UNIVERSITÀ DEGLI STUDI
DI GENOVA

DEPARTMENT OF EXPERIMENTAL MEDICINE (DIMES)

PhD COURSE IN EXPERIMENTAL MEDICINE

CURRICULUM BIOCHEMISTRY

CYCLE XXXI

LONG NONCODING RNA EPR CONTROLS EPITHELIAL CELL PROLIFERATION BY COORDINATING *CDKN1A* TRANSCRIPTION AND mRNA DECAY IN RESPONSE TO TGF-β

PhD STUDENT: Martina Rossi

TUTORS: Dr. Roberto Gherzi, Dr. Paola Briata

COURSE COORDINATORS: Prof. Giambattista Bonanno, Prof. Santina Bruzzone

ACADEMIC YEARS 2015-2018

INDEX

1. INTRODUCTION 5

2. BACKGROUND OF THE STUDY 8

2.1 RNA: Messenger, Enzyme, Regulator 8

2.1.1 RNA in Gene Expression Regulation 9

2.2 Non-coding RNAs: Form, Function, Physiology 10

2.2.1 LncRNA functions in gene expression 11

2.2.2 LncRNAs as chromatin regulators 12

2.2.3 LncRNAs as modulators of Transcription Factors 13

2.2.4 Post-transcriptional regulation by lncRNAs 14

2.2.5 LncRNAs able to encode proteins and peptides: Challenges and Ambiguities 15

2.2.6 LncRNAs and Cancer 16

2.3 Epithelial To Mesenchymal Transition 18

2.3.1 Molecular processes underlying EMT 19

2.3.2 Role of TGF-β in EMT 20

2.4 The Two faces of TGF-β1 in Breast Cancer 22

2.5 KHSRP 22

3. AIM OF THE STUDY AND SUMMARY 26

4. RESULTS 27

4.1 Identification and characterization of EPR, an epithelial cell type- enriched lncRNA. 27

4.2 EPR encodes a small polypeptide. 35

4.3 EPR regulates gene expression in NMuMG cells. 41

4.4 EPR regulates *Cdkn1a* gene expression and cell proliferation. 45

4.5 EPR regulates TGF-β-dependent *Cdkn1a* gene expression at different levels. 49

4.6 EPR affects both SMAD3-dependent *Cdkn1a* gene transcription and KHSRP-regulated mRNA decay. 53

4.7 EPR overexpression reduces breast cancer cell proliferation. 59

5. DISCUSSION 66

6. METHODS 70

6.1 Cell lines 70

6.2 Antibodies 70

6.3 Plasmids 71

6.4 Cell transfection 72

6.5 Scratch wound closure assay 72

6.6 Immunofluorescence 73

6.7 Orthotopic 4T1 injection in BALB/c mice 73

6.8 Electrophoretic mobility shift assay 74

6.9 Isolation of RNA from cytoplasmic, nucleoplasmic, and chromatin fractions 74

6.10 RNA preparation, quantitative RT-PCR, analysis of newly-synthesized transcripts, and of mRNA decay 75

6.11 Ribonucleoprotein complexes immunoprecipitation (RIP) assays 75

6.12 Protein identification by MALDI-TOF mass spectrometry (MS) analysis 76

6.13 RNA deep-sequencing (RNA-Seq) 77

6.14 Analysis of h.EPR (LINC01207) expression in human samples 77

6.15 Chromatin Isolation by RNA Purification (ChIRP) 78

6.16 ChIP-qPCR 79

6.17 Sucrose-gradient fractionation and polysome profiling 79

6.18 Analyses of cell cycle distribution and quantification of S phase cells by flow cytometer 80

6.19 G1 phase cell sorting 81

6.20 Quantification of cell proliferation by high-content image analysis 82

6.21 Quantification of cell proliferation by crystal violet staining 82

6.22 Clonogenic and anchorage-independent cell growth assays 82

6.23 Quantification and statistical analysis of RNA-Seq 83

6.24 Quantification and statistical analysis of transcript differential expression analysis 83

6.25 Venn Diagram 84

6.26 Gene Ontology and Pathway Enrichment 84

6.27 Protein alignment 84

6.28 Supplementary Table 1 85

7. REFERENCES 87

1. INTRODUCTION

Human transcriptome analysis has revealed the existence of a surprisingly high number of noncoding RNAs that have been classified in multiple families based on their size and biogenesis. Long non-coding RNAs (lncRNAs) are defined as transcripts longer than 200 nucleotides transcribed by RNA polymerase II and commonly originated from intergenic regions. LncRNAs can be capped, spliced, and polyadenylated and usually show limited protein coding potential (Jandura & Krause, 2017; Kopp & Mendell, 2018 and literature cited therein).

LncRNAs are emerging as a fundamental aspect of biology due to their ability to reprogram gene expression and influence distinct cellular functions including cell fate determination, cell cycle progression, apoptosis, and aging (Jandura & Krause, 2017; Kopp & Mendell, 2018). Their expression is usually tissue restricted, developmentally regulated, and can change under specific pathological conditions. Many lncRNAs influence hallmarks of cancer such as uncontrollable proliferation, evasion of cell death, as well as metastasis formation and it has been suggested that lncRNAs can act as oncogenes or tumor suppressors —either directly or indirectly— by interfering with different pathways (Long et al., 2017; Arun et al., 2018).

From a mechanistic point of view, lncRNAs may influence the function of transcriptional complexes, modulate chromatin structures, function as scaffolds to form ribonucleoprotein (RNP) complexes as well as act as decoys for proteins and micro-RNAs (miRNAs) (Long et al., 2017; Kopp & Mendell, 2018). Thus, lncRNA-mediated control of gene expression may take place at transcriptional and/or post-transcriptional levels (Gupta et al., 2010; Yoon et al., 2012; Geisler & Coller, 2013; Cao et al., 2017; Long et al., 2017).

Recently, lncRNAs have been described as important components of the Transforming Growth Factor β (TGF-β) signaling pathway (Yuan et al., 2014; Dhamija & Diederichs, 2016). TGF-β belongs to a large family of structurally related cytokines that regulate growth, survival, differentiation, and migration of many cell types including mammary gland epithelial cells (ref. Shi & Massague (2003) for a review). TGF-β activates membrane kinase receptors and induces phosphorylation of cell-specific SMAD proteins that, in complex with the common SMAD4, accumulate into the nucleus to regulate gene expression at different levels (ref. Budi et al. (2017) for a recent review).

In our previous studies, we showed that the multifunctional RNA binding protein KHSRP acts as a regulatory hub that conveys extracellular stimuli resulting into gene expression changes due to its ability to interact with several molecular partners (Briata et al., 2016). KHSRP is able to post-transcriptionally regulate gene expression by promoting decay of unstable mRNAs, favoring maturation of select miRNAs from precursors, and controlling alternative splicing events (Briata et al., 2016).

Recently, we reported that KHSRP controls the alternative splicing of a cohort of pre-mRNAs that encode regulators of cell adhesion and motility —such as CD44 and FGFR2— favoring their “epithelial type” exon usage, and that miRNA-mediated KHSRP silencing is required for TGF-β-induced Epithelial-to-Mesenchymal Transition (EMT) in immortalized NMuMG mammary gland cells (Puppo et al., 2016). Further, we found that Resveratrol - a natural polyphenolic compound endowed with anti-inflammatory, anti-proliferative, as well as pro-apoptotic activities- prevents TGF-β-dependent KHSRP down-regulation thus shifting Cd44 and Fgfr2 pre-mRNA alternative splicing from the mesenchymal-specific to the epithelial-specific isoform (Moshiri et al., 2017).

Our previous observation that the lncRNA H19 interacts with KHSRP and affects its mRNA decay promoting function (Giovarelli et al., 2014) prompted us to identify additional KHSRP/lncRNAs interactions endowed with regulatory potential.

2. BACKGROUND OF THE STUDY

2.1 RNA: Messenger, Enzyme, Regulator

The central dogma of molecular biology suggested that DNA contains the information to encode proteins and that three different types of RNA rather passively convert this code into polypeptides. Specifically, messenger RNA (mRNA) carries the protein blueprint from a cell's DNA to the ribosomes which drive protein synthesis. Transfer RNA (tRNA) carries the appropriate amino acids into the ribosomes for inclusion in the new protein and the ribosomes themselves consist largely of ribosomal RNA (rRNA) molecules (Clancy, 2008). However, since the structure of DNA was first described, scientists have learned that RNA does much more than simply playing a role in protein synthesis. For example, many types of RNA have been found to be catalytic - that is, they carry out biochemical reactions just like enzymes do (Clancy, 2008). The condensation of amino acids in the peptidyl transferase center of the ribosome is catalyzed by the major RNA component of the large subunit and the splicing of mRNA in eukaryotes is catalyzed by the U2-U6 small nuclear RNAs found within the splicing speckles. (Wilson and Lilley, 2015).

Further, several RNA molecules have been found to play complex regulatory roles in cell (Clancy, 2008). As a group, these RNAs are frequently referred to as regulatory RNAs, and, in eukaryotes, they have been further classified into a number of subcategories. Regulatory noncoding RNAs (ncRNAs) can be divided into two main different classes, small RNAs and long noncoding RNAs (lncRNAs). lncRNAs are usually defined as poly-adenylated RNA transcripts several hundred nucleotides long while the smallRNAs are in the range of 20-30 nucleotides.

RNA is in fact emerging as a major component of the regulatory circuitry that underpins the development and physiology of complex organisms. Defects in certain

RNAs have been implicated in a number of important human diseases, including heart disease, cancer, stroke and many others (Clancy, 2008).

2.1.1 RNA in Gene Expression Regulation

Protein production is regulated in multiple, diverse ways which all act in a controlled and dynamic manner in what we collectively call ‘gene expression regulation’. RNA plays an active role at different levels of gene expression regulation.

There are many steps in the pathway leading from DNA to protein and we now know that all of them can in principle be regulated (Alberts et al (2007) for a complete review; Fig I)

A cell can control the proteins it makes by:

- controlling when and how often a given gene is transcribed (transcriptional control)
- controlling the splicing and processing of RNA transcripts (RNA processing control)
- controlling the localization of the RNA molecules (RNA transport and localization control)
- controlling the way mRNAs are translated at the ribosome sites (translational control)
- Controlling the stability of certain mRNA molecules (mRNA degradation control)
- selectively activating, inactivating, degrading, or locating specific protein molecules after they have been produced (protein activity control).

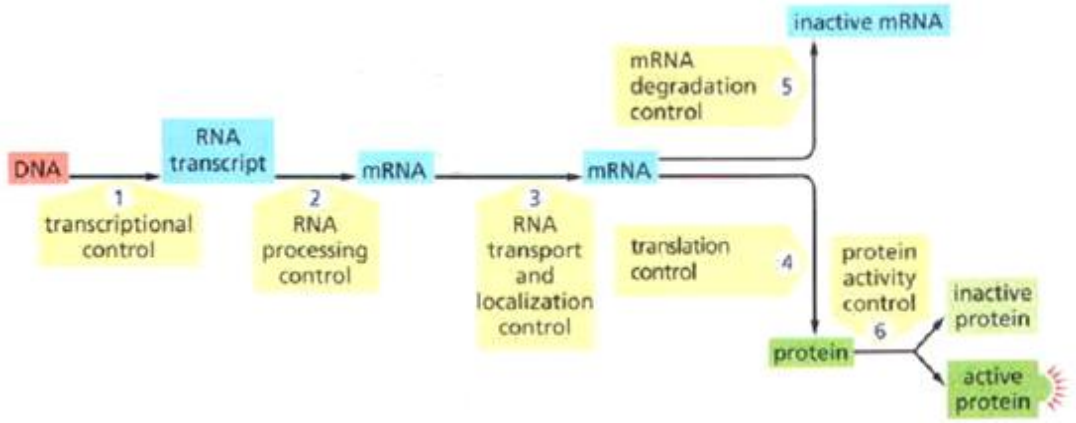


Fig. I. Six steps at which eukaryotic gene expression can be controlled

(Alberts et al., 2007).

In general, regulatory RNAs can control gene expression at the transcriptional and post-transcriptional level (Kaikkonen et al., 2011).

One widely known example of RNA-mediated transcriptional regulation is the role of X-inactive specific transcript (Xist) in X-chromosome inactivation. The large Xist transcript is expressed on the inactive chromosome and not on the active one and the chromosome coating by Xist is essential for inactivation.

MicroRNAs (miRNAs) are major members of regulatory RNAs playing an important role at the post-transcriptional level of gene expression. These short ncRNAs bind to a specific target messenger RNA with a complementary sequence in order to induce its degradation or block protein translation (Bartel, 2004, for a review).

The continual discovery of new regulatory RNA species suggests that we are only just beginning to understand their complexity and functions.

2.2 Non-coding RNAs: Form, Function, Physiology

Following the sequencing of the human genome, focus has shifted toward understanding gene function. In 2005, the FANTOM (Functional Annotation of Mammals)

Consortium determined that the genome harbors more non-coding genes than coding genes (Carninci et al., 2005). The ENCyclopedia Of DNA Elements (ENCODE) Consortium, revealed that only 2-3-% of the mammalian genome is transcribed into protein-coding RNAs (Birney et al., 2007). The other 98% has been considered for a long time as inactive material, regions of several mega-bases without any function, so-called "junk DNA" (Taft et al., 2007). However, the overwhelming development of high-density micro-arrays and high-throughput sequencing technologies, as well as bioinformatics analyses, has revealed that non-coding transcripts are major regulators of gene expression. The examples given above regarding the lncRNA Xist and the miRNAs show the importance of noncoding species in modulating gene expression (for other references see Rinn & Chang, 2012; Sole et al., 2015; De Andres-Pablo et al., 2017). "Non-coding RNA" is a broad definition that encompasses all types of RNA that lack empirical evidence of translation into protein although recent evidence revealed that certain short Open Reading Frames (ORF) of ncRNAs are actually translated into small peptides (Slavoff et al., 2013). NcRNAs come in many forms, perform diverse functions, and play critical roles in health and disease. Tens of thousands of ncRNAs have been identified so far, but we are still at the earliest stages of understanding their regulation and functions. Such knowledge will be critical in order to harness the full potential of non-coding RNAs for the diagnosis and treatment of human diseases (Delás & Hannon 2017).

2.2.1 LncRNA functions in gene expression

LncRNAs are highly heterogeneous molecules and have a substantial functional versatility that relies on different structures and interactions. Among the variety of mechanisms involved in gene expression regulation, many lncRNAs have been shown to interact with chromatin-modifying complexes (Rinn & Chang, 2012; Bierhoff et al., 2014;

Marchese & Huarte, 2014; Engreitz et al., 2016), while others have been shown to interfere with the transcriptional machinery (Prasanth et al., 2005; Clemson et al., 2009; Sunwoo et al., 2009). Furthermore, some lncRNAs mediate key steps in post-transcriptional regulation of gene expression (Yoon et al., 2013; Quinn & Chang, 2016).

2.2.2 LncRNAs as chromatin regulators

A large repertoire of lncRNAs is able to interact with chromatin-modifying complexes and these lncRNA-containing complexes can promote either selective repression or activation of genes (Gendrel & Heard, 2014; Marchese & Huarte, 2014; Morlando et al., 2014; Meller & Joshi, 2015). For instance, several lncRNAs have been shown to recruit histone H3K4 methyltransferases in order to promote activation of gene expression (Wang et al., 2011; Cabianca et al., 2012; Grote et al., 2013). Others bind DNA methyltransferases, such as DNMT1 and DNMT3b, and repress transcription by promoting DNA methylation (Mohammad et al., 2010; Schmitz et al., 2010; Johnsson et al., 2013). Finally, many lncRNAs have been shown to interact with the polycomb repressive complex 2 (PRC2), which catalyses generation of the H3K27me3 silencing mark (Rinn et al., 2007; Khalil et al., 2009; Shore et al., 2012; Grote et al., 2013; Marin-Bejar et al., 2013; Kotzin et al., 2016; Wang et al., 2016). PRC2 is the most studied chromatin complex with respect to the functional role of lncRNAs in the epigenetic regulation of gene expression (Davidovich & Cech, 2015).

The activity of several noncoding RNAs is tightly connected to their own site of transcription. These lncRNAs remains in the proximity of their own locus and associate with complexes that determine a local regulation of gene expression (*cis*-regulation) (Engreitz et al., 2013). Other lncRNAs, instead, act as a platform for trans-chromosomal interactions. In this context, the cooperative association of lncRNAs with other RNAs and

proteins is significant for acting at distant genomic loci (*trans*-regulation) (Hacisuleyman et al., 2014).

2.2.3 LncRNAs as modulators of Transcription Factors

Among the proteins that can be modulated by lncRNAs are the transcription factors - the key players of transcriptional regulation (Marchese et al., 2017). In the most canonical model, gene expression control is thought to be mediated by transcription factors, whose activation is usually regulated by signalling pathways and whose DNA-binding ability is associated with sequence specificity (Marchese et al., 2017). Several lncRNAs have been reported to bind transcription factors in order to regulate gene expression. As an example, a lncRNA, named 'damage-induced noncoding RNA' (*DINO*), was identified in the regulation of the DNA-damage-induced p53 response (Schmitt et al., 2016). The results support the notion that p53 can bind simultaneously DNA and RNA and ascribe to the lncRNA *DINO* a regulatory role in p53-mediated gene expression (Schmitt et al., 2016).

The possibility that a lncRNA can bind and regulate transcription factors or chromatin remodeling complexes confers numerous advantages to the cell (Marchese et al., 2017). For instance, lncRNAs are known to be highly cell and tissue specific (Cabili et al., 2011; Cabili et al., 2015) which means that, without changing the transcriptional machinery, cell- and tissue-specific regulation of gene expression could be achieved (Marchese et al., 2017). LncRNA-dependent regulation could also be considered in terms of cost-effectiveness, as RNAs are energetically less expensive to be produced in comparison with proteins. Also, lncRNAs are more rapidly produced than proteins in response to stimuli and this could confer faster cellular responses. Moreover, they can act locally at their site of transcription, whereas proteins need to be translated from their encoding RNAs in the cytoplasm and be returned to the nucleus (Marchese et al., 2017).

2.2.4 Post-transcriptional regulation by lncRNAs.

Although the majority of characterized lncRNAs are directly involved in transcriptional regulation, different lncRNAs have been implicated in the splicing, turnover, export or translation of mRNAs, as well as in the stability and post-translational modification of proteins (Huarte, 2015). An example of the latter is the lncRNA Breast Cancer Anti-estrogen Resistance 4 (BCAR4), the expression of which correlates with metastatic advanced breast cancer and anti-estrogen resistance (Godinho et al, 2011; Xing et al., 2014). BCAR4 promotes the acetylation of Smad nuclear interacting protein 1 (SNP1) and serine/threonine-protein phosphatase 1 regulatory subunit 10 (PNUTS), thus activating the hedgehog/GLI2 transcriptional program and promoting cell migration (Xing et al., 2014; Huarte, 2015).

Alternative splicing is another mechanism by which lncRNAs can influence post-transcriptional regulation and cancer progression (Huarte, 2015). Indeed, many mRNA-splicing isoforms are developmentally regulated, and they are preferentially re-expressed in tumors. (Huarte, 2015). The lncRNA Zeb2-NAT controls translation and therefore function of the ZEB2 protein coding gene through regulation of Zeb2mRNA splicing. Splicing inhibition allows translation of ZEB2, a transcriptional repressor of E-cadherin.

Further, some lncRNAs function as RNA sponges, basing their activity on sequence-specific interactions with other RNAs (Salmena et al., 2011). For instance, the cytoplasmic lncRNA Phosphatase and TENsin homolog Pseudogene 1 (PTENP1) binds to microRNAs that otherwise would target PTENmRNA, reducing its expression and tumor suppressor activity (Poliseno et al., 2010). This crosstalk between lncRNAs, miRNAs and mRNAs highlights the complex fine-tuning of RNA networks. However, these dynamics are still far from being completely understood (Huarte, 2015).

2.2.5 LncRNAs able to encode proteins and peptides: challenges and ambiguities

Increasing evidence underlines a series of intriguing translational events from lncRNAs which were previously considered to lack protein coding potential (Hube et al., 2006; Kondo et al., 2010; Ulveling et al., 2011). Recent studies also suggest that products derived from such novel translational events display important regulatory functions in many fundamental biological and pathological processes.

The first report of a bifunctional RNA was the human *Steroid Receptor Activator* (*SRA*). Originally, *SRA* was characterized as a ncRNA able to co-activate the steroid hormone receptors (Lanz et al., 1999). Remarkably, *SRA* transcripts have now been shown to encode a functional protein (SRAP) which appears to act antagonistically to SRA RNA (Chooniedass-Kothari et al., 2006).

A study by Nelson et. al (2016) described a mammalian muscle-specific lncRNA that encodes a peptide of 34 amino acids, named Dwarf Open Reading Frame (DWORF). DWORF localizes to the sarcoplasmic reticulum (SR), where it enhances the activity of the calcium pump SERCA by displacing SERCA inhibitors (Nelson et al., 2016). Further, Anderson et al. (2015) discovered a conserved micropeptide, named myoregulin (MRLN), encoded by a skeletal muscle-specific RNA annotated as a putative long noncoding RNA. MRLN interacts directly with SERCA and inhibits calcium uptake into the Sarcoplasmic Reticulum. These studies identified micropeptides as important regulators of muscle physiology and highlighted the possibility that additional peptides are encoded by many RNAs currently annotated as noncoding (Anderson et al., 2015; Nelson et al., 2016). More recently, Yu et al. (2017) showed that the lncRNA encoding MRLN, named Linc-RAM, promotes myogenic differentiation in a MRLN-independent manner (Yu et al., 2017).

2.2.6 LncRNAs and Cancer

It is increasingly evident that many of the genomic mutations in cancer reside inside regions that do not encode proteins and these regions are often transcribed into long noncoding RNAs (lncRNAs) (Huarte, 2015). The recent application of next-generation sequencing to a growing number of cancer transcriptomes has indeed revealed thousands of lncRNAs whose aberrant expression is associated with different cancer types. Among the few that have been functionally characterized, several have been linked to malignant transformation (Fig II).

One of the best characterized lncRNA is HOX-antisense intergenic RNA (HOTAIR), which plays a critical role in chromatin dynamics and gene silencing, and appears to be misregulated in a variety of cancers. HOTAIR regulates genes involved in cell cycle progression, cell migration, Epithelial to Mesenchimal Transition (EMT), metastasis, and tumor progression (Bhan & Mandal, 2015). Various reports have demonstrated that the expression of HOTAIR is significantly upregulated in ER-positive breast tumors and correlates with metastasis and progression (Gupta et al., 2010; Lu et al., 2012, Sorensen et al., 2013). Studies by Sorensen et al. showed that patients with low HOTAIR expression had better survival than those with high HOTAIR expression. (Sorensen et al., 2013).

The expression of the lncRNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) was also identified early as a prognostic parameter for lung cancer survival. MALAT1 expression has now been associated with malignancy in multiple types of tumors, including liver, breast and colon, suggesting that it has a general role in cell proliferation. MALAT1 is a nuclear lncRNA, highly conserved across mammals and extremely abundant in many cell types. It has been linked to the regulation of alternative mRNA splicing and the modulation of the epigenetic machinery, and it is known to be associated with nascent pre-mRNA of actively transcribed genes. Despite being the subject

of intensive study, the role of MALAT1 in cancer is not yet fully understood (Huarte, 2015).

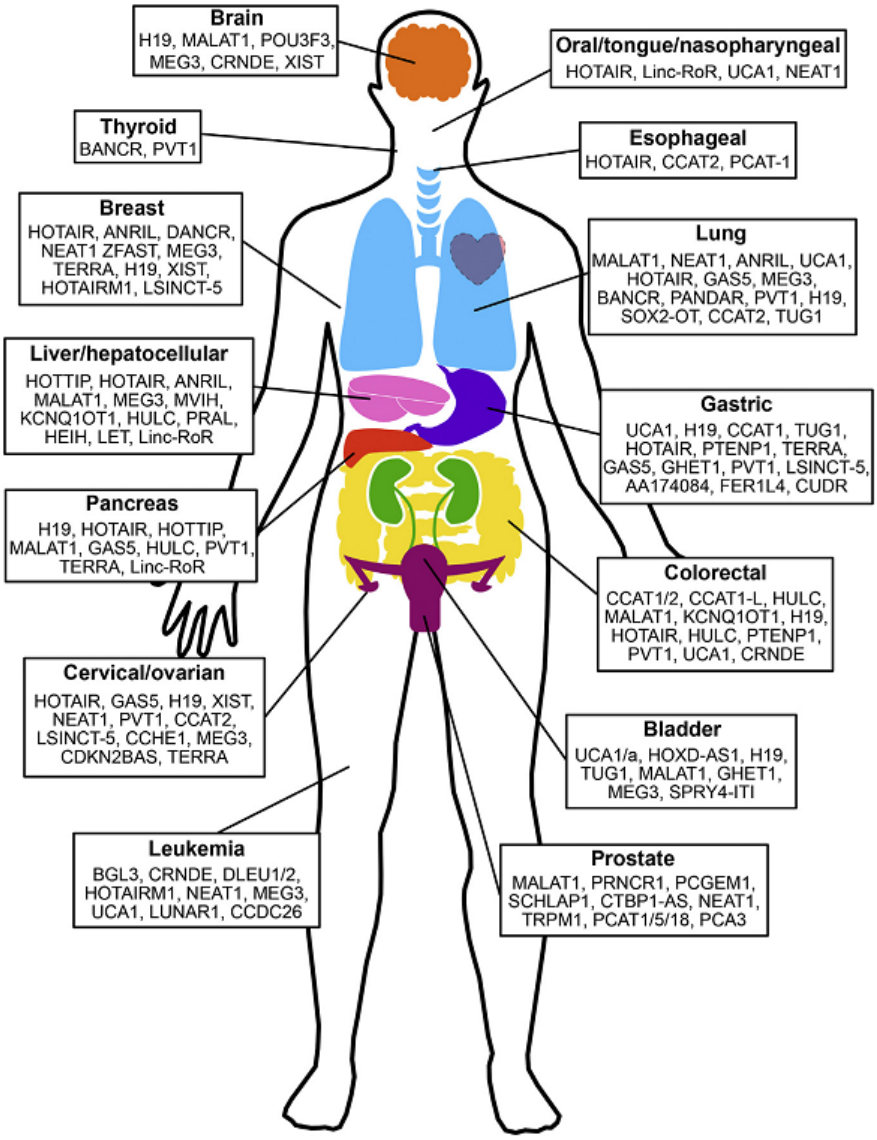


Fig II. LncRNAs associated with various cancer types.

M.M. Balas, A.M. Johnson, Non-coding RNA Research (2018)

2.3 Epithelial To Mesenchymal Transition

Epithelial to Mesenchymal Transition (EMT) is a biological process which allows epithelial cells to lose their characteristics and acquire a mesenchymal phenotype. It has recently been proposed that, rather than being a binary process, EMT occurs through distinct intermediate states: from epithelial to completely mesenchymal states, passing through intermediate hybrid states (Pastushenko et al., 2018). Through EMT, cells gain the ability to migrate and resist apoptosis as well as the potential to enter stem cell-like states (De Craene and Berx, 2013; Ye and Weinberg, 2015). Phenotypical hallmarks of EMT include morphological changes from a cobblestone-like epithelial shape to a spindle-like fibroblast phenotype, loss of epithelial CDH1 (also known as E-cadherin) at cell junctions, and increased expression of mesenchymal CDH2 (also known as N-cadherin). The opposite trans-differentiation process, mesenchymal-to-epithelial transition (MET), is also possible (De Craene and Berx, 2013; Ye and Weinberg, 2015).

Molecular key events in EMT include activation of transcription factors, expression of specific cell-surface proteins, reorganization of the cytoskeleton, production of proteases degrading the extracellular matrix (ECM), and expression of specific microRNAs. In many cases, these factors can be used as biomarkers to demonstrate the passage of a cell through EMT. (Fig III). (Kalluri and Weinberg, 2009).

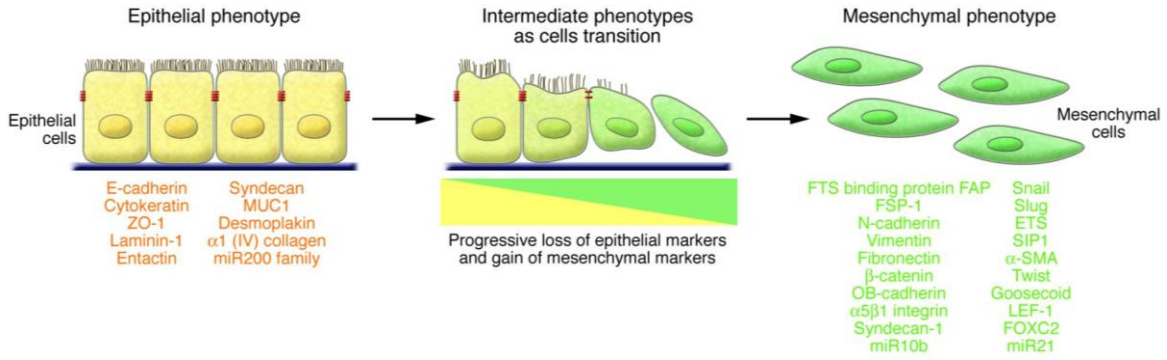


Fig.III. EMT process and molecular markers

Kalluri and Weinberg, The Journal of clinical investigation, 2009

EMT is involved in many physiological and pathological processes. During development, EMT allows implantation, embryo formation, cell differentiation and organ development. Moreover, in tissue repair, EMT generates fibroblasts and other related cells that contribute to wound healing (Kalluri and Weinberg, 2009).

Many studies show that the activation of an EMT program also plays an important role in tumor progression with cells acquiring the ability to invade and metastasize (Brabletz et al., 2001; Thiery, 2002; Fidler & Poste, 2008; Yang & Weinberg, 2008). In secondary tumors, carcinoma cells may undergo MET and revert to an epithelial phenotype during the colonization process (Zeisberg et al., 2005).

Recent studies revealed an unexpected role of EMT in cancer chemoresistance (Fischer et al., 2015; Zheng et al., 2015), Authors showed that suppressing EMT in mice models had no effect on the number of circulating tumor cells or the overall frequency of metastasis. However, suppressing EMT enhanced the sensitivity of the tumors to chemotherapy drugs, suggesting a role for EMT in cancer drug resistance. (Fischer et al., 2015; Zheng et al., 2015).

2.3.1 Molecular processes underlying EMT

EMT-inducing signals derived from the tumor microenvironment, such as TGF β, HGF, EGF, PDGF, are responsible for the induction of a series of transcription factors driving the EMT process. Once expressed and activated, each transcription factor can act pleiotropically to activate and orchestrate the EMT program. Upon the initiation of EMT, cell junctions are relocalized and/or degraded. The dissolution of tight junctions during EMT is accompanied by decreased claudin and occludin expression, and the diffusion of zonula occludens 1 (ZO1; also known as TJP1) from cell–cell contacts. During the destabilization of adherens junctions, the epithelial cadherin CDH1 is cleaved at the

plasma membrane and subsequently degraded. The downregulation of CDH1 is balanced by the increased expression of the mesenchymal neural cadherin CDH2 that alters cell adhesion. During EMT cells lose their apical-basal polarity and acquire a front-rear polarity. The cortical actin cytoskeleton is reorganized and cells acquire directional motility. EMT progression also requires disruption and reorganization of cell-Extracellular Matrix (ECM) interactions. Further, secretion of proteases, like the metalloproteinases MMP2 and MMP9, enable ECM degradation and tissues invasion (Kalluri & Weinberg, 2009, Lamouille, 2014, and literature cited in there).

EMT progression is also associated with differential splicing of nascent RNAs. Alternative splicing modulation in EMT is well illustrated by changes in the isoform expression of the adhesion protein CD44 and FGFR2. Indeed, during EMT there is a switch from epithelial to mesenchymal splice isoforms of these molecules (Lamouille, 2014 and literature in it).

2.3.2 Role of TGF-β in EMT

Among the signaling pathways that initiate EMT, TGF-β family has a predominant role. TGF-β family comprises three TGF-βs, two activins, several bone morphogenetic proteins (BMPs) and other ligands that all act through binary combinations of transmembrane kinase receptors. TGF-β1 is able to induce EMT in development, wound healing, fibrosis and cancer (Lamouille et al., 2014).

Binding of TGF-β to cell surface receptor enables the “type II” TGF-β family receptors to phosphorylate and activate the “type I” transmembrane kinases which then phosphorylate the C termini of the intracellular signaling effectors SMADs. The receptor-activated SMAD2 and/or SMAD3 combine with SMAD4 to form trimeric SMAD complexes. Following their translocation into the nucleus, SMAD complexes combine with

DNA-binding transcription factors at regulatory gene sequences and activate or repress transcription (Lamouille et al., 2014).

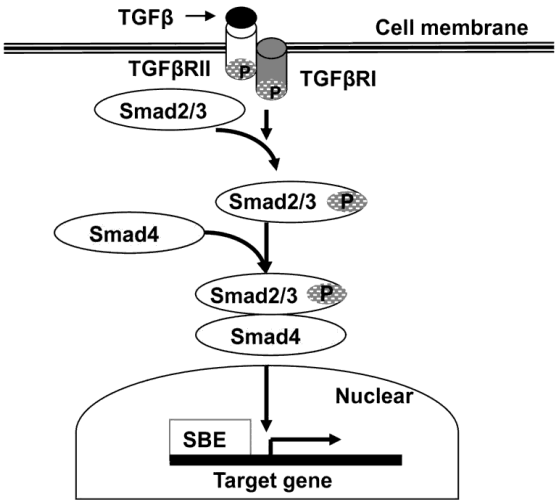


Fig IV Schematic of Smad mediated TGFβ signaling pathway.

TGFβ ligand binds to TGFβRII/TGFβRI receptors leading to phosphorylation of Smad2/3. Phosphorylated Smad2/3 binds to Smad4 to form a protein complex that undergoes nuclear translocation and regulates the expression of TGFβ target genes through binding to the Smad-binding element (SBE).

Han and Wang; Cell & Bioscience 2011

In response to TGF-β, SMAD complexes activate the expression and increase the activity of EMT transcription factors like SNAIL1, a master regulator of EMT. SMAD3-SMAD4 also cooperates with SNAIL1 in response to TGF-β and thus relays the TGF-β-activated repression of genes encoding, among others, the E-cadherin and the Occludin (Vincent et al., 2009).

2.4 The Two faces of TGF-β1 in Breast Cancer

In normal physiological conditions, TGF-β1 is a potent inhibitor of growth of epithelial cells. TGF-β1 primarily targets G1 events by regulating the expression of several genes promoting cell cycle arrest. Specifically, TGF-β1 mainly activates the transcription of cyclin-dependent kinases inhibitors such as p21 (encoded by the CDKN1A gene) and p15 (encoded by CDKN2B gene) (Zarzynska, 2014 and references in it).

The role of TGF-β1 in pathological conditions, specifically in breast cancer, has been shown to be multifaceted, dependent on the tumor stage. In early stages of breast cancer TGF-β1 inhibits epithelial cell cycle progression and promotes apoptosis, showing tumor suppressive effects. In late stages of breast cancer, TGF-β1 changes its action into a tumor promoter and several studies have shown a link between TGF-β1 treatment and increased tumor progression, higher cell motility, cancer invasiveness and metastasis (Zarzynska, 2014, and references in it). This is known as "TGF-β1 paradox" and it is still objects of deep investigation (Taylor et al., 2010; Parvani et al., 2011). Elucidation of the molecular mechanisms responsible for conferring oncogenic activities of TGF-β1 might provide new therapeutic opportunities to alleviate metastatic progression and disease recurrence of breast cancer.

2.5 KHSRP

The single-stranded nucleic acid binding protein KHSRP (K Homology -type splicing regulatory protein) modulates RNA life and gene expression at multiple levels and controls important cellular functions such as proliferation, differentiation and EMT. (Briata, 2016 for a review).

On the structural level, KHSRP is a 75-KDa protein organized in three distinct regions (Fig V): a central region that includes four KH domains responsible for nucleic acid

binding and two N- and C- terminal regions, respectively, that contain several sites for post-translational modification as well as protein interaction motifs. This modular structure and the resulting binding flexibility allow the protein to interact with multiple targets in order to regulate gene expression (García-Mayoral et al., 2007).

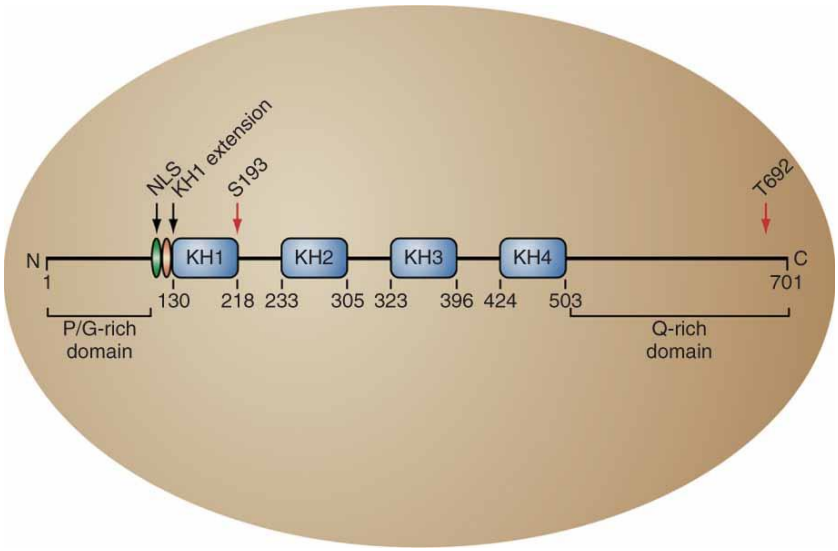


Fig. V. Schematic representation of the primary structure of human KHSRP

Gherzi et al., WIREs RNA, 2010

KHSRP plays an important role in post-transcriptional gene expression regulation, primarily by modulating mRNA decay and miRNA biogenesis (Gherzi et al., 2010).

mRNA stability varies considerably from one mRNA species to another and plays an important role in regulating gene expression levels. Differential mRNA decay rates are determined by specific cis-acting elements within the mRNA molecule. KHSRP promotes mRNA degradation by AU-Rich Element (ARE)-mediated decay (AMD). The ARE elements are found in the 3'untranslated region (3'UTR) of transcripts encoding cytokines, transcription factors, cell cycle regulators and proto-oncogenes (Gu et al., 2002; Lin et al., 2009). Biochemical and functional analyses showed that KHSRP binds to ARE elements and then recruits the exosome and other enzymatic complexes required for mRNA

degradation (Khabar, 2005; Eberhardt, 2007; Hao & Baltimore, 2009). Extracellular signals are able to control mRNA decay rates by targeting KHSRP. At least two signaling pathways, the MAPK p38 and Akt/PKB, have been shown to target decay-promoting ARE-binding proteins including KHSRP. The analysis of KHSRP primary structure revealed the presence of several potential phosphorylation sites, with two of them proved to be involved in regulation of KHSRP function in AMD. This finding strongly suggests that post-translational modification of KHSRP by phosporylation plays an important role in mRNA decay (Briata et al., 2015 and literature in it).

Expanding the array of KHSRP functions on RNA metabolism, the protein has been proved to promote maturation of select miRNAs from precursors. KHSRP binds to the terminal loop (TL) of a cohort of miRNA precursors and interacts with both Drosha and Dicer complexes in nuclear and cytoplasmic compartments, respectively. MiRNAs whose maturation is favored by KHSRP exert important functions in controlling cell proliferation, immune response, metabolic homeostasis, and cell fate determination. For example, KHSRP knockdown reduces the expression of mature let-7a and significantly upregulates cell proliferation in human bone osteosarcoma epithelial cells U2OS (Trabucchi et al., 2009).

We have previously shown that KHSRP plays an important role also in mRNA splicing (Puppo et al., 2016). Specifically, KHSRP maintains the epithelial identity of mammary gland cells by controlling the alternative splicing of specific pre-mRNAs in addition to favoring maturation of select miRNAs. KHSRP belongs to a ribonucleoprotein complex that includes hnRNPA1, and the two proteins cooperate in promoting epithelial-type exon usage of select pre-mRNAs. We showed that TGFβ-induced EMT in mammary gland cells requires silencing of KHSRP while sustained KHSRP expression limits TGF-β-dependent induction of EMT factors and cell migration. On the contrary, KHSRP

knockdown in untreated cells mimics TGF-β-induced EMT. Genome-wide sequencing analyses revealed that KHSRP controls (1) the levels of mature miR-192-5p, a microRNA that targets a group of EMT factors, and (2) the alternative splicing of a cohort of pre-mRNAs related to cell adhesion and motility including *Cd44* and *Fgfr2* (Puppo et al., 2016).

In the laboratory where I conducted my thesis, it has been previously described the interaction of KHSRP with a lncRNA and its consequences in gene expression regulation. KHSRP directly binds to H19 lncRNA in undifferentiated multipotent mesenchymal C2C12 cells and this interaction favors KHSRP-mediated decay of labile mRNAs such as myogenin. During the early steps of myogenesis, modulation by specific signaling pathways induces KHSRP dismissal from H19 and, as a consequence, myogenin mRNA is stabilized. Concurrently, KHSRP is repurposed to promote maturation of myogenic miRNAs thus favoring myogenic differentiation (Giovarelli et al., 2014).

3. AIM OF THE STUDY AND SUMMARY

Studies described above prompted us to search for potential lncRNAs interacting with KHSRP and whose expression/function is modulated during EMT.

We identified a previously uncharacterized mammalian lncRNA expressed in epithelial tissues that we termed EPR (after Epithelial Program Regulator). EPR came to our attention due to its ability to interact with KHSRP and to counteract TGF-β-induced EMT. EPR contains an open reading frame (ORF) that is translated into a small peptide localized at epithelial cell junctions. We found that EPR regulates a large set of target transcripts independently of the peptide biogenesis. Our studies revealed that EPR interacts with chromatin, regulates *Cdkn1a* gene expression by affecting both its transcription and mRNA decay, and controls cell proliferation in both immortalized and transformed mammary gland cells as well as in a mouse model of orthotopic transplantation. We propose that EPR enables epithelial cells to control proliferation by modulating waves of gene expression in response to stimuli.

4. RESULTS

4.1 Identification and characterization of EPR, an epithelial cell type-enriched lncRNA.

This study was initiated in an attempt to identify lncRNAs which are able to interact with KHSRP and whose expression is regulated by TGF-β in immortalized murine mammary gland NMuMG cells. To this end, we leveraged RNA-sequencing (RNA-Seq) and anti-KHSRP RNP complexes Immunoprecipitation followed by RNA sequencing (RIP-Seq) analyses performed in untreated or TGF-β-treated NMuMG cells. TGF-β treatment significantly reduced or increased the levels of 110 and 194 lncRNAs, respectively (|log2 fold changes| >2.0, $p<0.01$; Fig. 1A, left panel) while RIP-Seq analysis showed that TGF-β modulates the interaction of KHSRP with 67 lncRNAs (|log2 fold changes| >2.0, $p<0.01$; Fig. 1A, right panel). Among a set of lncRNA candidates of potential interest in EMT, we focused on the previously uncharacterized BC030870 (ENSMUSG00000074300, located on mouse chromosome 8 and transcribed in reverse orientation) that we renamed EPR. RIP analysis followed by qRT-PCR as well as band-shift analysis confirmed that EPR directly interacts with KHSRP (Fig. 1B). TGF-β induced a slight increase of EPR levels followed by rapid down-regulation (Fig. 1C) that accounts for the reduced interaction between KHSRP and EPR upon a 6 h treatment (Fig. 1A, right panel). SMAD complexes are major effectors of TGF-β-dependent transcriptional regulation (Budi et al., 2017) and our ChIP-qPCR showed that SMAD3 interacts with EPR promoter. SMAD3 binding to EPR promoter is enhanced by a 1 hour TGF-β treatment while falls below basal level after 6 hours (Fig. 1D).

EPR was expressed during embryonic development (Fig.1E) and in epithelial tissues of adult mice with a prevalence in the gastro-intestinal tract, lung, kidney and mammary

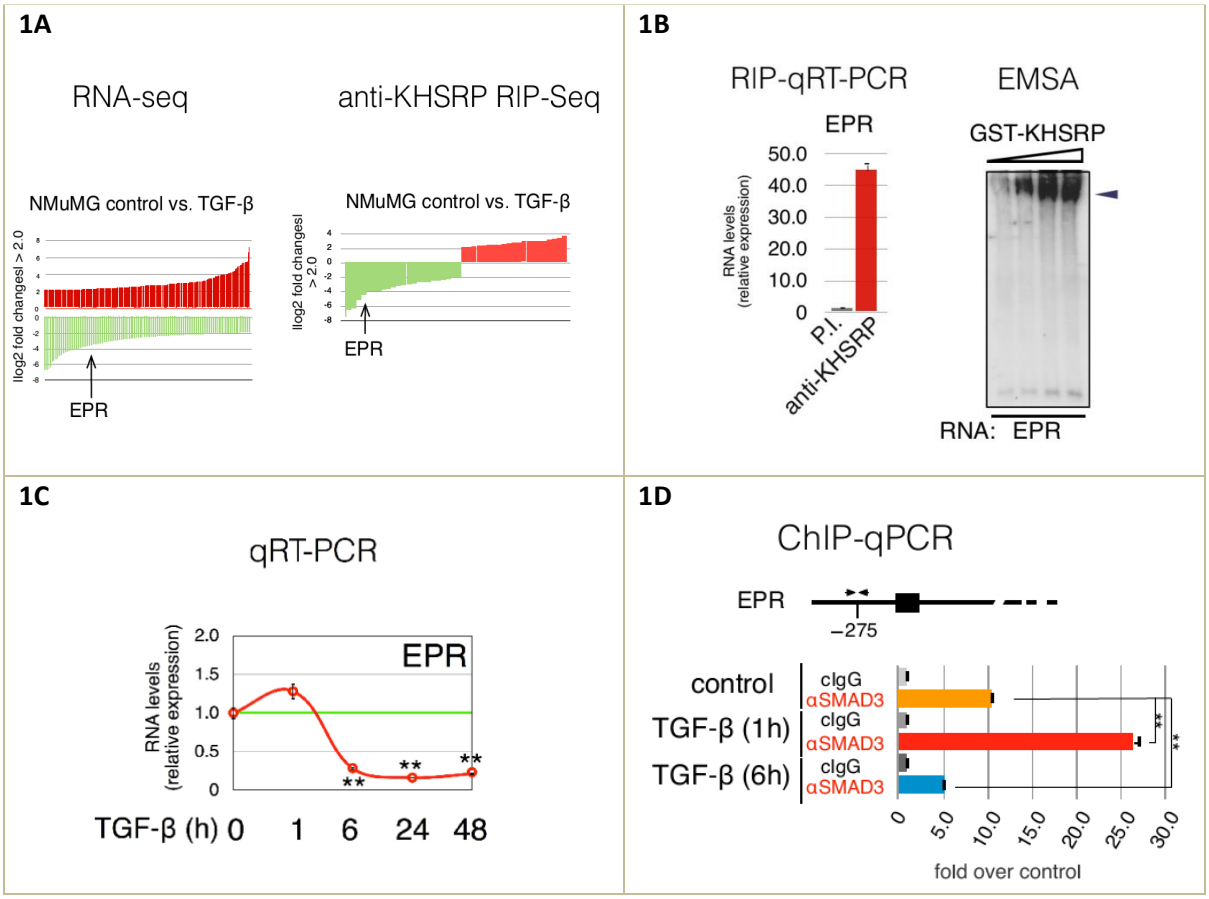
gland (Fig. 1F). EPR was polyadenylated, spliced and almost equally distributed in the cytoplasm, nucleoplasm and chromatin of NMuMG cells (Fig. 1G, 1H). LINC01207 (a.k.a. SMIM31, located on chromosome 4 and transcribed in forward orientation; hereafter indicated as h.EPR) is the human orthologue of EPR and displays superimposable epithelial tissue-enhanced expression (as evaluated through the Human BodyMap 2.0 data from Illumina; Fig. 1I). Bioinformatics analysis performed on RNA-Seq data derived from different subpopulations of normal breast cells isolated by FACS analysis from reduction mammoplasty specimens (Pellacani et al., 2016), revealed that h.EPR is expressed exclusively in differentiated luminal cells of mammary gland (Fig. 1J).

In order to investigate the potential role of EPR in TGF-β-induced EMT, we decided to counteract its TGF-β-dependent down-regulation by stably overexpressing the lncRNA in NMuMG cells (overexpression was 3-12-fold compared to the respective mock cells [empty vector-transfected], in different transfectant pools). The levels of transfected EPR were not affected by TGF-β treatment (data not shown). EPR overexpression prevented TGF-β-dependent morphological and gene expression changes by hindering down-regulation of epithelial markers (*Cdh1*, *Ocln*) and induction of mesenchymal markers (*Fn1*, *Fstl1*, *Zeb2*, *Adam12*) (Fig. 1K, 1L). Strikingly, EPR overexpression changed the levels of epithelial and mesenchymal markers and induced a cobblestone-like cell morphology even in the absence of TGF-β treatment (Fig 1M, 1N, 1O). As expectable, EPR overexpression in NMuMG cells significantly limited their migratory potential both in the absence and in the presence of TGF-β treatment (Fig. 1P). Conversely, transient silencing of EPR down-regulated the mRNA levels of epithelial markers while enhanced the levels of mesenchymal markers (Fig. 1Q). Further, transient EPR silencing rescued the gene expression changes induced by stable overexpression of the lncRNA (Fig. 1R). A link between the expression of h.EPR and EMT-related factors was also observed in human

normal breast samples where h.EPR positively correlated with epithelial markers —such as CDH1 and OCLN— and negatively correlated with mesenchymal markers such as VIM and SNAI1 (Fig. 1S).

In conclusion, the name EPR that we assigned to lncRNA BC030870 (after Epithelial Program Regulator) is consistent with its enriched expression in epithelial cells and with the up-regulation of epithelial markers and down-regulation of mesenchymal markers induced by its overexpression.

FIG.1 EPR displays epithelial expression and antagonizes TGF-β-induced EMT in mammary gland cells.



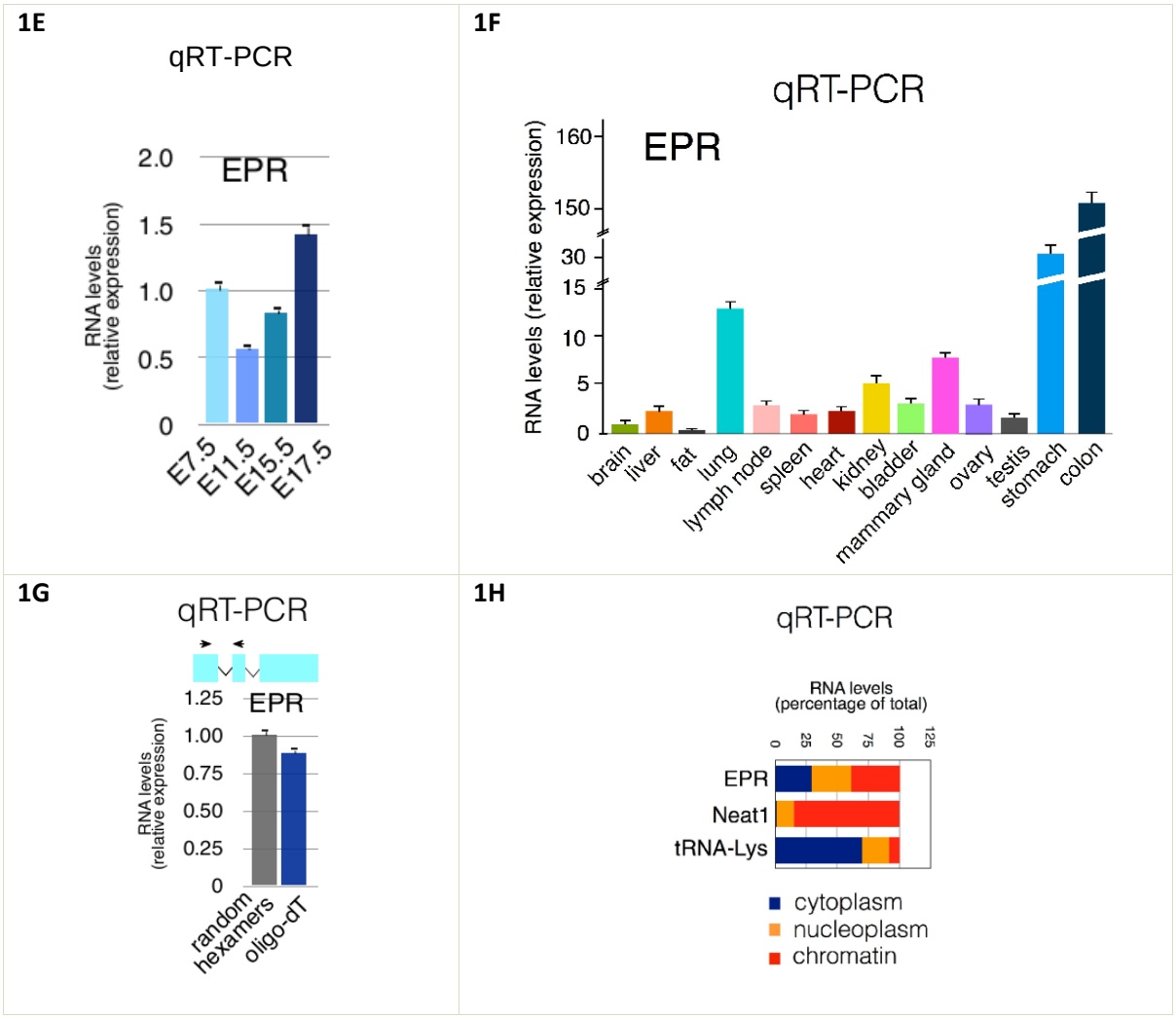
1A. RNA-Seq experiments indicated that TGF-βtreatment (24 hours) modulates (|log2 fold 8 changes| >2.0, p<0.01) the expression of 304 lncRNAs in NMuMG cells. Arrow points to EPR.

1B. RIP-Seq analysis revealed that the interaction of KHSRP with 67 lncRNAS is modulated (|log2 fold changes| >2.0, p<0.01) by TGF-βtreatment (6 hours). Arrow points to EPR. Total extracts were prepared from NMuMG cells and immunoprecipitated as indicated. RNA was purified from immunocomplexes and analyzed by qRT-PCR to quantitate EPR levels. Electrophoretic mobility shift assays (EMSA) performed to detect the interaction between highly purified recombinant KHSRP (50–300 nM) with ^{32}P-labeled RNA corresponding to nucleotides 246-407 of EPR. A representative autoradiogram is displayed. Arrowhead points to the KHSRP-EPR complex.

1C. Quantitative RT-PCR (qRT-PCR) analysis of EPR in NMuMG cells serum-starved (2% FBS, 16h) and either treated with TGF-β (10 ng/ml) for the indicated times or untreated (time 0).

1D. Chromatin prepared from NMuMG cells serum-starved and either treated with TGF-βfor the indicated times or untreated (control) was immunoprecipitated using either normal rabbit IgG (cIgG) or affinity-purified anti-SMAD3 rabbit polyclonal antibody. The association of SMAD3 with EPR promoter (schematic on the top) was quantitated by qPCR using specific primers (indicated as arrowheads).

The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).



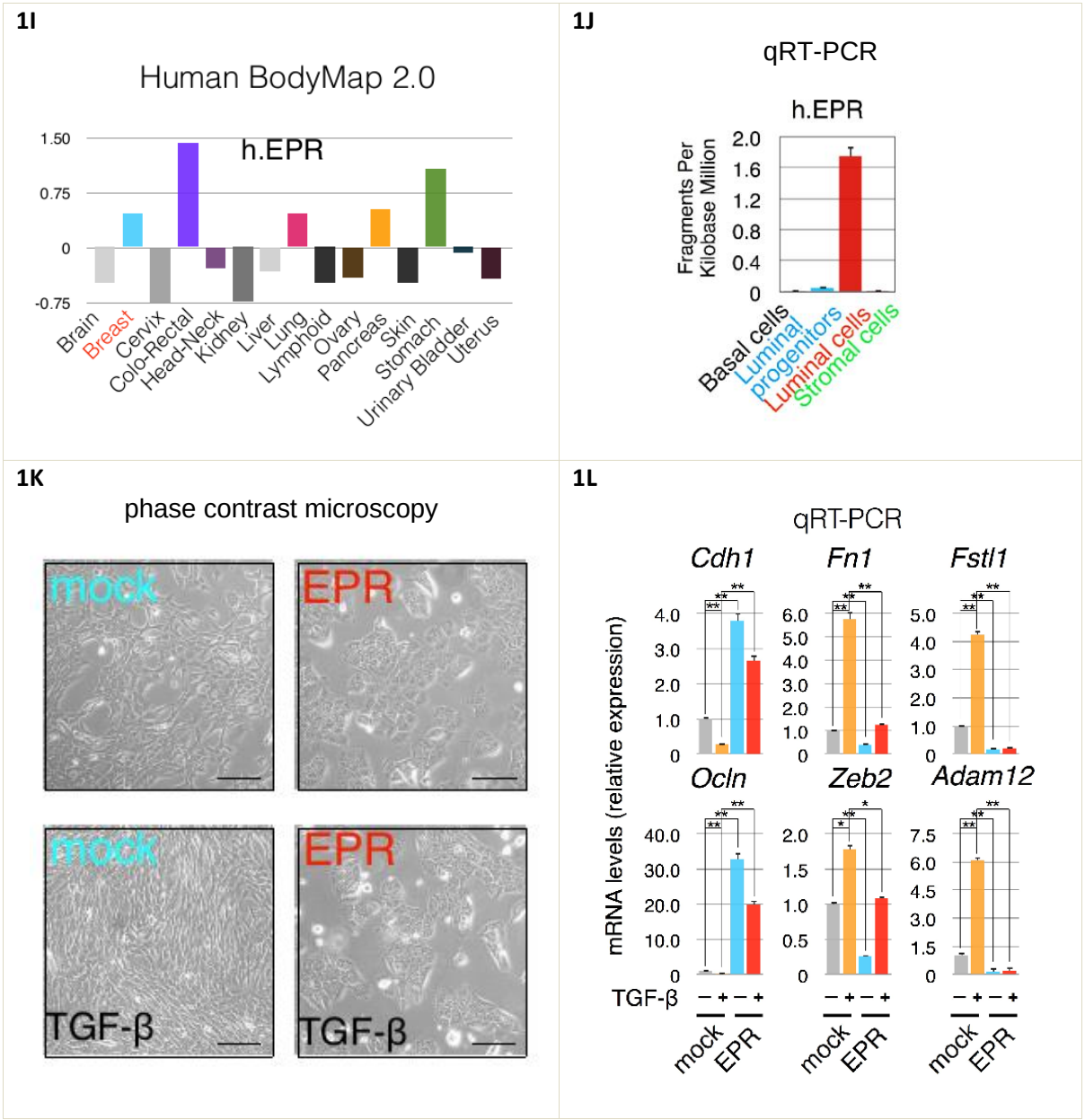
1E. qRT-PCR analysis of EPR expression in total RNA extracted from mouse embryos at the indicated gestational ages.

1F. qRT-PCR analysis of EPR in the indicated mouse tissues**.**

1G. qPCR analysis of EPR expression evaluated using a set of primers spanning exons 1 and 2 of EPR sequence in RNA samples from NMuMG cells retro-transcribed using either random hexamers or oligo-dT (as indicated). The position of primers that have been used is indicated in the schematic on the top**.**

1H. NMuMG cells were fractionated and RNA was prepared from cytoplasm, nucleoplasm, and chromatin and analyzed by qRT-PCR to quantitate the indicated RNAs.

The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).



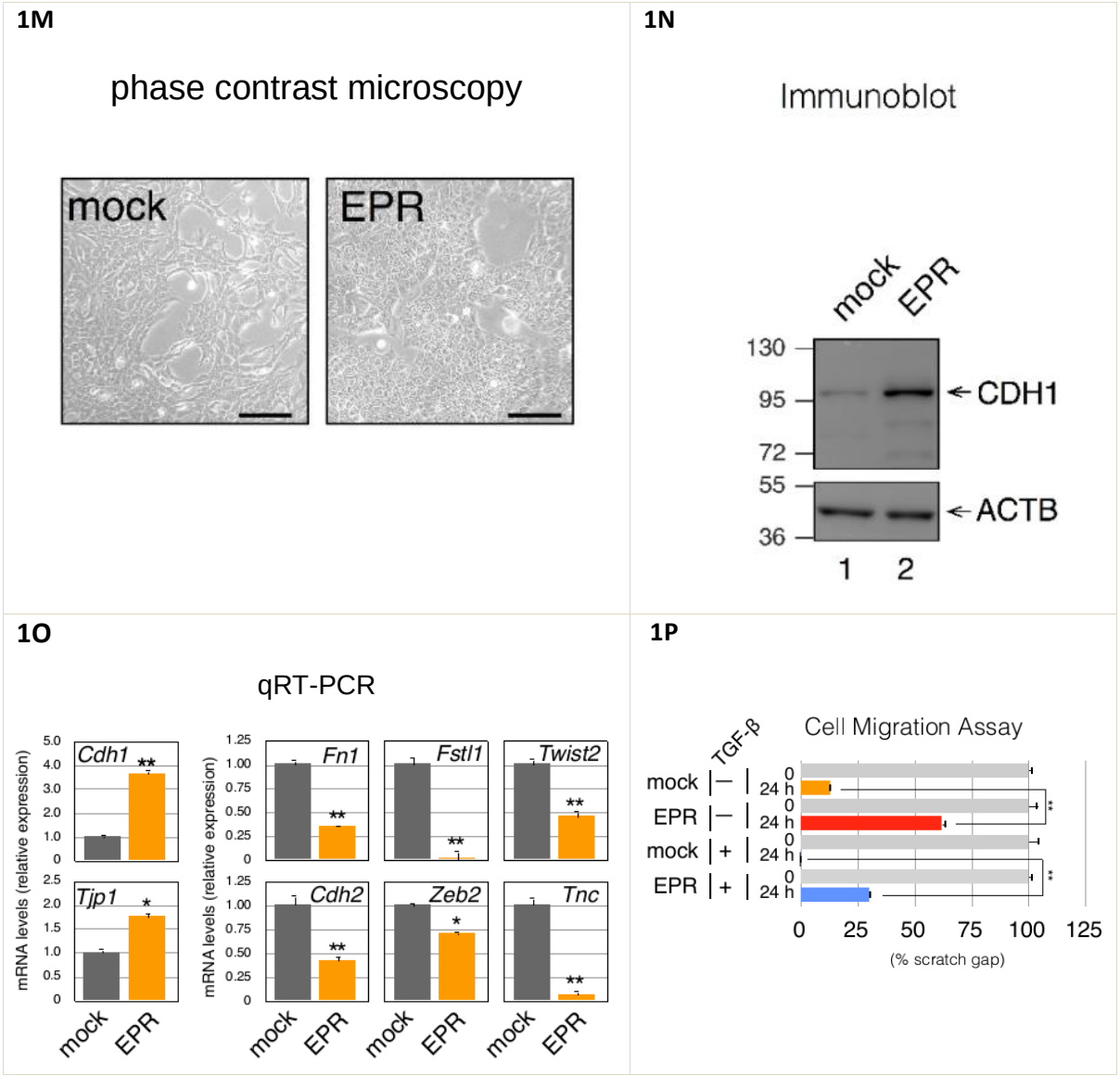
1I. The expression levels of h.EPR in multiple human tissue were extracted from the Illumina Human Body Map 2.0 data. Data were median centered to distinguish the tissues with high expression of h.EPR. Breast has been highlighted in red.

1J. qRT-PCR analysis of h.EPR in normal human breast cells isolated from reduction mammoplasty specimens

1K. Phase contrast microscopy of either mock or EPR-overexpressing (EPR) NMuMG cells that were serum-starved and then either treated with TGF-β or untreated for 24 hr. Scale bars: 100 μm.

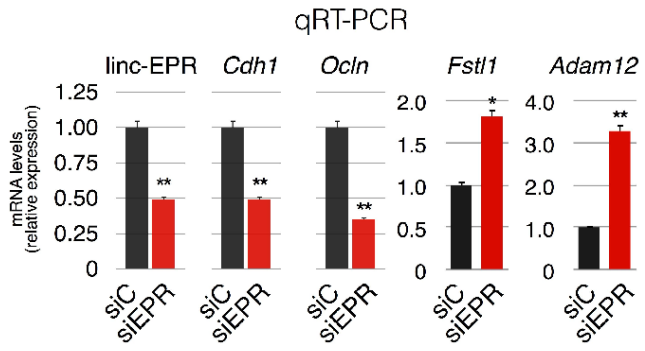
1L. qRT-PCR analysis of the indicated transcripts in either mock or EPR-overexpressing NMuMG cells serum-starved and either treated with TGF-β (+) for 24h or untreated (-).

The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).

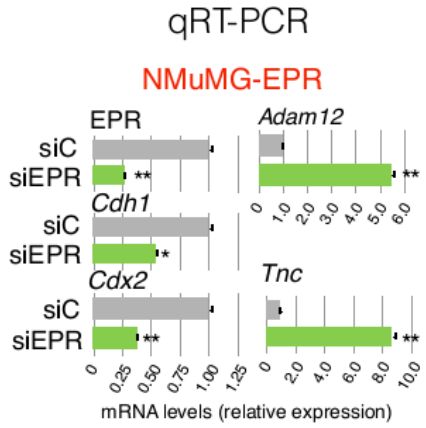


1M. Phase contrast microscopy of mock or EPR-overexpressing (EPR) NMuMG cells. Scale bars: 100 μm.
1N. Immunoblot analysis of total cell extracts from either mock or EPR-overexpressing NMuMG cells. The indicated antibodies were used. The position of molecular mass markers is indicated on the left. Representative gels are shown.
1O. qRT-PCR analysis of the indicated transcripts in either mock or EPR-overexpressing NMuMG cells.
1P. Scratch wound healing assays. Scratch wounds were introduced into confluent monolayers of either mock or EPR-overexpressing NMuMG cells. Cultures were either treated with TGF-β(for 24 hours (+)) or left untreated (-). Cultures were photographed and the width of the wound was measured at 0 hr and 24 hr after the scratch was made. The percentage of the scratch gap area for each culture condition was plotted.
Values of qRT-PCR and cell migration assays experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).

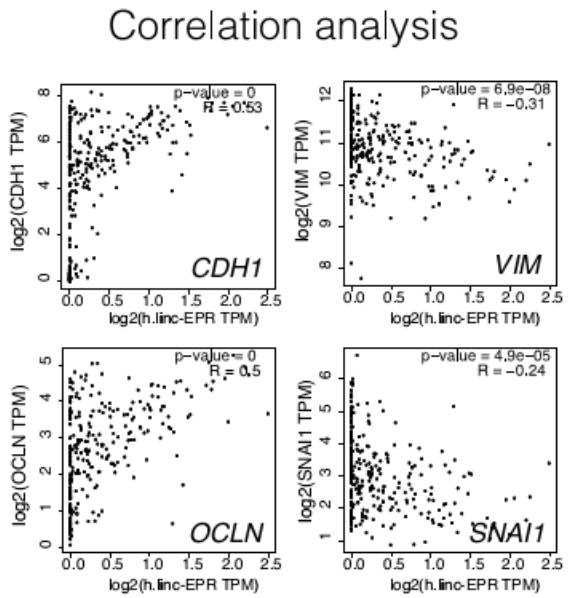
1Q



1R



1S



1Q. qRT-PCR analysis of the indicated transcripts in NMuMG cells transiently transfected with either control siRNA (siC) or siRNA designed to silence EPR-expression (siEPR).

1R. qRT-PCR analysis of the indicated transcriptsin either control siRNA (siC) or siEPR-transiently transfected (siEPR) NMuMG cells stably overexpressing EPR.

1S. Pearson correlation analysis between the expression levels of h.EPR and the indicated human transcript as calculated by the GEPIA website (http://gepia.cancer-pku.cn) using as input breast normal samples from the TCGA dataset and normal mammary gland tissues. Expression levels are plotted as Transcripts per million (TPM, log-scale). The p-value of Pearson correlation coefficient (R) and the p-value are presented.

The values of qRT-PCR shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).

4.2 EPR encodes a small polypeptide.

A few recent reports show that certain lncRNAs contain short ORFs that can be translated into peptides endowed with regulatory functions (Matsumoto et al., 2017; Huang et al., 2017; Andrews & Rothnagel, 2014; Makarewich & Olson, 2017 - for reviews). The analysis of EPR sequence revealed the presence of a 213 nucleotide-long ORF potentially encoding a 71-amino acid polypeptide that, interestingly, corresponds to the lncRNA region that displays the highest identity with the human orthologue (Fig. 2A). The putative polypeptide sequence is well conserved among mammalian species and *in silico* methods identified a conserved α-helical transmembrane domain while a predicted second α-helix was found in the putative cytosolic domain (Fig. 2B). Importantly, polysome fractionation followed by qRT-PCR analysis revealed that EPR localizes to actively translating polysomes (Fig 2C).

To investigate whether EPR ORF is translated, we inserted a FLAG tag at its 3' end and transiently transfected the resulting construct into HEK-293 cells (Fig. 2D, left). As shown in Fig. 2D (right), the ORF was translated into a short polypeptide of the expected molecular mass. To unambiguously prove the existence of the endogenous small EPR-encoded peptide (EPRp), the ORF was expressed in bacteria and the resulting peptide was purified and utilized as immunogen to generate a rabbit polyclonal antibody. Polyclonal anti-EPRp recognized a recombinant polypeptide transiently expressed in HEK-293 cells (Fig. 2E) and, most importantly, a ~ 8 KDa polypeptide in mouse gastrointestinal tract organs and breast (Fig. 2F). In keeping with EPR downregulation upon TGF-β treatment, the expression of the EPRp was downregulated in response to treatment with TGF-β for 24 h (Fig. 2G).

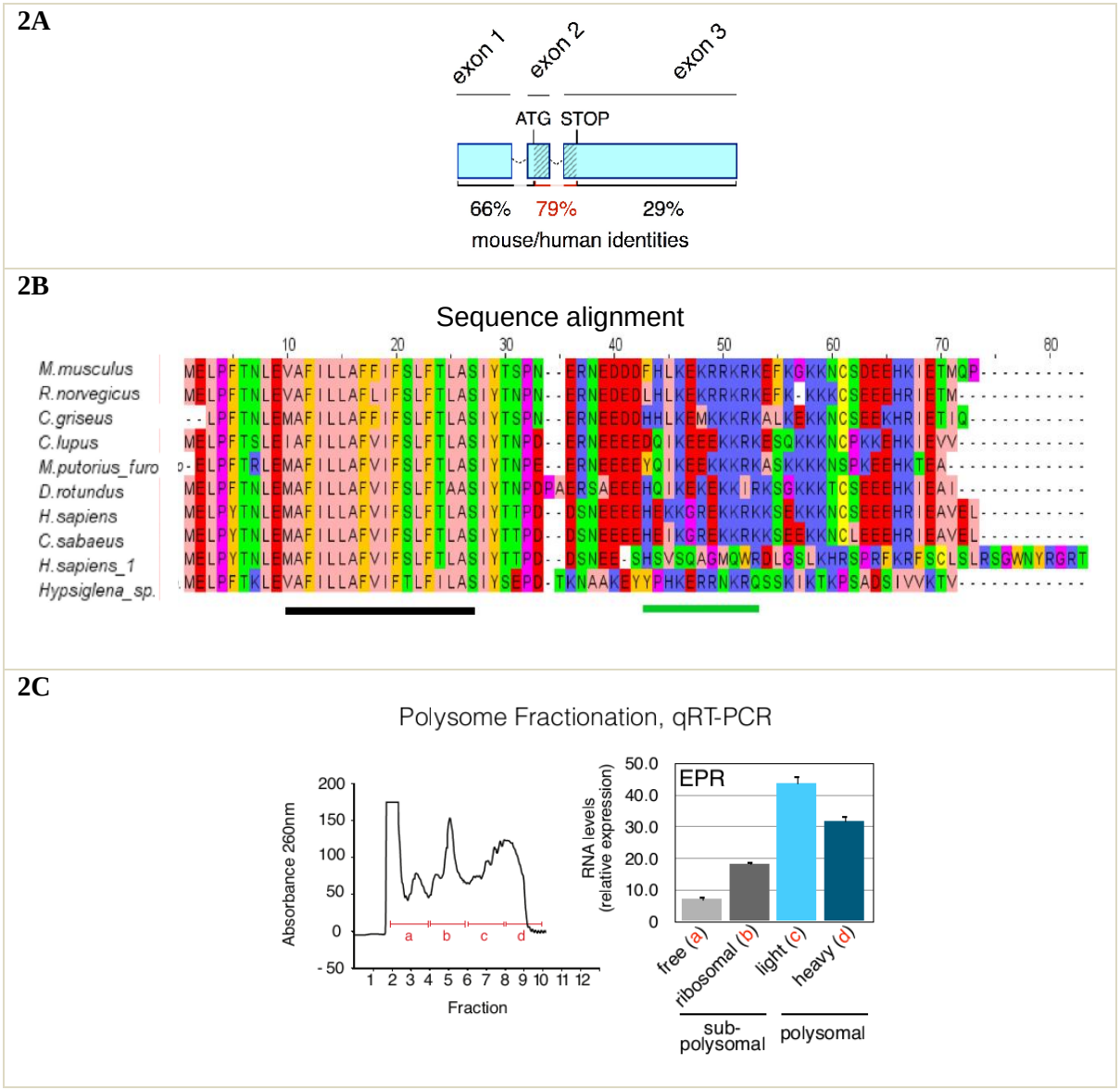
In order to identify the molecular partners of EPRp, we performed immunoaffinity purification of proteins interacting with EPR in NMuMG cells. Mass spectrometry (MS)

analysis of co-immunoprecipitating proteins separated by SDS-PAGE (Fig. 2H) revealed an enrichment in junctional and cytoskeletal proteins (Fig. 2I). Co-immunoprecipitation experiments confirmed that EPRp interacts with the tight junction proteins TJP1 (ZO-1) and CGN (Cingulin), with the tight and adherens junction protein CGNL1 (Paracingulin) as well as with the actin-associated proteins CTTN (Cortactin) and MYH9 (epithelial myosin-II) (Fig. 2J).

To establish whether the association of EPRp with junctional proteins reflects its subcellular localization, we performed immunofluorescence experiments in NMuMG cells stably transfected with either EPRp-FLAG or with a construct in which the second codon of the ORF —encoding glutamic acid, E, of EPRp— was mutagenized in order to obtain a STOP codon (see also below, EPRSTOPE-FLAG). Specific localization of FLAG signal at cell-cell junctions labelled by the junctional marker CGN was detected in cells stably expressing EPRp-FLAG (arrows in Fig. 2K) while no junctional FLAG labeling was detected in mock-transfected cells or in cells expressing the point mutant version unable to produce the peptide. CGN labeling was wavy and discontinuous in mock transfected cells and in cells expressing EPRSTOPE-FLAG, whereas it was linear and uninterrupted in cells expressing EPRp-FLAG, suggesting that EPRp overexpression promotes epithelial junction assembly and reorganization of the junction-associated actin cytoskeleton. A weak diffuse cytoplasmic staining observed in NMuMG cells expressing EPRp-FLAG might reflect EPRp interaction with cytoskeletal proteins (Fig. 2K).

On the basis of these results, we conclude that an ORF present in EPR is translated into a small peptide that is well conserved among species and that displays a junctional localization in epithelial cells.

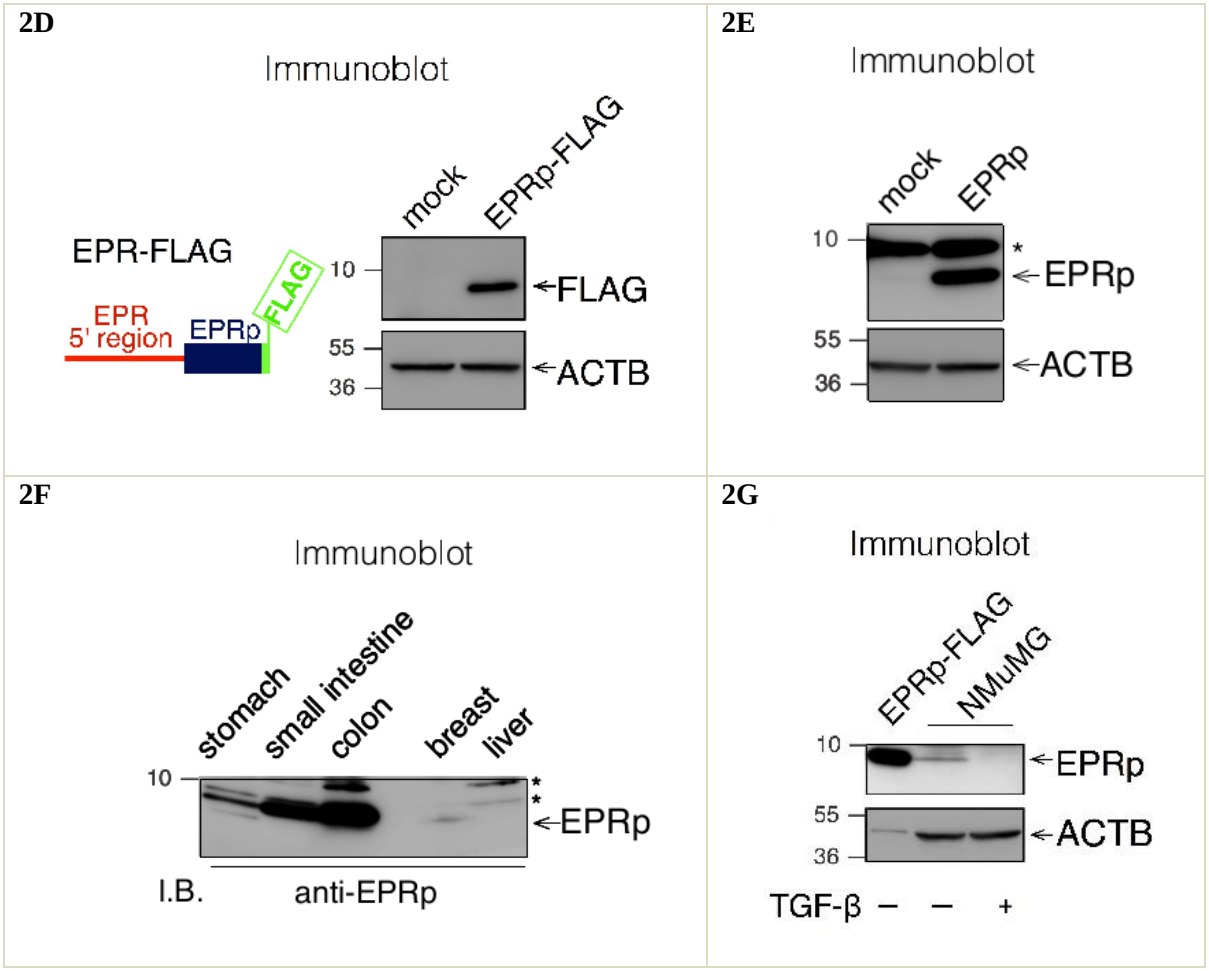
Figure 2. A small peptide (EPRp) originates from EPR and displays junctional localization in NMuMG cells.



2A. Schematic of the exon-intron structure (not in scale) of EPR, a putative ORF is dashed. Percentage of human/mouse identity in the putative ORF and in its flanking RNA regions is presented.

2B. Alignment of the predicted mammalian amino acids sequence encoded by the putative ORF present in EPR and in its orthologues. The position of the predicted transmembrane α-helix is shown as a solid black bar while the position of an additional cytoplasmic α-helix is shown as a solid green bar.

2C. Left panel. Representative example of polysome profiles obtained from cytoplasmic lysates of NMuMG cells fractionated by sucrose gradients. The relative absorbance of the various fractions is indicated. Four pools of fractions corresponding to free RNA + 40S (a), 60S + 80S (b), light polysomes (c), heavy polysomes (d) were prepared and processed for RNA extraction and qRT-PCR. Right panel. Relative expression of EPR across the four pools of sucrose gradient, cytoplasmic lysate fractions. The average relative expression and the standard error of three replicates are plotted.



2D. Left, diagram of the FLAG-fusion construct used for transfection (EPRp-FLAG); right, immunoblot analysis of total cell extracts from either mock or EPRp-FLAG-transfected HEK-293 cells.

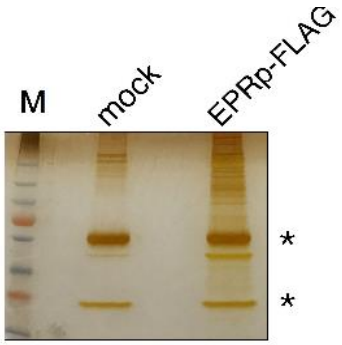
2E. Immunoblot analysis of total cell extracts from either mock or EPR-transfected HEK-293 cells. Asterisks mark nonspecific immunoreactivity.

2F. Immunoblot analysis of total extracts from the indicated mouse tissues. Asterisks mark nonspecific immunoreactivity.

2G. Immunoblot analysis of total cell extracts from NMuMG cells serum-starved and either treated with TGF-β (+) for 24h or untreated (-); extracts from EPRp-FLAG-transfected HEK-293 cells (EPRp-FLAG) represent a positive control.

2H

Silver staining after SDS PAGE

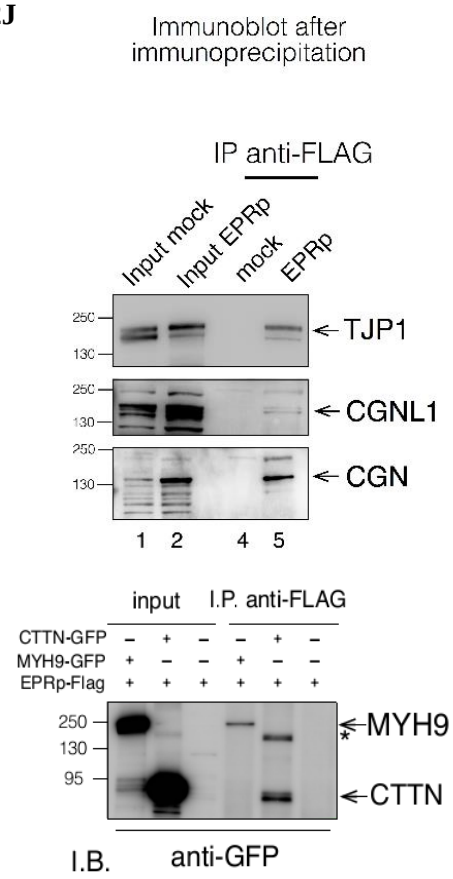


2I

Mass Spectrometry

Mame	Mol.Mass	Peptides	Name	Mol.Mass	Peptides
PLEC	534	5	MYO6	146	30
ITPR3	304	31	LMO7	131	20
SPTAN	285	51	MYO1B-E	132-116	13/14/17/15
FLUBA/B	280/278	23/23	PPP1R12A	115	10
SPTBN1	274	68	RAI14	109	8
MPRIP	256	6	CTNND1	105	5
SVIL	253	67	CTNNA1	100	5
MYO7A	250	17	JUP	82	6
MYOF	235	21	GSN	81	14
MYO18A	232	84	CALD1	60	16
MYH10/14	233/229	115/28	LAD1	59	12
MYH9	226	248	CTTN	57	5
ARHGEF17	222	16	ACTR2	45	11
MYLK	214	7	ARPC1B	41	5
MYO5B/A	213/212	9/108	TMOD3	39	12
AFDN	205	21	VCAM1	38	5
PPL	204	10	ARPC2	34	5
SIPA1L3/1	197/195	10/28	CAPZA1/2	33	6/5
TJP1	195	34	TRPM1	33	24
CLTC	192	25	CAPZB	31	16
IQGAP1	191	35	TRPM3	29	17
SCRIB	174	28	TRPM4	28	14
CGNL1	148	5			

2J

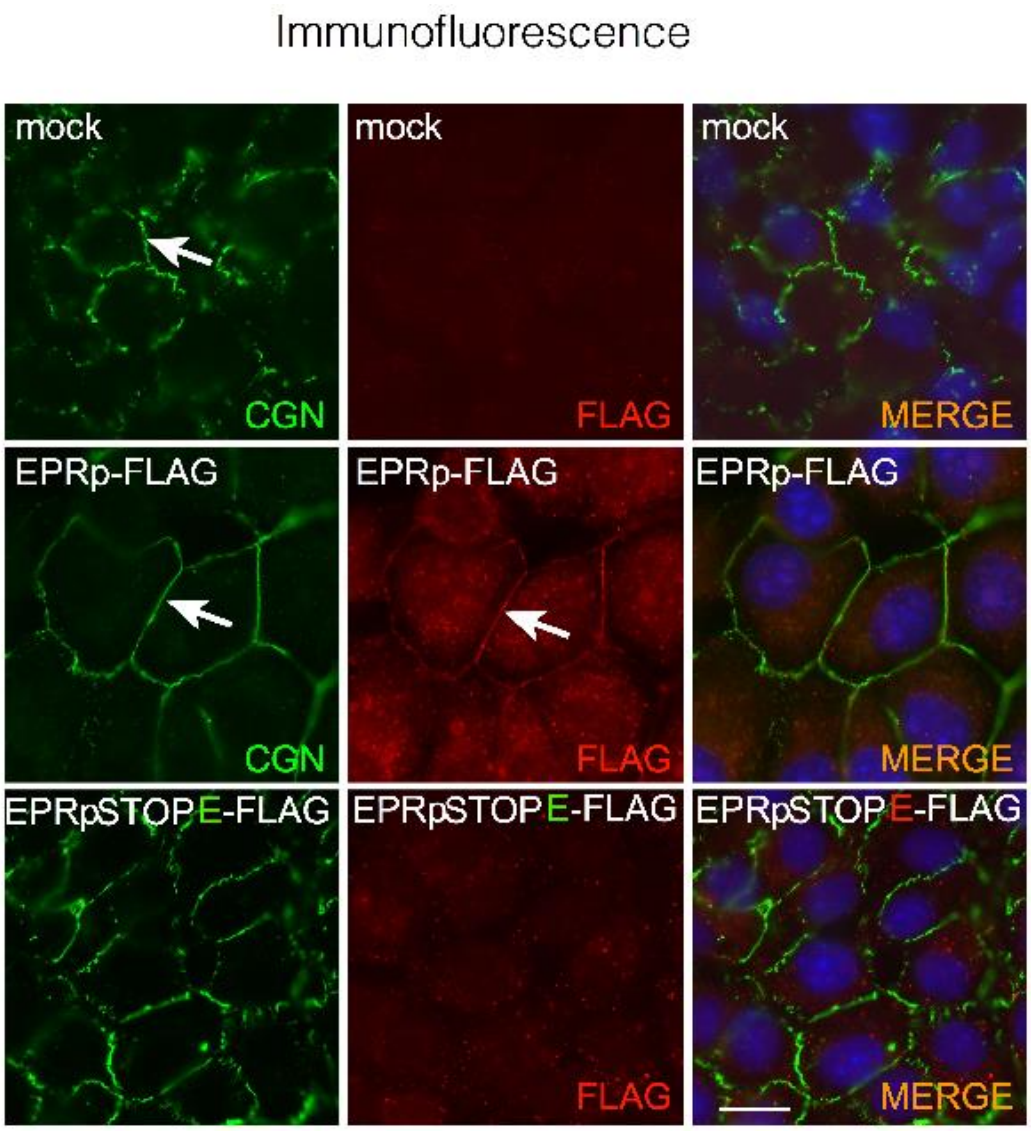


2H. SDS-PAGE analysis of total cell extracts from either mock or EPRp-FLAG- overexpressing (EPRp-FLAG) NMuMG cells immuno-purified using anti-FLAG monoclonal antibody. A representative silver-stained gel is shown. Asterisks indicate the position of immunoglobulin heavy and light chains.

2I. List of proteins that specifically interact with EPRp identified by mass spectrometry after immunoaffinity purification and SDS-PAGE analysis in NMuMG cells stably overexpressing FLAG-tagged EPRp.

2J. TOP PANEL: Co-immunoprecipitation of FLAG-tagged EPRp and distinct junctional proteins (as indicated) in total extracts from NMuMG cells stably transfected with EPRp-FLAG. BOTTOM PANEL: Co-immunoprecipitation of FLAG-tagged EPRp and distinct cytoskeletal proteins (as indicated) in total extracts from HEK-293 cells transiently co-transfected with EPRp-FLAG and the indicated GFP-fused proteins. Asterisks mark nonspecific immunoreactivity. For Immunoblots, the indicated antibodies were used; the position of molecular mass markers is indicated on the left and representative gels are shown.

2K



2K Immunofluorescence analysis of either mock or EPRp-FLAG- orEPRpSTOPE-FLAG- stably transfected NMuMG cells cultured to confluence, to allow formation of cell-cell junctions. Arrows point to CGN and EPRp-FLAG junctional localization.

4.3 EPR regulates gene expression in NMuMG cells.

We set out to investigate the function(s) of EPR in NMuMG cells. First, in order to answer the question whether the phenotypic changes that we observed by overexpressing EPR were caused by the lncRNA *per se*, the peptide or both, we performed transcriptome-wide RNA-Seq analyses in mock cells as well as in NMuMG cells overexpressing either EPR or a point-mutant version unable to produce the peptide (EPRSTOPE, for details see above and Fig. 3A). Bioinformatics analyses of RNA-Seq data revealed a vast rearrangement of the transcriptome as a consequence of both EPR and EPRSTOPE overexpression (data not shown). Gene ontology (GO) analysis of RNA-Seq results revealed the enrichment of terms related to epithelial morphogenesis, cell motility, cell migration, and epithelial cell proliferation among the top regulated categories (data not shown). Representative examples of transcripts either up-regulated or down-regulated by both EPR and EPRSTOPE are shown in Fig. 3B). In keeping with the sequence conservation between EPR and h.EPR, the overexpression of the human lncRNA in murine NMuMG cells yielded gene expression changes superimposable to those obtained by overexpressing the murine lncRNA (Fig. 3C).

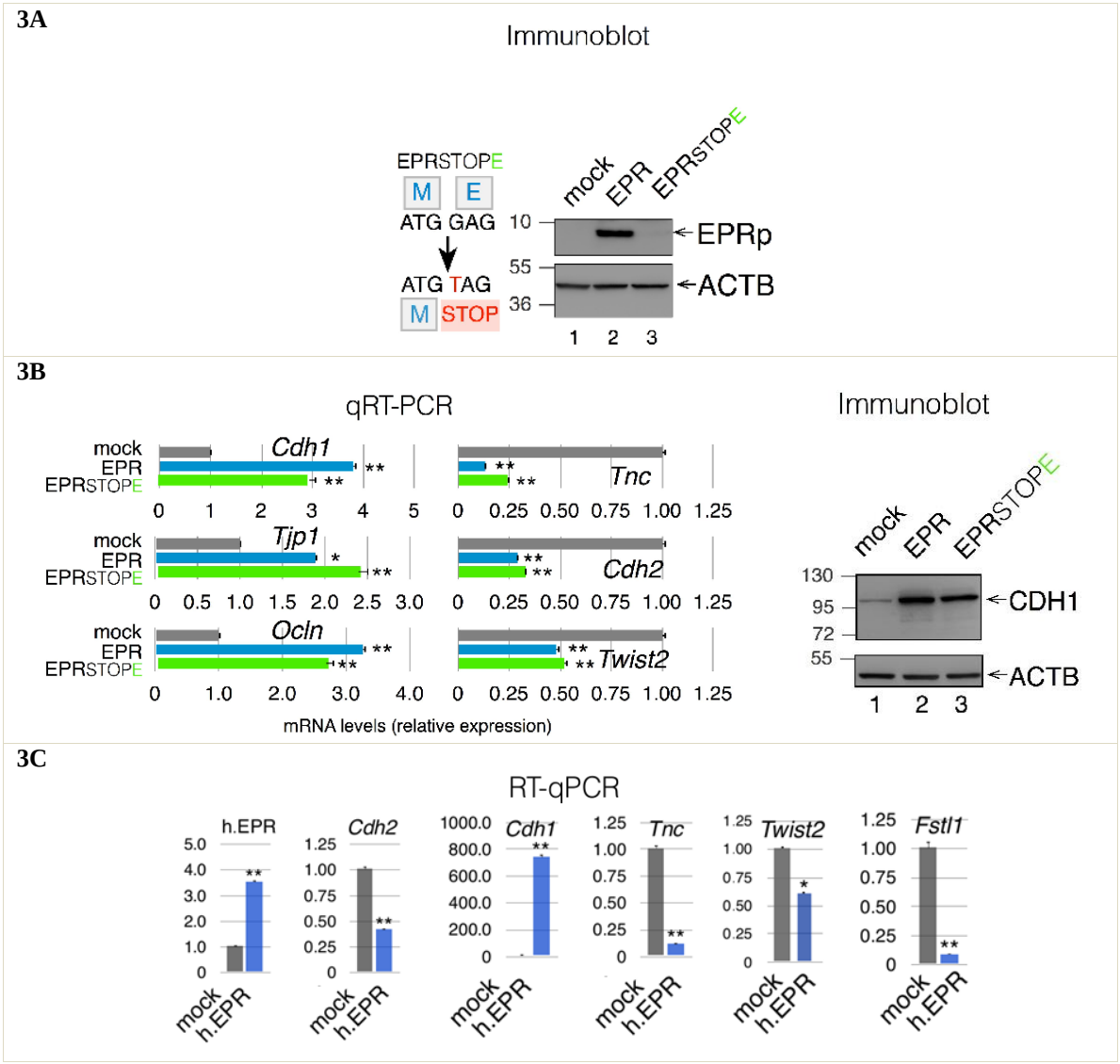
Interestingly, overexpression of either EPR or EPRSTOPE caused largely overlapping gene expression changes when compared to mock cells (Fig. 3D, left panel). When we directly compared gene expression changes induced by EPR or EPRSTOPE by applying stringent statistical criteria, we noticed that only a relatively small group of genes displayed expression changes dependent on the presence of EPR (Fig. 3D, right panel). The analysis of three independent NMuMG transfectant pools overexpressing EPRSTOPE, followed by qRT-PCR-based validation, allowed us to further restrict the number of transcripts whose levels are affected by EPR overexpression in a peptide-dependent way (Fig. 3E). These include transcripts encoding a calcium-dependent cell adhesion protein

(*Pcdh19*), two ion transporters (*Slc9a2*, *Scl39a4*), a cytokine receptor (*Fgfr2*) as well as a modulator of membrane transport and actin dynamics (*Anxa6*). Further, analysis of an additional EPR mutant (referred to as EPRSTOPM in which the start codon has been mutagenized to a STOP codon, see below) confirmed the restricted number of gene expression changes that can be ascribed to the peptide translation (Fig. 3F).

Considering the emerging evidence that some lncRNAs act locally (in *cis*) to regulate the expression of nearby genes, we investigated this possibility and RNA-Seq analysis revealed that the expression levels of genes proximal to EPR (*Palld, Cpe, Sc4mol, Klhl2, Tmem192, Tma16, Naf16, Nat2* and *Pssd,* localized over 8 MB of chromosome 8) are unaffected by the almost complete EPR down-regulation that occurs in NMuMG cells treated with TGF-β for 24 hours (Paola Briata et al., unpublished observation).

Altogether, transcriptome-wide analyses showed a EPR-dependent wide rearrangement of the transcriptome in NMuMG cells with relatively restricted effects on gene expression attributable to the peptide. Thus, we decided to focus our further studies on the peptide-independent functions of EPR.

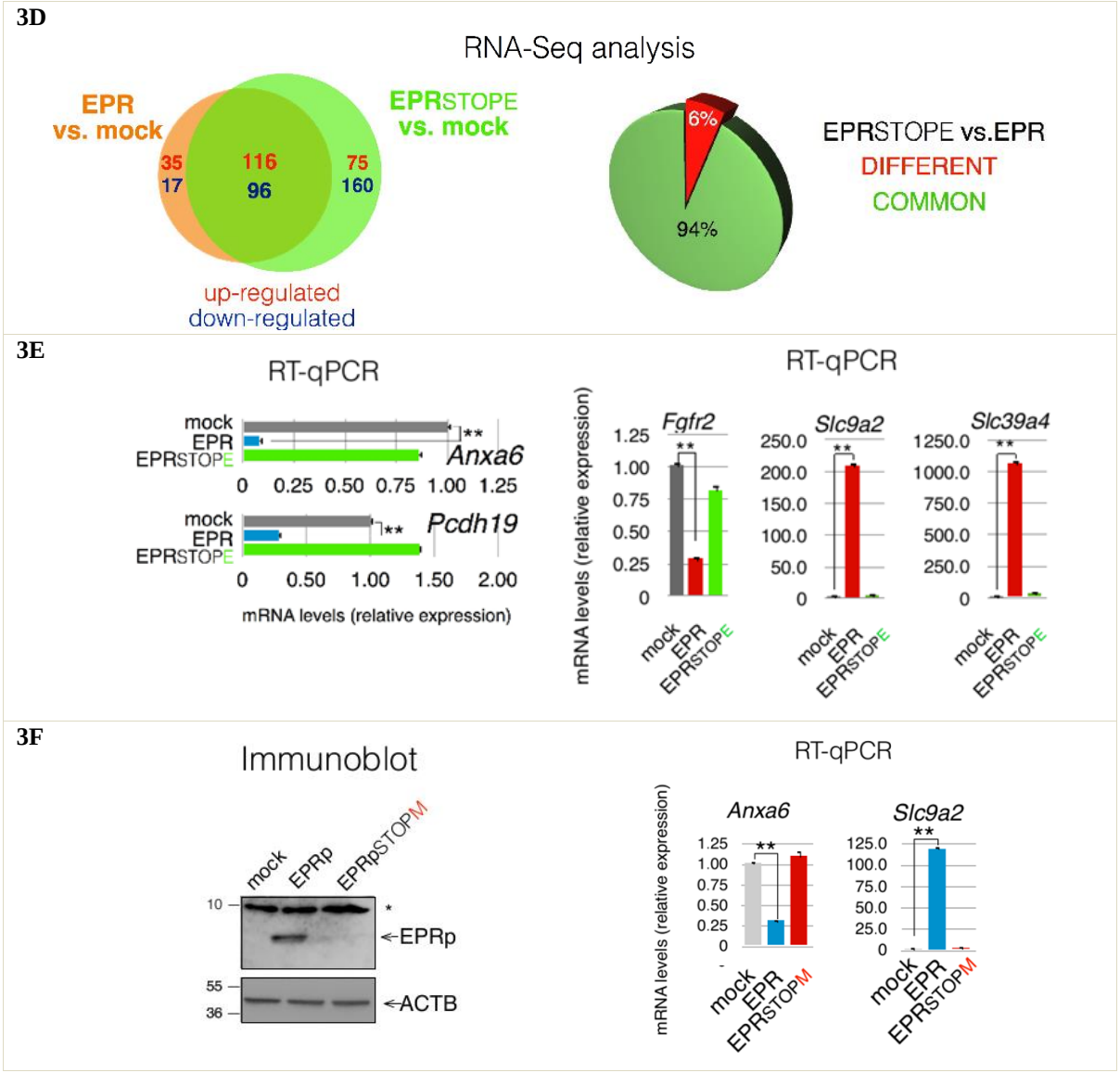
Figure 3. EPR overexpression reshapes NMuMG cells transcriptome independently of EPRp translation.



3A. Schematic of the point mutation that was introduced in the second codon of the EPR ORF to generate the EPRSTOPE construct (left panel). HEK-293 cells were transiently transfected with either EPR or EPRSTOPE constructs and the translation of EPR was assessed by immunoblotting (right panel).

3B. qRT-PCR analysis of the indicated transcripts in either mock, EPR- or EPRSTOPE- expressing NMuMG cells. Immunoblot analysis of total cell extracts from either mock, EPR- or EPRSTOPE-overexpressing NMuMG cells.

3C. qRT-PCR analysis of the indicated transcripts in either mock or human EPR (h.EPR) overexpressing NMuMG cells The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test). For Immunoblots, the indicated antibodies were used; the position of molecular mass markers is indicated on the left and representative gels are shown.



3D. Left panel, TopTable Venn diagram of the limma one-factor contrast analysis showing the overlap between transcripts either upregulated (numbers in red) or downregulated (numbers in blue) in NMuMG cells stably over-expressing either EPR or EPRSTOPE compared with mock cells. Only transcripts displaying |log2 fold expression difference| >2.5 ($p<0.0001$) were included in the graph and in further comparisons. Right panel, Pie diagram showing the percentage of gene expression changes common to NMuMG cells overexpressing EPR and EPRSTOPE as well as the percentage of changes that can be specifically attributed to the EPR translation (limma one-factor interaction contrast).

3E. qRT-PCR analysis of the indicated transcripts in either mock, EPR- or EPRSTOPE- overexpressing cells.

3F. Immunoblot analysis of HEK-293 cells transiently transfected with either the empty vector, EPR or EPRSTOPM (right panel). qRT-PCR analysis of the indicated transcripts in either mock, EPR or EPRSTOPM- overexpressing NMuMG cells (left panel). The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *$p< 0.01$,**$p < 0.001$ (Student's t test).

4.4 EPR regulates *Cdkn1a* gene expression and cell proliferation.

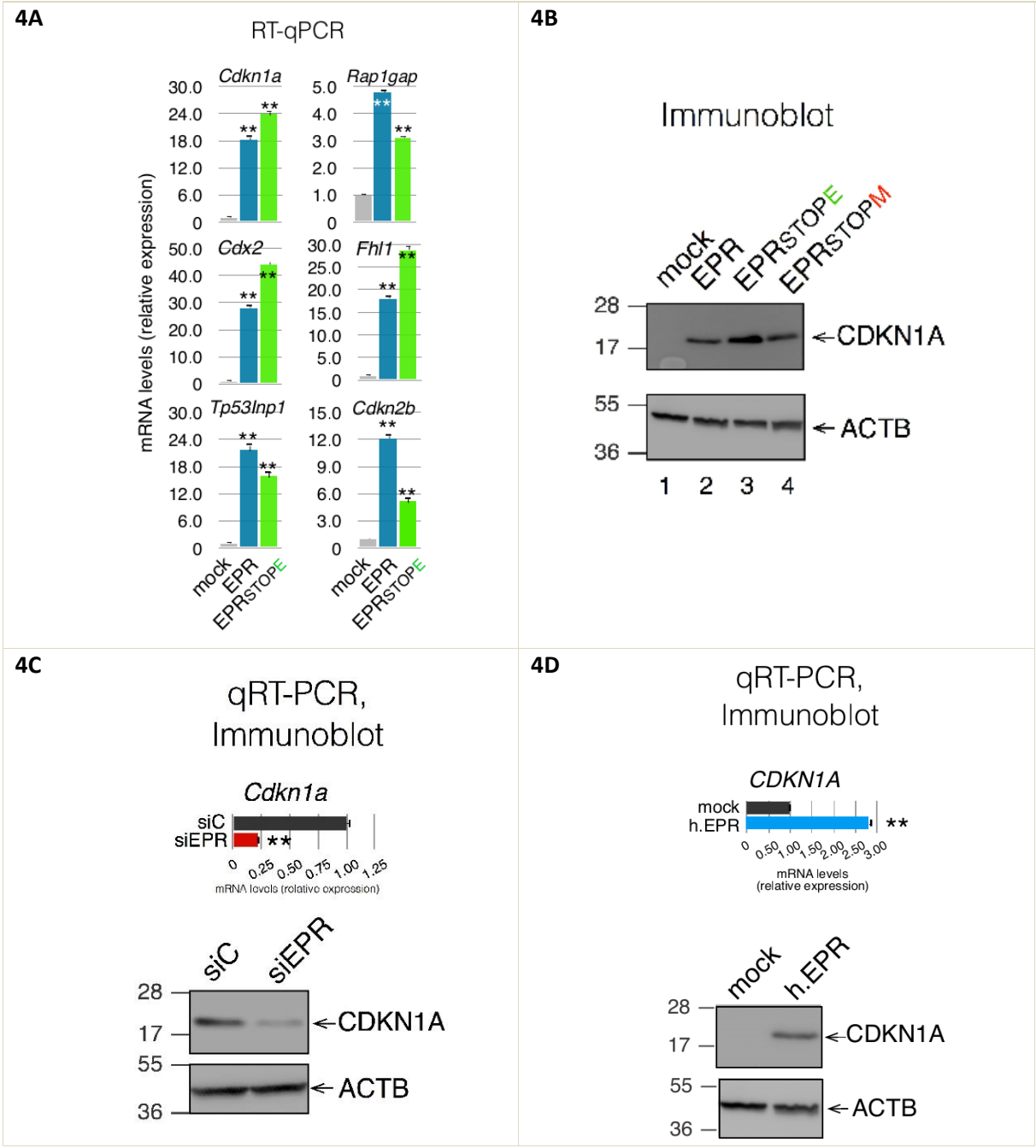
Among the GO terms significantly enriched by both EPR and EPRSTOPE we identified the category “regulation of epithelial cell proliferation”. Indeed, both EPR and EPRSTOPE overexpression significantly affected the levels of a group of transcripts belonging to this category including the cyclin-dependent kinase inhibitor *Cdkn1a* (a.k.a. $p21^{WAF1/Cip1}$) (Fig. 4A). Immunoblots presented in Fig. 4B show that overexpression of either EPR or EPRSTOPE or EPRSTOPM strongly enhanced CDKN1A levels. Conversely, EPR silencing strongly reduced CDKN1A expression (Fig. 4C). As expectable, CDKN1A levels were enhanced by overexpression of the human orthologue of EPR (Fig. 4D). Most important, we found that overexpression of either EPR or EPRSTOPE as well as of h.EPR strongly reduces cell proliferation rate in NMuMG cells (Fig. 4E).

Similar results were obtained by analyzing both cell proliferation and gene expression changes in different NMuMG cell pools transfected with EPR cloned in various vector backbones (data not shown). Cell cycle analysis demonstrated a relevant increment of cells arrested in G1 in the case of NMuMG cells transfected with either EPR or EPRSTOPE in comparison to mock cells (Fig. 4F). To exclude the possibility that gene expression changes that we prevously observed (Fig. 3B) might be dependent on the EPR-induced G1 arrest, we sorted cells in the G1 phase and analyzed gene expression changes by qRT-PCR. Data presented in Fig. 4G indicate that both EPR and EPRSTOPE induced gene expression changes in G1-enriched cells are superimposable to those observed in the total cell population (compare Fig. 4G with Fig. 3B).

Together, these results provide evidence that modulation of EPR levels regulates *Cdkn1a* gene expression and affects cell proliferation in NMuMG cells. Given the role of CDKN1A in promoting cell cycle arrest in response to many stimuli —including TGF-β

(Abbas & Dutta, 2016)—, we decided to focus our further mechanistic studies on the role of EPR in TGF-β-dependent regulation of *Cdkn1a* gene expression.

Fig 4. EPR regulates *Cdkn1a* gene expression and cell proliferation.

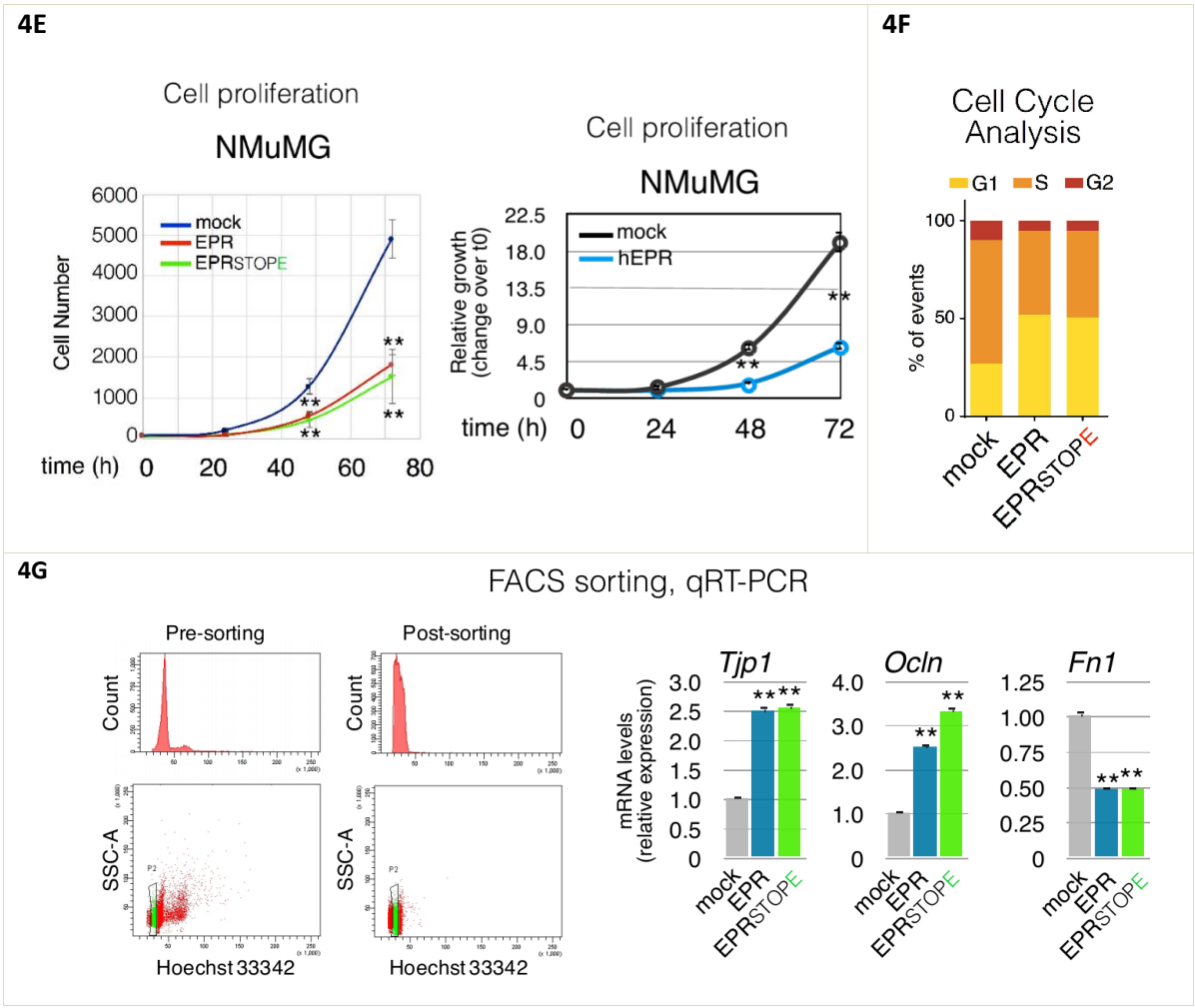


4A. qRT-PCR analysis of the indicated transcripts in either mock, EPR or EPRSTOPE overexpressing NMuMG cells.

4B. Immunoblot analysis of total cell extracts from mock, EPR-, EPRSTOPE- or EPRSTOPM-overexpressing NMuMG cells.

4C . qRT-PCR analysis of *Cdkn1a* mRNA levels in control siRNA (siC) or siEPR-transfected NMuMG cells (upper panel). Immunoblot analysis of total cell extracts from control siRNA (siC) or siEPR-transfected NMuMG cells (lower panel).

4D. qRT-PCR of Cdkn1a mRNA levels in either mock or human EPR (h.EPR)-overexpressing NMuMG cells (top panel). Immunoblot analysis of mock or human EPR (h.EPR) overexpressing NMuMG cells (bottom panel). Values of qRT-PCR are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance:*p<0.01,**p<0.001 (Student's t test).



4E. Left:Proliferation analysis (OperettaCLS High-Content Analysis System) of mock, EPR- or EPRSTOPE- overexpressing NMuMG cells. Right: cell proliferation analysis (analyzed by Crystal violet staining) of either mock or human EPR (h.EPR) -overexpressing NMuMG cells.

4F. Cell cycle distribution in cultures grown at similar density (near 90% confluence). Cell cycle was analyzed by flow cytometry after double staining with EdU and Propidium Iodide of mock, EPR- or EPRSTOPE- overexpressing NMuMG cells. Bars plot the relative proportion of cells in G1, S, and G2 phases for each cell line.

4G. G1 cell sorting of mock, EPR or EPRSTOPE-overexpressing NMuMG cells. Cells were stained by Hoechst 33342 and processed using a BD Aria II cytometer. The pre-sorting cell cycle profile (top left), gating (green area in the dot plot, bottom left), and purity checked by post-sorting sample re-run (center panels) are shown for one replicate of the mock sample; SSC-A is for side scatter area. qRT-PCR analysis of the indicated transcripts in either mock, EPR or EPRSTOPE-overexpressing cells sorted in the G1 phase of the cell cycle (right panels). Values of qRT-PCR and cell proliferation experiments shown are averages (±SEM) of three independent experiments performed in triplicate.. Statistical significance: **p < 0.001 (Student's t test). For Immunoblots, the indicated antibodies were used; the position of molecular mass markers is indicated on the left and representative gels are shown.

4.5 EPR regulates TGF-β-dependent *Cdkn1a* gene expression at different levels.

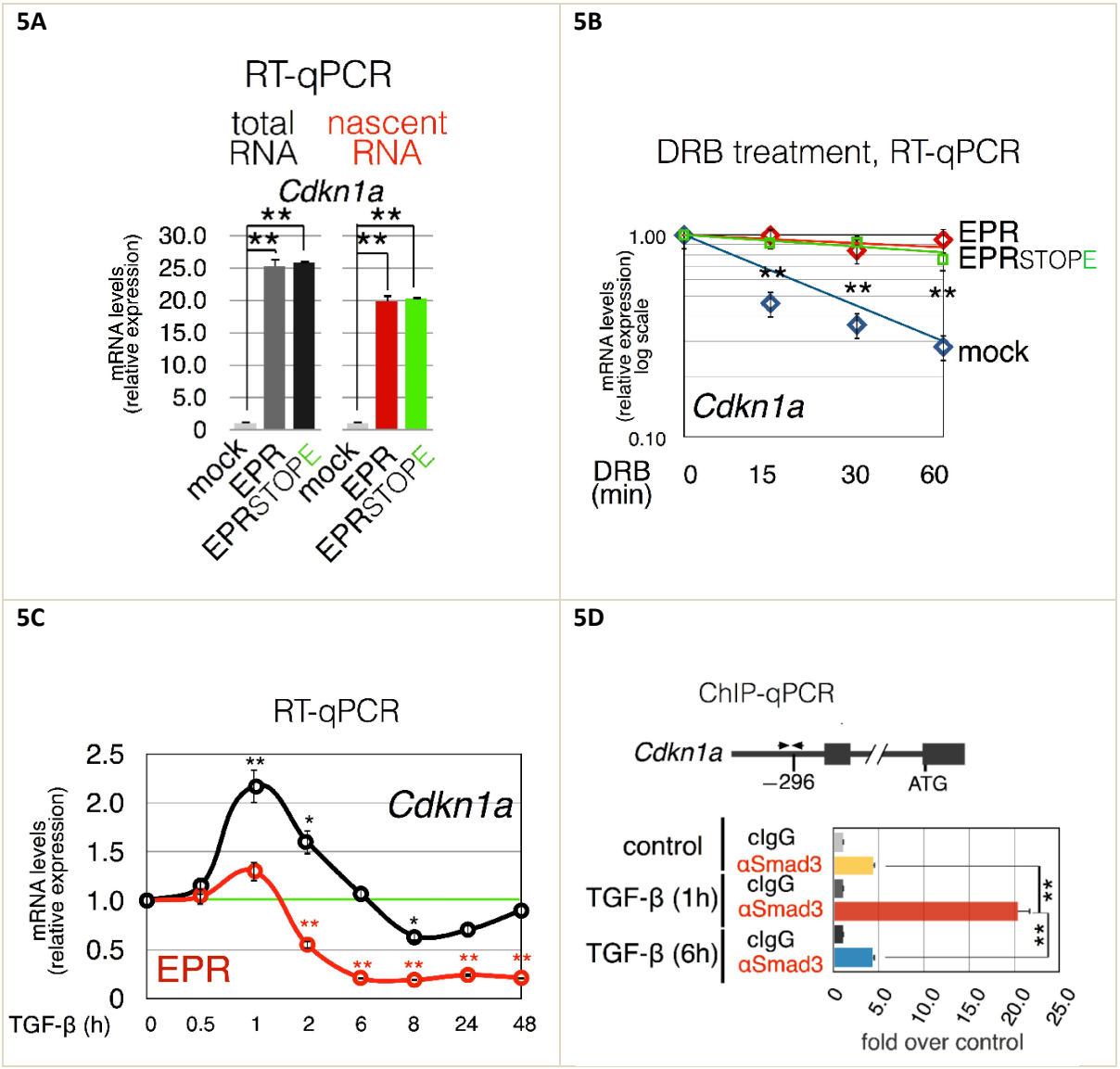
Analysis of newly synthesized transcripts revealed that EPR overexpression strongly enhances *Cdkn1a* transcription (Fig. 5A) and the kinetic analysis of mRNA decay indicated that EPR overexpression induces also a significant stabilization of *Cdkn1a* mRNA (Fig. 5B).

TGF-β signaling promotes tissue growth and morphogenesis during embryonic development while, as tissues mature, many cell types gain the ability to respond to TGF-β with growth arrest that is primarily due to imbalance of G1 events (Siegel & Massagué, 2003). As similarly reported in other cell types (Seoane et al., 2004; Pardali et al., 2005), treatment of NMuMG cells with TGF-β for 1 hour caused a rapid induction of *Cdkn1a* gene expression that was followed by return to baseline levels after 6 hours (Fig. 5C). The observation that *Cdkn1a* return to baseline levels matches EPR downregulation (Fig. 5C) and that EPR overexpression strongly enhances *Cdkn1a* levels, prompted us to hypothesize a role for EPR in the TGF-β-dependent modulation of *Cdkn1a* gene expression. Our ChIP-qPCR assays showed that TGF-β treatment for 1 hour stimulates the binding of SMAD3 to *Cdkn1a* promoter that returns to basal levels after 6 hours (Fig. 5D, see also Seoane et al., 2004). TGF-β-dependent control of *Cdkn1a* mRNA decay was never investigated in detail but, considering that cells often achieve rapid changes of gene expression by integrating gene transcription control with regulated mRNA decay (Komili & Silver, 2008; Braun & Young, 2014), we addressed the possibility that TGF-β could affect *Cdkn1a* mRNA decay. Fig. 5E showed that *Cdkn1a* mRNA stability is unaffected by 1 hour of TGF-β treatment (upper panel) but is reduced when the treatment is prolonged up to 6 hours (lower panel). Thus, the TGF-β-dependent rapid fluctuations of *Cdkn1a* expression depend on the regulation of both transcription and mRNA decay in NMuMG cells.

Our hypothesis that EPR plays a role in the regulation of TGF-β-dependent CDKN1A expression was supported by the evidence that EPR silencing abrogates *Cdkn1a* mRNA induction upon TGF-β treatment for 1 hour (Fig. 5F) while its overexpression enhances *Cdkn1a* levels and blunts its rapid modulation by TGF-β (Fig. 5G).

Together, our results indicate that EPR plays a dual role in TGF-β-dependent *Cdkn1a* gene expression control.

Figure 5. Dual regulation of *Cdkn1a* gene expression by EPR.



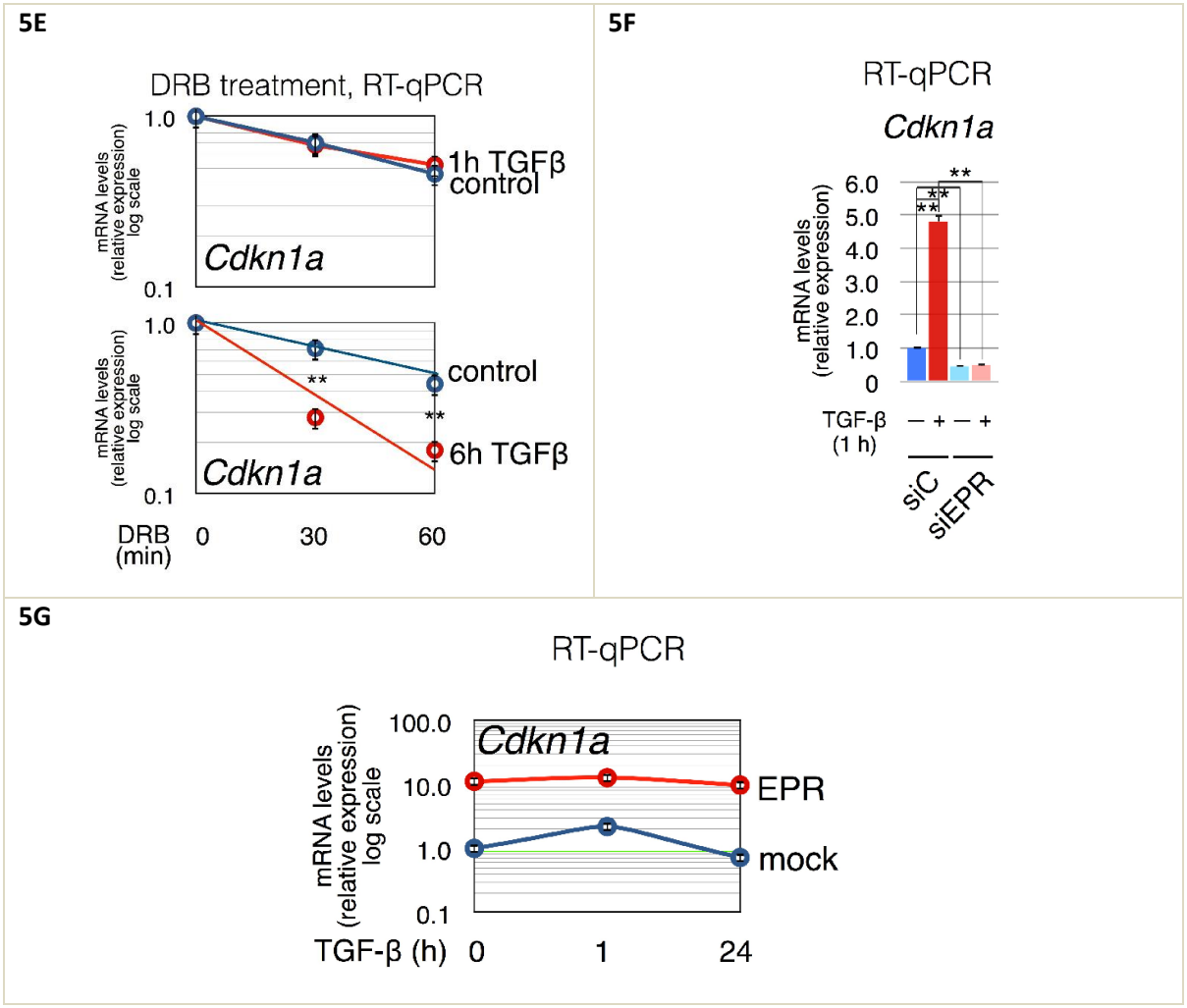
5A. qRT-PCR analysis of either total (left) or nascent (right) *Cdkn1a* transcript in either mock, EPR- or EPRSTOPE- overexpressing NMuMG cells.

5B. Either mock, EPR- or EPRSTOPE- overexpressing NMuMG cells were treated with 100 μM 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) for different times (as indicated). *Cdkn1a* gene expression was analyzed by qRT-PCR.

5C. qRT-PCR analysis of *Cdkn1a* (black line) and EPR (red line) in NMuMG cells serum-starved (2% FBS, 16h) and either treated with TGF-β (10 ng/ml) for the indicated times or untreated (time 0).

5D. Chromatin prepared from NMuMG cells serum-starved and either treated with TGF-β for the indicated times or untreated (control) was immunoprecipitated using either normal rabbit IgG (cIgG) or affinity-purified anti-SMAD3 rabbit polyclonal antibody. The association of SMAD3 with *Cdkn1a* promoter (schematic on the top) was quantitated by qPCR using specific primers (indicated as arrowheads).

The values of both qRT-PCR and qPCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).



5E. NMuMG cells were serum-starved (2% FBS, 16h) and either treated with TGF-β (10 ng/ml) for either 1 hour (top panel) or 6 hours (bottom panel) or left untreated (control in both panels). Subsequently, cells were treated with 100 μM DRB for the indicated times and total RNA was isolated and analyzed by qRT-PCR to quantitate *Cdkn1a* mRNA levels. Please note the slight, reproducible difference in the decay kinetic of *Cdkn1a* mRNA between non-transfected and mock-transfected NMuMG cells.

5F. qRT-PCR analysis of *Cdkn1a* in NMuMG cells transiently transfected with either control siRNA (siC) or siRNA designed to silence EPR-expression (siEPR), serum-starved (2% FBS, 16h) and then either treated with TGF-β (10 ng/ml) for 1 hour (+) or left untreated (-).

5G. qRT-PCR analysis of *Cdkn1a* expression in either mock (blue line) or EPR overexpressing (red line) NMuMG cells serum-starved (2% FBS, 16h) and either treated with TGF-β (10 ng/ml) for the indicated times or left untreated (time 0). Please note the logarithmic scale of the Y axis. The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).

4.6 EPR affects both SMAD3-dependent *Cdkn1a* gene transcription and KHSRP-regulated mRNA decay.

We investigated the molecular mechanism(s) by which EPR regulates *Cdkn1a* gene transcription. The evidence of enhanced RNA-Pol II occupancy and reduced presence of the H3K27me3 repressive mark at the *Cdkn1a* promoter in EPR-overexpressing cells (Fig. 6A) together with our finding that EPR is present in the chromatin fraction (see Fig. 1H), prompted us to explore the possibility that EPR affects *Cdkn1a* transcription through direct interaction with its promoter region. Chromatin Isolation by RNA Purification (ChIRP)-Seq experiments (Paola Briata., Gabriele Bucci., Ettore Zapparoli et al., unpublished observation) as well as ChIRP-qPCR experiments revealed the direct interaction of EPR with *Cdkn1a* promoter (Fig. 6B). The interaction of EPR with *Cdkn1a* promoter is not significantly affected by a 1 hour TGF-β treatment (Fig. 6C).

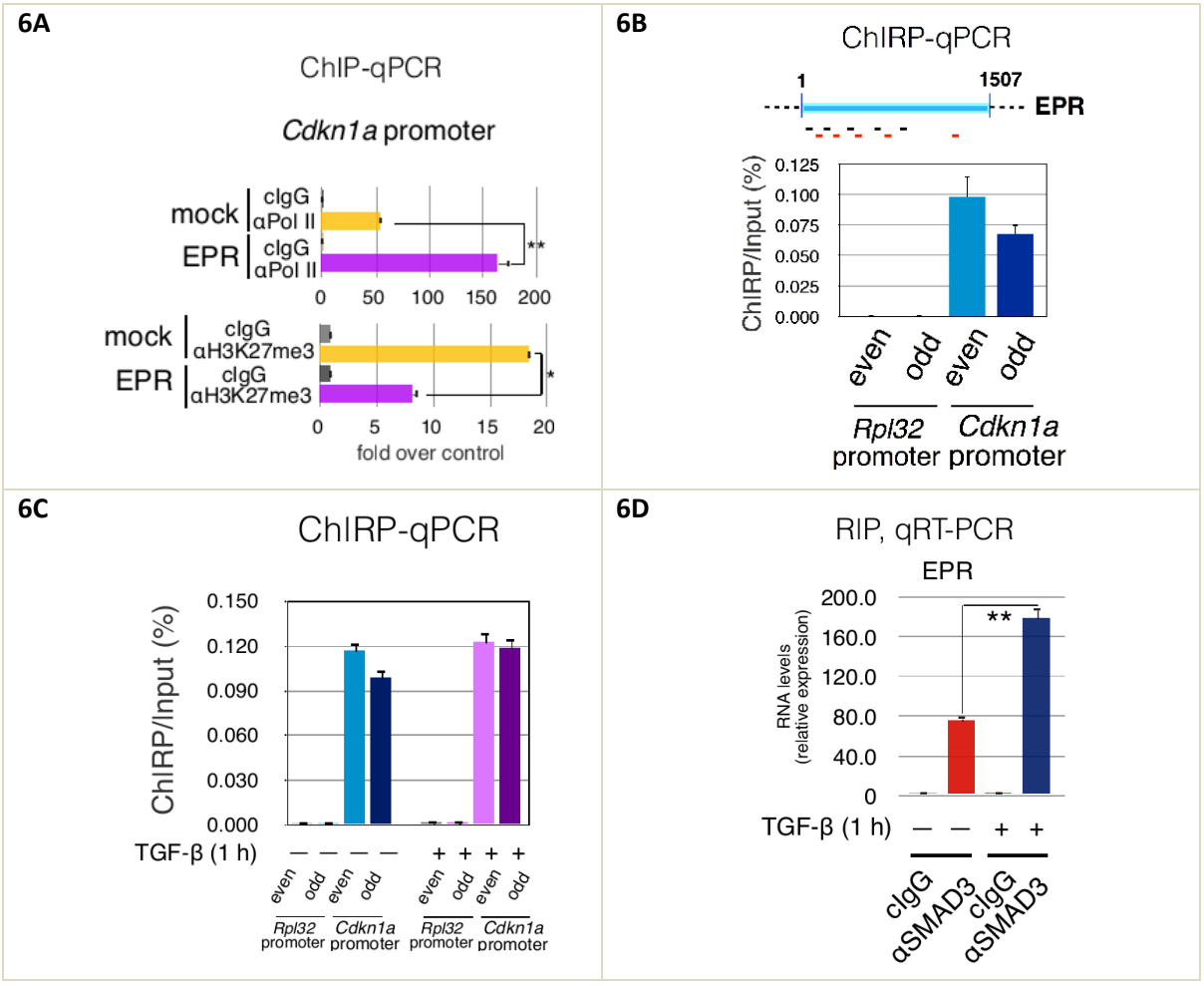
RIP experiments showed that SMAD3 interacts with EPR and the interaction is enhanced by treatment with TGF-β for 1 hour (6D). In keeping with growing evidence suggesting that the interaction between lncRNAs and specific transcription factors can affect gene expression (Long et al., 2017; Li et al., 2017), ChIP-qPCR experiments showed that EPR overexpression enhances SMAD3-*Cdkn1a* promoter association and abrogates its modulation by TGF-β thus preventing the dismissal of SMAD3 from the *Cdkn1a* promoter upon TGF-β treatment for 6 hours (Fig. 6E). These effects are reproduced by overexpression of EPRSTOPE (Fig. 6E). Notably, EPR overexpression favored SMAD3-*Cdkn1a* promoter interaction not only in TGF-β-treated NMuMG cells but also in untreated cells (Fig. 6E). Although SMAD proteins rapidly accumulates into nuclei upon TGF-β treatment (Reguly & Wrana, 2003) a certain amount of SMAD3 is present in the nuclei of untreated cells (Schmierer & Hill, 2005; Fig. 6F) thus suggesting that EPR overexpression

favors the recruitment of SMAD3 molecules already present in the nucleus of cells not treated with TGF-β.

In keeping with results shown in Fig. 5B, we found that EPR overexpression prevents *Cdkn1a* mRNA destabilization induced by a treatment with TGF-β for 6 hours (Fig. 6G). Our original observation that EPR interacts with KHSRP, a factor able to promote rapid decay of select labile mRNAs in many cell types (Briata et al., 2013), prompted us to explore whether KHSRP regulates *Cdkn1a* mRNA decay in NMuMG cells. Indeed, KHSRP silencing induced *Cdkn1a* mRNA accumulation and prevented its rapid degradation (Fig. 6H, 6I) while transient KHSRP overexpression in NMuMG cells stably expressing EPR promoted *Cdkn1a* mRNA destabilization (Fig. 6H lower panel). KHSRP is predominantly nuclear in NMuMG cells (Puppo et al., 2016) and we found that mature *Cdkn1a* mRNA is abundant in nuclear fractions of these cells where it undergoes rapid decay and is stabilized by EPR overexpression as measured by two independent techniques (6J). These findings suggested that EPR might interfere with the ability of KHSRP to interact with *Cdkn1a* mRNA. RIP experiments presented in Fig.6K show that KHSRP association with *Cdkn1a* mRNA was strongly reduced by EPR and EPRSTOPE overexpression. Based on these results, we investigated whether EPR down-regulation, that is physiologically obtained by TGF-β-treatment, affects the interaction of KHSRP with *Cdkn1a* mRNA. RIP analyses indicated that upon TGF-β treatment (6 hours) KHSRP-EPR is abrogated while KHSRP interaction with *Cdkn1a* mRNA is increased (Fig. 6L).

Altogether, these results suggest that EPR controls *Cdkn1a* expression by sustaining its transcription and impairing its mRNA decay in response to TGF-β.

Figure 6. EPR associates with *Cdkn1a* promoter and affects both SMAD3-dependent *Cdkn1a* gene transcription and KHSRP-regulated *Cdkn1a* mRNA decay.

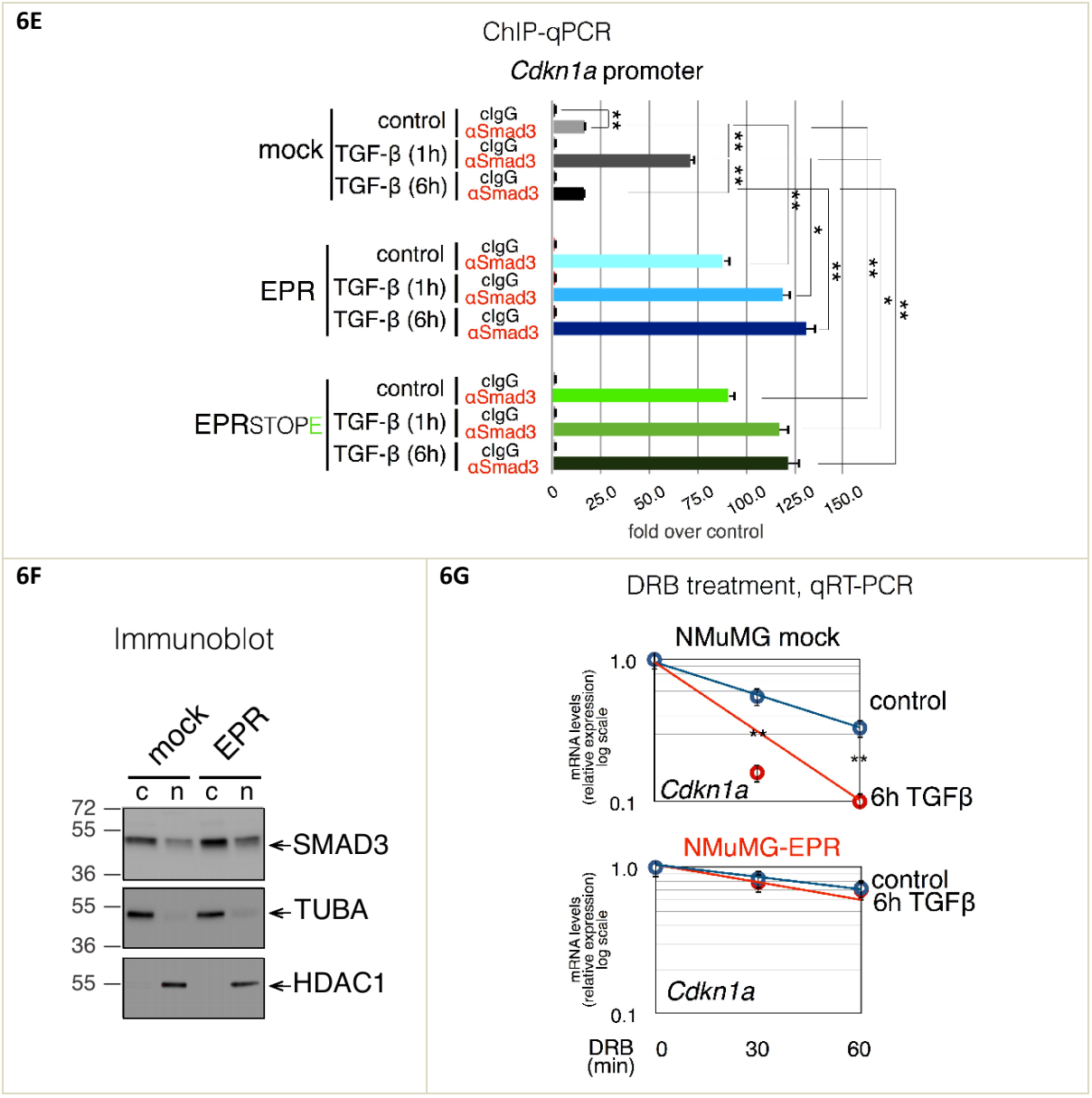


6A. Top panel:chromatin prepared from either mock or EPR-overexpressing NMuMG cells was immunoprecipitated using either normal mouse IgG (cIgG) or mouse monoclonal anti- PolII antibody. The association of PolII with Cdkn1a promoter was verified by qPCR using specific primers. Bottom panel, chromatin prepared from either mock or EPR-overexpressing NMuMG cells was immunoprecipitated using either normal rabbit IgG (cIgG) or affinity purified rabbit anti-H3K27me3 antibody. The association of H3K27me3 with Cdkn1a promoter was verified by qPCR using specific primers.

6B. ChIRP analyses performed using NMuMG cell lysates and either even or odd EPR probe sets. Top panel is a schematic of EPR and shows the location of biotinylated odd (black lines) and even (red lines) tiling oligonucleotides used for ChIRP. Both input and purified DNA were analyzed by qPCR using primers designed on *Rpl32* (negative control) or *Cdkn1a* promoters. Values are averages (±SEM) of three independent experiments performed in triplicate.

6C. NMuMG cells were serum-starved (2% FBS, 16h) and treated with TGF-β(10 ng/ml) for 1 hour (+) or left untreated (-). ChIRP analyses was performed using cell lysates and either even or odd EPR probe sets. Both input and purified DNA were analyzed by qPCR using primers designed on Rpl32 (neg. control) or Cdkn1a promoters.Values are averages (±SEM) of three independent experiments performed in triplicate.

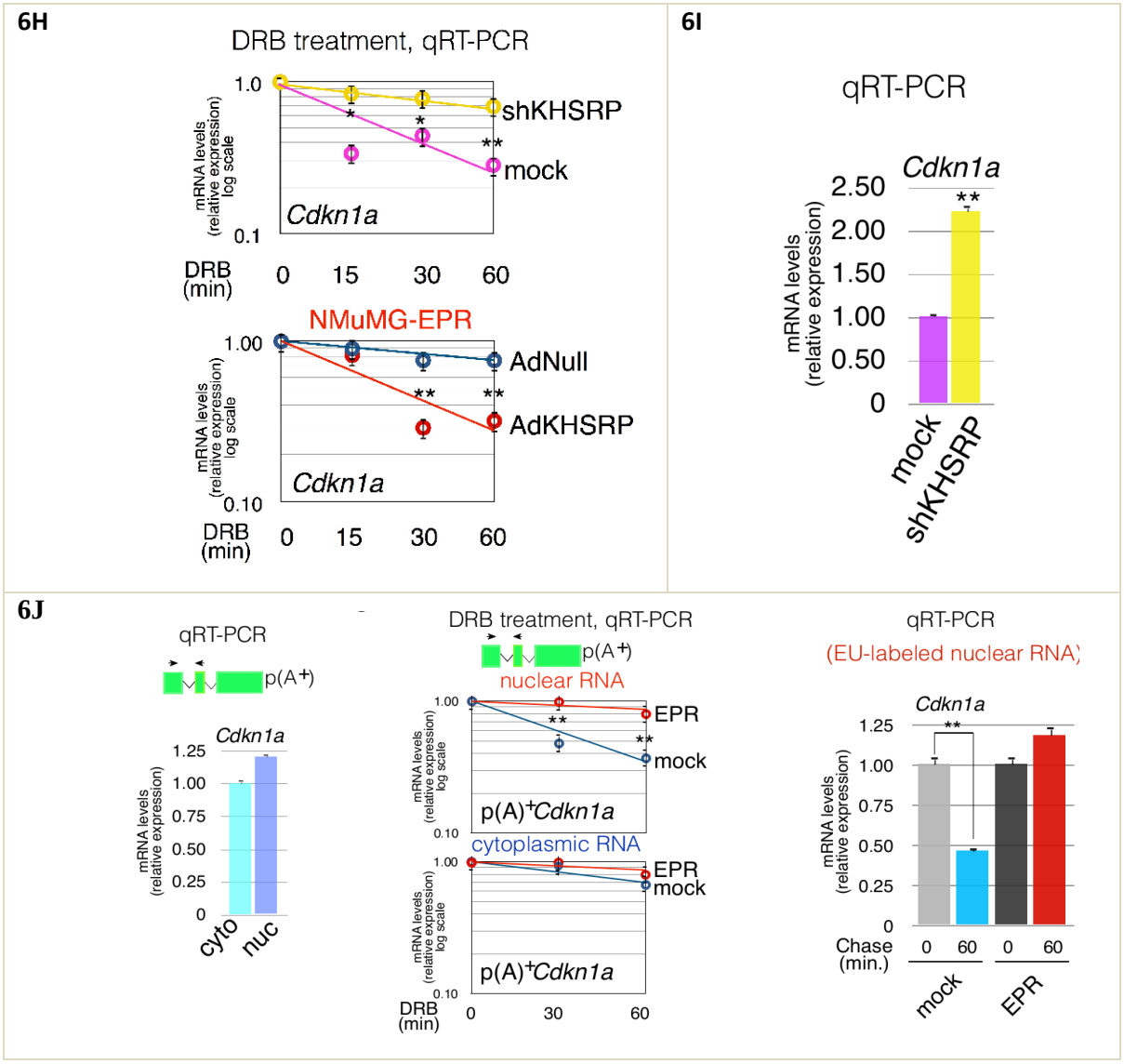
6D. NMuMG cells were serum-starved (2% FBS, 16h) and treated with TGF-β (10 ng/ml) for 1 hour (+) or left untreated (-). Total extracts were prepared and immunoprecipitated as indicated. RNA was purified from immunocomplexes and analyzed by qRT-PCR to quantitate EPR levels.



6E. Chromatin was prepared from either mock, EPR- or EPRSTOPE- overexpressing NMuMG cells serum-starved (2% FBS, 16h) and either treated with TGF-β (10 ng/ml) for the indicated times or left untreated (control). Chromatin was immunoprecipitated using either normal rabbit IgG (cIgG) or affinity-purified anti-SMAD3 rabbit polyclonal antibody. The association of SMAD3 with *Cdkn1a* promoter was verified by qPCR using specific primers.

6F. Immunoblot analysis of either cytoplasmic (c) or nuclear (n) cell extracts from either mock or EPR overexpressing NMuMG cells. The indicated antibodies were used; the position of molecular mass markers is indicated on the left and representative gels are shown.

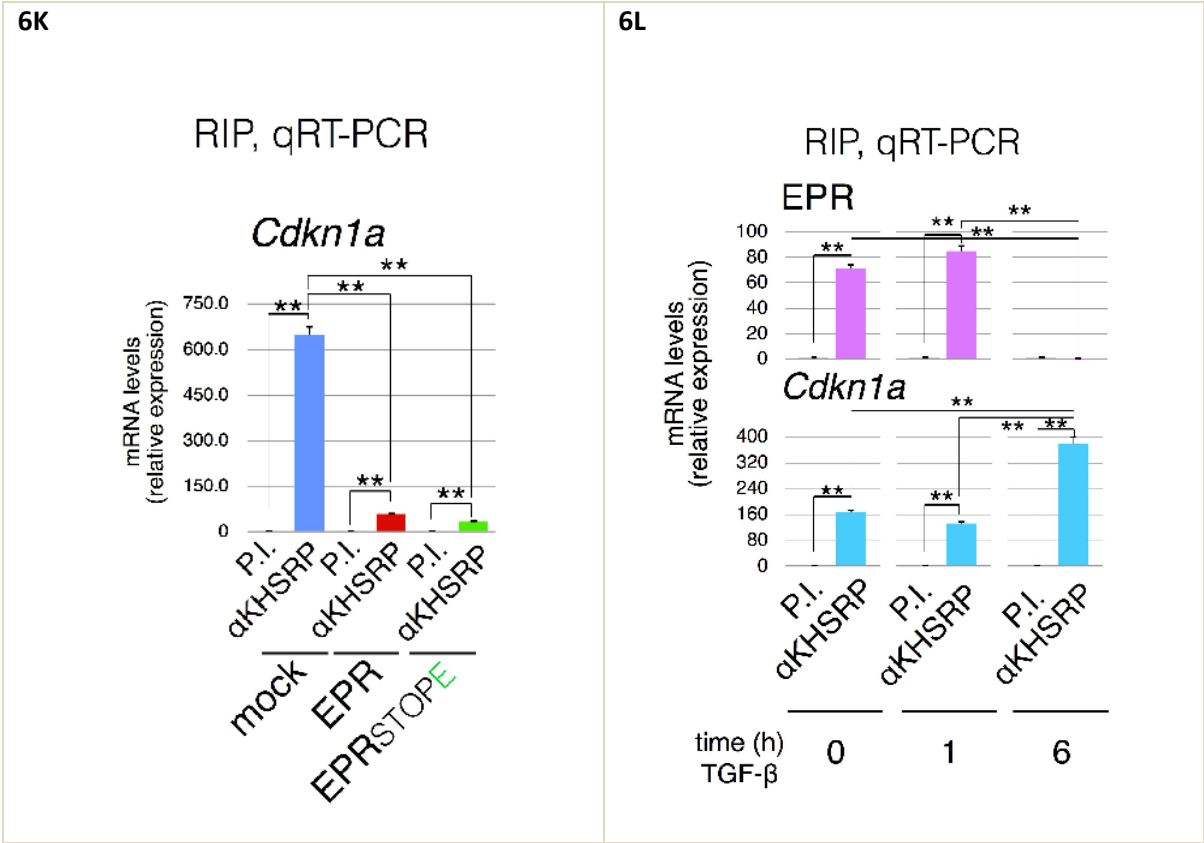
6G. Either mock or EPR-overexpressing NMuMG cells were serum-starved (2% FBS, 16h) and either treated with TGF-β (10 ng/ml) for 6 hours or left untreated (control). Subsequently, cells were treated with 100 μM DRB for the indicated times and total RNA was isolated and analyzed by qRT-PCR to quantitate *Cdkn1a* mRNA levels. Values of qRT-PCR experiments are averages (±SEM) of three independent experiments in triplicate. Statistical significance: *p< 0.01, **p 137 < 0.001 (Student's t test).



6H. Top panel, mock or shKHSRP NMuMG cells were treated with 100 µM DRB. Total RNA was isolated at different times and analyzed by qRT-PCR to quantitate *Cdkn1a* mRNA levels. Bottom panel, NMuMG cells stably overexpressing EPR were infected with either control (AdNull) or KHSRP-expressing (AdKHSRP) adenoviral vectors for 24 hours and then treated with 100 µM DRB. Total RNA was isolated at different times (as indicated) and analyzed by qRT-PCR to quantitate *Cdkn1a* mRNA levels.

6I. qRT-PCR analysis of total RNA isolated from either mock or shKHSRP NMuMG cells. Please note that NMuMG mock cells used for these experiments differ from those presented throughout this report and have been described in Puppo et al. (2016).

6J. Left: qRT-PCR analysis of polyadenylated and spliced Cdkn1a mRNA isolated from either cytoplasmic (cyto) or nuclear (nuc) extracts of NMuMG cells. Center: mock or EPR overexpressing cells were treated with 100µM DRB. Nuclear and cytoplasmic RNAs were isolated at different times an analyzed by qRT-PCR to quantitate polyadenylated and spliced Cdkn1a mRNA levels.. Right: Run-off analysis of EU-labeled nuclear RNA prepared fromeither mock-transfected (mock) or EPR- overexpressing (EPR) NMuMG cells (chase time 0 and 60 minutes). Values of qRT-PCR and qPCR experiments are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance:*p< 0.01,**p 137 < 0.001 (Student's t test).



6K. Total extracts from either mock, EPR- or EPRSTOPE- overexpressing NMuMG cells were immunoprecipitated as indicated. RNA was purified from immunocomplexes and analyzed by qRT-PCR to quantitate *Cdkn1a* mRNA levels.

6L. Total extracts were prepared from NMuMG cells serum-starved (2% FBS, 16h) and either treated with TGF-β (10 ng/ml) for the indicated times or left untreated (time 0) and immunoprecipitated as indicated. RNA was purified from immunocomplexes and analyzed by qRT-PCR to quantitate EPR and *Cdkn1a* mRNA levels. The values of both qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).

4.7 EPR overexpression reduces breast cancer cell proliferation.

We investigated EPR in murine breast cancer cell lines and observed that its expression is severely reduced when compared with immortalized NMuMG cells (Fig. 7A). Similarly, the expression of h.EPR was below detection levels in highly aggressive human breast cancer cell lines (Matteo J. Marzi and Francesco Nicassio, unpublished observation). H.EPR could be detected in about 75% of breast cancer primary samples (780/1043 cases from The Cancer Genome Atlas (TCGA) database (Cancer Genome Atlas Network, 2012) (7B left) and, according to PAM50 molecular subtype classification, it was more expressed in Luminal A and Her2 tumors while it was almost absent in Basal-like tumors, the most frequent subtype of triple negative breast cancers (Bianchini et al., 2016) (Fig. 7B right).

Based on these observations, we decided to express either EPR or EPRSTOPE in triple-negative mesenchymal-like breast cancer cells, such as murine 4T1 and human MDA-MB-231 cell lines, respectively.

EPR and EPRSTOPE overexpression in 4T1 cells (Fig. 7C) resulted in a strong induction of *Cdkn1a* gene expression (Fig. 7D, 7E left panel) as well as in a significant reduction of clonogenic potential, cell proliferation, and anchorage-independent cell growth (Fig 7E right panel, 7F). EPR overexpression in 4T1 cells also down-regulated mesenchymal factors such as *Cdh2* and *Adam12* (Fig. 7G). Very similar results were observed by overexpressing either human or murine EPR in human MDA-MB-231 cells (Fig. 7H, 7I).

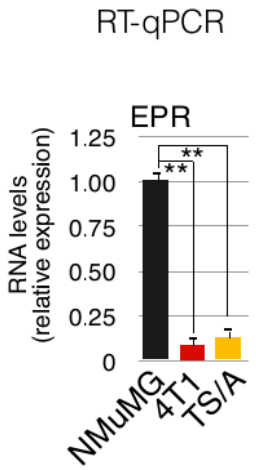
Finally, to interrogate the activity of EPR on cell proliferation control *in vivo*, we orthotopically injected either mock or EPR-expressing 4T1 cells into syngenic BALB/c mice. In concordance with our observations in cultured cells, EPR expression resulted in a remarkable reduction of tumor volume after 10 days (Fig. 7J, left panel). A significant

reduction of the tumor mass was still evident and statistically significant also at two weeks after the transplant when mice were sacrificed (Fig. 7J right panel; Fig. 7K).

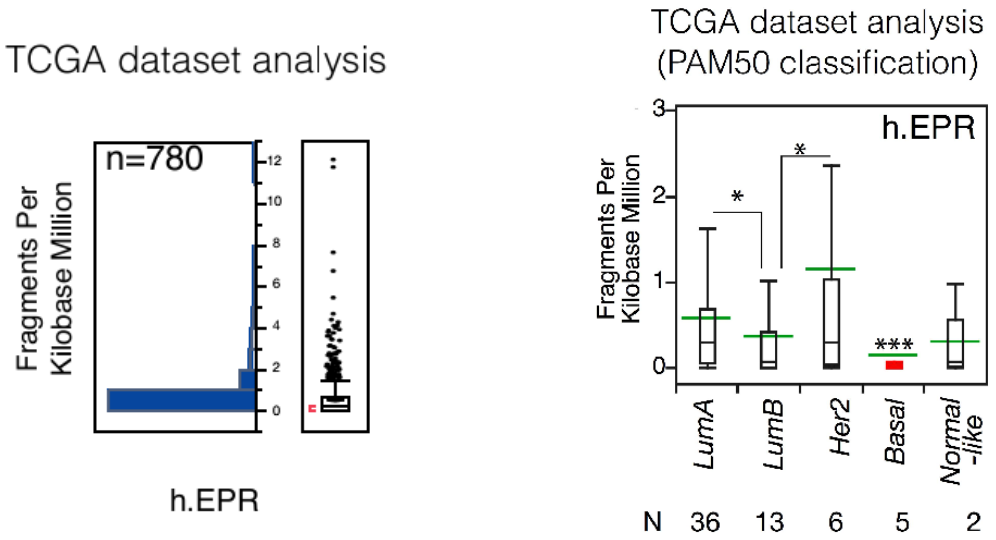
Altogether, our results indicate that EPR overexpression modulates cell proliferation and epithelial/mesenchymal markers levels in breast cancer cells and restrains cell proliferation in transplanted mice.

Figure 7. Anti-proliferative effect of EPR when expressed in transformed mammary gland cells.

7A

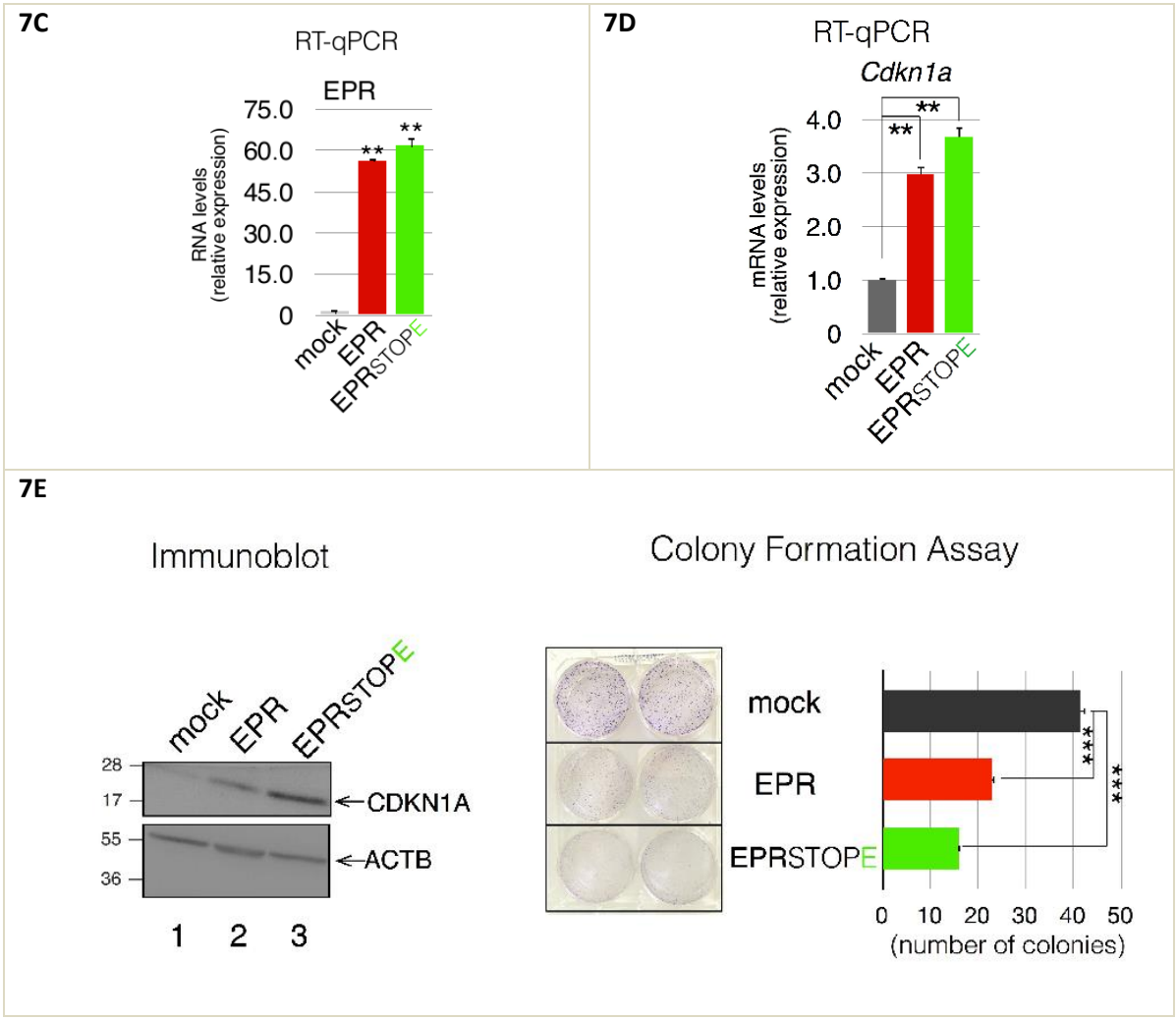


7B



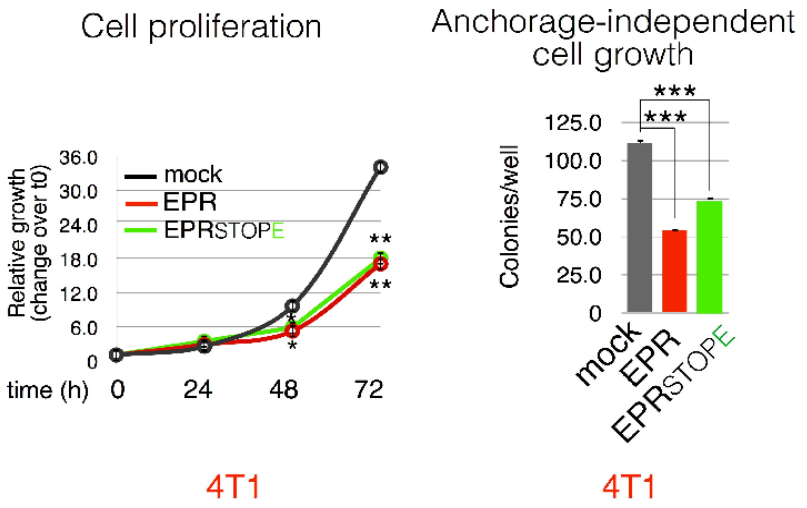
7A. The expression of EPR was quantitated by qRT-PCR analysis in the indicated cell types. values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: **p < 0.001 (Student's t test).

7B. Left Panel: distribution of h.EPR in the TCGA Breast Cancer (BRCA) dataset. Right panel:Box-plot shows the expression of h.EPR in the TCGA Breast Cancer (BRCA) dataset annotated according to PAM50 molecular subtype classification. The number of samples in each subtype is also presented. Asterisks mark significant values (Wilcoxon's test; * p<0.05, *** p<0.001).

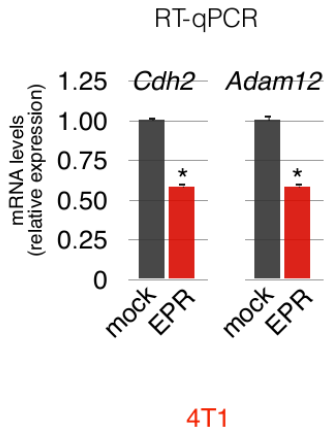


7C. qRT-PCR analysis of EPR levels in mock, EPR or EPRSTOPE-overexpressing 4T1 cells.
7D. qRT-PCR analysis of *Cdkn1a* mRNA levels in either mock, EPR- or EPRSTOPE- overexpressing 4T1 cells (left panel).
7E. Immunoblot analysis of either mock, EPR- or EPRSTOPE- overexpressing 4T1 cells (left panel).
Right panel: Either mock, EPR- or EPRSTOPE- overexpressing 4T1 cells were seeded at low density in 6-well plates, and colony-formation assays were performed after 4 days. A representative (of four independent) plate is shown. The values of colony number quantitation are averages (±SEM) of four independent experiments performed in duplicate. Statistical significance: ***$p < 0.00001$ (Student's t test).
The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: **$p < 0.001$ (Student's t test).

7F

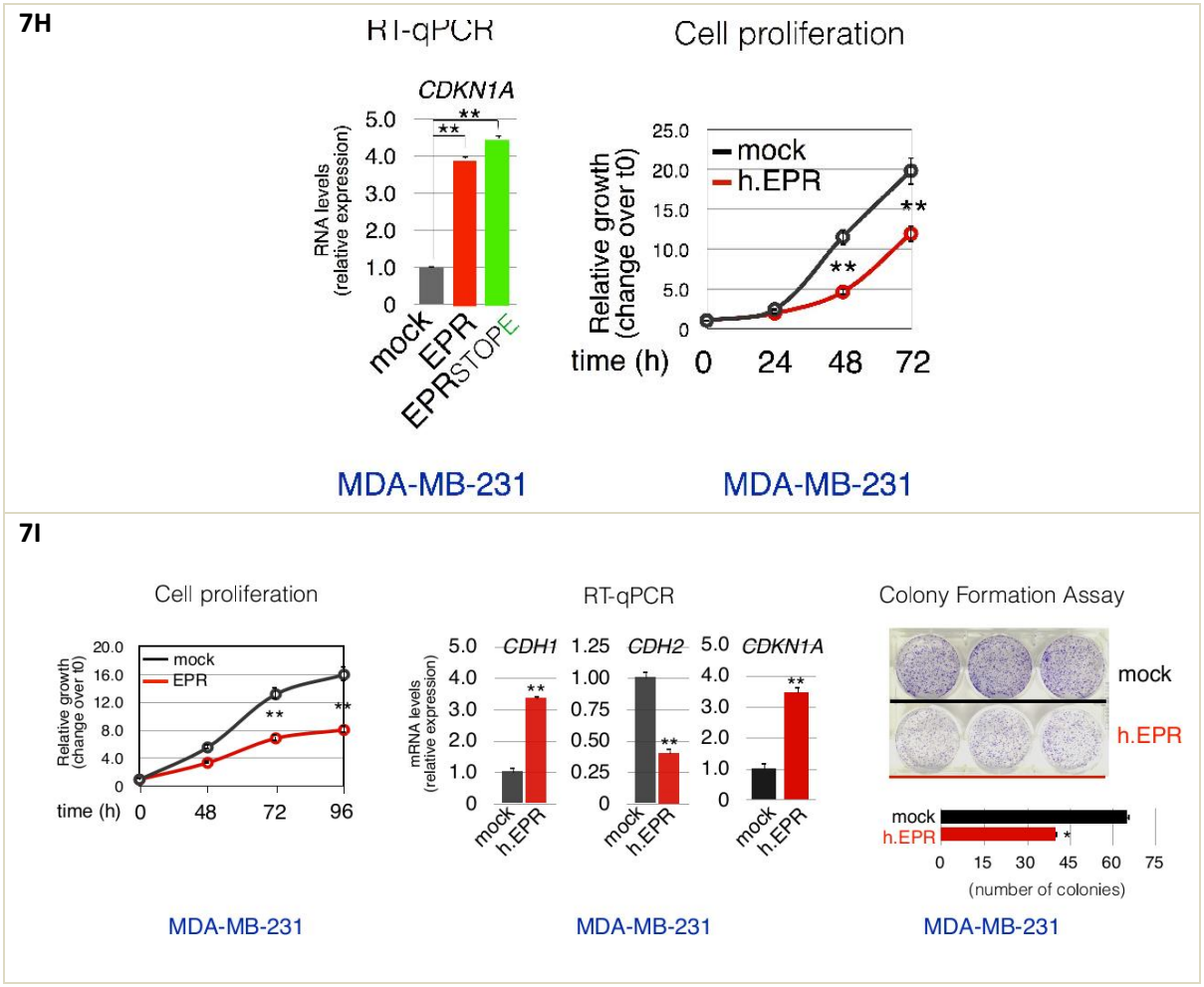


7G



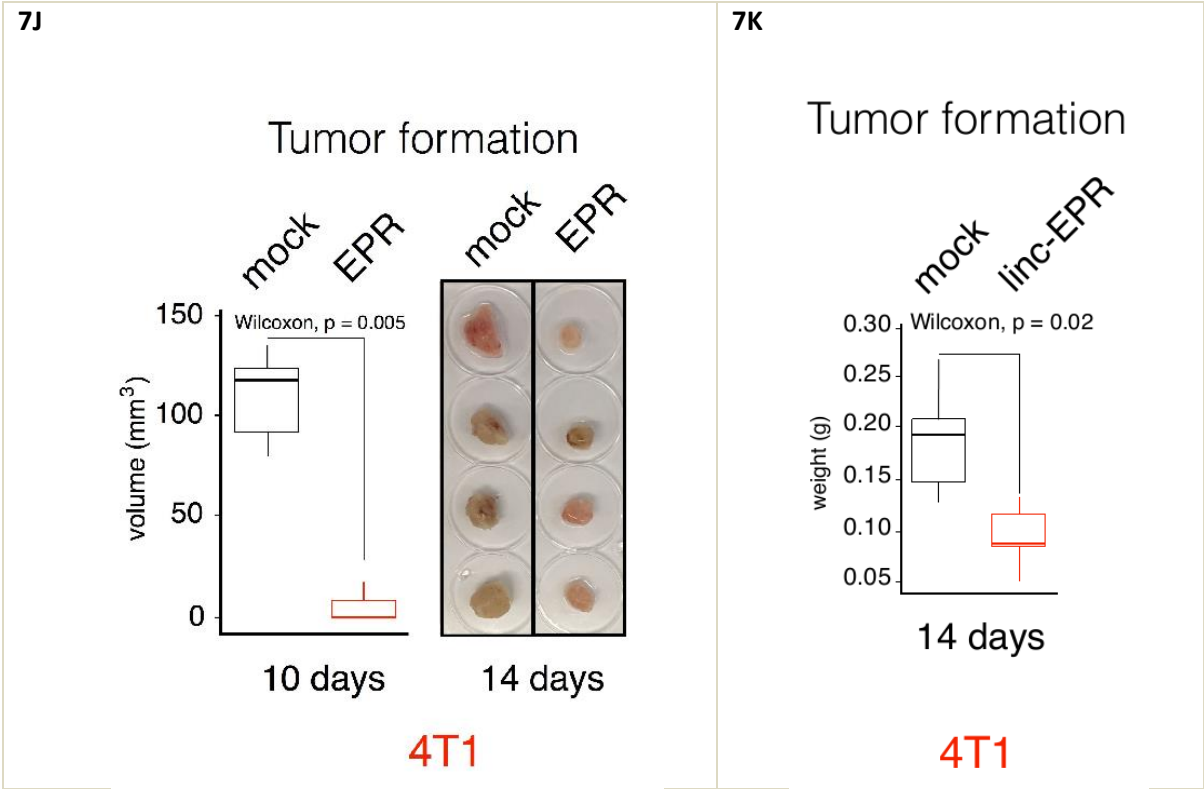
7F. Left: Cell proliferation analysis (quantitated using Crystal violet staining) of either mock, EPR- or EPRSTOPE- overexpressing 4T1 cells. Right: Either mock, EPR- or EPRSTOPE- overexpressing 4T1 cells were cultured in soft agar for 21 days and phase contrast micrographs were taken at 10 X magnification. The values are averages (±SEM) of four independent experiments performed in triplicate. Statistical significance: ***$p < 0.00001$ (Student's t test).

7G. qRT-PCR analysis of the indicated transcripts in either mock-transfected (mock) or EPR-overexpressing (EPR) 4T1 cells. The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: **$p < 0.001$ (Student's t test).



7H. Left panel: qRT-PCR analysis of *CDKN1A* mRNA levels in either mock, EPR- or EPRSTOPE-overexpressing MDA-MB-231 cells. Right panel: Cell proliferation analysis (quantitated using Crystal violet staining) of either mock or h.EPR-overexpressing MDA-MB-231 cells.

7I. Left: Cell proliferation analysis (using Crystal violet staining) of either mock or EPR-overexpressing MDA-MB-231 cells. Center: qRT-PCR analysis of the indicated transcripts in either mock-transfected (mock) or human EPR-overexpressing (h.EPR) MDA-MB-231 cells. Right: Either mock or h.EPR-overexpressing MDA-MB-231 cells were seeded at low density in 6-well plates and colony-formation assays were performed after 6 days. A representative (of three independent) plate is shown. The values of qRT-PCR experiments shown are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: **p < 0.001 (Student's t test). The values of cell proliferation experiment are averages (±SEM) of three independent experiments performed in triplicate. Statistical significance: *p< 0.01, **p < 0.001 (Student's t test).



7J. Tumor volume (n = 6 tumors/group) was measured by digital caliper assessment ten days after injection of either mock or EPR-expressing 4T1 cells in BALB/c mice. Box plot analysis of tumor volume is shown (left panel). Data were analyzed in R version 3.4.3, using the statistic method Wilcoxon Unpaired Test as implemented in `stat_compare_means` in 'ggplot 2.2.1'. Images of the tumors at the end of the experiment, two weeks after injection (right panel).

7K. Tumors (n = 6 tumors/group) were removed and weighted at the end of the experiment (two weeks). Box plot analysis of tumor weight is shown. Data were analyzed in R version 3.4.3, using the statistic method Wilcoxon Unpaired Test as implemented in `stat_compare_means` in 'ggplot 2.2.1'.

5. DISCUSSION

In this thesis I report on the initial functional characterization of the long intergenic non-coding RNA EPR well conserved among mammalian species and expressed in select epithelial tissues including differentiated luminal cells of human breast. The levels of EPR are rapidly downregulated by TGF-β/SMAD signaling in immortalized mammary gland cells and its sustained expression largely reshapes the transcriptome inducing epithelial traits while preventing the acquisition of mesenchymal markers upon TGF-β treatment. Remarkably, EPR overexpression enhances the levels of the cyclin-dependent kinase inhibitor CDKN1A and strongly reduces cell proliferation in both immortalized and transformed mammary gland cells as well as in transplanted mice.

EPR is almost equally distributed in chromatin, nucleoplasm and cytoplasm of epithelial cells and the cytoplasmic component associates with polysomes where a small peptide (EPRp) is translated. EPRp interaction with a variety of cytoskeletal and junctional proteins accounts for its junctional localization. However, the analysis of the phenotype that we observed in cells overexpressing EPR mutants unable to originate the peptide clearly indicates that the vast majority of gene expression changes that we describe here are independent of the peptide biogenesis.

In this project, I investigated how the lncRNA molecule *per se* controls gene expression and I focused our studies on the EPR-dependent regulation of CDKN1A that functions as both a sensor and an effector of multiple anti-proliferative signals and promotes cell cycle arrest in response to TGF-β (Abbas & Dutta, 2009). In NMuMG cells, TGF-β induces an early wave of *Cdkn1a* expression due, in part, to an increased SMAD complex-dependent gene transcription while a prolonged treatment causes the return of *Cdkn1a* levels to the baseline. Our data suggest that *Cdkn1a* promoter-bound EPR recruits

SMAD3 —that accumulate into the nucleus upon TGF-β treatment for 1 hour— to induce rapid gene transcription. In parallel, EPR interacts with KHSRP limiting its association with *Cdkn1a* mRNA and this results in its stabilization. I propose that EPR down-regulation upon 6 hours of TGF-β treatment causes SMAD3 dismissal from *Cdkn1a* promoter that results in a return of *Cdkn1a* transcription to basal levels and, in parallel, enables KHSRP to destabilize the *Cdkn1a* transcript. Thus, EPR-regulated molecular events shape the rapid wave of *Cdkn1a* expression in response to TGF-β. The evidence that in cells overexpressing EPR CDKN1A is abundant even in the absence of TGF-β treatment can be explained by EPR ability to recruit SMAD3 molecules already present in cell nuclei to *Cdkn1a* promoter region as well as to prevent KHSRP-induced *Cdkn1* mRNA degradation.

Data presented in this thesis indicate that EPR couples *Cdkn1a* transcriptional regulation with mRNA decay control. Indeed, the integration of transcription and mRNA decay provides a kinetic boost to a series of processes that would be otherwise slower and less efficient. Our data strengthen the idea that coupling transcription to mRNA decay enables cells to rapidly modulate waves of gene expression in response to a variety of stimuli thus achieving optimal mRNA homeostasis (Komili & Silver, 2008; Braun & Young, 2014).

Since the decay of mature mRNAs is generally thought to occur in the cytoplasm, our finding that spliced and polyadenylated *Cdkn1a* mRNA is abundant in the nucleus of NMuMG cells, where it undergoes regulated decay, might be somehow surprising. However, the possibility that mature mRNAs accumulate in the nuclei of mammalian cells has been described (Gondran et al., 1999; Bahar *et al*., 2015) and a previous report has suggested that *Cdkn1a* mRNA undergoes degradation by the nuclear exosome in mammalian cells (Singer et al., 2012). Further, a report about mature RNA degradation

controlled by nuclear histone acetyltransferases and deacetylases provocatively suggested the possibility that mRNA degradation pathways might operate also in the nucleus (Sharma et al., 2016).

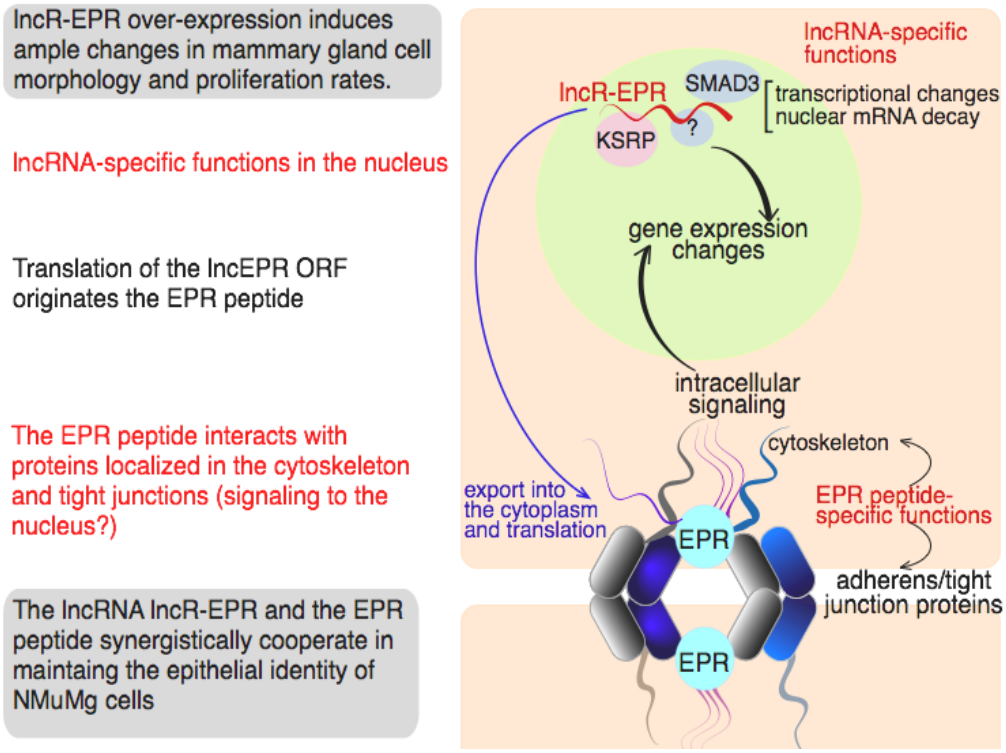
The TGF-β cytostatic program in epithelial cells involves, among other molecular events, the induction of *Cdkn1a* and *Cdkn2b* but cancer cells utilize any opportunity to circumvent TGF-β ability to inhibit cell proliferation. Inactivating mutations in the TGF-β Type-II receptors and SMAD4 have been described in tumors even though cancer cells can lose the cytostatic responsiveness due to defects downstream to SMAD factors. Under these conditions, transformed cells may exploit other functional properties of TGF-β turning this pathway into a proliferative signal. TGF-β phenotypic switch from a tumor suppressor to a tumor inducer is known as the "TGF-β paradox" and represents an unanswered question concerning its physio-pathological role (Principe et al., 2014). I hypothesize that the absence of EPR/h.EPR, that occurs in breast cancer cell lines and in certain breast cancers, may contribute to the loss of TGF-β ability to restrain cell proliferation while may enable the cytokine to sustain their carcinogenic potential.

The notion that lncRNAs are devoid of coding potential has been recently challenged by reports demonstrating that translated short peptides are responsible for the biological functions of the respective lncRNA (Matsumoto et al. 2017; Huang et al. 2017; Anderson et al., 2015; Nelson et al. 2016). However, Yu et al. (2017) reported that linc-RAM, a lncRNA that *per se* enhances myogenic differentiation by interacting with MYOD, is the transcript encoding myoregulin, a small peptide previously reported as a mediator of muscle performance through inhibition of the pump activity of SERCA (Anderson et al., 2015). In this case, RNA and peptide functions cooperate in modulating muscle physiology. Intriguingly, in our model the lncRNA *per se* is responsible for most of the gene expression changes while the peptide might be responsible for specific functions

related to its cytoskeletal/junctional localization. I hypothesize that the peptide could participate in multiprotein complexes serving as permeability barriers and/or implicated in the establishment of apico-basal polarity as well as in the transmission of signals to the cell interior. Further studies will be needed to clarify the specific functions of the peptide but we can predict that EPR and the peptide might synergize in executing epithelial cell specific programs (Figure 8).

EPR was discovered as a highly-regulated lncRNA in NMuMG mammary gland cells and, by exploiting the TGF-β-responsiveness of these cells, we were able to clarify some details of its molecular functions. How EPR influences normal physiology and disease in tissues undergoing repeated rounds of proliferation/differentiation, such as the gastro-intestinal tract where it is highly expressed, will represent an important area of future research.

Fig. 8. Distinct as well as complementary functions of EPR as noncoding RNA and EPR peptide



6. METHODS

6.1 Cell lines

Murine immortalized NMuMG cells (ATCC, no. CRL-1636) were cultured in DMEM plus 10% FBS and 10 μg/ml bovine insulin (Sigma-Aldrich), 4T1 mouse mammary gland cancer cells (obtained from ATCC, no. CRL-2539) and TS/A mouse mammary gland cancer cells (obtained from Dr. C. Bagni, VIB Center for the Biology of Disease) were cultured in DMEM/F12 plus 10% FBS, human mammary gland adenocarcinoma cells MDA-MB-231 (obtained from DSMZ, Germany, through Dr. G. Fronza, authenticated by STR DNA profiling) were cultured in DMEM plus 5% FBS, and human HEK-293 cells (ATCC, no. CRL-1573) were cultured in DMEM plus 10% FBS. NMuMG cells were maintained in DMEM supplemented with 2% for 16 hours prior to the addition of 10 ng/ml human recombinant TGF-β1 purchased from R&D Systems.

6.2 Antibodies

Anti-EPR polyclonal rabbit antibody was generated by injecting rabbits with recombinant purified EPR expressed in *E. Coli* using the pQE-EPR at Cambridge Research Biochemicals (Billingham, Cleveland, UK). Anti-CDH1 goat polyclonal antibody (sc-31020), anti-CDKN1A mouse monoclonal antibody (sc-6246), and anti-HDAC1 rabbit polyclonal antibody (sc-7872) were from Santa Cruz; anti-TJP1 rabbit polyclonal antibody (ab96587), anti-SMAD3 rabbit polyclonal antibody (ChIP grade ab28379), anti-GFP rabbit polyclonal antibody (ChIP grade ab90) were from Abcam; mouse monoclonal anti-FLAG (F1804), mouse monoclonal anti-TUBA (DM1) and mouse monoclonal anti-ACTB (AC-74) were from Sigma Aldrich. Mouse monoclonal anti-RNA Polymerase II (clone CTD4H8) and rabbit polyclonal antibody to H3K27me3 (CS200603) were from Millipore.

Rabbit polyclonal anti-CGN serum (C532) against a purified recombinant 50 kDa C-terminal fragment of chicken cingulin as well as anti-CGNL1 rabbit polyclonal antibody (20893) were raised at the University of Geneve.

6.3 Plasmids

Plasmid EPR was obtained by inserting the sequence from nucleotide 1 to 1487 of murine BC030870 into pBICEP-CMV-2 vector (Sigma-Aldrich); plasmid h.EPR was obtained by inserting the sequence from nucleotide 4 to 1126 of human LINC01207 into pBICEP-CMV-2 vector; plasmids EPRMUTE and EPRMUTM were obtained by Site-Directed Mutagenesis of plasmid EPR using the QuikChange II mutagenesis kit (Agilent Technologies) and the oligonucleotides:

5′-CACCGTTAGTCTTCCATGTAGCTACCATTC-3′

5′-CACCGTTAGTCTTCCTAGTAGCTACCATTC-3′, respectively.

Plasmids EPR-FLAG and EPRSTOPE-FLAG were generated by inserting the sequence from nucleotide 1 to 560 of murine BC030870 obtained by PCR and Flagged at its 3’ (either wild-type or mutagenized as above) into pIRES1*hyg* vector. Plasmids GFP-mouse cortactin (#26722, CTTN-GFP) and CMV-GFP-human NMHC II-A (#11347, MYH9-GFP) were obtained from Addgene. Plasmid pQE-EPR was obtained by inserting the sequence from nucleotide 345 to 560 of murine BC030870 into into pQE-30 vector (Qiagen).

For inserts obtained by RT-PCR, the Pfu DNA Polymerase (Promega) was used. The inserts cloned in all constructs were sequenced on both strands (BMR Genomics, Padova, Italy).

6.4 Cell transfections

NMuMG, 4T1, and MDA-MB-231 cells were transfected with Lipofectamine 2000 (ThermoFischer) while HEK-293 cells were transfected with Attractene transfection Reagent (Qiagen). NMuMG, 4T1, and MDA-MB-231 cells stably transfected with recombinant pBICEP-CMV-2-based vectors were maintained in selective medium containing 800 μg/ml, 350 μg/ml, and 750 μg/ml G418 (Sigma-Aldrich), respectively. NMuMG cells stably transfected with recombinant pIRES1*hyg*-based vectors were maintained in selective medium containing 600 μg/ml Hygromycin B (Sigma-Aldrich). Specific mock cells were generated (for every cell line and every plasmid backbone) by transfecting the corresponding empty vector in each cell type. Mock cells were subjected to a selection procedure identical to the other transfectants. siRNAs utilized to knock-down murine EPR (5′—GAGCAAAAGAGAAUGCUUA—3′) were purchased from Thermo Fisher. Stable KHSRP knock-down in NMuMG cells was obtained using previously described silencing sequences and pSuper-Neo (Oligoengine) according to manufacturer's instructions (Puppo et al, 2016). The adenoviral vectors pAdCMVnull (AdNull) and pAdKHSRP (full-length human KHSRP cDNA cloned into an Adenoviral-Type 5 backbone) were purchased from Vector Biolabs (Puppo et al., 2016).

6.5 Scratch wound closure assay

Either mock or EPR-overexpressing NMuMG cells were cultured in six-wells plates up to confluence and pretreated for 2 hours with 5 μg/ml Mitomycin C (Sigma-Aldrich). A wound was scratched into monolayers and cells were cultured for up to 48 hours in the presence of 5 μg/ml Mitomycin C. Images were taken using an Olympus CKX41 microscope and analyzed using the ImageJ 1.49r package (http://imagej.nih.gov/ij/index.html). Average distance of wound obtained from six

microscopic fields was used for the calculation of percent wound healed. Experiments were performed three times in triplicate.

6.6 Immunofluorescence

Either mock NMuMG cells or stable transfectants overexpressing either EPR-FLAG or EPRSTOPE-FLAG were plated on glass coverslips in 24-well plates (60,000 cells/well). Immunofluorescence was carried out 2 days after plating essentially as reported in Spadaro et al., (2017). Rabbit polyclonal anti-cingulin antiserum (C532) was used at a 1/5000 dilution while anti-FLAG antibody (F1804, Sigma) was used at a 1/500 dilution. Secondary antibodies were diluted in IF buffer and incubated for 30 minutes at 37°C, Alexa488 anti-rabbit (711-545-152, Jackson Laboratory) dilution 1/400, Cy3 anti-mouse (715-165-151, Jackson Laboratory) dilution 1/400. Pictures were taken using a Zeiss Axiophot widefield fluorescent microscope (X-Cite 120Q mercury lamp light source, Excelitas Technologies; retiga EXi, cooled mono 12-bit, Qimaging camera; 63x oil objective; Openlab software). Images were imported into ImageJ to split and merge channels, cropped and adjusted for resolution and for intensity level range using Photoshop (scale bar = 10 μm).

6.7 Orthotopic 4T1 injection in BALB/c mice

BALB/c 8–10-week-old female mice (Envigo) were anesthetized using 100 mg/kg ketamine and 10 mg/kg xylazine intra peritoneal. Eye lubricant was applied, hair around the abdominal and inguinal fat pads were trimmed and the skin was sterilized. With the aid of magnifying surgical loupes, a small incision of less than 3 mm was made externally and caudally to the fourth nipple with the tip of micro-dissecting scissors. The fourth mammary gland fat pad below was located and 100 μl of a suspension of either mock or EPR-

expressing 4T1 cell were injected. uccessful injection is confirmed by the swelling of the tissue. The incision was then sutured. All procedures involving animals have been approved by Institutional Animal Welfare Body (O.P.B.A.) and respect the national current regulations regarding the protection of animals used for scientific purpose (D. Lvo March 4^{th}, 2014, n. 26, legislative transposition of Directive 2010/63/EU of the European Parliament and of the Council of September 22, 2010 on the protection of animals used for scientific purposes). Tumor length and width were measured using a digital caliper at day 10 post-injection and tumor volume was calculated using the formula: volume = (length x $(width)^2/2$). Mice were euthanized after two weeks and tumor masses were removed, weighted and photographed.

6.8 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSA) were performed as detailed in Trabucchi et al., (2009). Production and purification of recombinant KSRP have been described previously in Briata et al., (2005). The labeled RNA was transcribed using Sp6 polymerase from a template generated by inserting into pCY vector a PCR product corresponding to nucleotides from 276 to 407 of murine BC030870.

6.9 Isolation of RNA from cytoplasmic, nucleoplasmic, and chromatin fractions

We followed the protocol recently published by Corey and coworkers (Gagnon et al., 2014) starting from 10 x 10^7 cells. Both cytoplasmic and nucleoplasmic RNAs were precipitated and washed with ice-cold 70% (vol/vol) ethanol prior to be dissolved in QIAzol Lysis Reagent (Qiagen) while the chromatin pellets were immediately dissolved in QIAzol. Ten µl of 0.5 M EDTA were added to all the samples in QIAzol that were heated

to 65 °C with vortexing until dissolved (~10 minutes). The preparation of RNA was continued as described below.

6.10 RNA preparation, quantitative RT-PCR, analysis of newly-synthesized transcripts, and of mRNA decay

Total RNA was isolated using either the miRNeasy mini kit (Qiagen) or QIAzol and retro-transcribed (50-100 ng) using Transcriptor Reverse Transcriptase (Roche) and either random hexamers or oligo-dT according to manufacturers' instructions. Quantitative PCR was performed using the Precision 2X QPCR master mix (Primer Design), and the Realplex II Mastercycler (Eppendorf) according to manufacturers' instructions. The sequence-specific primers utilized for PCR reactions are listed in Supplementary Table 1. In order to analyze gene expression changes among the pool of nascent mRNAs, we adopted the Click-iT Nascent RNA Capture kit (ThermoFischer) and performed experiments according to manufacturer's instructions. NMuMg cells were pulsed with 0.5 mM 5-ethynyl Uridine (EU) for 1 hour. In order to analyze mRNA decay we either blocked transcription by treating cells with 100 μM 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB, Sigma-Aldrich) and isolating total RNA at different intervals of times or performing EU labeling-based pulse chase experiments labeling cells with 0.2 mM EU for 16 hours, removing the culture medium, and chasing cells for 1 hour. RNA was prepared, "clicked", retrotranscribed, and analyzed by qRT-PCR according to Click-iT Nascent RNA Capture kit instructions.

6.11 Ribonucleoprotein complexes immunoprecipitation (RIP) assays

RIP assays were performed as previously described (Giovarelli et al., 2014) with minor modification. Briefly, total cell lysates were immunoprecipitated with Dynabeads

(Thermo Fisher) coated with protein A/protein G and pre-coupled to specific antibodies at 4^{o} C overnight. Pellets were washed three times with a buffer containing 50mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.5% Triton X-100, 1X Complete (Roche). Total RNA was prepared from immunocomplexes using the QIAzol Lysis Reagent, retro-transcribed, and amplified by qPCR as described above. The primer sequences are detailed in Supplementary Table 1.

6.12 Protein identification by MALDI-TOF mass spectrometry (MS) analysis

Total extract from either mock or EPR-FLAG (10 X 10^7 cells) were immunoprecipitated using anti-FLAG antibody-coupled Dynabeads. Immunoprecipitated material was analyzed by SDS-PAGE followed by silver staining. Protein identification was performed as a service at the Functional Proteomic Unit of IFOM (Milano, Italy; Drs. Angela Cattaneo and Angela Bachi). Bands of interest from SDS-PAGE were excised from gels, reduced, alkylated and digested overnight with bovine trypsin (Roche, Milan, Italy), as described (Shevchenko et al., 1996). One μl aliquots of the supernatant were used for mass analysis using the dried droplet technique and α-cyano-4-hydroxycinnamic acid as matrix. Mass spectra were obtained on a MALDI–TOF Voyager-DE STR mass spectrometer (Applied Biosystem). Alternatively, acidic and basic peptide extraction from gel pieces after tryptic digestion was performed and the resulting peptide mixtures subjected to a single desalting/concentration step before MS analysis over Zip-TipC18 (Millipore Corporation). Spectra were internally calibrated using trypsin autolysis products and processed via Data Explorer software. Proteins were unambiguously identified by searching a comprehensive non-redundant protein database of the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) and the Mass

Spectrometry protein sequence DataBase (MSDB, http://msdn.microsoft.com/en-us/library/ms187112.aspx), selected by default using in house the software programs ProFound v4.10.5 and Mascot v1.9.00, respectively. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 3 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

6.13 RNA deep-sequencing (RNA-Seq)

High-quality RNA was extracted from either mock, EPR-, or EPRSTOPE-overexpressing NMuMG cells (biological triplicates for each experimental condition), and a total of nine libraries were prepared using standard Illumina TrueSeq SBS PE 200 cycles protocol and sequenced on HiSeq2500. Image analysis and base calling were performed using the HiSeq Control Software and RTA component from Illumina. This approach yielded between 68 and 77 milions of reads that were further processed. Raw data have been published on the GEO archive under the Accession GSE113178 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113178).

6.14 Analysis of h.EPR (LINC01207) expression in human samples

Meta-analysis of RNA-Seq data of h.EPR in normal samples was performed by searching for h.EPR expression in different subpopulations of FACS-sorted normal breast cells (Pellacani et al., 2016) and in different human organs through either the Expression Atlas (https://www.ebi.ac.uk/gxa/home) or the GEPIA web server (http://gepia.cancer-

pku.cn). h.EPR expression in breast cancer samples was analysed using TCGA data, deriving PAM50 classification from the original publication (Schmierer & Hill, 2005).

6.15 Chromatin Isolation by RNA Purification (ChIRP)

Chromatin Isolation by RNA Purification (ChIRP) was performed according to the protocol published by Chu et al. (2012) with minor modifications. Briefly, 2.5 X 10^7 NMuMG cells were crosslinked in 20 ml of 1% glutaraldehyde in PBS at room temperature for 10 minutes on an end-to-end rotator. After glutaraldehyde quenching and repeated washes, cell pellets were weighted and resuspended in 1.0 ml of complete Lysis Buffer (50 mM Tris-Cl pH 7.0, 10 mM EDTA , 1% SDS, 1 X Complete, 500U RNAse inhibitor) per each 100 mg of cell pellet. Cell suspensions were sonicated for 90 min (power set to 70%) and the sonicated cell lysate was centrifuged at 16,100 x g at 4º C for10 minutes. Lysates were divided into two 1 ml aliquots, transferred into polypropylene tubes, mixed with 2 ml Complete Hybridization Buffer (750 mM NaCl , 1% SDS, 50 mM Tris-Cl, pH 7.0, 1 mM EDTA, 15% formamide, 1 X Complete, 1000 U RNAse Inhibitor) and hybridized with 1 µl (100 pmol) of either EVEN or ODD pools of 20-mer 3' Bio-TEG DNA oligonucleotides designed with single-molecule FISH online designer (Stellaris) (see Supplemental Table 1), respectively. Hybridization was carried out at 37º C for 4 hours under continuous shaking. Seventy µl of pre-washed C-1 magnetic beads (Thermo Fisher) were added to each hybridization mixture for 30 minutes at at 37º C under continuous shaking. Beads were immobilized and washed four times for 5 minutes ar 37º C with shaking (wash buffer: 2 X NaCl and Sodium citrate (SSC), 0.5% SDS, 1 X Complete). While one aliquot (10% of the material) was utilized for RNA extraction, the remaining 90% was subject to DNA purification by incubating two times each bead pellet with 150 µl Complete DNA Elution Buffer (50 mM $NaHCO_3$, 1% SDS, 25 µg/ml RNAseA, 100 U/ml

RNAseH) for 30 min at 37º C with shaking. Eluted DNA was incubated with Proteinase K (1mg/ml final dilution) for 45 min at 50º C with shaking, extracted with Phenol/Chlorophorm/Isoamylalchool, Ethanol-precipitated, and aliquots were analyzed by qPCR.

6.16 ChIP-qPCR

The Magna ChIP A/G (Millipore, catalog # 17-10085) kit was used in combination with anti-Pol II, anti-H3K27me3, and anti-SMAD3 antibodies (and the corresponding control IgG).

6.17 Sucrose-gradient fractionation and polysome profiling

Experiments were performed as described (Zaccara et al., 2014). NMuMG cells (~70% confluence) were treated with cycloheximide (0.1 mg/ml) for 5 min at 37°C, washed twice with PBS supplemented by 0.01 mg/ml cycloheximide, scraped in PBS 1X with 0.01 mg/ml cycloheximide, pelleted by centrifugation, lysed in 500 µl of ice-cold Lysis Buffer (Salt Solution 1X, 1% Triton-X100, 1% NaDeoxycholate, 0.2 U/µl RNase Inhibitor, 1mM DTT, 0.01 mg/ml cycloheximide), centrifuged for 5 minutes at 13000 rpm at 4°C, and supernatants were loaded onto sucrose gradients. One ml fractions were collected monitoring the absorbance at 260nm using a Density Gradient Fractionation System by Teledyne ISCO with sensitivity set to 0.2. Using the profile of the 260 nm absorbance, fractions corresponding to free ribosomal subunits (40S and 60S) and monosomes (80S, considered as not translating), separately from fractions corresponding to light polysomes (2-5 ribosomes) and heavy polysomes (>6 ribosomes) were pooled together and processed for RNA extraction and RNA was quantified by Nanodrop (Thermo Fisher).

6.18 Analyses of cell cycle distribution and quantification of S phase cells by flow cytometer

NMuMg cells (either mock or EPR- or EPRSTOPE- transfected) were seeded in 6-well plates. For the analysis by the Cycletest^{TM} Plus DNA Kit (BD Medical Technology), cells are detached by trypsinization and centrifuged in Eppendorf tubes at 300 Xg for 5 min at room temperature. Supernatant is removed and the pellet is resuspended in 1 ml 1x PBS followed by centrifugation. Cells are then resuspended in PBS and counted using Countess® Automated Cell Counter and the cell concentration is adjusted to $7x10^5$ cells/ml using the same buffer. The DNA staining procedure is performed using 0.5 ml of cell suspension ($7x10^5$ cells). Cells are pelleted by centrifugation (400 Xg for 5 min at RT). After carefully removing the supernatant, cells are mixed in Solution A (provided by the kit, containing trypsin in a spermine tetrahydrochloride buffer for digestion of cell membranes and cytoskeleton), without using a vortex. 200 µl of solution B (provided by the kit, containing trypsin inhibitor and ribonuclease A in citrate-stabilizing buffer, to inhibit the trypsin activity and to digest the RNA) is gently added and the sample is incubated for 10 min at RT, followed by the addition of 200 µl of cold solution C (provided by the kit, containing Propidium Iodide and spermine tetrahydrochloride in citrate stabilizing buffer). The sample is incubated in the dark and on ice for 10 min and then filtered by cell strainer caps and analyzed by flow cytometer (BD FACSAria II). Data on at least 10000 events for sample were processed using ModFit LT 4 software. The experiment was repeated two times.

To estimate more precisely the fraction of cells in S phase, the Click-iT^{TM} Plus EdU Flow Cytometry Assay (Invitrogen) was used. EdU (10 mM stock in DMSO) was added directly to the culture medium at the 20 µM final concentration and incubated for 40 minutes. Cells were then harvested by trypsinization and washed using 3 ml of PBS

containing 1% BSA. Pellets are resuspended in PBS + 1% BSA, counted using Countess® Automated Cell Counter and $1.5x10^6$ cells are transferred to flow tubes, washed again with 3 ml of PBS containing 1% BSA, pelleted by centrifugation followed by removal of the supernatant. Cells are resuspended in 100 µl of Click-iT™ fixative mixing well with a pipette and incubated for 15 min at RT in the dark. Cells are the washed as performed in the previous step, resuspended and incubated for 10 minutes in 100 µL of 1X Click-iT™ saponin-based reagent. Samples are then processed for the Click-iT™ reaction, preparing the Click-iT™ Plus reaction cocktail according to the manufacturer guidelines, adding 0.5 ml of it to each sample, to reach a final volume of 600 µl containing $1.5x10^6$ cells and incubating for 30 minutes at room temperature, in the dark. Cells are then washed once using 3 ml of 1X Click-iT™ saponin-based reagents, pelleted and resuspended in 600 µl of the same solution to which the Propidium Iodide staining solution is added to stain DNA. Propidium Iodide solution contains 50 µg/ml PI and 100 µg/ml RNAse. Samples are then analyzed by flow cytometer (counting 20000 events, BD FACSAria II). As controls, cell aliquots incubated with EdU and processed by the same protocol, but skipping the Click-iT™ reaction or the PI staining, or both.

6.19 G1 phase cell sorting

Cells (mock, EPR, EPRSTOPE) were harvested by trypsinization when they reached ~90% confluence, washed once in 1X PBS and resuspended in DMEM without serum at a concentration of 10^6 cells/ml. Hoechst 33342 (Thermo Fisher Scientific) was added to the media at the concentration of 10µg/ml and cells were incubated for 1 hour at 37°C. Cells were then centrifuged to remove Hoechst-containing media and resuspended in 1X PBS.

Sorting was performed by BD Aria II cytometer (BD Bioscience) using a 100µm nozzle and setting a gate on the population of cells in G1. At least 90K events for every

sample were sorted in 1X PBS at room temperature. After sorting, purity was assessed by re-running the samples. Sorted cells were pelleted and immediately stored at -80C. RNA was extracted and analyzed by qRT-PCR as described above.

6.20 Quantification of cell proliferation by high-content image analysis

The proliferation of NMuMG cells (either mock or EPR- or EPRSTOPE- transfected) was quantified using Operetta High-Content Imaging System, acquiring images at different time points by digital phase contrast with a 40x objective. Images were analyzed using Harmony® High Content Imaging and Analysis Software. 500 cells were seeded in 96 well plates in triplicates. Pictures were taken at different time points, by automatically acquiring eight fields for each well. Data were analyzed in Excel and plotted as average and standard deviations of replicates.

6.21 Quantification of cell proliferation by crystal violet staining

For some experiments cell proliferation was assessed by crystal violet staining. At the indicated time after plating, cells were fixed (10% formalin) and stained (0.1% crystal violet) with crystal violet solution. After two washes with water, crystal violet staining was measured by spectrophotometer at a wave-length of 590 nm.

6.22 Clonogenic and anchorage-independent cell growth assays

For the clonogenic assays, cells were plated in triplicate on six-well plates at 500 cells per well and left to growth for 4-6 days. Cells were fixed and stained with crystal violet solution. Anchorage-independent cell growth assays was assessed according to the protocol published by Borowicz et al. (2014) with minor modifications. Briefly, 2500 cells were seeded in 0.3% Top Agar in complete medium and placed on a layer of 0.5% of

Bottom Agar in 12-well plates. Each cell line was seeded in sextuplicates and fed every three days. After 21 days cells were colored with crystal violet and photographs were taken.

6.23 Quantification and statistical analysis of RNA-Seq

Raw FASTQ reads were trimmed at the ends to remove low quality calls with FASTX [http://hannonlab.cshl.edu/fastx_toolkit]. Paired-End reads were aligned to indexed mm10 genome with STAR (v 2.3.0e_r291).

To quantify expression levels mapped reads were counted from BAM files with HTSeq counts version 1.2.1, in "intersection-strict" mode, feature type "exon" and id attribute "gene_name" against reference annotation Ensembl GRCm38.74.

6.24 Quantification and statistical analysis of transcript differential expression analysis

In addition to gene-level analysis with STAR-HTSeq, the transcript abundance was further re-estimated using an alignment-free approach based on Kallisto 0.43.1 software, using Gencode Mouse vM15 transcripts as reference.

Abundance files were imported in R.3.1.1 with TxImport.1.2.0 with option "txOut=TRUE" to quantify alternatively spliced transcripts. edgeR_3.16.5 and limma_3.30.13 were used to log2 transform transcripts count in Count Per Million (cpm). Only transcript with >= 1 cpm in at least 3 samples were retained. Cpm were transformed by library size and normalized by mean variance with limma-voom. Statistics and log-ratio were calculated with limma-eBayes, by fitting data to a single-factor linear-model with three different levels (mock, EPR, EPRSTOPE).

6.25 Venn Diagram

We kept differentially expressed transcripts when the observed Bayesian statistic was significant (Benjamini and Hochberg corrected p-value < 0.01; logFC > | 0.5 |). The functions *limma-vennDiagram* and *pheatmap* were used to cluster and visualize the significant genes.

6.26 Gene Ontology and Pathway Enrichment

Significant transcripts were summarized at gene level (*tximport-summarizeToGene*), annotated by Gene Ontology and enriched by statistically over-represented term with the *EnrichR* web-application using a nominal p-value threshold of p<0.01. The *EnrichR* p-values refer to the Fisher Exact Test statistics, which is a proportion test that assumes a binomial distribution and independence for probability of any gene belonging to any set.

6.27 Protein alignment

Multiple alignment of mammalian EPR sequences was conducted by using the ClustalW2 package (ebi.ac.uk).

6.28 SUPPLEMENTARY TABLE 1. List of primers used for qRT-PCR and qPCR analyses as well as of 3' Bio-TEG oligonucleotides used for ChIRP experiments.

Primer name(qRT-PCR)	Forward	Reverse
mmu.Rpl32	CTGGCCATCAGAGTCACCAA	TGCACACAAGCCATCTACTCA
mmu.linc-EPR	AACTGCGCAACCTAGTCCTG	GCTCCATGGAAGACTAACGG
mmu.Cdh1	GTGAGGACGACTAGGGGACT	CACACTCAGGGAAGGAGCTG
mmu.Cdh2	CGAGAGGCCTATCCATGCTG	CCCAATATCCCCAGGGTGTG
mmu.Cdkn1a	TGTCGCTGTCTTGCACTCTG	ACCAATCTGCGCTTGGAGTG
mmu.Adam12	TGGGGATGTGCCTCTTCAAC	ATTCGTGCATTCCTCCGGTT
mmu.Zeb2	GTCTCTGCAAGTGCCATCCT	ACTGACACGGGTGCTTCAAA
mmu.Fstl1	ACGCTCCCACCTTCGCCTCT	GTCACCAGCGAGAGCGCCAG
mmu.Fn1	GCAGGAAAGTCACCCAGACA	CTGTGGGAGGGGTGTTTGAA
mmu.Tnc	AACGGACTGCCCACATCTCA	TTCCGGTTCAGCTTCTGTGGTA
mmu.Ocln	TGGCAAGCGATCATACCCAG	AGGAATCTCCTGGGCCACTT
mmu.Twist2	ACCCAGGTTCACATCCTCCT	TGAGGAGATGAGGGCACAGA
mmu.Anxa6	GTGTGTGCTTTAAAGGGCGG	TCAGATTGGACGGTGACAGC
mmu.Tjp1	GGACACCAAAGCATGTGAGC	AGGGTAAGGCATTCCTGCTG
mmu.Cdkn2b	GCAGCTGGATCTGGTCCTTGAGC	CCCCTGCAGCAGCAGCTTGT
mmu.Pcdh19	TCAAAAGCAGCTCCACCTTCA	GCAGTATCGCAGTACAGGCT
mmu.Slc9a2	GCGAGTCTCTGCTCAACGAT	ACGTCTACGGTCTGGATGGT
mmu.Slc39a4	GGGCCGTGTGAAAAGTGTCT	GCAGCAGCAGCACTAAGGTA
mmu.Rap1gap	CCACCTGGTATTCTCGCTCAA	GGTCCGACACTTGGTCCTAA
mmu.Cdx2	CTGGACAAGGACGTGAGCAT	ACTGCGGAGGACTGACAAA
mmu.Fhl1	GACAAGTTCTGCGCCAACAC	TTGTCGTGCCAGTAGCGATT
mmu.Tp53Inp1	TAGCTTTGCCCCTACTCCCT	GCCAAACAAATTCTGACGGCT
mmu.Fgfr2	CACGACCAAGAAGCCAGACT	CTCGGCCGAAACTGTTACCT
mmu.Neat1	GGGACTTGTGGGAGAAAGCA	TTCCAGGCACAATCCTCACC
hsa.linc-EPR	CAGACTCACGGGTTCCTTCC	CTCCATGAACCACCACCGAA
hsa.*CDH1*	CAGGTGTGCACAGAAAACCG	CCCCCTTTAGGGCCACATTT
hsa.*CDH2*	AGTGGTGGTGAGCAGGACTA	ACCTCCACCATACATGTCAGC
hsa.*TNC*	CCCTGAGGGAGCAATGTACTG	CAGCCACATCCTTCAGTGCT
hsa.*ADAM12*	CCGGTGTTCCCCACTGTATC	AGCTGGGTTCCCTTTTGTGT
hsa.*FSTL1*	AAATGCAGCTCCCTGTCCAA	CCCTCCTCCCATAGTGTCCA
hsa.*CDKN1A*	ACTCTCAGGGTCGAAAACGG	CTTCCTGTGGGCGGATTAGG
tRNA_lys	GCCCGGATAGCTCAGTCG	CCCGAACAGGGACTTGAAC
Primers for ChIP	Forward	Reverse
mmu.Cdkn1aP1	GGTGATTTTGGCGTCCACAC	TCCAAGGGACCAAGGGAGAT
mmu.EPR P1	GACCACAGGCACATACGTCA	ATATACAACCCGGAGCGAGC
mmu.Rpl32P1	ACATTTGCTCAACCAACCGC	GCCCTAAGTGGAAGCCCAAT
3' Bio-TEG oligos (ChIRP)	ODD	EVEN

Linc-EPR 26	GGAACCAGGAAGTGGTATGAA	
Linc-EPR 95		GTGAATCTGTACTCTCAAGG
Linc-EPR 175	TGCAAATGCTGACTTCACGG	
Linc-EPR 257		CTGGTTAGACAGTCTTTCTC
Linc-EPR 341	TGAATGGTAGCTCCATGGAA	
Linc-EPR 421		AGGGCTGGTGTAGATGGAAG
Linc-EPR 580	ATTCTTAGGGGAGCGACTGTG	
Linc-EPR 640		GCGATTATGATCTGAAGGCT
Linc-EPR 720	CAGCTAGTGCTCATAATTGG	
Linc-EPR 1441		ATAGAAGCAGCAGAGACATGG

7. REFERENCES

Abbas, T. & Dutta, A. p21 in cancer: intricate networks and multiple activities. *Nat. Rev. Cancer*. **9**, 400-414 (2009).

Anderson, D.M. *et al*. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595-606 (2015).

Andrews, S.J. & Rothnagel, J.A. Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet*. **15**, 193-204 (2014).

Arun, G., Diermeier, S.D. & Spector, D.L. Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends Mol. Med*. **24**, 257-277 (2018).

Bahar Halpern, K. *et al*. Nuclear Retention of mRNA in Mammalian Tissues. *Cell Rep*. **13**, 2653-62 (2015).

Balas, M. M., & Johnson, A. M. Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-coding RNA Research* (2018).

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

Berretta, J., & Morillon, A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO reports* **10 (9),** 973-982 (2009).

Bhan, A., & Mandal, S. S. . LncRNA HOTAIR: A master regulator of chromatin dynamics and cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* **1856 (1)**, 151-164 (2015).

Bianchini, G., Balko, J.M., Mayer, I.A., Sanders, M.E. & Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat. Rev. Clin. Oncol*. **13**, 674-690 (2016).

Bierhoff H., Postepska-Igielska A., Grummt I. Noisy silence: non-coding RNA and heterochromatin formation at repetitive elements. *Epigenetics* **9**, 53–61 (2014).

Birney E., Stamatoyannopoulos J.A., Dutta A., Guigo R. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* (2007).

Borowicz, S. *et al*. The soft agar colony formation assay. *J. Vis. Exp*. (**92**), e51998 (2014).

Brabletz, T., *et al.* Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc Natl Acad Sci U.S.A.* **98**:10356-10361 (2001).

Braun, K.A. & Young, E.T. Coupling mRNA synthesis and decay. *Mol. Cell. Biol*. **34**, 4078-4087 (2014).

Briata, P. *et al*. p38-dependent phosphorylation of the mRNA decay-promoting factor KSRP controls the stability of select myogenic transcripts. *Mol. Cell* **20**, 891-903 (2005).

Briata, P., Chen, C.Y., Ramos, A. & Gherzi, R. Functional and molecular insights into KSRP function in mRNA decay. *Biochim. Biophys. Acta* **1829**, 689-694 (2013).

Briata, P., Bordo, D., Puppo, M., Gorlero, F., Rossi, M., Perrone-Bizzozero, N., & Gherzi, R. Diverse roles of the nucleic acid-binding protein KHSRP in cell differentiation and disease. *Wiley Interdisciplinary Reviews: RNA* **7(2)**, 227-240 (2016).

Budi, E.H., Duan, D. & Derynck, R. Transforming Growth Factor-β Receptors and Smads: Regulatory Complexity and Functional Versatility. *Trends Cell Biol*. **7**, 658-672 (2017).

Cabianca D.S., Casa V., Bodega B., Xynos A., Ginelli E., Tanaka Y., Gabellini D. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819–31 (2012).

Cabili M.N., Trapnell C., Goff L., Koziol M., Tazon-Vega B., Regev A., Rinn J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**, 1915–27 (2011).

Cabili M.N., Dunagin M.C., McClanahan P.D., Biaesch A., Padovan-Merhar O., Regev A., Rinn J.L., Raj A. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16**,20 (2015).

Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 . (2012).

Cao, L., Zhang, P., Li, J. Wu, M. LAST, a c-Myc-inducible long noncoding RNA, cooperates with CNBP to promote CCND1 mRNA stability in human cells. *Elife* **6**, pii: e30433 (2017).

Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., *et al.* The transcriptional landscape of themammalian genome. *Science* **309**,1559–1563 (2005).

Chen, C.Y. *et al*. Nucleolin and YB-1 are required for JNK-mediated interleukin-2 mRNA stabilization during T-cell activation. *Genes Dev*. **14**, 1236-1248 (2000).

Chooniedass-Kothari S., Emberley E., Hamedani M.K., Troup S., Wang X., *et al.* The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett* **566**, 43–47 (2004).

Chooniedass-Kothari S., Hamedani M.K., Troup S., Hube F., Leygue E. The steroid receptor RNA activator protein is expressed in breast tumor tissues. *Int J Cancer* **118**,1054–1059 (2006).

Chu, C., Quinn, J. & Chang, H.Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp*. **25**, pii: 3912 (2012).

Clancy, S. RNA Functions. *Nature Education* **1**:102 (2008).

Clemson C.M., Hutchinson J.N., Sara S.A., Ensminger A.W., Fox A.H., Chess A., Lawrence J.B. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* **33**, 717–26 (2009).

Collins L.J., Schonfeld B., Chen X.S. The Epigenetics of Non-Coding RNA. In T. Tollefsbol (Ed.), Handbook of epigenetics: the new molecular and medical genetics (pp. 49-61). London: Academic. (2012).

Davidovich C. & Cech T.R. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA* **21**:2007–22 (2015).

De Andres-Pablo, A., Morillon, A., Wery, M. LncRNAs: lost in translation or licence to regulate? *Current genetics* **63(1)**, 29-33 (2017).

De Craene, B., & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. *Nature Reviews Cancer* **13(2)**, 97 (2013).

Delás, M. J., & Hannon, G. J. lncRNAs in development and disease: from functions to mechanisms. *Open Biology* 7, 170121 (2017).

Dhamija, S. & Diederichs, S. From junk to master regulators of invasion: lncRNA functions in migration, EMT and metastasis. *Int. J. Cancer*. **139**, 269-280 (2016).

Di Ruscio A., Ebralidze A.K., Benoukraf T., Amabile G., Goff L.A., Terragni J., Figueroa M.E., De Figueiredo Pontes L.L., Alberich-Jorda M., Zhang P., *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–6 (2013).

Engreitz J.M., Pandya-Jones A., McDonel P., Shishkin A., Sirokman K., Surka C., Kadri S., Xing J., Goren A., Lander E.S., *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**:1237973 (2013).

Engreitz J.M., Ollikainen N., Guttman M. Long non–coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* **17**, 756–70 (2016).

Fidler, I.J., Poste, G. The "seed and soil" hypothesis revisited. *Lancet Oncol* **9**:808 (2008).

Fischer K.R., Durrans A., Lee S., Sheng J., Li F., Wong S.T., Choi H., El Rayes T., Ryu S., Troeger J. *et al* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).

Gagnon, K.T., Li, L., Janowski, B.A. Corey, D.R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc*. **9**, 2045-2060 (2014).

García-Mayoral, M. F., Hollingworth, D., Masino, L., Díaz-Moreno, I., Kelly, G., Gherzi, R., Ramos, A. The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure* **15(4)**, 485-498 (2007).

Geisler, S. & Coller, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell. Biol*. **14**, 699-712 (2013).

Gendrel A.V., Heard E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu Rev Cell Dev Biol* **30**, 561–80 (2014).

Gherzi, R., Chen, C. Y., Trabucchi, M., Ramos, A., Briata, P. The role of KSRP in mRNA decay and microRNA precursor maturation. *Wiley Interdisciplinary Reviews: RNA* **1(2)**, 230-239 (2010).

Giovarelli, M. et al. H19 long noncoding RNA controls the mRNA decay promoting function of KSRP. *Proc. Natl. Acad. Sci. USA* **111**, E5023–E502 (2014).

Godinho, M., Meijer, D., Setyono-Han, B., Dorssers, L.C. & van Agthoven, T. Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. *J. Cell. Physiol* **226**, 1741–1749 (2011).

Gondran, P., Amiot, F., Weil, D. & Dautry, F. Accumulation of mature mRNA in the nuclear fraction of mammalian cells. *FEBS Lett*. **458**, 324-328 (1999).

Grote P., Wittler L., Hendrix D., Koch F., Wahrisch S., Beisaw A., Macura K., Blass G., Kellis M., Werber M., Herrmann B.G. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* **24**, 206–14 (2013).

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., ... & Wang, Y. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464** (7291), 1071 (2010).

Hacisuleyman E., Goff L.A., Trapnell C., Williams A., Henao-Mejia J., Sun L., McClanahan P., Hendrickson D.G., Sauvageau M., Kelley D.R., *et al.* Topological

organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* **21**,198–206 (2014).

Hammond, S. M. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS letters* **579(26)**, 5822-5829 (2005).

Hansen, T.B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).

Hendrickson G.D., Kelley D.R., Tenen D., Bernstein B., Rinn J.L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol* **17**, 28 (2016).

Huang, J.Z. *et al*. A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell* **68**, 171-184.e6 (2017).

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nature medicine* **21**(11), 1253.

Hube F., Guo J., Chooniedass-Kothari S., Cooper C., Hamedani M. K., Dibrov A. A., *et al.* Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol* **25**, 418–428 (2006).

Jandura, A. & Krause, H.M. The New RNA World: Growing Evidence for Long Noncoding RNA Functionality. *Trends Genet*. **33**, 665-676 (2017).

Johnsson P., Ackley A., Vidarsdottir L., Lui W.O., Corcoran M., Grander D., Morris K.V. A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat Struct Mol Biol* **20**, 440–6 (2013).

Kaikkonen M.U., Lam M.T.Y., Glass C.K. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research* **90**, 430–440 (2011).

Kalluri, R., & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* **119(6)**, 1420-1428 (2009).

Kang Y., Chen C.R., Massagué J. A self-enabling TGFβ response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells *Mol Cell* **11**:915–926 (2003).

Khalil A.M., Guttman M., Huarte M., Garber M., Raj A., Rivea Morales D., Thomas K., Presser A., Bernstein B.E., van Oudenaarden A., *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* **106**:11667–72 (2009).

Komili, S. & Silver, P.A. Coupling and coordination in gene expression processes: a systems biology view. *Nat. Rev. Genet*. **9**, 38-48 (2008).

Kondo T., Plaza S., Zanet J., Benrabah E., Valenti P., Hashimoto Y., *et al.* Small peptides switch the transcriptional activity of shavenbaby during Drosophila embryogenesis. *Science* **329**, 336–339 (2010).

Kopp, F. & Mendell, J.T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**, 393-407 (2018).

Kotzin J.J., Spencer S.P., McCright S.J., Kumar D.B., Collet M.A., Mowel W.K., Elliott E.N., Uyar A., Makiya M.A., Dunagin M.C., *et al*. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature* **537**, 239–43 (2016).

Lamouille, S., Xu, J., & Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nature reviews Molecular cell biology*, **15(3)**, 178 (2014).

Lanz R.B., McKenna N.J., Onate S.A., Albrecht U., Wong J., *et al.* A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17–27 (1999).

Li, X.L. *et al*. Long Noncoding RNA PURPL Suppresses Basal p53 Levels and Promotes Tumorigenicity in Colorectal Cancer. *Cell Rep*. **20**, 2408-2423 (2017).

Long, Y. Wang, X., Youmans, D.T. & Cech, T.R. How do lncRNAs regulate transcription? *Sci. Adv*. **3**, eaao2110 (2017).

Lu, L., Zhu, G., Zhang, C., Deng, Q., Katsaros, D., Mayne, S. T., ... & Benedetto, C. Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer. *Breast cancer research and treatment* **136(3)**, 875-883 (2012).

Magny, E. G., Pueyo, J. I., Pearl, F. M., Cespedes, M. A., Niven, J. E., Bishop, S. A., & Couso, J. P.. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341(**6150), 1116-1120 (2013).

Makarewich, C.A. & Olson, E.N. Mining for Micropeptides. *Trends Cell Biol*. **27**, 685-696 (2017).

Marchese F.P., Huarte M. Long non-coding RNAs and chromatin modifiers: their place in the epigenetic code. *Epigenetics* **9**, 21–6 (2014).

Marchese, F. P., Raimondi, I., & Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome biology* **18**(1), 206 (2017).

Marin-Bejar O., Marchese F.P., Athie A., Sanchez Y., Gonzalez J., Segura V., Huang L., Moreno I., Navarro A., Monzo M., *et al.* Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol* **14**, R104 (2013).

Matsumoto, A. *et al.* (2017). mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* **541**, 228-232 (2017).

Meller V.H., Joshi S.S., Deshpande N. Modulation of chromatin by noncoding RNA. *Annu Rev Genet* **49**, 673–95 (2015).

Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).

Mohammad F., Mondal T., Guseva N., Pandey G.K., Kanduri C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* **137**, 2493–9 (2010).

Morlando M., Ballarino M., Fatica A., Bozzoni I. The role of long noncoding RNAs in the epigenetic control of gene expression. *ChemMedChem* **9**, 505–10 (2014).

Moshiri, A., Puppo, M., Rossi, M., Gherzi, R. & Briata, P. Resveratrol limits epithelial to mesenchymal transition through modulation of KHSRP/hnRNPA1-dependent alternative splicing in mammary gland cells. *Biochim. Biophys. Acta* **1860**, 291-298 (2017).

Nelson, B.R. *et al.* A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* **351**, 271-275 (2016).

Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. **75**, 4646-4658. (2003).

Nishimura G., *et al.* δEF1 mediates TGF-β signaling in vascular smooth muscle cell differentiation. *Dev Cell* **11**:93–104 (2006).

Pardali, K., Kowanetz, M., Heldin, C.H. & Moustakas, A. Smad pathway-specific transcriptional regulation of the cell cycle inhibitor p21(WAF1/Cip1). *J. Cell. Physiol*. **204**, 260-272 (2005).

Parvani, J. G., Taylor, M. A., & Schiemann, W. P. Noncanonical TGF-β signaling during mammary tumorigenesis. *Journal of mammary gland biology and neoplasia* **16(2)**, 127-146 (2011).

Pastushenko I *et al.*, Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463-468 (2018).

Pellacani, D. *et al*. Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. *Cell Rep*. **17**, 2060-2074 (2016).

Poliseno, L. *et al*. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).

Postigo A.A. Opposing functions of ZEB proteins in the regulation of the TGFβ/BMP signaling pathway. *EMBO J* **22**:2443–2452 (2003).

Prasanth K.V., Prasanth S.G., Xuan Z., Hearn S., Freier S.M., Bennett C.F., Zhang M.Q., Spector DL. Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249–63 (2005).

Principe, D.R. *et al*. TGF-β: duality of function between tumor prevention and carcinogenesis. *J. Natl. Cancer Inst*. **106**, djt369 (2014).

Puppo, M. *et al*. miRNA-Mediated KHSRP silencing rewires distinct post-transcriptional programs during TGF-β-induced Epithelial-to-Mesenchymal Transition. *Cell Rep*. **16**, 967-978 (2016).

Quinn, J..J &. Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet***17**, 47–62 (2016).

Reguly, T. & Wrana, J.L. In or out? The dynamics of Smad nucleocytoplasmic shuttling. *Trends Cell Biol*. **13**, 216-220 (2003).

Riley K.J. & Maher L.J. p53 RNA interactions: new clues in an old mystery. *RNA* **13**, 1825–33 (2007).

Rinn J.L., Kertesz M., Wang J.K., Squazzo S.L., Xu X., Brugmann S.A., Goodnough L.H., Helms J.A., Farnham P.J., Segal E., Chang H.Y. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**:1311–23 (2007).

Rinn J.L. & Chang H.Y. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**,145–66 (2012).

Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P.P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353–358 (2011).

Schmierer, B. & Hill, C.S. Kinetic analysis of Smad nucleocytoplasmic shuttling reveals a mechanism for transforming growth factor beta-dependent nuclear accumulation of Smads. *Mol. Cell. Biol*. **25**, 9845-9858 (2005).

Schmitt A.M., Garcia J.T., Hung T., Flynn R.A., Shen Y., Qu K., Payumo A.Y., Peres-da-Silva A., Broz D.K., Baum R., *et al.* An inducible long noncoding RNA amplifies DNA damage signaling. *Nat Genet* **48**,1370–6 (2016).

Schmitt, A.M & Chang, H.Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**, 452-463 (2016).

Schmitz K.M., Mayer C., Postepska A., Grummt I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* **24**, 2264–9 (2010).

Seoane, J., Le, H.V., Shen, L., Anderson, S.A. & Massagué J. Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell* **117**, 211-223 (2004).

Sharma, S. *et al*. Acetylation-Dependent Control of Global Poly(A) RNA Degradation by CBP/p300 and HDAC1/2. *Mol. Cell* **63**, 927-938 (2016).

Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem*. **68**: 850–858 (1996).

Shi, Y. & Massague, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685–700 (2003).

Shore A.N., Kabotyanski E.B., Roarty K., Smith M.A., Zhang Y., Creighton C.J., Dinger M.E., Rosen J.M. Pregnancy-induced noncoding RNA (PINC) associates with polycomb repressive complex 2 and regulates mammary epithelial differentiation. *PLoS Genet* **8**, e1002840 (2012).

Siegel, P.M. & Massagué, J. Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. *Nat. Rev. Cancer* **3**, 807-821 (2003).

Singer, S. *et al*. Nuclear pore component Nup98 is a potential tumor suppressor and regulates posttranscriptional expression of select p53 target genes. *Mol. Cell* **48**, 799-810 (2012).

Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z. & Saghatelian, A. Peptidomic discovery of short open reading frame–encoded peptides in human cells. *Nature chemical biology* **9(1)**, 59 (2013).

Sole C., Nadal-Ribelles M., de Nadal E., Posas F. A novel role for lncRNAs in cell cycle control during stress adaptation. *Current genetics* **61**, 299–308 (2015).

Sorensen, K. P., Thomassen, M., Tan, Q., Bak, M., Cold, S., Burton, M., ... & Kruse, T. A. Long non-coding RNA HOTAIR is an independent prognostic marker of metastasis in estrogen receptor-positive primary breast cancer. *Breast cancer research and treatment* **142(3),** 529-536 (2013).

Spadaro, D. *et al*. Tension-dependent stretching activates ZO-1 to control the junctional localization of its interactors. *Curr. Biol*. **27**, 3783-3795 (2017).

Sunwoo H., Dinger M.E., Wilusz J.E., Amaral P.P., Mattick J.S., Spector D.L. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* **19**, 347–59 (2009).

Taft, R. J., Pheasant, M., & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29(3)**, 288-299 (2007).

Taylor, M. A., Parvani, J. G., & Schiemann, W. P. The pathophysiology of epithelial-mesenchymal transition induced by transforming growth factor-β in normal and malignant mammary epithelial cells. *Journal of mammary gland biology and neoplasia* **15(2)**, 169-190 (2010).

Thiery, J.P. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**, 442-454 (2002).

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A. & al. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459 (7249)**, 1010 (2009).

Ulveling D., Francastel C., Hubé F. When one is better than two: RNA with dual functions. *Biochimie* **93**, 633–644 (2011).

Vincent T., *et al.* A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-β mediated epithelial-mesenchymal transition *Nature Cell Biol.* **11**:943–950 (2009).

Wang K.C., Yang Y.W., Liu B., Sanyal A., Corces-Zimmerman R., Chen Y., Lajoie B.R., Protacio A., Flynn R.A., Gupta R.A., *et al*. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–24 (2011).

Wang Z., Zhang X.J., Ji Y.X., Zhang P., Deng K.Q., Gong J., Ren S., Wang X., Chen I., Wang H., *et al.* The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. *Nat Med* **22**, 1131–9 (2016).

Wilson, T. J., & Lilley, D. M. J. RNA catalysis—is that it? *RNA* **21(4)**, 534–537. (2015).

Xing, Z. *et al*. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* **159**, 1110–1125 (2014).

Yang, J., Weinberg, R.A. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* **14**:818-829 (2008).

Ye, X., & Weinberg, R. A. Epithelial–mesenchymal plasticity: a central regulator of cancer progression. *Trends in cell biology* **25(11),** 675-686 (2015).

Yoon, J.H. *et al*. LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* **47**, 648-655 (2012).

Yoon J.H., Abdelmohsen K, Gorospe M. Posttranscriptional gene regulation by long noncoding RNA. *J Mol Biol* **425**, 3723–30 (2013).

Yu, X., Zhang, Y., Li, T., Ma, Z., Jia, H., Chen, Q., Zou, X.. Long non-coding RNA Linc-RAM enhances myogenic differentiation by interacting with MyoD. *Nature communications* **8**, 14016 (2017).

Yuan, J.H. *et al*. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* **25**, 666-681 (2014).

Zaccara, S., Tebaldi, T., Pederiva, C., Ciribilli, Y., Bisio, A. & Inga A. p53-directed translational control can shape and expand the universe of p53 target genes. *Cell Death. Differ*. 21, 1522-34 (2014).

Zarzynska, J. M. Two faces of TGF-beta1 in breast cancer. *Mediators of inflammation*, (2014).

Zeisberg, M., Shah, A.A., Kalluri, R. Bone morphogenic protein-7 induces mesenchymal to epithelial transition in adult renal fibroblasts and facilitates regeneration of injured kidney. *J Biol Chem* **280**:8094-8100 (2005).

Zheng X., Carstens J.L., Kim J., Scheible M., Kaye J., Sugimoto H., Wu C.C., LeBleu V.S. and Kalluri R. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).

AUTHOR CONTRIBUTION.

I performed most of the experiments. Dr. Roberto Gherzi and Dr. Paola Briata conceived the study and performed some experiments (ChiP and ChiRP experiments). Dr. Gabriele Bucci (San Raffaele Hospital, Milano) performed the bioinformatics analysis of the RNA-Seq data. Dr. Dario Rizzotto (CIBIO, University of Trento) performed polysome profiling, analyses of cell cycle distribution, and high-content image analysis. Professor Alberto Inga (CIBIO, University of Trento) analyzed and discussed the data obtained in his laboratory and discussed the revision strategy with me, Dr. Briata and Dr. Gherzi.