.** (**A**) qRT-PCR analysis of *H19* from hearts, spleens and kidneys of WT and *HMGA1P7* transgenic mice. (**B**) qRT-PCR analysis of *Igf2* from hearts, spleens and kidneys of WT and *HMGA1P7* transgenic mice. (**C**) Western blot analysis of Igf2 from heart, spleen and kidney of WT and *HMGA1P7* transgenic mice. (**D**) Left Panel, qRT-PCR analysis of *H19* and *Igf2* from control and *HMGA1P7* overexpressing NIH3T3 cells. Right Panel, Western blot analysis of Igf2 from control and *HMGA1P7* overexpressing NIH3T3 cells. The results are reported as the mean of values. The error bars represent mean ± SE; *P < 0.05 **P < 0.01 ***P < 0.001 (t test).

## Discussion

lncRNAs are involved in regulating the complexity of biological processes with specific regulatory mechanisms, thereby, attracting considerable research interest[38]. We have previously isolated and characterized two pseudogenes, *HMGA1P6* and *HMGA1P7*, for the *HMGA1* gene and demonstrated that they act as decoys for

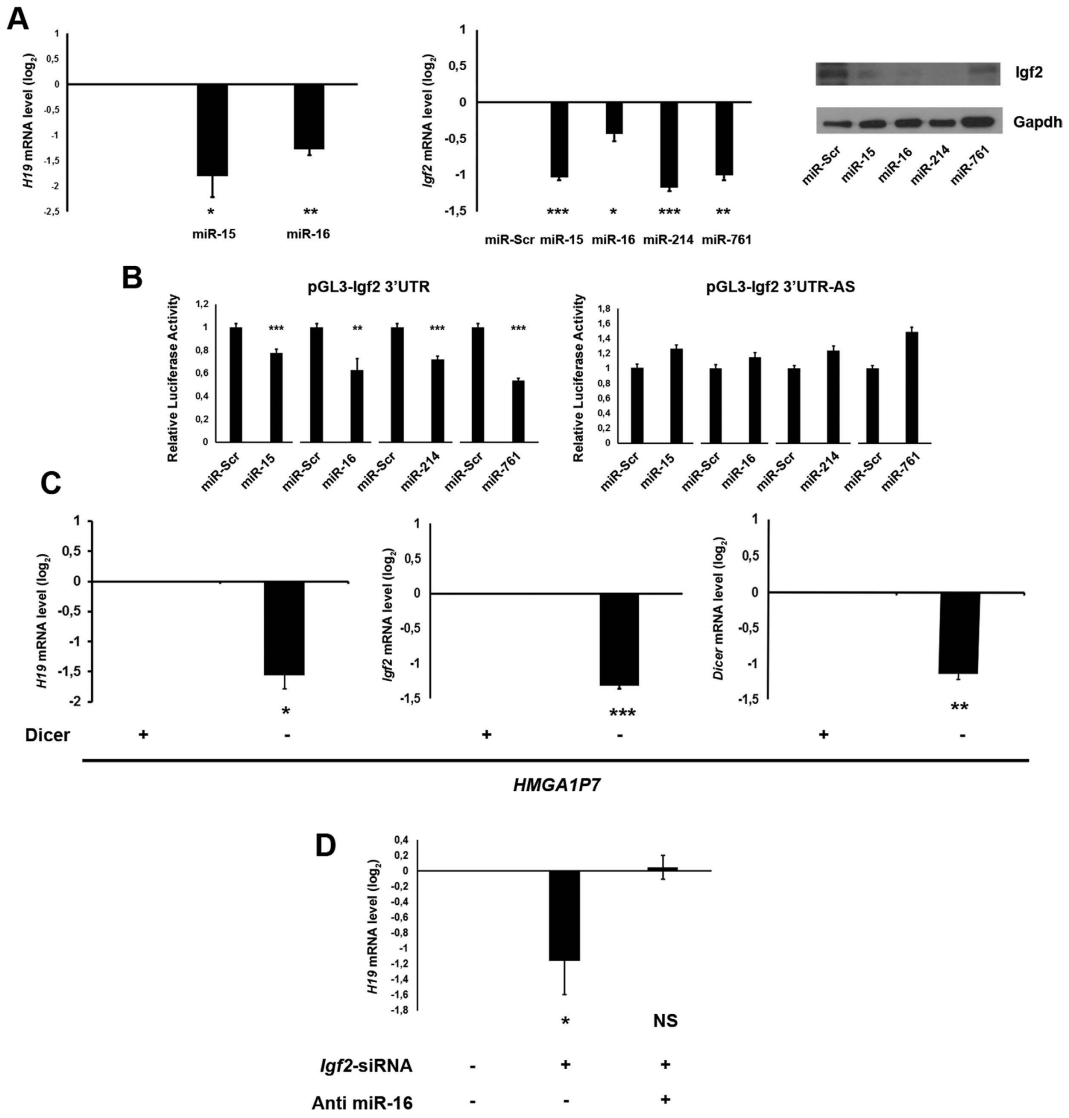



**Figure 4. *H19* and *Igf2* are targeted by *HMGA1P7*-targeting miRNAs.** (**A**) qRT-PCR analysis of *H19* (left Panel) and *Igf2* (middle Panel) mRNA from the NIH3T3 cells transfected with scrambled-oligonucleotide, miR-15, miR-16, miR-214 and miR-761. Right Panel, Western blot analysis of Igf2 in the same samples as in the middle panel. (**B**) *Igf2* was cloned into the pGL3 control vector. Relative luciferase activity in HEK293 cells transiently transfected with miR-15, miR-16, miR-214, miR-761 and a control scrambled oligonucleotide. (**C**) *H19* and *Igf2* mRNA levels 24 h after the transfection of *HMGA1P7* in scrambled oligonucleotide or siRNA-Dicer NIH3T3 transfected cells. (**D**) qRT-PCR analysis of *H19* mRNA levels from the NIH3T3 cells transfected with siRNA-control, siRNA-*Igf2* alone or in in combination with the Anti miR-16 oligonucleotide.

*HMGA1*-targeting miRNAs[17]. In fact, their overexpression enhances HMGA1 protein levels whereas their knocking down results in the reduction of *HMGA1* mRNA and protein amounts. Moreover, their decoy activity protected the synthesis of other proteins involved in carcinogenesis[17,18]. In this study, we used RNA-seq technology to identify additional mRNAs differentially expressed in MEFs transgenic for *HMGA1P7*.

We found that the expression of several genes were influenced by *HMGA1P7* including also genes involved in cancer progression such as *Col6a3*, *Mki67*, *H19*, *Igf2, Gpx3* and *Leprel1*[30–37]. Indeed, oncomine analyses and tissue-microarray immunohistochemistry showed overexpression of COL6A3 in colorectal carcinomas that was significantly and directly correlated with Dukes stage, T stage, stage, recurrence and smoking status and then with a poor prognosis[30]. The MKi-67 protein (also known as Ki67) is a cellular marker for proliferation. Ki-67 protein is expressed during all active phases of the cell cycle ($G_1$, S, $G_2$, and mitosis), but is absent from resting cells ($G_0$)[32]. *GPX3* gene codes for the Glutathione peroxidase 3, also known as plasma glutathione peroxidase (GPx-P), the variations in activity of GPX1, GPX2, and GPX3 isoforms may be associated with the development of cancers, for example, prostate cancer or even colon cancer[39]. Leprecan-like 1 is a potential tumor suppressor gene since it has been demonstrated to be downregulated in the hepatocarcinoma tissues and its overexpression inhibits cancer cell proliferation and colony formation through regulation of the cell cycle by downregulation of cyclins[40].

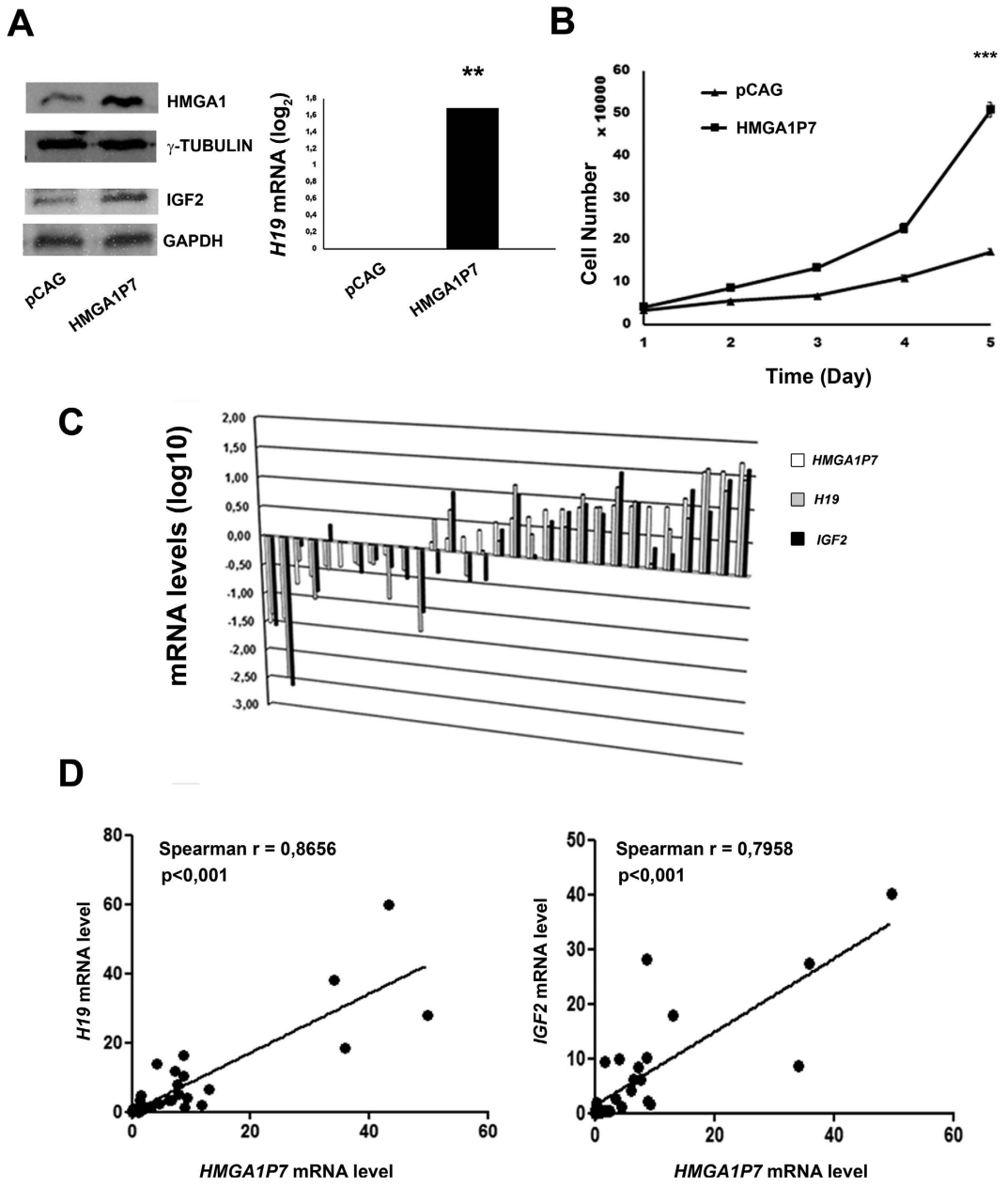



**Figure 5. *H19* and *IGF2* expression positively correlates with *HMGA1P7* in breast cancer.** (**A**) Left Panel, Western blot analysis of HMGA1 and IGF2 protein levels in control and *HMGA1P7*-overexpressing MCF7 cells. Right Panel, qRT-PCR analysis of *H19* expression of control and *HMGA1P7*-overexpressing MCF7 cells. (**B**) MCF7 cell proliferation of control and *HMGA1P7*-overexpressing cells (**C**) qRT-PCR analysis in tumor and normal breast tissues. The fold change indicates the relative change in expression levels between tumor samples and normal samples, assuming that the value of normal sample is equal to 1. (**B**) Correlation analysis of *HMGA1P7* versus *H19* and *HMGA1P7 versus IGF2* are shown. The Spearman's rank correlation coefficient is shown.

Deregulation of *H19* noncoding gene was found in many tumors such as hepatocellular and bladder cancer[22,23]. Finally, *IGF2* overexpression is widely reported in pediatric and adult tumors[27], and several studies involve *IGF2* as a key factor leading to cancerogenesis[27–29].

Among the most deregulated genes, we selected and studied *H19* non-coding gene and *Igf2,* that share several miRNAs with *HMGA1P7*. Here, we report the ceRNA relationship between *HMGA1P7*, *H19* and *Igf2*. We demonstrate that *HMGA1P7* overexpression increases *H19* and *Igf2* levels inhibiting their mRNA suppression by miRNAs that target *HMGA1P7* gene, namely, miR-15, miR-16, miR-214, and miR-761. Interestingly, preliminary results show an analogous ceRNA connection between *H19, Igf2* and *HMGA1P6,* supporting the oncogenic role of the both *HMGA1* pseudogenes.

Finally, we show that expression of *HMGA1P7* significantly correlates with *H19* and *IGF2* levels in human breast cancer, suggesting the upregulation of *HMGA1P7* may increase *H19* and *IGF2* expression by a ceRNA mechanism then contributing to cancer progression. Interestingly, the oncogenic role of *HMGA1P7* is also supported by the development of malignant hematological neoplasias in *HMGA1P7* transgenic mice (manuscript in preparation).

Then, the data reported here confirm the oncogenic role of the *HMGA1P7* pseudogene that is exerted by the increased expression through a ceRNA mechanism of HMGA1 and other cancer-related genes. Future studies need, however, to characterize other genes regulated by the HMGA1 pseudogenes and thereby better define the mechanisms by which they can contribute to cancer progression.

## Materials and Methods

**Cell culture and transfections.** MEFs and MCF7 were cultured in DMEM supplemented with 10% foetal calf serum (Thermo Fisher Scientific Inc). NIH3T3 cells were maintained in DMEM supplemented with 10% calf serum (Thermo Fisher Scientific Inc), glutamine and antibiotics. MycoAlert (Lonza) was regularly used to test that cells were not infected by mycoplasma. Lipofectamine plus reagent was used to transfect the cells (Thermo Fisher Scientific Inc) according to the manufacturer's instructions. The transfected cells were selected in a medium containing geneticin (Sigma). Transfection efficiency was tested for each experiment by assessing GFP signal. To inhibit *Dicer* and *Igf2* expression, small interfering RNAs and corresponding scramble small interfering RNAs were designed and used as suggested by the manufacturer (RIBOXX).

**RNA-sequencing.** RNA samples were initially checked for quality and quantity using a Bioanalyzer with the total RNA Pico chip (Agilent Technologies, Inc) and a Qubit® with RNA Assay Kit (Thermo Fisher Scientific Inc) respectively.

Spike-In Mix 1 and Spike-In Mix 2, each containing the full complement of 92 polyadenylated transcripts from the ERCC plasmid reference library, were added to samples.

mRNA was selected from total RNA preparation using MicroPolyA Purist kit (Ambion, Inc).

SOLiD™ Total RNA-Seq Kit (Life Technologies Corporation) was used to convert RNA transcripts into a cDNA library, starting from low input amounts of poly(A) RNA, for analysis on the 5500 Genetic Analysis System. First of all, mRNA was incubated in a thermal cycler at 95 °C for 10 minutes to fragment the RNA by chemical hydrolysis. The RNA digested was hybridized and ligated with Solid specific adaptors. Two rounds of size selection using Agencourt® AMPure® XP Reagent were performed to increase the percentage of library inserts that were in the desired size range, >150 bp. The purified DNA was amplified and barcoded by 18 PCR cycles to enable sequencing of all the samples in a single multiplexed SOLiD System sequencing run.

The yield and size distribution of the amplified DNA libraries was assessed running the samples on an Agilent® 2100 Bioanalyzer™ Instrument with the DNA HS Kit following the manufacturer's instructions and Qubit® dsDNA HS kit.

Ninety nanograms of each library was pooled together and one E80 emulsion was prepared following SOLiD® EZ Bead™ System. About 400 millions of pooled templated beads were deposited on a 4 lanes of a 6 lanes slide and the sequencing was performed up to a read length of 50 bp, based on 5500 Genetic Analysis System Run sequencer protocol.

**Bioinformatic Analysis.** Four samples were analysed: two from WT and two from *HMGA1P7* transgenic MEFs. The comparison performed was WT versus transgenic, two biological replicates for condition.

Sequencing reads in SOLiD "xsq" format were mapped against the reference genome (UCSC GRC38/mm10); reference gene structure was Refseq from the refGene.txt file of the UCSC genome browser FTP site; the mapping software was the Whole Transcriptome Analysis module from the Lifescope 2.5.1 Genomic Analysis Software analysis suite from Applied Biosystems/ThermoFisher Scientific.

A filter file, containing 6415 sequences (sequencing adaptors; barcodes; tRNAs; rRNAs; rRNA fragments; repetitive sequences; ERCC RNA sequences) was used (1) to filter the transcripts for non-significant reads and (2) to quantify the absolute expression using the External RNA Controls Consortium (ERCC) RNA Spike-In Mix.

The genome-mapped reads were then correlated with Refseq genes and the resulting gene-associated read counts were analysed with a Genomnia proprietary procedure based on the Bioconductor library edgeR[41]. The chosen limit for evaluating differential expression was 5 counts per millions in at least half of the examined samples. The normalization procedure used was the standard for edgeR (TMM). Genes were called differentially expressed when the comparison was evaluated with a FDR $< 0.05$. Absolute gene expression was evaluated from the read counts in RPKM (Reads per kilo base per million mapped reads).

Primary gene annotation was performed using the Bioconductor libraries biomaRt and GOstats, while functional clustering of the genes was performed using the DAVID functional annotation web site (https://david.ncifcrf.gov/).

**Mouse embryo and tissue samples.** The use of mouse embryos and tissues and the experiments performed in this study were approved by the Ministero della Salute; the methods and experiments were carried out in accordance with the approved guidelines by the Ministero della Salute.

**Human breast tissue samples.** Normal and neoplastic human breast tissues were obtained from surgical specimens and immediately frozen in liquid nitrogen. 32 breast samples were collected at the Institute of Pathology, University of Basel, Switzerland. The tumor samples were frozen until required for RNA extraction. The use of human tissues and the experiments performed in this study were approved by the Institute of Pathology, Molecular Pathology Division, University of Basel; the methods and experiments were carried out

in accordance with the approved guidelines by the University of Basel. We declare that informed consent for the scientific use of biological material was obtained from all patients.

**RNA extraction and quantitative reverse transcription PCR.** Total RNA was extracted from cells or tissues with TRIsure (Aurogene) according to the manufacturer's instructions. For mRNA trascripts detection, we reverse transcribed total RNA from samples by using the QuantiTect Reverse Transcription Kit (Qiagen), and then Real-time PCR was performed by using Power SYBR Green PCR Master Mix (Bio-Rad) and the following primers:

*HMGA1P7*-Fw 5′-gctccttctcggctcctc-3′
*HMGA1P7*-Rev 5′-gcttgggcctcttttatgg-3′
*mIgf2* Fw 5′-cctccttacccaacttcaggt-3′
*mIgf2* Rv 5′-aagagatgagaagcaccaacatc-3′
*mh19* Fw 5′-atgtcttcatttctccctatagcc-3′
*mh19* Rv 5′-gtcatcctcgccttcagtg-3′
*mG6pd*-Fw5′-cagcggcaactaaactcaga-3′
*mG6pd*-Rev 5′-ttccctcaggatcccacac-3′
*mCol6a3* Fw 5′-ggaggtgtacaggaagttccac-3′
*mCol6a3* Rev 5′-gactgagccgtcaaagagga-3′
*mMki67* Fw 5′-gctgtcctcaagacaatcatca-3′
*mMki67* Rev 5′-ggcgttatcccaggagact-3′
*mGpx3* Fw 5′-gtgaacggggagaaagagc-3′
*mGpx3* Rev 5′-tgagcccaggagttctgc-3′
*mLeprel1* Fw 5′-tggaccctctttaccgagaa-3′
*mLeprel1* Rev 5′-tgatccaagatggcaatcac-3′
*hActin Fw* 5′-ccaaccgcgagaagatga-3′
*hActin Rv* 5′-ccagaggcgtacagggatag-3′
*hH19* Fw 5′-ttacttcctccacggagtcg-3′
*hH19* Rv 5′-gagctgggtagcaccatttc-3′
*hIGF2* Fw 5′-gctggcagaggagtgtcc-3′
*hIGF2* Rv 5′-gggattcccattggtgtct-3′

The $2^{-\Delta\Delta CT}$ formula was used to calculate the differential gene expression, and described elsewhere[42].

**Plasmid and miRNA oligonucleotides.** For transfection of miRNA oligonucleotides, cells were transfected with 50 nmol/ml of miRNA precursors or with a control no-targeting scrambled oligonucleotides (Thermo Fisher Scientific Inc) using siPORT neoFX Transfection Agent (Thermo Fisher Scientific Inc). For transfection of Anti miR-16 oligonucleotides, cells were transfected with 50 nmol/ml of Anti miR-16 or with a control no-targeting scrambled oligonucleotides (Thermo Fisher Scientific Inc).

For *Igf2* luciferase reporter construct (pGL3-*Igf2*), the miRNA seed sequence conteining fragment of *Igf2* gene (ENSMUST00000000033) was amplified by using the primers:

*Igf2* Fw 5′-aatttctagacccaaaatctcacttttccc-3′
*Igf2* Rev 5′-aatttctagagatggcccataggtgtgctc-3′.

The amplified fragment was cloned into pGL3-Control luciferase reporter vector (Promega).

All the generated vectors were confirmed by sequencing. The Renilla luciferase vector (pRL-CMV), for transient transfection efficiency, was purchased from Promega.

**Protein extraction, western blotting and antibodies.** Protein extraction and Western blotting were performed as previously described[43,44]. The primary antibodies used were anti-IGF2 (#32592) from Sabbiotech; anti-GAPDH (sc-32233) and anti-γ-Tubulin (sc-17787) from Santa Cruz Biotechnology. Blots were visualized by using the Western blotting detection reagents (Thermo Fisher Scientific Inc).

**Dual-luciferase reporter assay.** For dual-luciferase reporter assay, $3 \times 10^5$ NIH3T3 cells were co-transfected in 6-well plates with the pGL3- *Igf2* or the pGL3-*H19* luciferase reporter vectors, together with the Renilla luciferase plasmid and miRNA precursors or a control no-targeting scrambled oligonucleotides (Thermo Fisher Scientific Inc), using siPORT neoFX Transfection Agent (Thermo Fisher Scientific Inc). The pRL-TK control vector expressing Renilla luciferase (Promega) was used for normalization of cell number and transfection efficiency. Luciferase activity was measured 48 hours after transfection using the Dual-Luciferase Reporter Assay System (Promega) with a Lumat LB 9507 apparatus (Berthold Technologies).

**Growth curve assay.** For each experimental point $3 \times 10^4$ cells were plated in a 60 mm plate. Cells were counted in triplicate for 5 days with Burker hemocytometer chamber.

**Statistical analysis.** Data were analyzed using a two-sided unpaired t test (GraphPad Prism, GraphPad Software, Inc.). Values of $P < 0.05$ were considered statistically significant. Regression analysis, correlation coefficients and statistical analysis were generated using GraphPad Prism, GraphPad Software, Inc.

## References

1. Welch, J. D., Baran-Gale, J., Perou, C. M., Sethupathy, P. & Prins, J. F. Pseudogenes transcribed in breast invasive carcinoma show subtype-specific expression and ceRNA potential. *BMC Genomics*. **16,** 113 (2015).
2. Zhang, Z. D., Frankish, A., Hunt, T., Harrow, J. & Gerstein, M. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol.* **11,** R26 (2010).
3. Jacq, C., Miller, J. R. & Brownlee, G. G. A pseudogene structure in 5S DNA of Xenopus laevis. *Cell.* **12,** 109–120 (1977).

4. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature.* **453,** 534–538 (2008).
5. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature.* **453,** 539–543 (2008).
6. Bier, A. *et al.* Connexin43 pseudogene in breast cancer cells offers a novel therapeutic target. *Mol Cancer Ther*. **8,** 786–793 (2009).
7. Chiefari, E. *et al.* Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. *Nat Commun*. **1,** 40 (2010).
8. Han, Y. J., Ma, S. F., Yourek, G., Park, Y. D. & Garcia, J. G. A transcribed pseudogene of MYLK promotes cell proliferation. *FASEB J*. **25,** 2305–2312 (2011).
9. Hawkins, P. G. & Morris, K. V. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription*. **1,** 165–175 (2010).
10. Johnsson, P. *et al.* A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat Struct Mol Biol*. **20,** 440–446 (2013).
11. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. **465,** 1033–1038 (2010).
12. Arvey, A., Larsson, E., Sander, C., Leslie, C. S. & Marks, D. S. Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol*. **6,** 363 (2010).
13. Fusco, A. & Fedele, M. Roles of HMGA proteins in cancer. *Nat Rev Cancer*. **7,** 899–910 (2007).
14. Esposito, F., Tornincasa, M., Chieffi, P., De Martino, I., Pierantoni, G. M. & Fusco, A. High-Mobility Group A1 Proteins Regulate p53-Mediated Transcription of *Bcl-2* Gene. *Cancer Res*. **70,** 5379–88 (2010).
15. Berlingieri, M. T. *et al.* Inhibition of HMGI-C protein synthesis suppresses retrovirally induced neoplastic transformation of rat thyroid cells. *Mol Cell Biol*. **15,** 1545–1553 (1995).
16. Berlingieri, M. T., Pierantoni, G. M., Giancotti, V., Santoro, M. & Fusco, A. Thyroid cell transformation requires the expression of the HMG1 proteins. *Oncogene*. **21,** 2971–2980 (2002).
17. Esposito, F. *et al.* HMGA1 pseudogenes as candidate proto-oncogenic competitive endogenous RNAs. *Oncotarget*. **5,** 8341–8354 (2014).
18. Esposito, F., De Martino, M., Forzati, F. & Fusco, A. HMGA1-pseudogene overexpression contributes to cancer progression. *Cell Cycle*. **13,** 3636–3639 (2014).
19. Esposito, F. *et al.* A HMGA1-pseudogene expression is induced in human pituitary tumors. *Cell Cycle*. **14,** 1471–1475 (2015).
20. De Martino, M., Forzati, F., Arra, C., Fusco, A. & Esposito, F. HMGA1-pseudogenes and cancer. *Oncotarget*. [Epub ahead of print] (2016).
21. Gabory, A., Jammes, H. & Dandolo, L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays*. **32,** 473–480 (2010).
22. Imad, J. *et al.* The H19 Non-Coding RNA Is Essential for Human Tumor Growth. *PLoS ONE*. **2,** e845 (2007).
23. Liu, C. *et al.* H19-derived miR-675 contributes to bladder cancer cell proliferation by regulating p53 activation. *Tumour Biol*. **37,** 263–270 (2015).
24. Zhou, X., Yin, C., Dang, Y., Ye, F. & Zhang, G. Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. *Sci Rep*. **5,** 11516 (2015).
25. Chen, J. S., Zhang, X. Q., Lv, J. M., Li, Y. & Liu, T. X. X. H19 serves as a diagnostic biomarker and up-regulation of H19 expression contributes to poor prognosis in patients with gastric cancer. *Neoplasma*. **63,** 223–230 (2016).
26. Liang, W. C. *et al.* The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. *Oncotarget*. **6,** 22513–22525 (2015).
27. Foulstone, E. *et al.* Insulin-like growth factor ligands, receptors, and binding proteins in cancer. *J Pathol*. **205,** 145–153 (2005).
28. Nielsen, H. M. *et al.* Copy number variations alter methylation and parallel IGF2 overexpression in adrenal tumors. *Endocr Relat Cancer*. **22,** 953–967 (2015).
29. Brouwer-Visser, J. & Huang, G. S. IGF2 signaling and regulation in cancer. *Cytokine Growth Factor Rev*. **26,** 371–377 (2015).
30. Qiao, J. *et al.* Stroma derived COL6A3 is a potential prognosis marker of colorectal carcinoma revealed by quantitative proteomics. *Oncotarget*. **6,** 29929–29946 (2015).
31. Xie, X., Liu, X., Zhang, Q. & Yu, J. Overexpression of collagen VI $\alpha$3 in gastric cancer. *Oncol Lett*. **7,** 1537–1543 (2014).
32. Li, L. T., Jiang, G., Chen, Q. & Zheng, J. N. Ki67 is a promising molecular target in the diagnosis of cancer. *Mol Med Rep*. **11,** 1566–1572 (2015).
33. Zhou, J. D. *et al.* GPX3 hypermethylation serves as an independent prognostic biomarker in non-M3 acute myeloid leukemia. *Am J Cancer Res*. **5,** 2047–2055 (2015).
34. Qi, X. *et al.* Clinical significance and therapeutic value of glutathione peroxidase 3 (GPx3) in hepatocellular carcinoma. *Oncotarget*. **5,** 11103–11120 (2014).
35. Shah, R. *et al.* The prolyl 3-hydroxylases P3H2 and P3H3 are novel targets for epigenetic silencing in breast cancer. *Br J Cancer*. **100,** 1687–1696 (2009).
36. Shetty, P. J. *et al.* Regulation of IGF2 transcript and protein expression by altered methylation in breast cancer. *J Cancer Res Clin Oncol*. **137,** 339–345 (2011).
37. Sun, H. *et al.* H19 lncRNA mediates 17$\beta$-estradiol-induced cell proliferation in MCF-7 breast cancer cells. *Oncol Rep*. **33,** 3045–3052 (2015).
38. Lieberman, J. *et al.* Noncoding RNAs and cancer. *Cell*. **153,** 9–10 (2013).
39. Zmorzyński, S., Świderska-Kołacz, G., Koczkodaj, D. & Filip, A. A. Significance of Polymorphisms and Expression of Enzyme-Encoding Genes Related to Glutathione in Hematopoietic Cancers and Solid Tumors. *Biomed Res Int*. **2015,** 853573 (2015).
40. Wang, J. *et al.* LEPREL1 Expression in Human Hepatocellular Carcinoma and Its Suppressor Role on Cell Proliferation. *Gastroenterol Res Pract*. **2013,** 109759 (2013).
41. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*. **40,** 4288–4297 (2012).
42. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. **25,** 402–408 (2001).
43. Esposito, F. *et al.* Aurora B expression in post-puberal testicular germ cell tumours. *J Cell Physiol*. **221,** 435–439 (2009).
44. Esposito, F. *et al.* The high-mobility group A1-estrogen receptor $\beta$ nuclear interaction is impaired in human testicular seminomas. *J Cell Physiol*. **227,** 3749–3755 (2012).


## Acknowledgements

This study has been supported by grants from: PNR-CNR Aging Program 2012–2014, POR Campania FSE 2007–2013 (CREMe), CNR Flagship Projects (Epigenomics-EPIGEN), PON 01-02782 (Nuove strategie nanotecnologiche per la messa a punto di farmaci e presidi diagnostici diretti verso cellule cancerose circolanti), Associazione Italiana per la Ricerca sul Cancro (AIRC IG 11477).

## Author Contributions

F.E. and A.F. conceived and designed the study. M.D.M. with help from F.E., F.F., M.M. and S.P. performed all experiments, and F.E., M.D.M., C.A., L.T. and A.F. analyzed most data. F.E. and A.F. wrote the manuscript with contributions from all authors.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article**: De Martino, M. *et al. HMGA1P7*-pseudogene regulates *H19* and *Igf2* expression by a competitive endogenous RNA mechanism. *Sci. Rep.* **6**, 37622; doi: 10.1038/srep37622 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.