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

Cristina Oliveira-Mateos1, Anaís Sánchez-Castillo1,2, Marta Soler1, Aida Obiols-Guardia1, David Piñeyro1, Raquel Boque-Sastre1,3, María E. Calleja-Cervantes1, Manuel Castro de Moura1, Anna Martínez-Cardús1, Raquel Boque-Sastre1,3, Maria E. Calleja-Cervantes1, Manuel Castro de Moura1, Anna Martínez-Cardús1, Teresa Rubio4, Joffrey Pelletier4, María Martínez-Iniesta5, David Herrero-Martín6, Oscar M. Tirado2,6, Antonio Gentilella4,7, Alberto Villanueva5, Manel Esteller1,2,8,9,10, Lourdes Farre5,11 & Sonia Guí1,10

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

1Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L’Hospitalet de Llobregat, Barcelona, Catalonia, Spain. 2Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Carlos III Institute of Health (ISCIII), Madrid, Spain. 3Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3AX Wales, UK. 4Laboratory of Cancer Metabolism, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L’Hospitalet de Llobregat, Barcelona, Catalonia, Spain. 5Program Against Cancer Therapeutic Resistance (ProCURE), ICO, IDIBELL, L’Hospitalet de Llobregat, Barcelona, Catalonia, Spain. 6Sarcoma Research Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L’Hospitalet de Llobregat, Barcelona, Catalonia, Spain. 7Department of Biochemistry and Physiology, Faculty of Pharmacy, University of Barcelona (UB), Barcelona, Catalonia, Spain. 8Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain. 9Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain. 10Josep Carreras Leukaemia Research Institute (JCI), Badalona, Barcelona, Catalonia, Spain. 11Laboratory of Experimental Pathology (LAPEX), Gonçalo Moniz Research Center, Oswaldo Cruz Foundation (CPQGM/FIOCRUZ), Salvador, Bahia, Brazil. Correspondence and requests for materials should be addressed to L.F. (email: mfarre@idibell.cat) or to S.G. (email: sguil@carrerasresearch.org)
he largest part of the mammalian genome is transcribed into RNA species with little or no coding potential, known as noncoding RNAs (ncRNAs). Although their biological roles are still largely unknown, a growing number of long noncoding RNAs (lncRNAs, a label arbitrarily assigned to transcripts longer than 200 nucleotides) display regulatory properties by acting at all levels in gene expression control (from epigenetic modifications and chromatin dynamics to the control of post-transcriptional messenger RNA stability and translation). In some cases, their key functions in normal homeostasis and development links the dysregulation of their expression with causal roles in cancer, and there are instances of lncRNAs involved in each of the cancer hallmarks, including sustained proliferative signaling and growth (e.g., ANRIL, lincRNA-p213, MEG34), invasion and metastasis (e.g., HULC5, MALAT16, HOTAIR7), resistance to cell death (e.g., PCGEM111), and replicative immortality (e.g., TERC12, TERRA13). Mechanisms of action include the interaction with other nucleic acids and/ or protein factors, which confers the ability to function as scaffolds, guides, decays, or allosteric regulators of several nuclear or cytoplasmic processes. In a growing number of examples, their roles intertwine with that of the better-studied miRNAs16, either by cooperating in their function17 or by impairment of the miRNA-mediated regulation18. The latest annotation in GENCODE estimates that up to 16,000 genes in the human genome correspond to lncRNAs, and a similar number is given to pseudogenes (https://www.gencodegenes.org/stats/current.html). Although some pseudogenes do code for proteins, the majority are thought to be lncRNAs owing to the accumulation of mutations in the definition of the open reading frames, and as such their biological functions include the ability to regulate gene expression similarly to lncRNAs19, and are thereby also involved in growth-regulatory roles in cancer.20

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

Results

RPSAP52 impacts on IGF2BP2 and let-7 in breast cancer cells. We have previously uncovered the positive impact of the expression of the pseudogene RPSAP52 on its sense, protein-coding gene HMGA2 (Fig. 1a). Both genes are generally expressed at low levels in differentiated normal tissues and overexpressed in a number of human cancers, including breast cancer, concomitant with a hypomethylation of the associated CpG island (Fig. 1b). In breast cancer patients, a positive correlation between the expression of both genes is observed (Fig. 1c), as it is also seen in the NCI60 panel of cell lines (Supplementary Fig. 1a). Other studies have reported that high HMGA2 expression predicts poor outcome in breast cancer patients. Since our observations indicate that knockdown of RPSAP52 results in a reduction in HMGA2 expression, we decided to look further into the molecular mechanism of RPSAP2-mediated regulation of the locus. A panel of breast cancer cell lines was used to confirm the presence of RPSAP52 transcript by semi-quantitative PCR (Fig. 1d). Surprisingly, most cell lines expressed the annotated RPSAP52 transcript (Rfseq NR_026825.2) together with an additional species that corresponds to the inclusion of a 104-nucleotides-long internal exon (Fig. 1d and Supplementary Fig. 1b). Evidence as to the presence of this alternative exon in the spliced transcript can also be found in the MiTranscriptome database4, which catalogs long polyadenylated RNA transcripts (www.mitranscriptome.org, with reference G018828[T081486]). The quantitative measurement of expression levels indicates that HMGA2 mRNA and the two isoforms of RPSAP52 are 2-3 orders of magnitude overexpressed when there is hypomethylation of the promoter-associated CpG island, as shown with Illumina’s HumanMethylation450 BeadChip analysis (Fig. 1e) and was confirmed by bisulfite sequencing at the nucleotide level (Supplementary Fig. 1c). Altogether, these observations confirm the coordinate expression of both genes and their silencing in hypomethylated conditions. RPSAP52 is annotated as a noncoding RNA in Refseq, but is labeled as coding in some coding potential calculator tools. Pseudogenes are more likely to give false positive results in programs such as PhyloCSF (since they are similar to their parental protein-coding, and PhyloCSF evaluates conservation to predict coding capacity). We thus conducted in vitro transcription/translation assays, which confirmed the absence of RPSAP52 coding potential (Supplementary Fig. 2a). However, analysis of RNA presence along sucrose gradients from MCF10A cells showed the presence of RPSAP52 transcripts in polysomal fractions, indicating a role in translation. Interestingly, a strong correlation in co-sedimentation of HMGA2 mRNA and RPSAP52 transcript was observed (Supplementary Fig. 2b-e). Indeed, further characterization of RPSAP52 transcripts showed that they are enriched in the cytoplasm (Fig. 1f) and polyadenylated (Fig. 1g), suggesting additional roles besides the ability to regulate HMGA2 transcription in the nucleus. In order to identify protein partners of RPSAP52 that could help characterize its activity, we performed RNA pull-down assays combined with mass spectrometry (MS). In vitro synthesized full-length RPSAP52 RNA was incubated in the presence of MCF10A extracts and the retrieved proteins were analyzed by SDS-PAGE. As shown in Fig. 2a, a protein band of ~70 kDa is specifically pulled-down by RPSAP52 RNA, but not by its antisense sequence or another unrelated RNA. This band was characterized by MS, which identified two proteins within the isolated fragment: the insulin-like growth factor 2 mRNA-binding protein 2 (IGF2BP2), also known as IMP2 (from which seven peptides were identified), and the heterogeneous nuclear ribonucleoprotein Q (HNRNPO), also known as SYNCRIIP (identified with six peptides) (Supplementary Fig. 3a). IGF2BP2, together with IGF2BP1 and IGF2BP3 partake of a family of RNA binding proteins that have been implicated in post-transcriptional control, including the regulation of mRNA localization, stability, and translation. Similar to HMGA2, although the expression of IGF2BPs is normally restricted to embryonic stages, they are re-expressed upon malignant transformation, playing roles in the maintenance of cancer stem cells and the promotion of tumor growth. Western blot with specific antibodies confirmed that both IGF2BP2 and HNRNPO are enriched in the RPSAP52 pull-down, and analysis of RPSAP52 truncates indicate that the two isoforms are able to bind to these two factors (Fig. 2b). In accordance with the pull-down results, the previously reported consensus binding site for IGF2BP2, the CAUH (H = A, C, U) motif28, is abundant along the two constitutive RPSAP52 exons, but absent on the alternative exon (Supplementary Fig. 3b), suggesting that the alternative splicing event does not impact on the affinity of the binding.

IGF2BP2 is a direct transcriptional target of HMGA2 and both proteins partake of a pro-proliferative axis that interweaves with the function of let-7 family of miRNAs. Both HMGA2 and IGF2BP2 mRNAs are direct targets of let-7, but IGF2BP proteins have been suggested to modulate let-7 action via
the formation of cytoplasmic mRNPs that would protect certain mRNAs from let-7 binding and repression. The most abundantly expressed members of the family are let-7a/b/e in MCF10A cells, and let-7a/d/f/g/i in Hs578T cells (Supplementary Fig. 3c). Interestingly, the levels of the mature form of these miRNAs were upregulated in RPSAP52-depleted cells, both in MCF10A and Hs578T clones stably expressing shRNAs (Fig. 2c and Supplementary Fig. 3g) and in cells transiently expressing three different locked nucleic acid (LNA)-based antisense oligonucleotides (ASOs) gapmers (Fig. 2e). RPSAP52 has a region...
of homology with other RPSA pseudogenes, but none was affected by our depletion strategy (Supplementary Fig. 3d). Also, given the possibility that some pseudogenes regulate parental gene expression, we analyzed the levels of RPSA protein in the RPSAP52-depleted cells, but no quantitative change was found (Supplementary Fig. 3e). Of note, gammer-mediated depletion of HMGA2 increased both RPSAP52 isoforms and resulted in a decrease in let-7 levels, suggesting that the negative regulation exerted by RPSAP52 on the miRNAs is not through HMGA2 pathway (Fig. 2e). Let-7 regulates IGF2BP2 mRNA and other members of IGF1 signaling pathway, among others IGFR1 and RAS. Downregulation of RPSAP52 with shRNAs or gammers reduces the amount of these proteins, in accordance with the increased let-7 levels (Fig. 2d, f and Supplementary Fig. 3g, h). LIN28A and LIN28B are the main negative regulators of let-7 biogenesis, through direct binding to either pre-let-7 and/or pri-let-7, but often only one of the two proteins is found expressed in human cancer cell lines. We could not detect LIN28B protein expression in MCF10A cells, and LIN28A was not altered upon RPSAP52 knockdown (Supplementary Fig. 3f), suggesting that changes in let-7 in MCF10A cells were not consequences of impaired biogenesis at the level of regulation by LIN28. However, RPSAP52-mediated regulation of IGF2BP2 protein levels is reverted by overexpression of LIN28B, indicating a convergence on the same regulatory network (Fig. 2g).

Next, the phenotypic impact of the altered control of the IGF2BP2/IGFR1/RAS pathway by RPSAP52 was tested both in vitro and in vivo. RPSAP52 has oncogenic-like features in vitro and in vivo. Upon RPSAP52 knockdown, all three breast cell lines tested (the non-transformed MCF10A and the tumorigenic HS78T and HCC1143 cells) proved to be significantly more proliferative in the sulforhodamine B (SRB) assay (Fig. 3a), and had a significantly lower percentage colony formation density than control cells (Fig. 3b). Interestingly, RPSAP52 depletion was also associated with a decreased migration potential (Fig. 3c). High levels of let-7 miRNAs often correlate with a lower capacity for self-renewal and pluripotency. Given the observed reduction in proliferation and migration following RPSAP52 knockdown, we next assessed the levels of markers of cell stemness (Fig. 3d). NANOG and OCT4 protein levels were decreased in RPSAP52-depleted cells, suggesting this IncRNA promotes features of cancer stem cells. This was further confirmed in soft-agar colony formation experiments, in which the measure of the anchorage-independent growth of the cells showed a significant decrease upon depletion of RPSAP52 (Fig. 3e).

For the in vivo approach, we next used tumor formation assays in nude mice. MCF10A and HS78T cells stably expressing either scrambled shRNAs or shRNAs against RPSAP52 were subcutaneously injected into mice, and the tumor formation and volume was monitored. Tumors originating from RPSAP52 knockdown cells had a significantly lower volume and weight at end point than control tumors, both for the non-tumorigenic and the tumorigenic cells (Fig. 3f, g). Importantly, the amount of RAS and IGF2BP2 proteins were markedly reduced in the excised tumors at end point, indicating that the in vitro findings were maintained in the in vivo context (Fig. 3f, g).

RPSAP52 regulates IGF2BP2/LIN28B/let-7 axis in sarcoma. We next attempted to determine whether RPSAP52-mediated regulation of proliferative pathways occurred in other cancer types. The analysis of the collection of human cancers available from the Cancer Genome Atlas (TCGA) indicates HMGA2 and RPSAP52 expression is specially increased in adenocortical carcinoma, mesothelioma and in the sarcoma samples available (as measured by Z-score, Supplementary Fig. 4a). TCGA RNA expression and DNA methylation data showed that RPSAP52 promoter hypermethylation was associated with transcript downregulation across sarcoma samples (Fig. 4a, upper panel; Pearson correlation, r² = 0.264, P-value = 4.764e-10). Of note, HMGA2 expression in the same samples shows a poorer correlation (Supplementary Fig. 4b, r² = 0.131, P-value = 2.582e-05). Both genes maintain a positive expression correlation (Fig. 4a, lower panel) and a difference of ~1–2 orders of magnitude in their relative expression (Supplementary Fig. 4c). This is in agreement with absolute quantification of RPSAP52 and HMGA2 transcripts in MCF10A and A673 cell lines, in which HMGA2 mRNA is ~2 orders of magnitude at higher levels (Supplementary Fig. 4d). Since the HMGA2-IGF2BP2-RAS pathway has been previously involved in the pathogenesis of embryonic rhabdomyosarcoma, we then assessed HMGA2 and RPSAP52 expression in a panel of cell lines derived from rhabdomyosarcoma and also Ewing’s sarcoma (Fig. 4b). Both RPSAP52 isoforms were abundantly expressed in most rhabdomyosarcoma cell lines, with just one order of magnitude higher HMGA2 expression in Rh28, Rh41, or CW9019 cells. In Ewing’s sarcoma, RPSAP52 was generally lowly expressed with the exception of A673 cell line. We thus focused on A673 cells to further characterize the molecular function of this IncRNA. As seen in MCF10A cells, RNA pull-downs confirmed the ability of RPSAP52 to interact with IGF2BP2 and SYNCRIP/HNRNPQ (Fig. 4c). Importantly, stable clones expressing two different shRNAs against both RPSAP52 isoforms resulted in a strong increase in let-7 family members (Fig. 4d, upper panel), even when HMGA2 levels were only...
Fig. 2 RPSAP52 interacts with IGF2BP2 and HNRNPQ and influences proliferative pathways in MCF10A cells. a RNA pull-down assay to detect RPSAP52-associated proteins. In vitro synthesized full-length RPSAP52 transcript (including the alternative exon) or control sequences (the antisense transcript and the unrelated Uc.160+ RNA) were tested. The proteins retrieved were analyzed by SDS-PAGE and the band of ~70 kDa indicated by the arrow was identified by MS as containing IGF2BP2 and HNRNPQ. b Western blot showing the association between RPSAP52 RNA and IGF2BP2 and HNRNPQ proteins. Different truncated fragments of RPSAP52 RNA (as shown in the upper diagram) were incubated in the presence of MCF10A total protein extracts and the pulled-down material was subject to western blot with specific antibodies. Total extract from the MCF10A cell lines was used as input control, and a reaction without RNA (beads) as negative control. c Stable knockdown of RPSAP52 results in upregulation of let-7 family of miRNAs. Total RNA from MCF10A clones constitutively expressing two different shRNAs against RPSAP52 (sh1 or sh4) was analyzed by RT-qPCR to assess HMGA2 mRNA, RPSAP52 transcripts and let-7 miRNAs levels. Graphs represent the mean ±SD of three independent RNA extractions. Two-tailed student t-test were used (*P < 0.05, **P < 0.01, ns = not significant). d Western blot to analyze IGF2BP2, IGF1R, and RAS protein levels upon transient overexpression of LIN28B protein in the background of RPSAP52 depletion. Source data are provided as a Source Data file.
moderately reduced (Fig. 4d, lower panel). In this case, RPSAP52 knockdown did not correlate with IGF2BP2 decrease, but with a marked reduction in LIN28B protein levels, which in contrast to breast cell lines, is abundantly expressed in A673 cells (Fig. 4e and Supplementary Fig. 4e). Also, while RAS levels were only partially reduced, downstream signaling was impaired, as observed by the decrease in p-ERK levels (Fig. 4e). In an in vivo setting, and similarly to the observations in breast cell lines, this results in a marked reduction in tumor formation when mice are subcutaneously injected with RPSAP52-depleted A673 cells (Fig. 4f).
Fig. 3 RPSAP52 displays oncogenic features in breast cancer cells. a Viability/cytotoxicity assays in MCF10A, Hs578T and HCC1143 clones. The experiment was performed three times and one representative graph is shown for each cell line. Values are mean ± SD of n ≥ 6 measurements. One-way ANOVA was used (***P < 0.001, ****P < 0.0001). b Effect of RPSAP52 silencing on colony formation ability. Representative plates are shown. Colonies were counted from three replicate plates and two independent experiments. Values are mean ± SD. Two-tailed unpaired t-test were used (**P < 0.01, ***P < 0.001, ****P < 0.0001). c Migration capacity of RPSAP52-depleted clones was monitored over 24 h (n = 5 replicates per condition), with higher cell index indicating higher migration. Values are mean ± SD. A two-tailed Mann-Whitney U test of data at end point was used (**P < 0.01). Inset: migration was also assessed with transwells. Scale bar = 100 µm. d Western blot analysis of NANG0G, OCT4, and SOX2 in RPSAP52-depleted cells. e The clonogenic ability was assessed with at least n = 12 replicates per condition. Values are mean ± SD. A two-tailed Mann-Whitney U test was used (**P < 0.001). F Growth-inhibitory effect of RPSAP52 knockdown in MCF10A mice xenografts. Upper graph: tumor volume (n = 10) was monitored over time. Mean values are shown ±SEM. Lower graph: tumors were excised and weighed at 77 days (**P < 0.001, two-tailed Mann-Whitney U test). Western blot was carried out from sh4 tumors since no material could be recovered from sh1 tumors, and the levels of RAS and IGF2BP2 proteins were analyzed. The photograph shows the relative size of all tumors extracted. Scale bar = 10 mm. Source data are provided as a Source Data file.

Interestingly, LIN28B protein reduction is only partially explained by a decrease in LIN28B mRNA levels (Supplementary Fig. 4f). This discrepancy, together with the interaction detected between RPSAP52 and IGF2BP2, and the presence of RPSAP52 along sucrose gradient’s heavy fractions, which correspond to translating poly-ribosomes (see text above and Supplementary Fig. 2d), prompted us to investigate the possibility that LIN28B levels were regulated by the lncRNA at the translational level.

RPSAP52 modulates IGF2BP2 binding to its mRNA targets. IGF2BP2 is a mRNA stability and translational regulator with some well-described targets, such as IGFBP3, NRAS, or HMGA1. The closely related IGF2BP1 protein has been shown to interact with LIN28B mRNA and increase LIN28B protein levels in ES-2 cells. In order to assay the interaction of IGF2BP2 with LIN28B mRNA in A673 cells, we carried out protein immunoprecipitation followed by RT-qPCR of the pulled-down RNA. We could confirm the interaction of IGF2BP2 with both RPSAP52 isoforms, with IGF2BP2 and NRAS mRNA, and importantly, with LIN28B mRNA. Of note, even though the RPSAP52 isoform lacking the alternative exon is more abundant in A673 cells, both transcripts were recovered in comparable amounts in IGF2BP2 immunoprecipitate, with a ~10-fold higher affinity of IGF2BP2 for RPSAP52 3′ > UTR RNA (Fig. 5a, see RT-qPCR). Further, LIN28B protein was not co-immunoprecipitated with IGF2BP2 protein (Fig. 5b), suggesting its putative regulation by IGF2BP2 is at the level of transcript. We next wanted to test the possibility that this binding is regulated by RPSAP52 presence. Interestingly, whereas binding to IGFR1 and IGF2BP2 mRNAs was not altered, binding of IGF2BP2 to LIN28B mRNA was reduced upon stable knockdown of RPSAP52 (Fig. 5c). This suggests that this lncRNA might regulate LIN28B post-transcriptionally through modulation of IGF2BP2 function. In view of this, we decided to characterize in a transcriptome-wide manner the IGF2BP2-RNA interactions with individual nucleotide resolution (iCLIP-seq) under control or RPSAP52-knockdown conditions. We identified 290,060 and 131,729 iCLIP-tags in control and RPSAP52-depleted A673 cells, corresponding to 3075 and 1639 peak regions, respectively (Supplementary Fig. 5a, b). As has been shown before, IGF2BP2 iCLIP tags were enriched in 3′UTRs, with ~60% of the iCLIP peaks falling within 3′UTRs in control cells. Remarkably, knockdown of RPSAP52 resulted in a specific decrease in the number of 3′UTR peaks revealed by iCLIP and an increase in intronic regions (Fig. 5d). Motif enrichment analysis around the crosslinking-induced truncation sites (CITS) indicate that the previously described CAUH (H = A, C, U) consensus binding site also ranks high in our iCLIP experiments, but is more enriched in the control samples (Supplementary Fig. 5c, d). We found 1775 and 810 peaks with the CAUH motif in the control and depleted sample, respectively, representing a statistically significant difference in occurrence (Fisher’s exact test P-value = 5.286e-08). In addition, the number of peaks with more than one CAUH motif was higher in the control cells (average number of motifs was 1.81 for control and 1.64 for depleted cells; Mann–Whitney U test, P-value = 1.786e-05). The full list of statistically significant iCLIP-seq peaks and CITS can be found in Supplementary Data 1. Differences in motif binding and the reduction in 3′UTR recognition results in a shortlist of 34 transcripts with differential IGF2BP2 iCLIP counts along their 3′UTR (Fig. 5e). GO enrichment analysis shows these genes belong to categories that may relate to IGF2BP2 involvement in cancer invasion and metastasis, including cell-substrate adhesion, spreading, and wound healing, as well as the canonical function for IGF2BP2 pathway, cellular glucose homeostasis (Fig. 5e). Of note, previous CLIP experiments for IGF2BP2 in pluripotent stem cells have revealed that cell adhesion is also the most significant GO category for CLIP-enriched 3′UTRs for IGF2BP1. This suggests that the levels of RPSAP52 have a dramatic impact on IGF2BP2 global role. In addition, top ten GO categories for genes with significant iCLIP peaks present on their 3′UTRs in the control sample correspond to signaling pathways and cell cycle progression, whereas none of these categories are enriched in the RPSAP52-depleted sample (Fig. 5f).

Previous CLIP-seq studies with IGF2BP2 had revealed binding sites on the 3′UTR of LIN28B mRNA in HEK293T cells, and we detected similar sites in our experimental setting and a tendency to decrease upon RPSAP52 depletion, although without any statistical power (Fig. 5g). For other validated IGF2BP2 targets, such as HMGA2, we also detected abundant iCLIP signal corresponding to direct binding of IGF2BP2 to its 3′UTR. In this case, binding is dramatically lost upon RPSAP52 knockdown (Fig. 5h), and a corresponding decrease in HMGA2 protein level is observed (Supplementary Fig. 5e). This is not a general phenomenon for all IGF2BP2 targets, since other well-characterized mRNA partners, such as HMGA1, NRAS, and IGFR1 maintain comparable iCLIP signals in both control and RPSAP52-depleted conditions (Supplementary Fig. 5f). Our results thus suggest a specific loss of IGF2BP2 affinity for particular mRNA targets. Several of the best characterized IGF2BP2 targets are regulated by let-7 (e.g., RAS, HMGA2…) but, interestingly, we could not find a differential presence of let-7 miRNA recognition motifs along the 3′UTRs of the immunoprecipitated mRNAs in control or depleted samples (Fisher’s exact test P-value = 0.8974). Also,
depletion of LIN28B does not impact on IGF2BP2 levels or its binding to mRNA targets, indicating that the regulation of RPSAP52 on IGF2BP2 does not proceed through LIN28B (Supplementary Fig. 5h–j). Both HMGA2 and LIN28B mRNAs are present at high levels upon RPSAP52 depletion (Fig. 4d and Supplementary Fig. 4f), and their half-lives are not substantially altered (Supplementary Fig. 5g), pointing to a decrease in their translation as a consequence of a diminished binding to IGF2BP2.

**RPSAP52 controls IGF2BP2 and mRNA distribution on polysomes.** To obtain direct evidence of the changes in translation efficiency for specific IGF2BP2 targets, we analyzed the

---

**Figure a** shows a scatter plot with regression line and R² value for RPSAP52 expression against HMGA2 expression. The p-value is also provided. **Figure b** displays RT-qPCR results with relative expression values for RPSAP52, HMGA2, and GAPDH. **Figure c** presents RNA immunoprecipitation and Western blotting results, highlighting IGF2BP2, HNRNPQ, RPSAP52, and HMGA2. **Figure d** illustrates relative expression results for A673 cells with let-7a, let-7b, and let-7e. **Figure e** shows tumor volume (mm³) and tumor weight (g) over time for A673 cells with sh1, sh4, scr E4, G8, B9, B11, and sh3 F9.
The detailed mechanism by which IncRNAs may contribute to altering the output of signal transduction pathways is largely
unexplored. Our previous work had shown that the transcribed pseudogene RPSAP52 enhances HMG2 transcription through the formation of an R loop structure.\textsuperscript{23} We have further explored the impact of RPSAP52 expression in cell physiology and propose a mechanism of action that also influences post-transcriptional regulation in the cytoplasm through the interaction with the RNA binding protein IGF2BP2. Regulation of IGF2BP2 expression or function by lncRNAs appears as a common theme in a number of lineage commitment programs, including adipocyte, cardiac or muscle differentiation.\textsuperscript{18–30} However, while other studies have reported lncRNAs that interact with IGF2BP2 and compete for its binding to target mRNAs (e.g., LncMyoD promotes muscle differentiation by outcompeting c-Myc and N-Ras mRNAs for IGF2BP2 binding\textsuperscript{18}), in the cancer setting that we are studying...
reexpression of RPSAP52 facilitates IGF2BP2 binding to a subset of mRNA targets, prominently HMGA2 and LIN28B mRNAs. Our data indicate that this is achieved through modulation of the binding affinity that IGF2BP2 has for particular 3′ UTRs and its distribution in large polyosomes. This is reminiscent of the mechanism of action of HIFIA-A52 in glioblastoma cell lines, where binding of an antisense transcript to IGF2BP2 and DHX9 stimulates expression of their target mRNAs and promotes adaption to hypoxic stress51. Thus, our working model is that by forming ternary complexes (IGF2BP2-RPSAP52-other mRNAs), RPSAP52 may influence the recruitment into ribonucleoprotein particles that dictate mRNA fate, and in particular enhance the translation of mRNAs that would otherwise be repressed by miRNAs. Binding by IGF2BP3 (another member of the IGF2BP family), for instance, has been associated with resistance to miRNA-dependent destabilization for many oncogenes, including HMGA2 and LIN28B2,3. We hereby describe a similar scenario for IGF2BP2, and IGF2BPs are thus emerging as key nodes that integrate lncRNA-mediated post-transcriptional regulation of gene expression and pro-proliferative and self-renewal axis.

An important added layer of regulation exerted by RPSAP52 is the influence on let-7 levels, which may be a consequence of the control on LIN28B translation efficiency, or (in those cells where LIN28B is absent, such as MCF10A), may derive from the altered levels observed in IGF2BP2 protein itself upon depletion of the pseudogene. In fact, in glioblastoma cells lacking LIN28, let-7 targets have been observed to be protected from miRNA-dependent silencing by the binding of IGF2BP2 to let-7 mRNA responsive elements27. This in turn may indirectly cause a decrease of let-7 levels, since miRNA turnover might also depend on binding to mRNA targets, with some previous evidence suggesting that target availability prevents miRNA decay23,34. The greater effect of directly inhibiting biogenesis versus indirectly influencing the turnover might explain why let-7 levels increase moderately in MCF10A upon RPSAP52 depletion (where LIN28B is not expressed and LIN28A is not altered, but IGF2BP2 levels decrease), and, by contrast, increase by almost one order of magnitude more in A673 cells (where expression of LIN28B protein is reduced) (compare Figs. 2c and 4d). Taken as a whole, RPSAP52 is a pseudogene with an important impact on a major tumor suppressor miRNA. While silencing of HMGA2 expression by let-7 has been reported before65, this is the first time that regulation of let-7 levels by transcripts originating from HMGA2 locus is proposed. Importantly, this effect does not proceed through HMGA2 itself, since depletion of HMGA2 expression with gampers actually increases RPSAP52 levels and consequently results in a decrease in let-7 family (Fig. 2e). This adds further complexity to the regulatory network, one hypothesis being that the tumorigenic cell activates an alternative pathway (increase of RPSAP52) to compensate for the loss of HMGA2 function.

Consistent with their convergent roles in the same pathway, low expression of HMGA2/RPSAP52 in differentiated cells and reexpression in cancer mirrors LIN28 levels, which is one of the key players in maintenance of the pluripotent state. Let-7 levels are maintained low in embryonic stem cells and certain primary tumors due to inhibition by LIN28 proteins, which are present at characteristically high levels in undifferentiated cells56,57. Of all tumor suppressor miRNAs, let-7 is the one whose loss is most frequently correlated with poor prognosis in meta-analysis reports38. Accordingly, LIN28A/B high expression is a marker of poor prognosis and more aggressive tumors in a variety of cancers, and their levels have also been associated with metastatic and drug-resistant cases39. Thus, regulation of LIN28B/let-7 balance is one important driver in cancer development.

An important aspect of this LIN28B/let-7 balance is their counteracting action on the stemness characteristics of cancer cells. Interestingly, LIN28B/let-7 signaling has been shown to regulate endogenous Oct4 and Sox2 expression by using ARID3B and HMGA2 as downstream effectors, and thereby regulate stemness properties in oral squamous cancer50. Also, the role of let-7 in antagonizing self-renewal and promoting differentiation has been established via targeting of Myc, Ras, and HMGA2 pathways61,62. In accordance with let-7 anti-pluripotency properties, we observe a decrease in NANOG and OCT4 levels as well as in clonogenicity upon RPSAP52 depletion (Fig. 3d, e), suggesting that RPSAP52 is an enhancer of stem cell characteristics. To date, few lncRNAs have been thoroughly described regarding their involvement in stemness, among them H19 (whose down-regulation reduces NANOG, OCT4, and SOX2 in glioma and breast cancer63) and lncRNA ROR (which inhibits proliferation of glioma stem cells by negatively regulating KLF464). In particular, H19 has recently been proposed to facilitate tumorigenesis through splicing of let-75. The regulatory mechanism used by RPSAP52, by contrast, targets let-7 family through modulation of LIN28B and/or target availability.

Taken together, we have observed that RPSAP52 (1) stimulates proliferative and self-renewal axes together with a reduction of let-7 levels, (2) promotes tumorigenic behavior in vitro and in vivo, and (3) is overexpressed in a number of human cancers and its expression is associated with worse outcome. This, together with its virtual absence in normal differentiated cells and embryonic expression pattern allows us to propose that RPSAP52 is an oncofetal pseudogene that enhances proliferative
and survival programs across several tumor types and whose expression in cancer can have important clinical implications. Indeed, RPSAP52 levels are more useful as biomarkers in sarcoma than HMGA2 mRNA levels, which do not seem to correlate well with protein levels (as suggested by our work and others66), probably due to the complex post-transcriptional regulation of HMGA2. The potential use of this pseudogene as an effective therapeutic target in human cancer will thus be the focus of future studies.

**Methods**

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

Fig. 6 IGF2BP2 is redistributed on polysome gradients upon RPSAP52 depletion. a–d Polysome profiles of A673 cells stably depleted of RPSAP52 (shRPSAP52, corresponding to sh4 B11 clone) or control cells (scr). HMGA2 mRNA (a), LIN28B mRNA (b), or NRAS mRNA (c) distribution across the gradient was evaluated in each fraction by RT-qPCR. For comparison, GAPDH mRNA distribution was also assessed (d). Graphs represent the mean ±SD of three replicates. The red and blue lines indicate absorbance at 260 nm for each fraction in control or depleted cells, respectively. e Total RNA from the same cells (n = 3) was analyzed by RT-qPCR. Mean values are shown ±SD. A two-tailed student t-test was used (*P < 0.05, **P < 0.01). f Protein extracted from the 20% of the polysome profile fractions shown in (a–d) were subjected to dot blot analysis with an anti-IGF2BP2 antibody (middle panel) or with anti-RPL5 antibody as control (lower panel). Proteins from 10% of fractions 1 and 2 were loaded together. Membranes were previously stained with Ponceau S (top panel) for loading control. Source data are provided as a Source Data file.
Bisulfite genomic sequencing. The Methyl Primer Express v1.0 software (Applied Biosystems) was used to design specific primers for the methylation analysis of HMG2A/RPSAP52 island (Supplementary Table 1). Genomic DNA (1 µg) was subjected to sodium bisulfite treatment using the EZ DNA Methylation-Gold kit (Zymo Research). For bisulfite genomic sequencing, 300–500 bp fragments were amplified using 1–2 µl of bisulfite-converted DNA with Immolase Taq polymerase (Bioline) for 42 cycles. The resulting PCR products were gel-purified with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) and then cloned into the pSpark® TA vector (Canvas) according to the manufacturer’s protocol. For all samples, 10 colonies were randomly chosen, the DNA was purified using NucleoSpin® 96 Plasmid (Macherey-Nagel) and sequenced by the 3730 DNA Analyzer (Applied Biosystems). After sequencing analysis with BioEdit v7.2.5 software, C nucleotides that remained unaltered were transformed into percentages of CpGs showing methylation.

Western blotting. Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8, 5 mM EDTA, 350 mM NaCl, 0.5% NP40, 10% glycerol, 0.1% SDS and phosphatase inhibitors), sonicated and centrifuged to recover the supernatant. The concentration

Overall survival
Years since diagnosis
Log rank; p = 0.027
HR (95% CI) = 0.46 (0.23–0.93); p = 0.031

Cell proliferation and metastasis

Gene expression

Western blotting. Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8, 5 mM EDTA, 350 mM NaCl, 0.5% NP40, 10% glycerol, 0.1% SDS and phosphatase inhibitors), sonicated and centrifuged to recover the supernatant. The concentration
was determined with the Pierce BCA Protein Assay Kit (#23227, ThermoFisher).
Proteins were boiled for 5 min with Laemmli buffer (2% SDS, 10% glycerol, 60 mM Tris-Cl pH 6.8, 0.01% bromphenol blue) plus 2% 2-mercaptoethanol as a loading buffer, and equal amounts of extracts were loaded onto Tris-Glycine-SDS gels. Proteins were transferred to a nitrocellulose membrane (Whatman, GE Healthcare) and incubated overnight at 4°C with primary antibodies diluted in 5% skimmed milk in PBS containing 0.1% Tween-20. The detected proteins were IGF2BP2 (H00010644-m1, Abcam, 1:4000), LIN28A (#8641, Cell Signaling, 1:750), IGF1R (#3027, Cell Signaling, 1:1000), ERK (a3854, Sigma, 1:20,000), NANOG (#4903, Cell Signaling, 1:2000), OCT4 (#2750, Cell Signaling, 1:1000), GAPDH and β-TUBULIN HRP (ab40742, Abcam, 1:5000). Poly(A) mRNA Purification Kit (#AM1921, Life Technologies). Equal amounts of RNA from each fraction were subject to RT-qPCR and the results were normalized taking into account the total quantity of RNA recovered from each fraction. To verify the nuclear and cytoplasmic fractionation of the mRNA, GAPDH and HMGA2 were analyzed on a Ion Trap Amazon Speed ETD (Bruker Daltonics, Bremen, Germany) with 10 mM DTT and alkylated with 35 mM iodoacetamide. Extracted peptides were digested with trypsin overnight. Bands were then washed with water, 50 mM PBS containing 0.1% Tween-20. The detected proteins were IGF2BP2 (H00010644-m1, Abcam, 1:4000), LIN28A (#8641, Cell Signaling, 1:750), IGF1R (#3027, Cell Signaling, 1:1000), ERK (#6495, Cell Signaling, 1:1000), p-ERK (#9101, Cell Signaling, 1:1000), LAMIN B1 (ab16048, Cellab, 1:1000), LIN28A (#6841, Cell Signaling, 1:750), α-TUBULIN HRP (ab40742, Abcam, 1:1000), HNRNPAQ (ab84946, Abcam, 1:1000), β-ACTIN HRP (a2854, Sigma, 1:2000), NAOAB #4903, Cell Signaling, 1:1000), OCT4 (#2750, Cell Signaling, 1:1000), SOX2 (#4195, Cell Signaling, 1:1000), NUCLEOLIN (#6031, Santa Cruz, 1:2000), HMGA2 (ab72726, Abcam, 1:1000), HISTONE H3 (ab7191, Abcam, 1:5000), RPSA (ab133645, Abcam, 1:1000), and RPL5 (A303-933A, Company Bethyl, 1:1000). After three washes with PBS containing 0.1% Tween-20, membranes were incubated for 1 h at room temperature in a bench-top shaker with the secondary antibodies conjugated to horseradish peroxidase anti-rabbit IgG (A0545, Sigma, 1:10,000) or anti-mouse IgG (NA9310, GE Healthcare, 1:5000). ECL reagents (Luminata-HRP; Merck-Millipore) were used to visualize the proteins.

RNA-biotin pull-down. Full-length RPSAP52 (including alternative exon) or truncated fragments, as well as the anti-sense version or the sequence corresponding to the unrelated Uc.160 + RNA, were biotin-labeled by standard in vitro transcription reactions and gel-purified. DNA templates for transcription were prepared by PCR with oligos described in Supplemental Table 1. The pull-downs were carried out with 10 pmol of each biotinylated RNA and 1 mg of total MCF10A or A673 protein extracts. Following incubation with the extract, each RNA was retrieved with 25 μl of Dynabeads® M-280 Streptavidin beads (#65305, Invitrogen) and washed in RIP buffer (150 mM KCl, 25 mM Hepes at pH 7.9, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 1 x protease inhibitor cocktail (Roche)). Binding proteins were released by boiling in SDS loading buffer and samples were run on a 4–12% gradient pre-cast Bis-Tris protein gels (Invitrogen) in MOPS buffer. After electrophoresis, the gels were either stained with SYPRO Ruby (Invitrogen) for band visualization and MS analysis or transferred to nitrocellulose membranes for western blotting.

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

Revers e pull-down. Endogenous IGF2BP2 was immunoprecipitated from A673 cell extracts. One milligram of total protein was incubated overnight with 2 μg of anti-IGF2BP2 polyclonal antibody (#H00010644-m1, Abnova) or control mouse IgG antibody (#127-371, Millipore) and 40 μl of Dynabeads® M-280 anti-mouse IgG beads (#1120D, ThermoFisher) in 1 ml of RIP buffer (150 mM KCl, 25 mM Hepes at pH 7.9, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 1 x protease inhibitor cocktail (Roche)). Beads were then washed three times with RIP buffer and 10% of the volume was boiled in the presence of Laemmli buffer for western blot analysis. The purified IgG complexes from the remaining 90% of beads was extracted by adding 1 ml of TRIzol® Reagent (15996-018, ThermoFisher). After phenol extraction and isopropanol precipitation, the final pellet was resuspended in 10 μl of H2O and re-rotor chromatized. cDNA was analyzed by 50 cycles of semi-quantitative RT-PCR or by RT-qPCR (Applied Biosystems 7900HT Fast Real-Time PCR System). Two micrograms of input RNA was processed in parallel to estimate pull-down efficiency.

Cell culture. MCF7, MCF10A, HCC1143, and A673 cell lines were purchased from ATCC. The remaining breast and sarcoma cell lines were obtained from Dr. Esteller and Dr. Tirado’s labs, respectively. Authenticity of the cell lines was routinely confirmed by STR profiling analysis done at qGenomics SL (Esplugues de Llobregat, Barcelona, Spain). All cell lines were routinely checked for mycoplasma contamination. Non-malignant MCF10A breast cells were grown in DMEM/Ham’s F-12 medium (#L0093-900, Biowest) supplemented with 20 mg ml−1 EGFR (#SRP3072, Sigma), 500 ng ml−1 hydrocortisone (#H8888, Sigma), and 10 μg ml−1 insulin (#I9278, Sigma). Ewing’s sarcoma A673 cells and breast cancer HCC1143 cell line were grown in RPMI-1640 medium with Glutamax (#I9278-010, Gibco). Human-derived lung cancer cell line H578T were cultured in DMEM with Glutamax (#31966-021, Gibco). All the media were supplemented with 10% fetal bovine serum (FBS) (#10270, Gibco), and the cells were grown at 37°C in a humidified atmosphere of 5% CO2 and 95% air.

Plasmid construction and transfections. Stable knockdown of RPSAP52 was achieved with the following sequences shRNA and shRNA target, respectively, 5′-TCTCTAAGCTCTTGCCAGT3′ and 5′-CAAGGTACCTTAAAGCAACA3′ sequences of RPSAP52 mRNA (both located on the last exon), whereas shRNA targets the 5′-GGCAAGACTGAGGACTA3′ sequence of RPSAP52 mRNA (both located on the last exon), whereas shRNA targets the 5′-GGCAAGACTGAGGACTA3′ sequence of RPSAP52 mRNA (both located on the last exon). These shRNAs were expressed by cloning oligos shRPSAP52-1 or shRPSAP52-2 into shRNA1, shRPSAP52-3 or shRPSAP52-4 into shRNA4, the RPSAP52-5 for shRPSAP52-3 into the BamHI and EcoRI sites of the vector pLVX-shRNA2 (Clontech). A scramble (scr) sequence was used as a control. For lentivirus-mediated depletion, HEK293T cells were transfected with pLVX-shRNA2-construction plus packaging plasmids with jetPRIME® (Polyplus-transfection) according to the manufacturer’s recommendations. The target cell line was infected with the supernatant containing viral particles 48 h post-transfection. ZGreen was used as a marker to visualize transductants by fluorescence microscopy, and these cells were selected by fluorescence-activated cell sorting (FACS) and plated to obtain stable clones. Antisense oligonucleotides (LNA® GapmeRs, #300600, Exiqon) target- ing HMGA2 mRNA (HMGA2, exon1 (RPSAP52 Ex) or the first intron (RPSAP52 RL1 and RL2) of RPSAP52 transcript were transcribed to a final concentration of 65 nM using HiPerfect (Qiagen). Cells were transfected 48 h later and collected 72 h after the second round of LNA treatment. A control LNA GapmeR (#300610, Exiqon) was used as mock transfection. For siRNA-mediated knockdown of LIN28B, cells were transfected with a 1:1 mix of two different siRNAs against LIN28B (#216387-216388, Ambion) and a negative control (C- ) (#AM4611, Ambion), using LipofectamineTM RNAiMAX Transfection Reagent.
RNA isolation and RT-qPCR analysis. Total RNA, including miRNAs, was extracted using the Maxwell® RSC instrument with the Maxwell® RSC miRNA Tissue kit (Promega) according to the manufacturer’s recommendations. For miRNA expression analysis, total RNA was reverse transcribed using the SuperScript® III Reverse Transcriptase (180800, Invitrogen). Real-time PCR reactions were performed in triplicate in an Applied Biosystems 7900HT Fast Real-Time PCR system, using 30–100 ng cDNA, 6 µl SYBR® Green Master Mix (Applied Biosystems), and 416 nM primers in a final volume of 12 µl for 38-well plates. All data were acquired and analyzed with the QuantStudio Design & Analysis Software v1.3.1 and normalized with respect to data were acquired and analyzed with the SuperScript® III Reverse Transcriptase (#180800, Invitrogen). Real-time PCR reactions were performed in triplicate in an Applied Biosystems 7900HT Fast Real-Time PCR system, using 30–100 ng cDNA, 6 µl SYBR® Green Master Mix (Applied Biosystems), and 416 nM primers in a final volume of 12 µl for 38-well plates. All data were acquired and analyzed with the QuantStudio Design & Analysis Software v1.3.1 and normalized with respect to the Universal cDNA Synthesis Kit II (#203301) for the RNA retrotranscription and the ExLent SYBR® Green master mix (203421) for the RT-qPCR in the LightCycler® 480 (Roche) with the LightCycler® 480 Software v1.5.0 SP4. To normalize the data, RNU6 and mlr-195 were used as endogenous control.

Actinomycin D treatment and RNA stability analysis. Control or RPSAP25-depleted A673 clones were treated with either 0.5% DMSO or 5 µg ml⁻¹ Actinomycin D (Sigma) for 9 h. Pellets of each condition and treatment were harvested at different times and RNA was extracted for the RT-qPCR experiments. All data were normalized with respect to GUSB as endogenous control and gene expression fold-changes induced by Actinomycin D were calculated relative to the control (DMSO) cells of each condition and time point. ε-FOS and GAPDH were used as controls of the experiment due to their short and long half-life, respectively.

SRR assay. Cell viability and proliferation were determined by the sulforhodamine B (SRR) assay. Cells were seeded in 96-well microplates in medium with 10% FBS, and the experiment started after 24 h of incubation at 37 °C and 5% CO₂. The optimal cell number (100 cells per well for MCF10A and 2000 cells per well for A673) was extracted for the RT-qPCR experiments. All data were normalized with respect to GUSB as endogenous control and gene expression fold-changes induced by Actinomycin D were calculated relative to the control (DMSO) cells of each condition and time point. ε-FOS and GAPDH were used as controls of the experiment due to their short and long half-life, respectively.

Colonogenicity assay. Transwell migration assay. The xCELLigence Real-Time system (ACEA Biosciences) was used with CIM-16 plates of 8 µm pore membranes. The lower chamber wells were filled with 160 µl of medium containing 10% FBS and the top chamber wells with 40 µl of serum-free medium. The two chambers were assembled to form two 100 mm wells separated by a 200 µm filter. Colonies were incubated for 24 h in serum-free medium, rinsed with PBS, trypsinized, and resuspended in medium supplemented with 10% FBS to inactivate the trypsin, followed by centrifugation and resuspension in serum-free medium. A total of 8 × 10⁵ cells were seeded onto the top chamber of CIM-16 plates and placed into the xCELLigence system for data collection after background measurement. The software RTCA 2.0 was set to collect impedance data every 15 min. The cell index represents the capacity for cell invasion, whereas the slope of the curve can be related to the cell invasion ability.

Transwell migration assay. Transwell Permeable Supports (3422, Cultek) with 8 µm pore polycarbonate membranes in 24-well plates were used to measure cell migration through 8 µm pore polycarbonate membranes in 24-well plates were used to measure cell migration through the transfected cells. The top well contained serum-free medium, rinsed with PBS, trypsinized, and resuspended in 10% FBS-containing medium to inactivate the trypsin, followed by centrifugation and resuspension in serum-free medium. A total of 1 × 10⁶ cells were seeded onto each transwell with 150 µl of serum-free medium and the transwells were placed in the wells of a 24-well plate with 500 µl of 10% FBS-containing medium. The chemotransactant promoted the migration of the cells from the upper part of the transwell to the lower part. After 24 h of incubation at 37 °C and 5% CO₂, the cells in the upper part of the membrane were removed with a cotton swab and several washes with 1× PBS. Cells in the lower part were fixed for 10 min with ice-cold 100% methanol. For the staining, cells were covered with 0.5% crystal violet in 25% methanol for 10 min. Transwells were washed several times with 1× PBS and air-dried. Membranes were then mounted on a slide for image acquisition.

Clonogenicity assay. The clonogenicity of RPSAP25-depleted MCF10A clones was tested in soft agar by using the CytoSelect 96-well Cell Transformation Assay Kit (Cell Biolabs, #CBA-130), following the manufacturer’s instructions. Briefly, a base agar layer was prepared by mixing equal volumes of 1.2% agar solution and 2 × DMEM/20% FBS medium in each well of a 96-well flat-bottom microplate. In total, 5000 wells were prepared in a top layer by mixing equal volumes of the cell suspension, 12% agar solution and 2 × DMEM/20% FBS (1:1:1), and incubated for 6 days after covering the solidified cell agar layer with 100 µl of DMEM-F12 medium plus supplements. The CyQuant GR dye was used to detect the lysed colonies and the data were normalized relative to the total number of colonies on the plates. Each experiment was performed in parallel with known amounts of the following recombinant proteins: LIN28B (ab134596, Abcam), IGF2BP2 (ab153107, Abcam), HNRNPQ (ab153089, Abcam), and GFRA1 (ab153089, Abcam). Western blot was performed with the following antibodies: anti-LIN28B (ab71415, Abcam, 1:1000), anti-IGF2BP2 (ab153107, Abcam), and anti-HNRNPQ (NB21009, Novus Biologicals, 1:1000). Band intensity was measured by densitometry with an iBright™ CL.1000 Imaging System (ThermoFisher).

ICLIP-seq. ICLIP-seq was performed on stable A673 clones. Approximately 8 × 10⁵ A673 cells stably expressing scrambled shRNA (scr) or shRNA-4 against RPSAP25 (clone B11) were crosslinked with 150 mJ cm⁻² total 254-nm irradiation in a Stratagene Stratalinker 2400. The same amount of non-crosslinked cells were used as controls. Cell lysates were treated with different concentrations (2 or 0.4 U ml⁻¹) of RNase R.
ICLIP computational analysis. Read quality was assessed using FastQC (v0.11.7) software (available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc). After sequencing, PhiX sequences were removed using BWA (Burrows-Wheeler Aligner, v0.7.17)67. All the pre-processing steps, peak calling, CITs calling, and annotations were performed using CTK (CLIP Tool Kit) (v1.0.9) software68, following the recommendations found in https://zhanglab.hcbi.columbia.edu/index.php/Plasmids/Plasmid_1418. As recommended by Miseq Reporter software (v3.2), alignments were subjected to stringent read filtering. Two independent experiments were performed for each condition. The four libraries were sequenced in two separated pools (#31 & #38) and data acquired with Miseq Reporter v2.6.3.2. Raw data can be downloaded from https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA468688.

Polysome profile analysis. MCF10A cells or A673 stable cells were plated in 130 mm dishes and treated with 100 µg ml⁻¹ cycloheximide (CHX) at 37 °C for 5 min. Cells were then washed with PBS supplemented with 10% FBS, pelleted and resuspended in 250 µl of hypotonic lysis buffer (1.5 mM KCl, 2.5 mM MgCl₂, 5 mM Tris-HCl pH 7.4, 1 mM DTT, 1% sodium deoxycholate, 1% Triton X-100, 100 µg ml⁻¹ CHX) supplemented with mammalian protease inhibitors (SIGMA) and RNase inhibitor (NEB) at a concentration of 100 U ml⁻¹, and left on ice for 5 min. Cell lysates were cleared of debris and nuclei by centrifugation for 5 min at 17,000 × g. Protein concentrations were determined by BCA assay and 500 µg of lysate were loaded on 10–50% sucrose linear gradients containing 80 mM NaCl, 5 mM MgCl₂, 20 mM Tris-HCl pH 7.4, 1 mM DTT, 10 µl ml⁻¹ RNase inhibitor buffered with 50 mM Tris HCl (pH 7.5) and 1% CHX. Gradients were centrifuged at 17,000 × g for 3.5 h at 217,290 × g. Gradients were analyzed on a BIOCOP gradient station and collected in 11 (MCF10A) or 13 (A673) fractions ranging from light to heavy sucrose. Fractions were supplemented with SDS at a final concentration of 1% and placed for 10 min at 65 °C. To each fraction was added 1 mg of firefly luciferase mRNA, followed by phenol–chloroform extraction and precipitation with isopropanol. Purified RNAs from each fraction were reconstituted and subjected to qPCRx. mRNA quantification was normalized to firefly mRNA.

IGFBP2 communoprecipitation and mass spectrometry analysis. Immunoprecipitation and sample digestion for mass spectrometry analysis: 1 mg of pre-cleaved protein extract from three replicates of control cells (scr) and cells depleted for RPSA5/2 (shl B11 clone) were immunoprecipitated overnight at 4 °C using 5 µl BSA-free Protein G beads (ThermoFisher, #704704-6901). After washing the beads twice with pre-scale buffer and incubated in 20 µl 1xNuPAGE buffer for 5 min at 37 °C prior to loading the supernatant on a 4–12% NuPAGE Bis-Tris gel (#NP0341BOX, ThermoFisher). The gel was run at 180 V for 50 min, and the protein–RNA complexes were transferred to a nitrocellulose membrane at 30 V for 1 h. The membrane was then autoradiographed through exposure to a film at –80 °C for 1 h. Regions of interest containing the IGF2BP2–RNA crosslinked products were cut out of the membrane and the RNA fragments were eluted in 1 µl of 100 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% SDS, 0.5% NaCl, 0.5% Triton X-100 and incubated with RNase-free Dounce homogenizer for 10 min at room temperature. The RNA was collected by centrifugation at 14,000 × g for 2 min at 4 °C. Both the supernatant and pellet were used for next step. Total RNA was extracted from each lane using RNeasy Mini kit (Qiagen) and 1 µg of total RNA was amplified using the BioRad iAMP iCLIP Kit. Enriched RNAs were puriﬁed and sequenced on an Illumina HiSeq X Ten platform at 2×100 bp. The resulting FASTQ reads were submitted to cBioPortal database (http://www.cbioportal.org/).

Primary tumors expression and methylation analysis. RNA expression and DNA methylation data from different tumor types was collected from The Cancer Genome Atlas (TCGA) for the breast (BRCA), lung (LUAD, LUSC), and colorectal (CRC) cancer types. The expression data were downloaded from the GEO database using TCGAbiolinks v2.9.273. Bioconductor package from the current GDC (Genomic Data Communities) harmonized database aligned against hg38 genome. Box plots analysis represent normalized expression and methylation values corresponding to TCGA COAD, LUSC, LUAD, THCA, and BRCA projects from the GDC data portal (https://portal.gdc.cancer.gov/). We use the Wilcoxon signed-rank test to compare differences between the groups. The association between chromatin marks and promoter activity was evaluated with a Pearson’s correlation. All the statistical analyses and graphical representations were performed using R v3.4.3. For primary samples, an average of HMGAC2 and RPSAP2 promoter activity was calculated (±2σ). For primary samples, the expression was considered as hypermethylated. NCIGCZ cell lines expression data were downloaded from cbioPortal database (http://www.cbioportal.org/).

http://www.biocompare.com/
the slides, the arrays were scanned by the Agilent Scanner G2505C. Agilent Feature Extraction software (version 11.0.1.1) was used to analyze acquired array images. Quantile normalization and subsequent data processing were performed using the GeneSpring GX v12.1.2 software (Agilent Technologies). Differentially expressed genes were identified through Fold Change filtering and Volcano filtering. Pathway analysis and GO Analysis were applied to determine the roles of these differentially expressed genes.

**Statistical analysis.** Bar graphics and statistical comparisons were obtained with the GraphPad Prism 8.1.2.2 software. Comparative analyses between different experimental groups were performed using t-student test and one-way ANOVA with Bonferroni’s or Dunnet’s post hoc tests for intergroup comparisons. For cancer patients’ samples, we used the Kaplan–Meier method for survival analysis and the log-rank test was used to analyze the differences between the groups. Cox regression method was used to analyze the independent prognostic importance of expression or methylation. Results of the univariate Cox regression analysis are represented by the hazards ratio (HR) and 95% confidence interval (CI). Results were considered significant if the P-value was <0.05 (*), <0.01 (**), <0.001 (***) or <0.0001 (****). Unless otherwise stated, data are presented as the mean ±SD.

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

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

Received: 29 August 2018 Accepted: 8 August 2019

Published online: 04 September 2019

**References**

1. Djebari, S. et al. Landscape of transcription in human cells. Nature 489, 101–108 (2012).
2. Engert, J. M., Ollikainen, N. & Gutmann, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat. Rev. Mol. Cell Biol. 17, 756–770 (2016).
3. Délás, M. J. & Hannon, G. J. LncRNAs in development and disease: from functions to mechanisms. Open Biol. 7, 170121 (2017).
4. Di Gesualdo, F., Capaccioli, S. & Lulli, M. A pathophysiological view of the long non-coding RNA world. Oncotarget 5, 10976–10996 (2014).
5. Holdt, L. M. et al. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene expression. Front. Cell. Biol. 3, 330 (2014).
6. Ying, L. et al. Downregulated MEG3 activates autophagy and increases cell proliferation in bladder cancer. Cell. Biol. 210, 2106–2123 (2018).
7. Poliseno, L. et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465, 1033–1038 (2010).
8. Degrawe, N., Suvà, M. L., Janiszewska, M., Riggi, N. & Stamenkovic, I. IMPs: an RNA-binding protein family that provides a link between stem cell maintenance in normal development and cancer. Genes Dev. 30, 2459–2474 (2016).
9. Yisraeli, J. K. VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. Biol. Cell. 97, 87–96 (2005).
10. Huarte, M. et al. A large intergenic noncoding RNA induced by p53 mediates transcriptional responses to DNA damage. Science 320, 910–915 (2008).
11. Li, Z. et al. Oncogenic NRAS, required for the pathogenesis of embryonic rhabdomyosarcoma, relies upon the HMGA2-IGF2BP2 pathway. Cancer Res. 73, 3041–3050 (2013).
12. Boque-Sastre, R. et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Prog. Natl Acad. Sci. USA 112, 5785–5790 (2015).
13. Iyer, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. 47, 199–208 (2015).
14. Wu, J. et al. Elevated HMG2a expression is associated with cancer aggressiveness and predicts poor outcome in breast cancer. Cancer Lett. 376, 292–296 (2016).
15. Ciechanowicz, P. et al. Anti-sense LncRNA CRABP1 provides high malignancy of transformed Ewing sarcoma. J. Cancer 6, 1517–1527 (2015).
16. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-cRNA, miRNA-nrRNA and protein-RNA interaction networks from large-scale CLIP-seq data. Nucleic Acids Res. 42, D92–D97 (2014).
17. Wu, Q. et al. Analysis of the mRNA-mRNA-IncRNA network in ER+ and ER− breast cancer cell lines. J. Cell. Mol. Med. 19, 2874–2887 (2015).
18. Arun, K. et al. Comprehensive analysis of aberrantly expressed lncRNAs and construction of ceRNA network in gastric cancer. Oncotarget 9, 18386–18399 (2018).
19. Johnsson, P., Morris, K. V. & Grander, D. Pseudogenes: a novel source of trans-acting antisense RNAs. Methods Mol. Biol. 1167, 213–226 (2014).
20. Poliseno, L. et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465, 1033–1038 (2010).
21. Xu, Y. T., Yao, Y. W. & Lin, J. Intronic regulatory element in cooperation with nuclear factor-kappaB. Nat. Cell. Biol. 19, 843–848 (2007).
22. Li, Z. et al. Oncogenic NRAS, required for the pathogenesis of embryonic rhabdomyosarcoma, relies upon the HMGA2-IGF2BP2 pathway. Cancer Res. 73, 3041–3050 (2013).
23. Busch, B. et al. The oncogenic triangle of HMG2a, LIN28B and IGF2BP1 antagonizes tumor-suppressive actions of the let-7 family. Nucleic Acids Res. 44, 3845–3864 (2016).
24. Degrange, N. et al. The RNA binding protein IMP2 preserves glioblastoma stem cells by preventing let-7 target gene silencing. Cell Rep. 15, 1634–1647 (2016).
25. Kim, J. H. et al. The human protein Pan2: an RNA-binding protein family that provides a link between stem cell maintenance in normal development and cancer. Genes Dev. 30, 2459–2474 (2016).
26. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–141 (2010).
27. Djebari, S. et al. Landscape of transcription in human cells. Nature 489, 101–108 (2012).
28. Guttman, M., Garber, M., Russmann, J., Odom, D. T. & Lander, E. S. A human-nature connects spatial maps of gene expression with regulatory anatomy. Cell 171, 1262–1270 (2017).
29. Ying, L. et al. Downregulated MEG3 activates autophagy and increases cell proliferation in bladder cancer. Mol. Biosyst. 9, 407–411 (2013).
30. Panzitt, K. et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology 132, 330–342 (2007).
31. Arun, G. et al. Differentiation of mammary tumors and reduction in metastasis upon Malat1 IncRNA loss. Genes Dev. 30, 34–51 (2016).
32. Gupta, R. A. et al. Long non-coding RNA HOXATIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071–1076 (2010).
33. Fu, X., Ravindranath, L., Tran, N., Petrovics, G. & Srivastava, S. Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PGEM1. DNA Cell Biol. 25, 135–141 (2006).
34. Venteicher, A. S. et al. Human telomerase holoenzyme protein required for telomere body localization and telomere syntheses. Science 323, 664–648 (2009).
35. Tu, Y. T., Yao, Y. W. & Lin, J. Intronic transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. Proc. Natl Acad. Sci. USA 111, 3377–3382 (2014).
36. Achour, C. & Aguilo, F. Long non-coding RNA and Polycomb: an intricate partnership in cancer biology. Front. Biosci. 23, 2106–2123 (2018).
37. Kepp, F. & Mendell, J. T. Functional classification and experimental dissection of long non-coding RNAs. Cell 172, 393–407 (2018).
45. Emori, M. et al. High expression of CD109 antigen regulates the phenotype of cancer stem-like cells/cancer-initiating cells in the novel epithelioid sarcoma cell line ESX and is related to poor prognosis of soft tissue sarcoma. PLoS One 8, e84187 (2013).

46. Fukushima, A. et al. Targeting PTPRZ inhibits stem cell-like properties and tumorigenicity in glioblastoma cells. Sci. Rep. 7, 5609 (2017).

47. Zeleniak, A. E., Huang, W., Brinkman, M. K., Fishel, M. L. & Hill, R. Loss of MTSS1 results in increased metastatic potential in pancreatic cancer. Oncotarget 8, 16473–16487 (2017).

48. Agarwal, E. et al. Role of Akt2 in regulation of metastasis suppressor 1 expression and colorectal cancer metastasis. Oncogene 36, 3104–3118 (2017).

49. Zhang, X. et al. Interrogation of nonconserved human adipose lincRNAs identifies a regulatory role of linc-ADAL in adipocyte metabolism. Sci. Transl. Med. 10, eaat5878 (2018).

50. Hosen, M. R. et al. Airn regulates Igf2bp2 translation in cardiomyocytes. Circ. Res. 122, 1347–1353 (2018).

51. Mineo, M. et al. The long non-coding RNA HIFIA-AS2 facilitates the maintenance of mesenchymal glioblastoma stem-like cells in hypoxic niches. Cell Rep. 15, 2500–2509 (2016).

52. Jensen, L. et al. IM3 RNPs safe houses prevent miRNA-directed HMGA2 mRNA decay in cancer and development. Cell Rep. 7, 539–551 (2016).

53. Chatterjee, S., Fasler, M., Büssing, I. & Grosshans, H. Target-mediated protection of endogenous microRNAs in C.elegans. Dev. Cell. 20, 388–396 (2011).

54. Pitchiaya, S., Heinicke, L. A., Park, J. I., Cameron, E. L. & Walter, N. G. Resolution subcellular miRNA trafficking and turnover at single-molecule resolution. Cell Rep. 19, 630–642 (2017).

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

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

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

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

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

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

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

62. Damianakis, A. I. et al. MicroRNAs let-7 expression in thyroid cancer: correlation with their depicted targets HMGA2 and SLC5A5. J. Cancer Res. Clin. Oncol. 142, 1213–1220 (2016).

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

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

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

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

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

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

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

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

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

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

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