nature cell biology

A regulated *PNUTS* mRNA to lncRNA splice switch mediates EMT and tumour progression

Simon Grelet[1], Laura A. Link[1], Breege Howley[1], Clémence Obellianne[1], Viswanathan Palanisamy[2,3], Vamsi K. Gangaraju[1,3], J. Alan Diehl[1,3] and Philip H. Howe[1,3,4]

The contribution of lncRNAs to tumour progression and the regulatory mechanisms driving their expression are areas of intense investigation. Here, we characterize the binding of heterogeneous nuclear ribonucleoprotein E1 (hnRNP E1) to a nucleic acid structural element located in exon 12 of *PNUTS* (also known as PPP1R10) pre-RNA that regulates its alternative splicing. HnRNP E1 release from this structural element, following its silencing, nucleocytoplasmic translocation or in response to TGFβ, allows alternative splicing and generates a non-coding isoform of *PNUTS*. Functionally the lncRNA-*PNUTS* serves as a competitive sponge for miR-205 during epithelial–mesenchymal transition (EMT). In mesenchymal breast tumour cells and in breast tumour samples, the expression of lncRNA-*PNUTS* is elevated and correlates with levels of ZEB mRNAs. Thus, *PNUTS* is a bifunctional RNA encoding both *PNUTS* mRNA and lncRNA-*PNUTS*, each eliciting distinct biological functions. While *PNUTS* mRNA is ubiquitously expressed, lncRNA-*PNUTS* appears to be tightly regulated dependent on the status of hnRNP E1 and tumour context.

Breast cancer in females and lung cancer in males are the most frequently diagnosed cancers and the leading cause of cancer death worldwide[1]. Although metastasis is the overwhelming cause of mortality in patients with solid tumours, the molecular and cellular mechanisms that drive tumour cells to become metastatic remain largely unknown[2–4].

Non-coding RNAs have recently emerged as key mediators of tumour progression through their regulation of both oncogenic and tumour-suppressive pathways[5,6]. Long non-coding RNAs (lncRNAs) have been implicated in cellular processes such as proliferation, apoptosis, migration and cell invasion and their dysregulated expression has been observed in various human cancers[7,8]. Despite these recent findings, the regulatory role of lncRNAs in mediating these cellular processes and in cancer development remains an area of active investigation and the subject of controversy[9,10].

Epithelial–mesenchymal transition (EMT) is a developmental process aberrantly reactivated during tumour progression of epithelial cells and contributes to resistance of both conventional and targeted therapies[11,12]. We have previously demonstrated that post-transcriptional regulation of gene expression plays an important role in EMT, especially during TGFβ-mediated EMT[13,14]. We described a transcript-selective translational regulatory pathway involving the binding of hnRNP E1 protein to a BAT structural element (for TGF-beta-activated translational element) located in the 3′-UTR of transcripts involved in EMT-related tumour progression[13,15]. In addition, hnRNP E1 protein was previously described to regulate other critical cellular processes such as transcription, messenger RNA stability, transport and splicing[16].

Alternative splicing regulates over 90% of multi-exon protein-coding genes in humans[17] and hnRNP E1 is well documented for its repressive role in this process. HnRNP E1 represses tumour cell invasion by inhibiting the alternative splicing of CD44 (ref. 18) and binds to the growth hormone receptor pseudoexon to prevent its usage, thus allowing expression of a functional protein[19].

Here we report the binding of hnRNP E1 to an alternative splicing site in the pre-RNA of *PNUTS* to control the generation of an alternative spliced isoform of *PNUTS* that we describe as a lncRNA involved in EMT-related tumour progression. The study reveals that the *PNUTS* pre-RNA transcript serves as a bifunctional RNA capable of generating *PNUTS* mRNA or lncRNA-*PNUTS* in an hnRNP-E1-dependent and cell-context-dependent manner.

RESULTS

The predicted lncRNA-*PNUTS* is upregulated following hnRNP E1 loss and during tumour cell progression

Using an hnRNP E1 knockdown-induced EMT model in NMuMG cells (Supplementary Fig. 1a), we performed an Affymetrix array analysis and identified *PNUTS* pre-RNA as downregulated following

[1]Department of Biochemistry, Medical University of South Carolina, Charleston, South Carolina 29425, USA. [2]Department of Oral Health Sciences, Medical University of South Carolina, Charleston, South Carolina 29425, USA. [3]Hollings Cancer Center, Medical University of South Carolina, Charleston, South Carolina 29425, USA. [4]Correspondence should be addressed to P.H.H. (e-mail: howep@musc.edu)

Received 1 March 2017; accepted 19 July 2017; published online 21 August 2017; corrected after print 13 November 2017; DOI: 10.1038/ncb3595

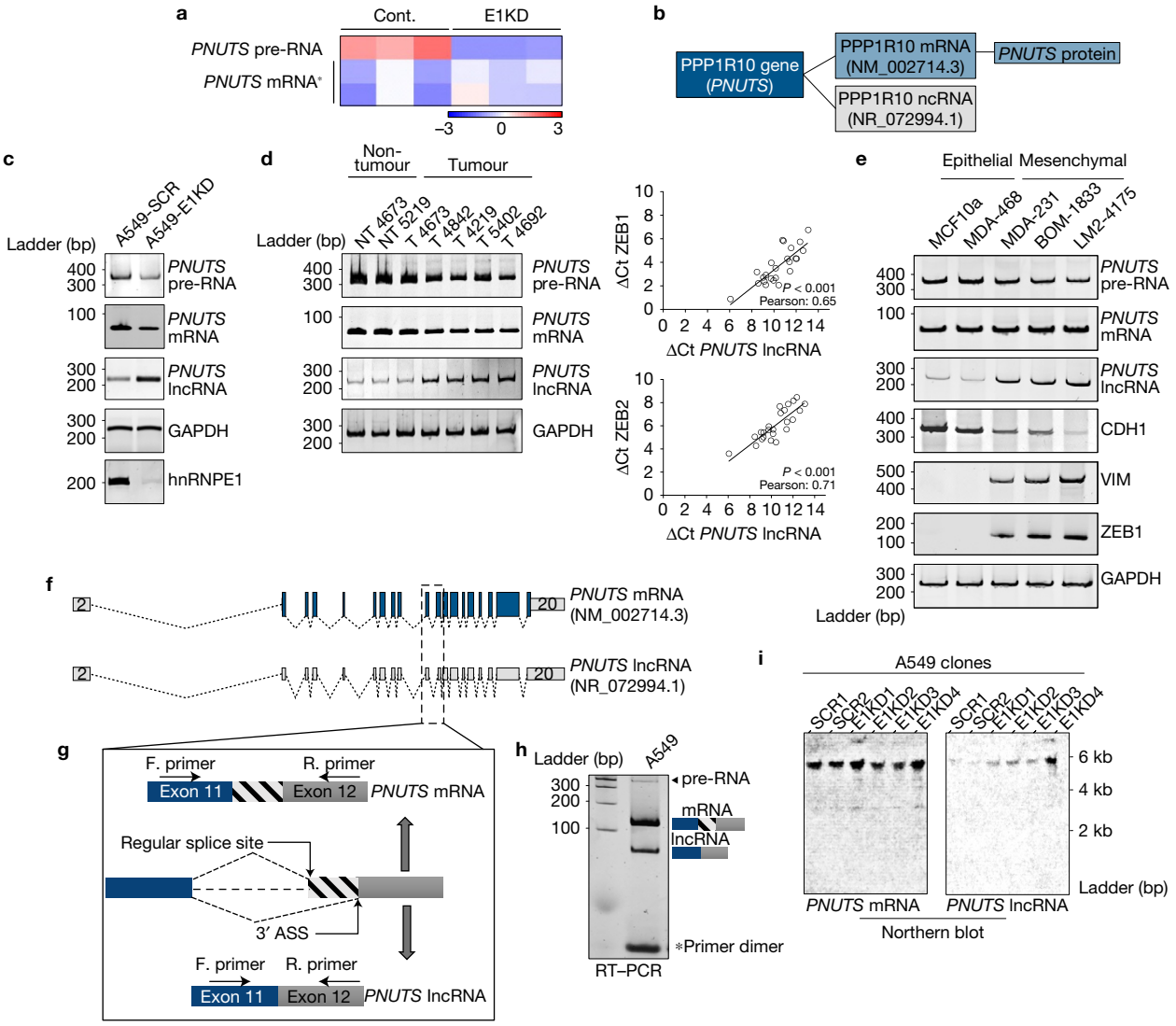


Figure 1 *PNUTS* alternative splicing occurs following hnRNP E1 loss and is increased in mesenchymal tumour cells. (**a**) Heat map of Affymetrix array data showing expression levels (log_2 fold) of either *PNUTS* pre-RNA or *PNUTS* mRNA in control (Cont.) or hnRNP E1 knockdown (E1KD) NMuMG cells. The data were generated from triplicate samples. *Two distinct probes were used to target the spliced *PNUTS* RNA. (**b**) NCBI database accession numbers of *PNUTS* mRNA and the *PNUTS* predicted lncRNA isoform in human. (**c**) Validation by RT–PCR analysis with primers specific to *PNUTS* isoforms of alternative *PNUTS* gene processing following hnRNP E1 knockdown in the human A549 cell line. (**d**) Left, *PNUTS* isoform expression levels analysed by RT–PCR in human breast tumour samples (T) or non-tumour counterparts (NT). Right, quantitative RT–PCR analysis of lncRNA-*PNUTS*, ZEB1 and ZEB2 expression in 24 human breast tumour samples. Relative expression levels of transcripts were calculated using the ΔCt method, normalizing to GAPDH. Correlations between transcript expression levels were evaluated using the Pearson correlation coefficient test. (Linear regression, df = 24–2, a Pearson score > 0.515 and $P < 0.05$ was considered as significant). Source data are available in Supplementary Table 2. (**e**) *PNUTS* isoform expression screening by RT–PCR analysis in MCF10a mammary gland epithelial cells and MDA-MB-468 breast cancer epithelial cells, or in the metastasis progression model of the MDA-MB-231 mesenchymal cell line (MDA-231, BOM-1833, LM2-4175). E-cadherin (CDH1) was used as an epithelial marker while vimentin (VIM) and ZEB1 were used as mesenchymal-cell-specific markers. (**f**) Map of *PNUTS* isoforms acquired by sequence alignment and drawn by using fancyGene online software. (**g**) Schematic representation of the alternative splicing region of the *PNUTS* variants (ASS, alternative splicing site). (**h**) RT–PCR amplification of the exon 11–exon 12 junction encompassing the predicted alternative splicing site using intron-flanking PCR primers as indicated in **g**. (**i**) Northern blot analysis of the expression levels of both the *PNUTS* mRNA and lncRNA isoforms in control (SCR) or hnRNP E1 knockdown (E1KD) A549 cell clones.

hnRNP E1 knockdown while the associated *PNUTS* mRNA remained relatively unaffected (Fig. 1a), suggesting a differential processing of *PNUTS* pre-RNA. Interestingly, the human *PNUTS* gene is described to encode two sequenced variants. While variant 1 encodes the well-characterized *PNUTS* mRNA, variant 2 has not been investigated and is predicted to be a lncRNA (Fig. 1b). We validated the differential processing of *PNUTS* pre-RNA by PCR with reverse transcription (RT–PCR) analysis with primers specific to *PNUTS* isoforms (Fig. 1c and Supplementary Fig. 6).

The biological significance of *PNUTS* pre-RNA differential processing is demonstrated in human breast tumour samples (Fig. 1d) and in breast cancer cell lines (Fig. 1e). We observed upregulation of the predicted lncRNA-*PNUTS* in breast tumour samples and a correlation between ZEB1/ZEB2 mesenchymal marker expression and

lncRNA-*PNUTS* (Fig. 1d) but not with the *PNUTS* mRNA (Supplementary Fig. 1b). We also observed a correlation between lncRNA-*PNUTS* expression and the epithelial/mesenchymal status of breast cancer cells. In the more mesenchymal MDA-MB-231 cell line, and its metastatic bone (BOM-1833) and metastatic lung (LM2-4175) derivatives[20], we observed increased expression of the predicted lncRNA-*PNUTS* correlating with the expression of the mesenchymal markers vimentin and Zeb1; whereas, in the more epithelial MCF10a and MDA-MB-468 cell lines expressing the epithelial marker E-cadherin, there was less expression of the predicted lncRNA (Fig. 1e).

The NCBI database predicts the generation of lncRNA-*PNUTS* as a result of the removal of 61 bases in the 5′-region of exon 12 leading to a break in the open reading frame of the transcript (Fig. 1f,g). By RT–PCR, using flanking primers (Supplementary Fig. 6), we demonstrated the existence of an alternative splice product of the expected size (Fig. 1h) and validated the alternative splicing model by sequencing (Supplementary Fig. 1c). The alternative splice site is also identified in the mouse (Supplementary Figs 1d,e), and northern blot validated the size of the full-length lncRNA-*PNUTS* and its upregulation following hnRNP E1 knockdown (Fig. 1i).

hnRNP E1 prevents the splicing of the lncRNA-*PNUTS* isoform by binding to a BAT structural element located at the alternative splice site

The alternative splice site in exon 12 of the *PNUTS* gene is also predicted *in silico* by the HSF finder[21] (Supplementary Fig. 2a). Interestingly, this alternative site has a higher consensus splice site value (91.74) than the regular splice site (79.38) used to generate the *PNUTS* mRNA (Supplementary Fig. 2a), suggesting the existence of an inhibitory mechanism of alternative splicing site utilization. Since hnRNP E1 is a known repressor of alternative splicing[18,19] and its knockdown results in upregulation of lncRNA-*PNUTS*, we postulated that it is an endogenous repressor of *PNUTS* pre-RNA splicing.

We previously described that hnRNP E1 binds to a consensus BAT element, consisting of a stem-loop structure with an asymmetric bulge located in the 3′-UTR of RNAs[13,22]. An analysis of the secondary structure of the *PNUTS* alternative splicing site in human and mouse sequences revealed the existence of a similar evolutionarily conserved BAT-like element encompassing the alternative splicing site (Fig. 2a and Supplementary Fig. 1f). We thus designed *PNUTS* BAT alternative splicing site RNA probes, wild type or mutated, to perform RNA electromobility shift assays and validated the direct and specific binding of hnRNP E1 to the structural element. Combinations of whole-cell lysates from A549 cells with the wild-type probe show a significant gel shift that is abolished by using either the mutant probe or whole-cell lysates prepared from A549 cells silenced for hnRNP E1 (Fig. 2b). Direct binding was further validated using recombinant hnRNP E1 protein (Fig. 2b). To test whether hnRNP E1 removal from the BAT alternative splicing site can mediate alternative splicing, we used two methods to induce its dissociation from RNA. First, on the basis of our earlier demonstration that TGFβ-induced Akt2 phosphorylation of hnRNP E1 leads to its loss of binding and release from the BAT element[13,15], we treated cells with TGFβ and observed alternative splicing occurring in both A549 and MDA-468 cells, generating the lncRNA-*PNUTS* isoform within 30 min and persisting for ∼3 h (Fig. 2c and Supplementary Fig. 1g). Second, since inhibition of transcription with RNA polymerase inhibitors results in cytoplasmic accumulation of many splicing factors[23], we treated cells with actinomycin D (ActD) and performed both immunofluorescence analysis and cell fractionation to investigate hnRNP E1 localization following transcriptional inhibition (Fig. 2d,e). We observed that ActD induced the release of hnRNP E1 from pre-RNAs and its nuclear/cytoplasmic shuttling, resulting in cytoplasmic accumulation (Fig. 2d,e). Concomitantly, cytoplasmic accumulation of hnRNP E1, in response to ActD, results in a strong induction of *PNUTS* alternative splicing that is not observed in hnRNP-E1-silenced cells (Fig. 2f). These results suggest that hnRNP E1 binding to the *PNUTS* pre-RNA alternative splicing site mediates an inhibitory effect on alternative splicing. To further validate this splicing model, we designed an antisense oligonucleotide (ASO) to the alternative splicing site to prevent its utilization. As shown in A549 cells (Fig. 2g), the ASO prevents, in a concentration-dependent manner, the alternative splicing induced by hnRNP E1 release following ActD treatment.

PNUTS alternative splicing product is a cytosolic and nuclear lncRNA

PNUTS mRNA encodes the PNUTS protein, which has a relative molecular mass of 99,000 (M_r 99K). However, *PNUTS* alternative splicing leads to a break at position Lys318 generating a downstream premature stop codon, potentially allowing the generation of a M_r 41K truncated protein. By immunoblot analysis, using an amino-terminal generated *PNUTS* antibody, we failed to detect a truncated expression product even following lncRNA-*PNUTS* overexpression in various cell lines (Supplementary Fig. 2b). Further, as analysed by polysome fractionation, lncRNA-*PNUTS* is observed only in the non-translating, monosomal fractions and not in the actively translating, polysomal fractions compared with *PNUTS* mRNA (Fig. 3a), underlining its non-translatability. Finally, endogenous lncRNA-*PNUTS* has a poly(A)+ tail (Fig. 3b) and is located in both the cytoplasmic and nuclear compartments as observed by cell fractionation, GFP tracking microscopy employing the MS2-Tag strategy and fluorescent *in situ* hybridization (FISH) analysis (Fig. 3c,d and Supplementary Fig. 2c).

lncRNA-*PNUTS* interacts with miR-205

Given the subcellular localization of lncRNA-*PNUTS*, we next explored its biological function as a presumed competing-endogenous RNA (ceRNA). By *in silico* analysis, we predicted 21 microRNAs (miRNAs) targeting at least five sites with a score higher than 0.6 (Supplementary Table 1). We focused on miRNA-binding sites most represented in the coding DNA sequence (CDS) of the cognate mRNA rather than in its 3′-UTR region to explore the intrinsic properties of the full-length lncRNA-*PNUTS* as a ceRNA. Among the ten miRNAs meeting this criterion, miR-205 was an obvious candidate due to its critical role in EMT and its high conservation among species[24]. We used quantitative real-time PCR to quantify the copy numbers of both miR-205 and lncRNA-*PNUTS* per cell since comparable levels are suggestive of ceRNA function (Fig. 3e). FISH analysis (Supplementary Fig. 2c) demonstrates co-localization of miR-205 with lncRNA-*PNUTS* and not *PNUTS* mRNA. This suggests a preferential interaction of miR-205 with the lncRNA isoform, which was further confirmed by the use of biotinylated antisense probes (Fig. 3g).

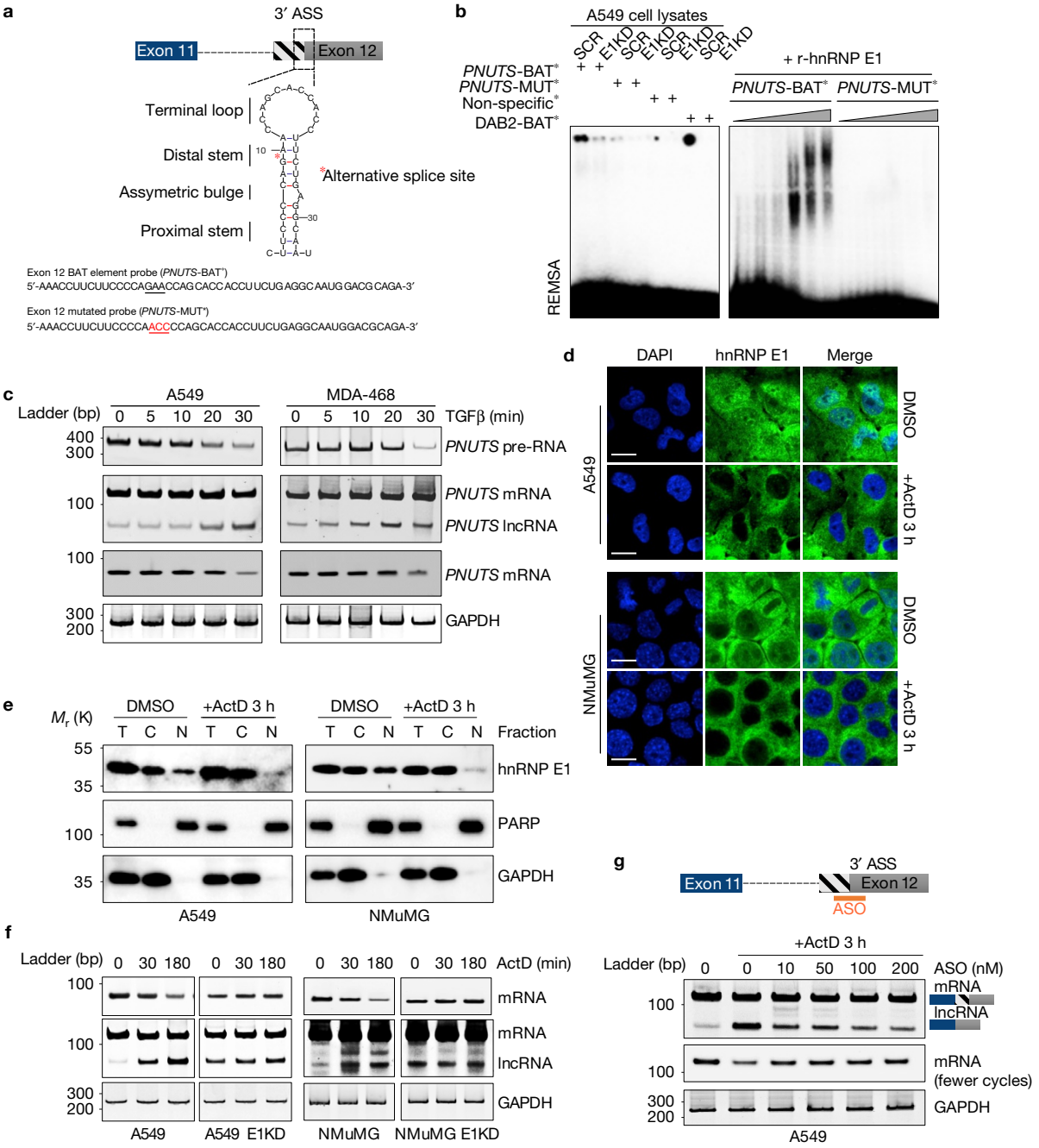


Figure 2 hnRNP E1 protein prevents *PNUTS* alternative splicing by its specific binding to a BAT structural element. (**a**) Secondary structure of the human *PNUTS* alternative splicing site as predicted by the Mfold algorithm ($\Delta G = -3.90$ kcal mol^{-1}). The underlined nucleotides coloured in red represent the mutant probe used for the RNA electromobility shift assay (REMSA) experiment in **b** and **c**. The red asterisk represents the exact alternative splicing site leading to the lncRNA-*PNUTS* isoform generation (ASS, alternative splicing site). (**b**) Left, REMSA experiment using either wild-type *PNUTS*-BAT or mutated *PNUTS*-MUT α-^{32}P-labelled *PNUTS* alternative splicing site probes combined with control (SCR) or hnRNP-E1-knockdown (E1KD) A549 cell lysates. The *PNUTS*-MUT probe was mutated by a nucleotide substitution to destroy its secondary structure. Non-specific and DAB2-BAT α-^{32}P-labelled probes were used as negative and positive controls respectively. DAB2-BAT corresponds to the BAT sequence located on the Dab2-3′-UTR already described to bind to hnRNP E1. Right, REMSA using a combination of *PNUTS*-BAT or mutated *PNUTS*-MUT α-^{32}P-labelled probes with increasing concentration of recombinant hnRNP E1 protein purified from *Escherichia coli* bacteria. (**c**) Time course experiment using RT–PCR analysis of *PNUTS* gene processing after addition of 5 ng ml^{-1} of TGFβ. (**d**) Confocal microscopy imaging of the hnRNP E1 nucleocytoplasmic shuttling by addition of 5 μg ml^{-1} of ActD for 3 h in A549 and NMuMG cell cultures. Scale bars, 10 μM. (**e**) Characterization of the nucleocytoplasmic transportation of hnRNP E1 following ActD treatment by using cell fractionation and subsequent western blot analysis of hnRNP E1 expression. To check the purity of the fractions, GAPDH and PARP were used as cytoplasmic (C) and nuclear (N) compartment markers respectively. (**f**) Time course experiment using RT–PCR analysis of *PNUTS* predicted lncRNA alternative splicing activation following addition of 5 μg ml^{-1} of ActD in control (Cont.) or hnRNP-E1-silenced (E1KD) A549 and NMuMG cells. (**g**) Inhibition of alternative splicing induced by ActD in A549 cells using an antisense oligonucleotide (ASO) targeting the alternative splicing site of *PNUTS*. GAPDH was used as a loading control. Unprocessed original scans of blots are shown in Supplementary Fig. 7.

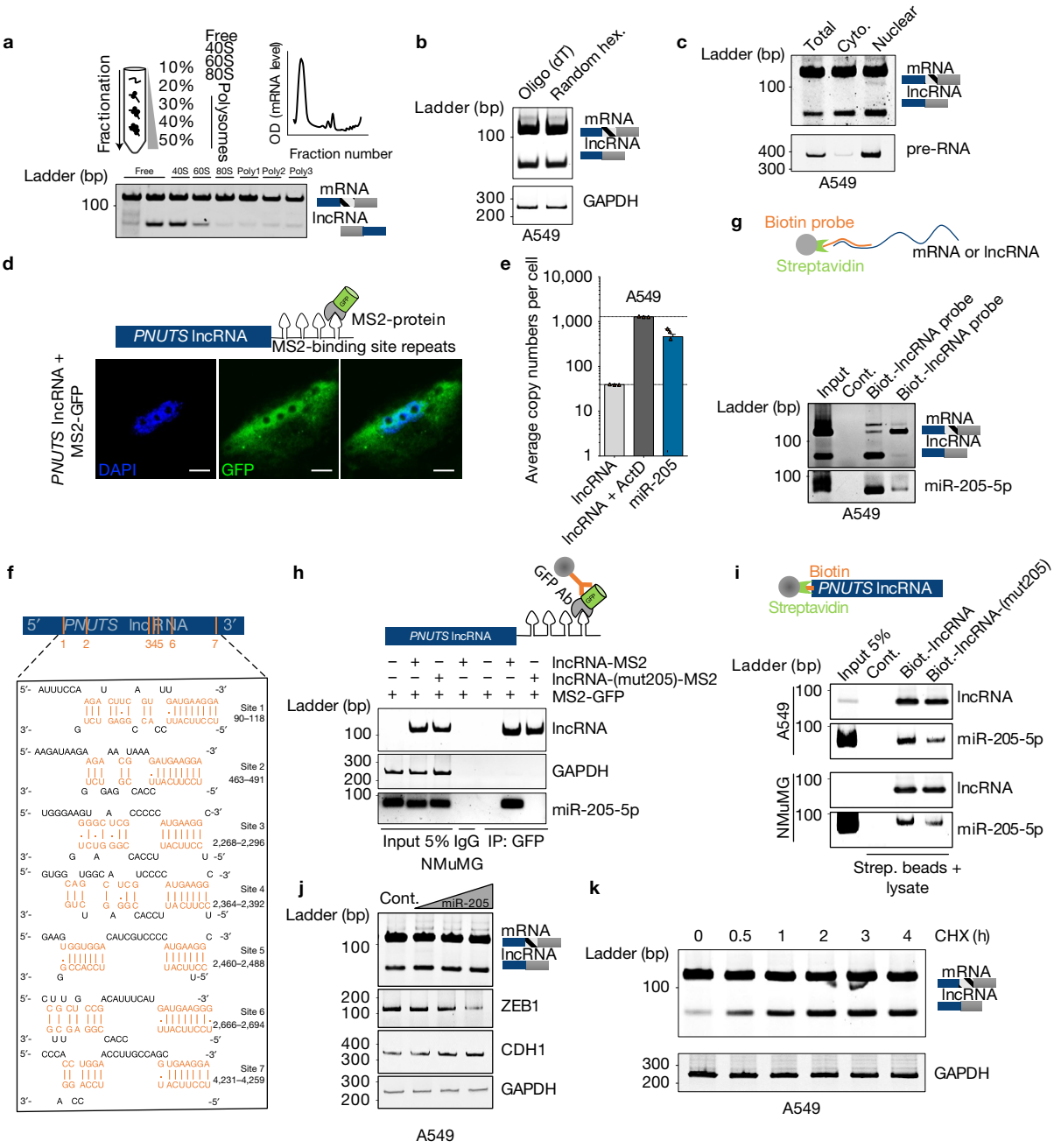


Figure 3 *PNUTS* alternative splicing product is non-coding and interacts with miR-205. (**a**) Polysome fractionation experiment of A549 cells followed by RT–PCR analysis of *PNUTS* mRNA and lncRNA-*PNUTS* expression in each fraction 40S, 60S and 80S represent ribosomal fractions and Poly1–3 the polysomal fractions. (**b**) RT–PCR analysis of *PNUTS* mRNA and lncRNA-*PNUTS* expression after the use of oligo-(dT) or random hexanucleotides as primers for initial reverse transcription reaction. (**c**) RT–PCR analysis of lncRNA-*PNUTS* expression in A549 cells. The total, cytoplasmic (Cyto.) and nuclear fractions are shown. *PNUTS* pre-RNA and *PNUTS* mRNA were used as endogenous controls to monitor the purity of the fractions. (**d**) Confocal microscopy imaging of subcellular localization of lncRNA-*PNUTS* using co-transfection of an MS2-tagged-RNA construct of lncRNA-*PNUTS* and a fused MS2-GFP protein construct. Scale bars, 5 μM. (**e**) The exact copy numbers of lncRNA-*PNUTS* (basal levels or levels following activation by ActD treatment for 3 h) and miR-205 were quantified with limiting-dilution quantitative RT–PCR. Data are shown as mean ± s.d., $n = 3$ independent experiments per condition. Source data are available in Supplementary Table 2. (**f**) *In silico* prediction of miR-205-binding sites located on lncRNA-*PNUTS*, obtained using the DIANA-microT web server. (**g**) Selective pulldown of either endogenous lncRNA-*PNUTS* or *PNUTS* mRNA isoforms by using antisense biotinylated probes followed by miRNA-specific RT–PCR analysis to detect endogenously associated miR-205 with lncRNA-*PNUTS* in A549 cells. (**h**) MS2-RIP followed by miRNA-specific RT–PCR analysis to detect the association of miR-205 with lncRNA-*PNUTS* in NMuMG cells. LncRNA-*PNUTS* and GAPDH expression were used as internal controls. (**i**) A549 and NMuMG cell lysates incubated with *in vitro*-transcribed biotin-labelled lncRNA-*PNUTS* were subjected to pulldown followed by miRNA extraction and analysis by RT–PCR. (**j**) A549 cells overexpressing lncRNA-*PNUTS* were transfected with an increasing concentration of a synthetic miR-205 mimic and the lncRNA expression was assessed by RT–PCR. ZEB1 and CDH1 were used to monitor the efficiency of miR-205 overexpression on the mesenchymal–epithelial transition process. (**k**) Time course experiment by using RT–PCR analysis of lncRNA-*PNUTS* levels following addition of 10 μg ml^{-1} cycloheximide in A549 cells. GAPDH was used as a loading control.

LncRNA-*PNUTS* harbours seven miR-205 sites, including one located in the 3′-UTR of the cognate *PNUTS* mRNA (Fig. 3f). To ensure that the part including the first six miR-205-binding sites is functionally active, we cloned this portion, either wild type or mutated for the miRNA-205-binding sites, into the MS2-TRAP vector and validated the specific binding by an MS2-tagged RNA affinity purification strategy and by avidin-affinity pulldown of cellular lysates (Fig. 3h,i). To investigate the decay mechanism(s) of lncRNA-*PNUTS*, we treated A549 cells with increasing concentrations of miR-205, which results in an expected decrease in ZEB1 and increase in E-cadherin expression levels, respectively; however, miR-205 levels have no significant impact on the level of lncRNA-*PNUTS* (Fig. 3j). Moreover, cycloheximide treatment of cells for up to 4 h considerably increases the expression of lncRNA-*PNUTS*, suggesting its sensitivity to nonsense-mediated mRNA decay (Fig. 3k).

lncRNA-*PNUTS* regulates EMT migration and invasion *in vitro* through its miR-205 interaction

Since lncRNA-*PNUTS* interacts with miRNA-205, a well-established regulator of ZEB proteins and of epithelial cell maintenance, we investigated its effects on ZEB expression and cell plasticity. We silenced endogenous lncRNA-*PNUTS* in mesenchymal and invasive MDA-231-LM2-4175 cells that express high levels of the lncRNA (Fig. 1e) to test whether lncRNA-*PNUTS* could modulate cell plasticity. LncRNA-*PNUTS* silencing led to a significant decrease in cell invasion correlating with reduced vimentin expression (mesenchymal marker) and re-expression of epithelial marker E-cadherin concomitant to morphological changes (Fig. 4a). To assess whether lncRNA-*PNUTS* generation is a prerequisite for TGFβ-mediated EMT, we treated A549 cells with ASO against the alternative splice site (Fig. 2g) and showed that it significantly impaired TGFβ-mediated EMT (Supplementary Fig. 3a,d,e). This was confirmed using short interfering RNA (siRNA) specifically targeting lncRNA-*PNUTS* to prevent either TGFβ-mediated or hnRNP-E1-knockdown-mediated EMT (Supplementary Fig. 3b,c). Moreover, the overexpression of lncRNA-*PNUTS* induced an EMT in both A549 and NMuMG as characterized by a morphological change from an epithelial-like, cobblestone phenotype to a more spindle-shaped mesenchymal phenotype, an E-cadherin/vimentin switch (Fig. 4b,c), and an accompanying increase in the levels of the EMT transcription factors, *ZEB1*, *ZEB2*, *SNAI1* and *SNAI2* (Fig. 4d). While the wild-type lncRNA-*PNUTS* induced an EMT associated with a downregulation of E-cadherin and upregulation of ZEB1, co-transfection with miRNA-205 as well as the overexpression of the miR-205-mutant form of the lncRNA-*PNUTS* abolished this effect (Fig. 4e,f). The lncRNA-*PNUTS* controls both migration and invasion (Fig. 4g,h) of A549 and NMuMG cells in a manner dependent on its miR-205-binding sites, and miR-205 overexpression is able to abolish this effect (Fig. 4h). These observations were further validated by immunofluorescence (Fig. 4i).

lncRNA-*PNUTS* controls the miR-205/ZEB/E-cadherin axis

Using luciferase reporter assays we next validated the regulation of the miR-205/ZEB/E-cadherin axis by lncRNA-*PNUTS*. We first confirmed the binding of miR-205 to lncRNA-*PNUTS* by cloning the S3-S6-miR-205-binding sites of lncRNA-*PNUTS* as a 3′-UTR of the luciferase CDS (Fig. 5a) and demonstrated that co-transfection with miR-205 in A549 and NMuMG cells reduces luciferase expression with the wild type (WT), but not the mutated S3-S6-miR-205 construct (Fig. 5a). Second, using a luciferase reporter construct whose stability is dependent on miR-205 binding (miR-205 microRNA recognition element (MRE)), we demonstrated that WT, but not the miR-205-mutated-lncRNA-*PNUTS*, decreases miR-205 bioavailability (Fig. 5b). Third, using constructs of *Renilla* fused to the ZEBs 3′-UTR (Fig. 5c and Supplementary Fig. 4), we observed that WT lncRNA-*PNUTS* stabilized ZEBs 3′-UTR and this effect was partially abolished using the miR-205-mutated lncRNA or reversed by co-transfection with miR-205 mimics (Fig. 5c and Supplementary Fig. 4). Importantly, no effects were observed on the ZEB1 3′-UTR mutated for its miR-205-binding site (Fig. 5c). Last, using an E-cadherin (CDH1) promoter luciferase construct that contains either wild-type or mutant ZEBs-binding sites (E-boxes), we observed an ∼50% decrease of luciferase activity following lncRNA-*PNUTS* overexpression. This downregulation was partially rescued by co-transfection with miR-205 or abolished by using either the mutated lncRNA-*PNUTS* or the CDH1 promoter mutated for its E-boxes (Fig. 5d).

LncRNA-*PNUTS* regulates tumour implantation, growth and metastasis

Given the role of miR-205 in regulating mammary stem cell fate and tumorigenesis through EMT[25], we investigated whether lncRNA-*PNUTS* contributes to these phenotypes. Utilizing the tridimensional, sphere-formation assay, we showed a significant increase in sphere formation of A549 and NMuMG cells induced by lncRNA-*PNUTS* dependent on its miR-205-binding sites (Fig. 6a,b). Using limiting dilutions of MDA-468 cells in an *in vivo* fat pad injection assay, we observed that lncRNA overexpression resulted in an ∼80-fold increase of tumour-initiating cell number compared with the control (Fig. 6c). Next, we FACS-sorted HMLE cells (Fig. 6d,e) using CD24/CD44 to determine whether lncRNA-*PNUTS* is upregulated in stem cells. Although the $CD24^{-}/CD44^{+}$ subpopulation (mesenchymal phenotype) is known to be highly enriched for tumour-initiating cells[26,27], we did not observe an upregulation of the lncRNA-*PNUTS* in this subpopulation compared with the $CD24^{+}/CD44^{-}$ subpopulation (epithelial phenotype; Fig. 6f). Nevertheless, overexpression of lncRNA-*PNUTS* in the epithelial subpopulation induced a significant decrease in the number of $CD24^{+}/CD44^{-}$ cells and revealed a minor subpopulation of cells harbouring the $CD24^{-}/CD44^{+}$ phenotype (Fig. 6g and Supplementary Fig. 5a). Additionally, silencing of lncRNA-*PNUTS* in MDA-231-LM2 cells led to increased expression of the CD24 epithelial marker (Supplementary Fig. 5b).

We next tested the effects of lncRNA-*PNUTS* on tumour progression *in vivo* and demonstrated that silencing of lncRNA-*PNUTS* in MDA-231-LM2 cells impairs tumour formation when injected orthotopically into mammary fat pads (Fig. 6h). Moreover, lncRNA-*PNUTS* also contributes to *in vivo* metastasis as we observed a significant decrease of lung colonization in MDA-231-LM2 cells silenced for lncRNA-*PNUTS* compared with their scrambled-control counterparts (Fig. 6i).

DISCUSSION

The discovery of lncRNAs as biologically relevant molecules has led to a rethinking of the central dogma of biology and unravelled new layers

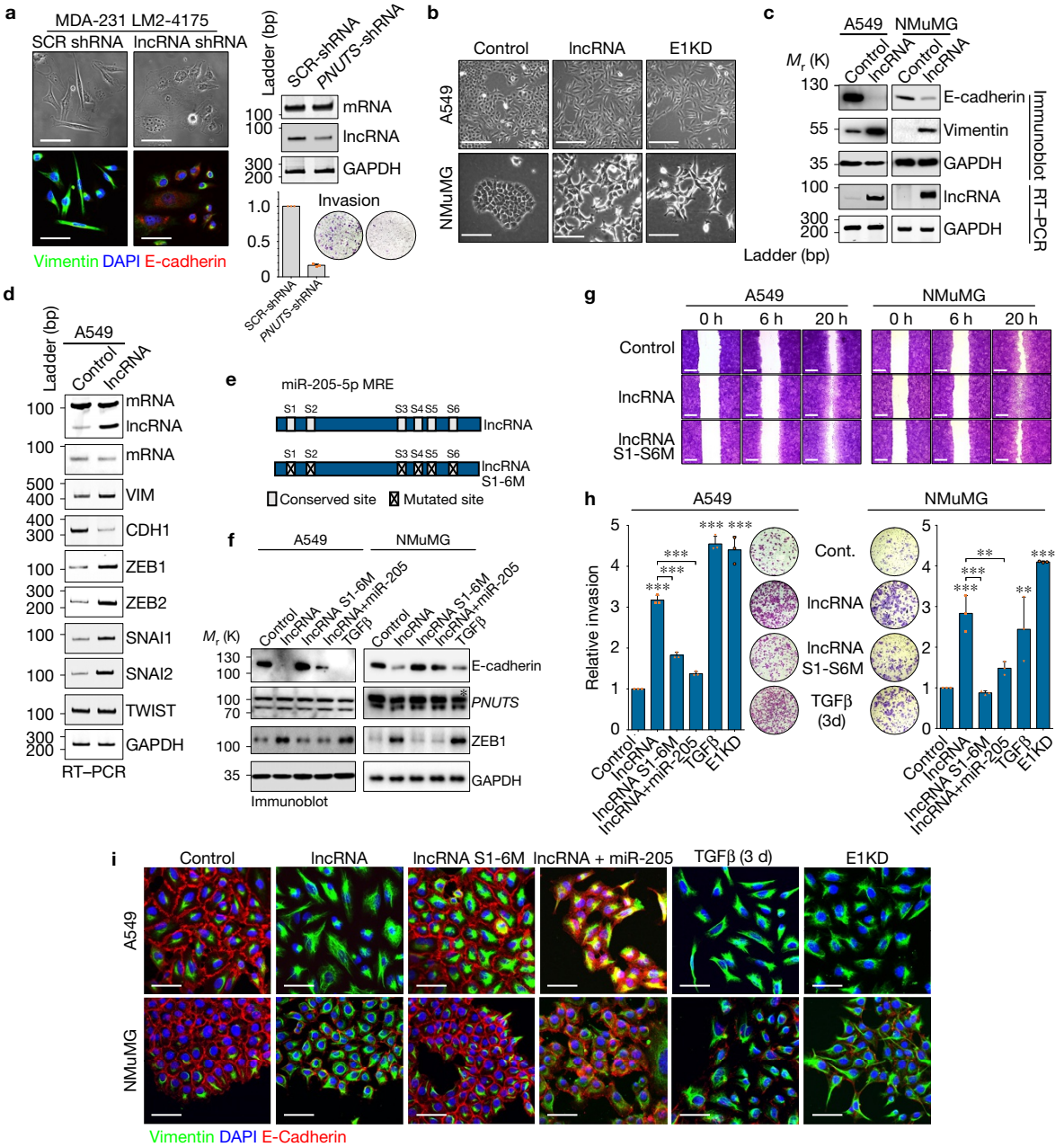


Figure 4 LncRNA-*PNUTS* regulates EMT and cell migration/invasion *in vitro*. (**a**) MDA-231-LM2-4175 cells stably silenced for lncRNA-*PNUTS* were analysed by immunofluorescence (left) using antibodies against vimentin (green), E-cadherin (red) and merged with DAPI (blue). Scale bars, 50 μM. lncRNA-*PNUTS* silencing was monitored by RT–PCR (right, top). Invasive capacities of control (SCR-shRNA) or lncRNA-*PNUTS*-silenced (*PNUTS* shRNA) cells were monitored in a modified Boyden chamber assay (right, bottom) (mean ± s.d., $n = 3$ independent experiments per condition). Source data are available in Supplementary Table 2. (**b**) A549 and NMuMG cells stably overexpressing lncRNA-*PNUTS* were analysed using bright-field microscopy. hnRNP E1 knockdown (E1KD) cells were used as controls. Scale bars, 100 μM. (**c**) Western blot (top) and RT–PCR (bottom) analysis of E-cadherin, vimentin and lncRNA-*PNUTS* in A549 and NMuMG cells overexpressing lncRNA-*PNUTS*. (**d**) RT–PCR analysis of several EMT-related transcription factors in A549 cells stably overexpressing lncRNA-*PNUTS*. (**e**) Schematic outlining the constructs used in this study for wild-type (lncRNA) or mutated (lncRNAS1-6M) lncRNA. (**f**) Western blot analysis of E-cadherin, *PNUTS* and ZEB1 protein expression in A549 and NMuMG cells overexpressing wild-type (lncRNA) or mutated (lncRNAS1-6M) constructs of lncRNA-*PNUTS* and treated or not with synthetic miR-205 mimic or TGFβ for 3 days. TGFβ was used as a positive control. *PNUTS protein band. (**g**) Wound-healing migration assay of control (Cont.), lncRNA-*PNUTS* (lncRNA) or mutated lncRNA-*PNUTS* (lncRNAS1-6M) A549 and NMuMG cell models. Scale bars, 400 μM. (**h**) Modified Boyden chamber invasion assay of wild-type (lncRNA) or mutated (lncRNAS1-6M) lncRNA-*PNUTS* overexpressing A549 and NMuMG cells pre-treated ± synthetic miR-205 mimic or TGFβ for 3 days. hnRNP-E1-knockdown (E1KD) and TGFβ-treated cells were used as a positive control (mean ± s.d., $n=3$ independent experiments per condition, ANOVA followed by *post hoc* Tukey's multiple comparisons test, $^{**}P < 0.01$; $^{***}P < 0.001$). Source data are available in Supplementary Table 2. (**i**) Confocal microscopy imaging of co-immunostaining of vimentin (green), E-cadherin (red) and merged with DAPI (blue) in A549 and NMuMG cells overexpressing wild-type (lncRNA) or mutated (lncRNAS1-6M) constructs of lncRNA-*PNUTS* and treated ± synthetic miR-205 mimic or TGFβ for 3 days. Scale bars, 50 μM. For all western blots and RT–PCRs, GAPDH was used as a loading control. Unprocessed original scans of blots are shown in Supplementary Fig. 7.

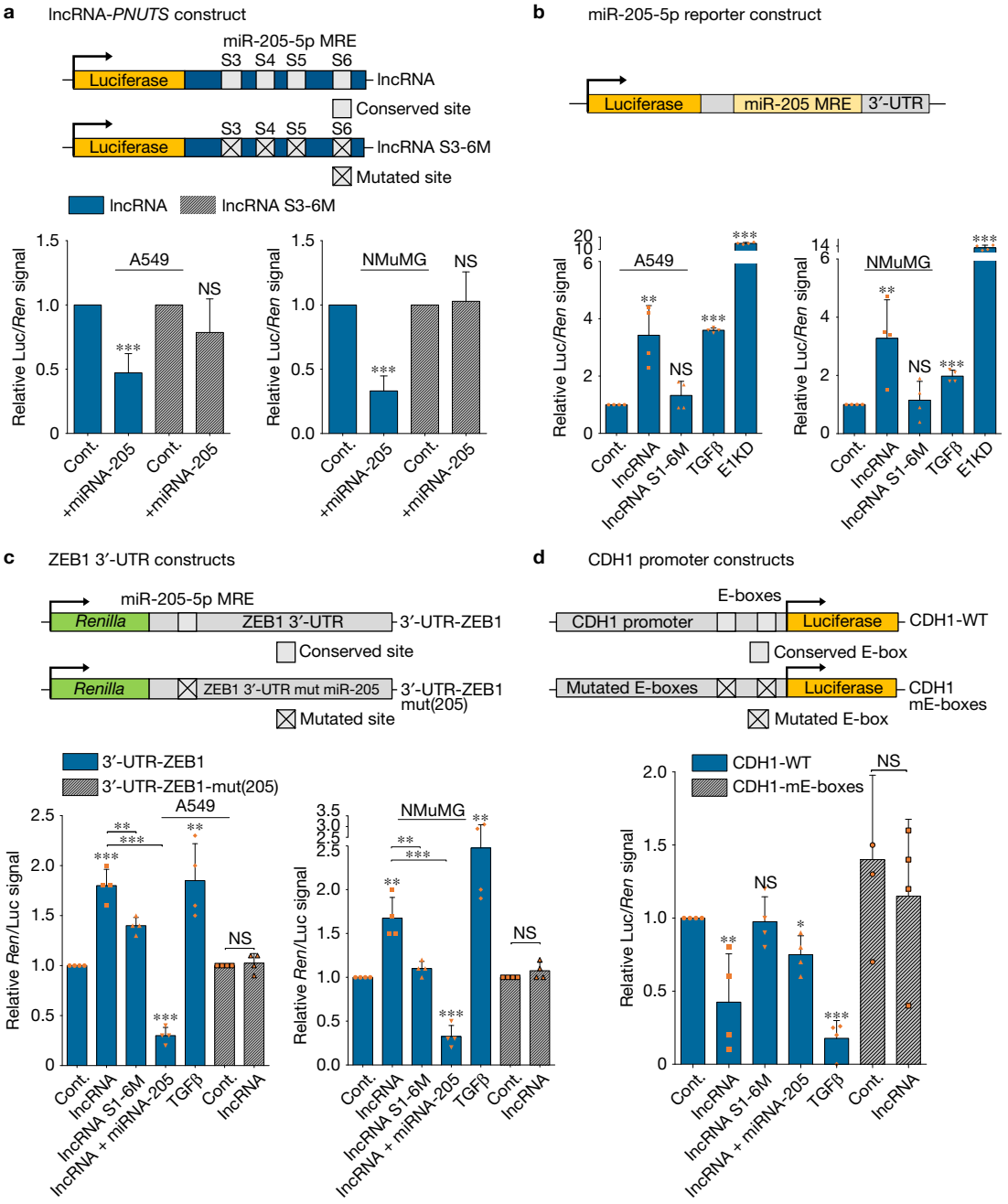


Figure 5 LncRNA-*PNUTS* controls the miR-205/ZEB/E-cadherin axis. (**a**) Dual-luciferase reporter assays to test the interaction between miR-205 and lncRNA-*PNUTS* (S3 to S6 region) by using a synthetic miR-205 mimic (+ miRNA-205) co-transfected with wild-type (Luc-lncRNA) or mutated (Luc-lncRNA-S3-S6M) constructs of lncRNA-*PNUTS* cloned into the 3′-UTR of the luciferase reporter gene. For each condition, assays were normalized to *Renilla* reporter gene expression (mean ± s.d., $n=7$ independent experiments per condition, two-tailed Student's *t*-test, $^{***}P<0.001$; NS, not significant). (**b**) Dual-luciferase reporter assay of miR-205 bioavailability in A549 and NMuMG cells overexpressing wild-type (lncRNA) or mutated (lncRNAS1-6M) constructs of lncRNA-*PNUTS*. TGFβ treatment and hnRNP E1 knockdown (E1KD) were used as internal controls. For each condition, assays were normalized to *Renilla* reporter gene expression (mean ± s.d., $n=4$ independent experiments per condition, two-tailed Student's *t*-test, $^{**}P<0.01$; $^{***}P<0.001$; NS, not significant). (**c**) The wild-type (*Ren*-3′-UTR-ZEB1) 3′-UTR of ZEB1 cloned into the 3′-UTR of the *Renilla* gene was transfected in A549 and NMuMG cells overexpressing wild-type (lncRNA) or mutated (lncRNAS1-6M) constructs of lncRNA-*PNUTS* and treated ± synthetic miR-205 mimic or TGFβ for 3 days. Mutated construct (*Ren*-3′-UTR-ZEB1-mut(205)) for the miR-205-binding site located in the 3′-UTR of ZEB1 was also used. TGFβ was used as a positive control. For each condition, assays were normalized to luciferase reporter gene expression (mean ± s.d., $n=4$ independent experiments per condition, two-tailed Student's *t*-test, $^{**}P<0.01$; $^{***}P<0.001$; NS, not significant). (**d**) The wild-type (prom-CDH1-WT) proximal promoter of E-cadherin driving the luciferase reporter gene expression was transfected in A549 cells overexpressing wild-type (lncRNA) or mutated (lncRNA S1-6M) constructs of lncRNA-*PNUTS* and treated or not with synthetic miR-205 mimics or TGFβ for 3 days. A mutated construct for both E2-boxes 1 and 3 (Prom-CDH1-mE-boxes) located on the promoter was also used. TGFβ was used as a positive control. For each condition, assays were normalized to *Renilla* reporter gene expression. (mean ± s.d., $n=4$ independent experiments per condition, two-tailed Student's *t*-test, $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$; NS, not significant).

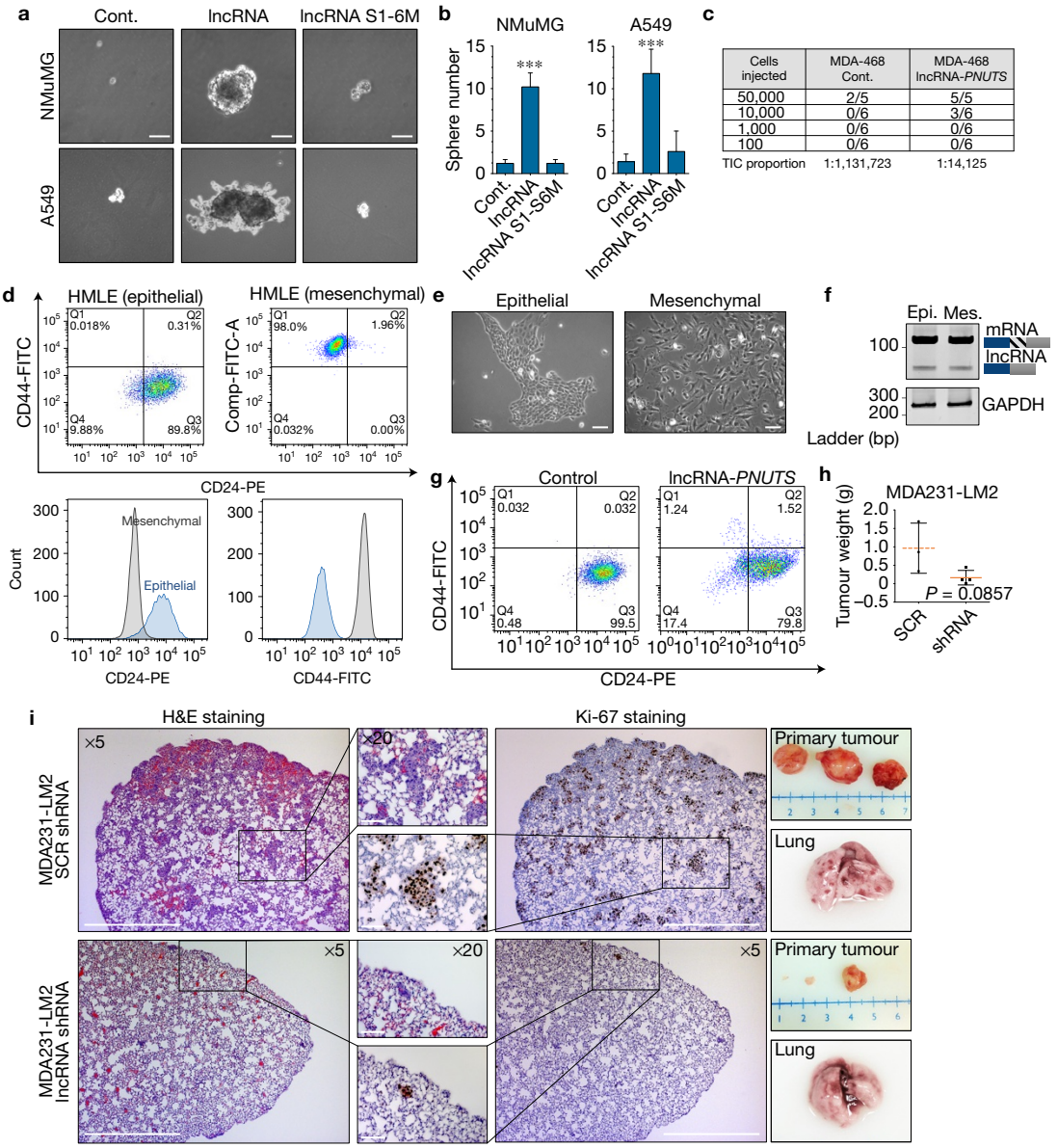


Figure 6 LncRNA-*PNUTS* promotes tumour initiation/growth and metastasis *in vivo*. (**a**) Bright-field microscopy pictures of *in vitro* mammosphere/oncosphere formation assay in NMuMG and A549 cells overexpressing empty vector (Cont.), lncRNA-*PNUTS* (lncRNA) or mutated lncRNA-*PNUTS* (lncRNAS1-6M). Scale bars, 20 μM. (**b**) Absolute quantification of the sphere numbers obtained in **a** (mean ± s.d., $n=5$ independent experiments, two-tailed Student's *t*-test, $^{***}P<0.001$; NS, not significant). Source data are available in Supplementary Table 2. (**c**) Number of tumours formed following limiting-dilution injection of control and lncRNA-*PNUTS*-overexpressing MDA-468 cells. MDA-MB-468 cells were injected into the mammary fat pads of 6–8-week-old female mice in limiting dilution. Tumour-initiating cell (TIC) number was determined using ELDA software[46]. The number of mice used for each condition is indicated. (**d**) Flow cytometry analysis of CD24/CD44 cell surface expression levels in the epithelial ($CD44^-/CD24^+$ sorted cells) and mesenchymal ($CD44^+/CD24^-$ sorted cells) HMLE subpopulations. (**e**) Cell morphology observed by phase-contrast microscopy. Scale bars, 50 μM. (**f**) RT–PCR analysis of lncRNA-*PNUTS* expression level in mesenchymal and epithelial sorted HMLE cells. (**g**) Flow cytometry analysis of the CD24/CD44 cell surface expression levels in the epithelial ($CD44^-/CD24^+$ sorted cells) HMLE subpopulation expressing empty vector (control) or overexpressing lncRNA-*PNUTS*. (**h**) Tumour weight of primary tumours obtained following mammary fat pad injection of MDA-231-LM2 expressing scrambled control (SCR) or lncRNA-*PNUTS* targeting shRNA (shRNA) in NOD/SCID mice (mean ± s.d., $n=4$ mice per condition, two-tailed Mann–Whitney test, $P=0.08570$). Source data are available in Supplementary Table 2. (**i**) Left, histopathological analysis of paraffin-embedded lung serial sections of mice injected in the mammary fat pad with MDA-231-LM2 expressing scrambled or lncRNA-*PNUTS* targeting shRNA. Haematoxylin and eosin (H&E) staining and immunostaining of Ki-67 protein was performed in serial lung sections to identify macro- and micro-metastases. Right, photographs of primary tumours and of a representative lung collected for each condition. Scale bars, 500 μM

of cellular and molecular complexities. It is now well established that many genes can encode both mRNA and ncRNA. For instance, a large number of miRNAs and most of the small nucleolar RNA (snoRNAs) are processed from spliced introns. Furthermore, circular RNAs can be processed from introns or back-spliced exons[28]. Moreover, the existence of large numbers of bifunctional RNAs whose isoforms are regulated by alternative splicing, as described herein, was previously theorized and predicted on the basis of genome-wide data mining of

alternative splicing events[29]. Here, we describe how under different cellular contexts, a gene, through alternative splicing, can encode for either an mRNA or a lncRNA and demonstrate the biological relevance of the generated lncRNA in targeting and sequestering miR-205 to ultimately regulate EMT (model Fig. 6e).

Since alternative splicing and generation of lncRNA-*PNUTS* is an early event in TGFβ-mediated EMT, we postulate that lncRNA-*PNUTS* operates as a transient inhibitor of miR-205 to allow for the temporal upregulation of ZEBs and subsequent regulation of downstream EMT events. Indeed, ZEBs proteins are reciprocally linked in a feedback loop with the miR-200 family, each strictly controlling the expression of the other[30]. In this way, a transient, but nevertheless, strong decrease in miR-205 bioavailability, sufficient to activate the ZEB proteins, would allow for transcriptional repression of the miR-200 family or other miRNAs such as miR-183 or miR-203, thereby further stabilizing ZEB proteins and reinforcing the EMT process. Furthermore, the transient nature of the lncRNA-*PNUTS* allows for its early regulation of the miR-205/ZEB axis during EMT but is not sufficient to sustain a decrease in *PNUTS* mRNA and protein expression, thus allowing independent functions of the isoforms. Moreover, the fact that the lncRNA-*PNUTS* is upregulated in tumour samples compared with their non-tumour counterpart despite its transient functional role might be the result of tumour heterogeneity with regard to TGFβ signalling[31,32] and that at any given time lncRNA-*PNUTS* is elevated in certain tumour cells. Finally, the alternative splicing of *PNUTS* is consistently accompanied by a decrease in the expression of the *PNUTS* pre-RNA. Since RNA splicing occurs co-transcriptionally, this decrease could be explained by the influence of transcriptional regulators pausing the RNAPII elongation complex to allow splicing to proceed on the alternative site[33].

LncRNA-*PNUTS* is localized in both the cytoplasm and nucleus (Fig. 3c,d and Supplementary Fig. 2c) and while its function as a ceRNA could be attributed to its cytoplasmic localization, its role in the nuclear compartment was not investigated herein. Nuclear biogenesis of lncRNA-*PNUTS* might explain its localization in the nucleus, although we speculate that it could also be involved in nuclear processes such as transcription or epigenetic regulation, as is the case for other previously described lncRNAs[34–36]. Moreover, we demonstrate that the transcriptional inhibitor ActD is a potent activator of the *PNUTS* alternative splicing through its effects on hnRNP E1 translocation from the nucleus to cytoplasm. Furthermore, lncRNA-*PNUTS* is upregulated by cycloheximide, a compound widely used to inhibit translation but also to test the sensitivity of RNAs to nonsense-mediated mRNA decay by inhibiting the first round of translation[37] and which is also known to induce transport of hnRNPs into the cytoplasm[38]. On the basis of these results, we expect that any anti-tumour agents whose pharmacological properties block transcription might activate the alternative splicing of *PNUTS*. Given the role of EMT in drug resistance and the contribution of miR-205 in chemotherapy sensitivity[39–42], it will be of interest to evaluate the contribution of lncRNA-*PNUTS* in EMT-mediated drug resistance mechanisms.

However, it is of note that both *PNUTS* isoforms share miR-205 sites, raising the obvious question as to why the *PNUTS* mRNA does not itself serve to sequester miR-205 and regulate EMT. We postulate that the location of miRNA-binding sites in the CDS of the cognate mRNA, relative to those in the lncRNA, may affect miRNA binding. As previously established, ribosomal hindrance could interfere with the ability of the miRNA to attach to its target site if it is located in the CDS[43] and this is supported by our data demonstrating preferential co-localization and binding of miR-205 with lncRNA-*PNUTS*. Additionally, the two *PNUTS* isoforms could have distinct secondary structures that might also explain the preferential binding of miR-205 to the lncRNA isoform. The fact that miR-205 binds weakly to the *PNUTS* mRNA, presumably to its site in the 3′-UTR, could also contribute to PNUTS protein stability during alternative splicing since the sponge activity of the lncRNA-*PNUTS* could counteract the inhibitory effect of the miR-205 initially occurring on the 3′-UTR of the *PNUTS* mRNA.

Collectively, our work confirms the key roles of lncRNAs and RNA-binding proteins in biological processes and human diseases. The study describes the generation and function of a lncRNA, and of an RNA-binding protein with which it associates, as key regulators of EMT and of the mesenchymal properties of tumour cells. Our *in vivo* data also demonstrate that modulation of lncRNA-*PNUTS* regulates the metastatic potential of tumour cells. Since most cancer-related mortalities result from metastatic disease and no mutations that are selective for metastases have been identified[3,44,45], it is imperative to identify potential metastatic mediators for prognostic and therapeutic benefit. As such, the identification of both hnRNP E1 and lncRNA-*PNUTS* provides two additional targets that could potentially serve as predictive markers of metastasis and/or chemoresistance, as well as effective targets for anti-metastatic therapies. □

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

The authors are grateful to M. Gooz for help with the microscopic analysis. We also thank Y. Shao for histopathological analysis and K. Gibbs (Charleston, South Carolina USA—Biorepository & Tissue Analysis Shared Resource at Medical University of South Carolina) for providing the human tumour samples. We thank M. Bidyut for his assistance with the cloning and K. Noguchi for his help with hnRNP E1 plasmid constructs and J. Isaacs for her precious help concerning hypoxia experiments. This work was supported in part by the Cell & Molecular Imaging, Small Animal Imaging, and the Biorepository & Tissue Analysis Shared Resources, Hollings Cancer Center, Medical University of South Carolina (P30 CA138313), and the Shared Instrumentation Grant S10 OD018113. This work was also supported by grants CA555536 and CA154664 from the National Cancer Institute to P.H.H.

AUTHOR CONTRIBUTIONS

The study was supervised by P.H.H. The conception and design were made by S.G. and P.H.H. The methodology was developed by S.G. and P.H.H. Acquisition and interpretation of the data was carried out by S.G. and P.H.H. The Affymetrix experiment and care of the mice was provided by B.H. Mice injections and dissection were performed by S.G., L.A.L. and C.O. The manuscript was written by S.G. and P.H.H. and all the authors contributed to its reviewing. V.P., V.K.G. and J.A.D. provided critical discussion.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://dx.doi.org/10.1038/ncb3595

1. Jemal, A. *et al*. Global cancer statistics. *CA. Cancer J. Clin.* **61,** 69–90 (2011).
2. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331,** 1559–1564 (2011).
3. Gupta, G. P. & Massagué, J. Cancer metastasis: building a framework. *Cell* **127,** 679–695 (2006).
4. Weigelt, B., Peterse, J. L. & van't Veer, L. J. Breast cancer metastasis: markers and models. *Nat. Rev. Cancer* **5,** 591–602 (2005).
5. 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).
6. Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. *Cell* **136,** 656–668 (2009).
7. Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152,** 1298–1307 (2013).
8. Yuan, J. *et al*. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* **25,** 666–681 (2014).
9. Leone, S. & Santoro, R. Challenges in the analysis of long noncoding RNA functionality. *FEBS Lett.* **590,** 2342–2353 (2016).
10. Thomson, D. W. & Dinger, M. E. Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.* **17,** 272–283 (2016).
11. Smith, B. N. & Bhowmick, N. A. Role of EMT in metastasis and therapy resistance. *J. Clin. Med.* **5,** 17 (2016).
12. Ye, X. & Weinberg, R. A. Epithelial–mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol.* **25,** 675–686 (2015).
13. Chaudhury, A. *et al*. TGF-β-mediated phosphorylation of hnRNP E1 induces EMT via transcript-selective translational induction of Dab2 and ILEI. *Nat. Cell Biol.* **12,** 286–293 (2010).
14. Grelet, S. *et al*. The human NANOS3 gene contributes to lung tumour invasion by inducing epithelial—mesenchymal transition. *J. Pathol.* **237,** 25–37 (2015).
15. Hussey, G. S. *et al*. Identification of an mRNP complex regulating tumorigenesis at the translational elongation step. *Mol. Cell* **41,** 419–431 (2011).
16. Chaudhury, A., Chander, P. & Howe, P. H. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: focus on hnRNP E1's multifunctional regulatory roles. *RNA* **16,** 1449–1462 (2010).
17. Wang, E. T. *et al*. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).
18. Zhang, T. *et al*. PCBP-1 regulates alternative splicing of the CD44 gene and inhibits invasion in human hepatoma cell line HepG2 cells. *Mol. Cancer* **9,** 72 (2010).
19. Akker, S. A. *et al*. Pre-spliceosomal binding of U1 small nuclear ribonucleoprotein (RNP) and heterogenous nuclear RNP E1 is associated with suppression of a growth hormone receptor pseudoexon. *Mol. Endocrinol.* **21,** 2529–2540 (2007).
20. Minn, A. J. *et al*. Genes that mediate breast cancer metastasis to lung. *Nature* **436,** 518–524 (2005).
21. Desmet, F.-O. *et al*. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37,** e67 (2009).
22. Brown, A. S., Mohanty, B. K. & Howe, P. H. Identification and characterization of an hnRNP E1 translational silencing motif. *Nucleic Acids Res.* **44,** 5892–5907 (2016).
23. Tennyson, C. N., Klamut, H. J. & Worton, R. G. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* **9,** 184–190 (1995).
24. Gregory, P. A. *et al*. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* **10,** 593–601 (2008).
25. Chao, C.-H. *et al*. MicroRNA-205 signaling regulates mammary stem cell fate and tumorigenesis. *J. Clin. Invest.* **124,** 3093–3106 (2014).
26. Mani, S. A. *et al*. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133,** 704–715 (2008).
27. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100,** 3983–3988 (2003).
28. Yang, L. Splicing noncoding RNAs from the inside out. *Wiley Interdiscip. Rev. RNA* **6,** 651–660 (2015).
29. Ulveling, D., Francastel, C. & Hubé, F. Identification of potentially new bifunctional RNA based on genome-wide data-mining of alternative splicing events. *Biochimie* **93,** 2024–2027 (2011).
30. Brabletz, S. & Brabletz, T. The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer? *EMBO Rep.* **11,** 670–677 (2010).
31. Li, Y. *et al*. Metastatic heterogeneity of breast cancer cells is associated with expression of a heterogeneous TGFβ-activating miR424-503 gene cluster. *Cancer Res.* **74,** 6107–6118 (2014).
32. Oshimori, N., Oristian, D. & Fuchs, E. TGF-β promotes heterogeneity and drug resistance in squamous cell carcinoma. *Cell* **160,** 963–976 (2015).
33. Kornblihtt, A. R. *et al*. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14,** 153–165 (2013).
34. 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).
35. West, J. A. *et al*. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* **55,** 791–802 (2014).
36. Xing, Z. *et al*. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* **159,** 1110–1125 (2014).
37. Carter, M. S. *et al*. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts *in vivo* is reversed by protein synthesis inhibitors *in vitro*. *J. Biol. Chem.* **270,** 28995–29003 (1995).
38. Laury-Kleintop, L. D., Tresini, M. & Hammond, O. Compartmentalization of hnRNP-K during cell cycle progression and its interaction with calponin in the cytoplasm. *J. Cell. Biochem.* **95,** 1042–1056 (2005).
39. Fischer, K. R. *et al*. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527,** 472–476 (2015).
40. Hu, Y. *et al*. miRNA-205 targets VEGFA and FGF2 and regulates resistance to chemotherapeutics in breast cancer. *Cell Death Dis.* **7,** e2291 (2016).
41. Li, J., Liu, H., Yu, J. & Yu, H. Chemoresistance to doxorubicin induces epithelial-mesenchymal transition via upregulation of transforming growth factor β signaling in HCT116 colon cancer cells. *Mol. Med. Rep.* **12,** 192–198 (2015).
42. Li, Q.-Q. *et al*. Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multidrug resistance and invasive potential in breast cancer cells. *Clin. Cancer Res.* **15,** 2657–2665 (2009).
43. Gu, S., Jin, L., Zhang, F., Sarnow, P. & Kay, M. A. Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.* **16,** 144–150 (2009).
44. Vanharanta, S. & Massagué, J. Origins of metastatic traits. *Cancer Cell* **24,** 410–421 (2013).
45. Wan, L., Pantel, K. & Kang, Y. Tumor metastasis: moving new biological insights into the clinic. *Nat. Med.* **19,** 1450–1464 (2013).
46. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347,** 70–78 (2009).

DOI: 10.1038/ncb3595

METHODS

Cell culture, antibodies, primers and reagents. NMuMG, A549, MCF7, CaCo-2, HMLE, MCF10a and MDA-MB-468 cells were obtained from the American Type Culture Collection (ATCC), and the MDA-231 progression model was graciously provided by J. Massagué (Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM, cat. no. SH30081.01, GE Healthcare Life Sciences) high glucose supplemented with 10% fetal bovine serum and 1% antibiotic/antimycotic solution (penicillin G, streptomycin, amphotericin B). MCF10a cells were obtained from ATCC and cultured in DMEM:F12 supplemented with 5% horse serum, 0.5 μg ml^{-1} hydrocortisone, 10 μg ml^{-1} insulin, 20 ng ml^{-1} epidermal growth factor, 100 ng ml^{-1} cholera toxin and 1% antibiotic/antimycotic. All cells were cultured in a 37 °C, 5% CO_2 incubator. TGFβ2 was a generous gift from Genzyme Corporation. Antibody dilutions, company names, catalogue numbers and clone numbers and their respective dilutions are listed below. Puromycin and G418 were purchased from InvivoGen. Cycloheximide and actinomycin D were purchased from Sigma. Site-directed mutagenesis was performed using a site-directed mutagenesis kit from Life Technologies as described by the manufacturer. Mouse monoclonal anti-E-cadherin (clone 4A2, cat. no. 14472, 1:2,000 dilution), rabbit monoclonal anti-E-cadherin (clone 24E10, cat. no. 3195, 1:2,000 dilution), rabbit monoclonal anti-vimentin (clone D21H3, cat. no. 5741, 1:2,000 dilution), rabbit monoclonal anti-ZEB1 (clone D80D3, cat. no. 3396, 1:1,000 dilution), rabbit monoclonal anti-Ki-67 (clone D2H10, cat. no. 9027, 1:1,600 dilution) and rabbit polyclonal anti-PARP (cat. no. 9542, 1:2,000 dilution) were purchased from Cell Signaling Technology. Mouse monoclonal anti-GAPDH (clone 6C5, cat. no. sc32233, 1:5,000 dilution) and mouse monoclonal anti-HSP90 (clone F-8, cat. no. sc-13119, 1:2,000 dilution) were purchased from Santa Cruz Biotechnology. Rabbit monoclonal anti-vimentin (clone EPR3776, cat. no. Ab92547, 1:2,000 dilution) was purchased from Abcam. Rabbit polyclonal anti-*PNUTS* (cat. no. 24450-1-AP, 1:1,000 dilution) was purchased from Proteintech. Mouse monoclonal anti-hnRNPE1 (clone 1G2 cat. no. H00005093-M01, 1:1,000 dilution) was purchased from Abnova. Mouse monoclonal anti-GFP (clones 7.1 & 13.1, cat. no. 11 814 460 001, 1:500 dilution) was purchased from Roche. The primers used in PCR analysis were purchased from IDT and are as follows:

3′-hCDH1-FTGCCGCCATCGCTTACACCA-5′; 3′-hCDH1-RCCACGCTGGGGTATTGGGGG-5′; 3′-hGAPDH-FTGATGACATCAAGAAGGTGGTGAAG-5′; 3′-hGAPDH-RTCCTTGGAGGCCATGTGGGCCAT-5′; 3′-hPNUTS-v1-FCCAAGCCCCTTTGAAGGGAAA-5′; 3′-hPNUTS-v1-RCTGGGGAAGAAGGTTTGGCTGT-5′; 3′-hPNUTS-v1-v2-flanking-FAAGTACTGTCACCTACGGCTGCC-5′; 3′-hPNUTS-v1-v2-flanking-RGGACGGTCTGCGTCCATTGC-5′; 3′-hPNUTS-v2-boundary-FGTACTGTCACCTACGGCTGCCAAGAAC-5′; 3′-hPNUTS-v2-boundary-RTGCCTTCCTCAGGCCATGTCA-5′; 3′-hPNUTS-v2-boundary2-FTGCCTTCCTCAGGCCATGTCA-5′; 3′-hPNUTS-v2-boundary2-RTGCTGGTTCTTGGCAGCCGT-5′; 3′-hSNAI1-FCCTCAAGATGCACATCCGAAG-5′; 3′-hSNAI1-RACATGGCCTTGTAGCAGCCA-5′; 3′-hSNAI2-FCCCACACATTACCTTGTGTTTGCAA-5′; 3′-hSNAI2-RCAAATGCTCTGTTGCAGTGAGG-5′; 3′-hTWIST-FGGACAAGCTGAGCAAGATTCAGA-5′; 3′-hTWIST-RGTGAGCCACATAGCTGCAG-5′; 3′-hVIM-FCAACGACAAAGCCCGCGTCG-5′; 3′-hVIM-RGCGCAGGGCGTCATTGTTCC-5′; 3′-hZEB1-FGGCAGAGAATGAGGGAGAAG-5′; 3′-hZEB1-RCTTCAGACACTTGCTCACTACTC-5′; 3′-hZEB2-FTCTCGCCCGAGTGAAGCCTT-5′; 3′-hZEB2-RGGGAGAATTGCTTGATGGAGC-5′; 3′-mPNUTS-v1-v2-flanking-FAGGTACTATCGCCGACTGCT-5′; 3′-mPNUTS-v1-v2-flanking-RGGGCGGTCCGTGTCCATGGG-5′.

Transfections. All cell transfections were carried out using 5 μg DNA (or the specified amount) per 8 ml of medium with cells at 70% confluence cultured in 100 mm plates. The transfection reagent Lipofectamine (ThermoFisher Scientific) was used according to the protocol provided by the manufacturer.

Transfection of small interfering RNA. Two specific sequences were designed across the new exon–exon junction generating the lncRNA-*PNUTS*. The sequences were submitted to a BLAST search against the human genome to ensure the specificity of the siRNA to the targeted sequence. Two corresponding scramble duplexes, which do not recognize any sequence in the human genome, were used as controls. The sense and antisense strands were then annealed to obtain duplexes with identical 3′ overhangs. For transfection of the siRNA duplexes, 75,000 cells were seeded in a six-well plates containing 2 ml of culture medium. Twenty-four hours after the seeding, the cells were transfected by phosphate calcium precipitation by adding in each well 200 μl of a mixture containing the siRNA duplexes, 140 mM NaCl, 0.75 nM Na_2HPO_4, 6 nM glucose, 5 mM KCl, 25 mM HEPES and 125 mM $CaCl_2$. Twenty-four hours after transfection, the cells were extensively washed with PBS and incubated for 48 h in culture medium before they were harvested for PCR with reverse transcription (RT–PCR) analyses and western blotting analyses.

ASO oligonucleotide design and usage. Antisense oligonucleotide (IDT) was designed against the splicing site used to generate the lncRNA-*PNUTS* to prevent its usage. The oligonucleotide consists of modified 2′-O-methyl phosphothioate oligonucleotide where each ribose and each phosphate group was modified by a 2′-O-methyl modification or a single sulfur, respectively (sequence: mG*mU*mG*mG* mU*mG*mC* mU*mG*mG* mU*mU*mC* mU*mG). Cells were transfected with the indicated amount of the oligonucleotide 2 h prior to treatment of the cells. For the reverse transcription step, the RNAs were pre-heated in the reaction mix (65 °C/5 min, 75 °C/2 min, 35 °C/30 s) prior to addition of the reverse transcriptase and RT reaction.

Polysome profiling. Cells were extracted in TMK_{100} buffer (10 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 1% (v/v) Triton X-100, 0.5% deoxycholate, 2 mM dithiothreitol, 100 μg ml^{-1} cycloheximide) and then supernatant was collected by centrifugation (14,000 r.p.m. (12,000*g*), 10 min). Cell extracts were layered onto sucrose gradients (10–50%) and centrifuged at 35,000 r.p.m. in a SW40Ti rotor for 3 h at 4 °C. Fractions were collected using a density gradient fractionation system (Teledyne Isco) and then RNA was isolated using Trizol. Monosomal and polysomal fractions were determined by analysis of 18S and 28S rRNA levels using denaturing agarose gel electrophoresis.

Modified Boyden chamber invasion assay. Invasion across a basement membrane was performed using BD BioCoat Matrigel Invasion Chambers (BD Biosciences) as per the manufacturer's instructions. Briefly, a total of 10^5 cells were placed in the upper compartment of the invasion chamber (BD BioCoat Matrigel Invasion Chamber, BD Biosciences) for 24 h at 37 °C. Non-invading cells were removed with a swab and the filters were then fixed in methanol and stained with crystal violet. Quantification of the invasion assay was performed by spectrophotometry after resuspension of the stain.

Western blot analysis. Western blot analysis was performed by standard SDS–PAGE. Whole-cell lysates were prepared from 2–5 × 10^6 cells in 300 μl of lysis buffer (20 mM Tris, pH 7.4, 1% Triton X-100, 10% glycerol, 137 mM NaCl, 2 mM EDTA, 1 mM Na_3VO_4 and protease inhibitors). Lysates were sonicated and clarified by centrifugation at 4 °C for 10 min in a Beckman tabletop microcentrifuge at maximum speed. Typically, 5–20 μg of whole-cell lysates were separated on 10 or 12% acrylamide minigels and transferred to Immobilon-P membrane (Millipore). The membrane was blocked for 1 h in wash buffer (PBS containing 0.1% Tween 20) containing 5% non-fat dry milk followed by an overnight incubation with primary antibody diluted in the same blocking buffer. After extensive washing, the blot was incubated with secondary antibody for 1 h in blocking buffer, washed, and processed using the ECL+ Western blotting detection system (Amersham Biosciences).

Immunofluorescence, FISH and imaging. For immunofluorescence, cells were fixed for 15 min in PBS containing 3.7% (w/v) paraformaldehyde, followed by permeabilization with 0.2% (w/v) Triton X-100. Cells were then incubated for 1 h in 3% BSA and incubated overnight with primary antibody at 4 °C. Then cells were incubated with secondary antibodies conjugated with Alexa Fluor (Life Technologies) at room temperature for 1 h followed by three washes with PBS before analysis with the FV10i confocal laser scanning microscope (Olympus).

For FISH analysis, cells were fixed for 15 min in PBS containing 3.7% (w/v) paraformaldehyde, then slides were incubated overnight at 37 °C in hybridization solution (10% formamide, 2× SSC, 10% dextran sulfate (w/v), 10 μM each probe, labelled with ATTO-488,590 and 649 respectively, IDT). Cells were then washed twice for 30 min at 37 °C with 10% formamide in 2× SSC. DAPI was applied during the second wash. Cells were then rinsed twice with 2× SSC before imaging in 2× SSC buffer.

Immunohistochemistry. Formalin-fixed, paraffin-embedded sections were deparaffinized in xylene, rehydrated in alcohol, and processed as follows. Sections were incubated with target retrieval solution (Dako) in a steamer for 45 min followed by 3% hydrogen peroxide solution for 10 min and protein block (Dako) for 20 min at room temperature. Sections were incubated overnight in a humid chamber at 4 °C with antibody against Ki-67 purchased from Cell Signaling Technology (clone D2H10, cat. no. 9027, 1:1,600 dilution) followed by biotinylated secondary antibody (Vector Laboratories) for 30 min and ABC reagent for 30 min. Immunocomplexes of horseradish peroxidase were visualized by DAB reaction (Dako), and sections were counterstained with haematoxylin before mounting. Micrographs of stained sections were taken using a Leica DMIL LED microscope with an Amscope camera and acquisition software.

Immunoprecipitation assays and biotin pulldown. MS2-TRAP immunoprecipitation assays were performed as described previously[46]. Immunoprecipitated RNA was isolated from beads by addition of Trizol, followed by RT–PCR as described above.

LncRNA-*PNUTS* biotinylation was performed using the Biotin RNA Labelling Mix (Sigma) and the T7 RNA polymerase (Promega) after PCR amplification of lncRNA-*PNUTS* vectors. Biotin pulldowns were performed by using antisense biotinylated probes (IDT) specific to *PNUTS* isoforms. RNA was isolated from beads by addition of Trizol, followed by mRNA or miRNA-specific RT–PCR analysis (Quantimir, System Biosciences).

RNA electromobility shift assays. Recombinant hnRNP E1 protein was prepared as previously described[25] and allowed to incubate for 30 min at 4 °C in RNA–protein binding buffer (40 mM Tris-HCl pH 7.5, 30 mM KCl, 1 mM $MgCl_2$, 0.01% NP40, 1 mM dithiothreitol). After binding, a loading buffer composed of 50% glycerol and bromophenol blue/xylene cyanol was added to samples. Samples were loaded into non-denaturing polyacrylamide gel, electrophoresed and autoradiographed.

Dual-luciferase reporter assays. Transient transfections were performed using XtremeGENE 9 DNA transfection reagent on 50,000 cells plated in a 24-well plate. At 24 h after transfection, the cells were lysed in 100 µl of passive lysis buffer and the firefly luciferase activity and the *Renilla* activity were determined with a luminometer using the Dual Luciferase Assay System (Promega) on 20 µl of lysate. For each experiment, either the firefly luciferase or *Renilla* activity was normalized to either the activity of the *Renilla* or firefly luciferase used as an internal control. The results were expressed as fold induction determined by normalizing each firefly luciferase or *Renilla* value to the internal control value and by dividing these normalized values with the mean normalized value of the corresponding reporter construct transfected with the empty expression vector.

Flow cytometry. Single-cell suspensions of cells were washed three times in PBS containing 1% BSA followed by incubation in 100 µl PBS/1% BSA containing anti-CD24 (PE) and anti-CD44 (FITC) antibody (BD Biosciences) for 2 h at room temperature. Cells were then washed three times in PBS containing 1% BSA and resuspended in 500 µl PBS. Samples were analysed using the BD LSRFortessa Analytic Flow Cytometer. FACS sorting of $CD44^+/CD24^-$ and $CD44^-/CD24^+$ HMLE populations was performed using the FACSAria II Cell Sorter and FACSDiva 6 software (BD Biosciences).

Microarray processing and analysis. Conversion of total RNA into labelled material, mouse genome 430 2.0 GeneChip hybridization, and post hybridization washing, staining and scanning were performed in accordance with Affymetrix protocols by the MUSC Proteogenomics Core Facility. Hybridization data were processed with Affymetrix Expression Console software to obtain normalized hybridization data (RMA algorithm) and detection scores (MAS5 algorithm). These data were imported into dChip software for hierarchical clustering and comparative analysis where a combination of fold change and Student's *t*-test (unpaired) was utilized to identify genes changing significantly for pairwise relationships. Pathway analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) and Molecular Signature Database (MSigDB) platforms. Raw data files were deposited in the NCBI Gene Expression Omnibus (GEO) repository as series GSE94637.

Statistics and reproducibility. Invasion assays were subjected to ANOVA statistical analysis followed by *post hoc* Tukey's multiple comparisons test. For luciferase reporter assays, statistical analysis was performed by two-tailed Student's *t*-test. Human tumour samples analysis were subjected to Pearson correlation score analysis ($df = 24–2$, a Pearson score > 0.515 and $P < 0.05$ was considered as significant). No statistical method was used to predetermine sample size and experiments were not randomized, and we were not blinded to allocation during experiments and outcome assessment. The representative images shown in Figs 1–4 and 6 and Supplementary Figs 1–3 and 5 are representative of at least two independent experiments performed with similar results, excepted for *in vivo* experiments and Supplementary Fig. 3d with only one repeat. All other experiments were repeated two or more times, as indicated in the legends. All the results are expressed as mean ± s.d. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Human samples. The use of human breast tumour tissues and database was approved by the Institutional Review Board for Human Research of the Medical University of South Carolina. Written informed consent from the donors for research use of tissue in this study was obtained prior to acquisition of the specimen. Samples were confirmed to be tumour or normal on the basis of pathological assessment.

Ethics statement. Animals were kept on a 12:12 h light–dark cycle and provided with food and water *ad libitum*. All experiments were performed according to approved protocols of the Institutional Animal Care and Use Committee (IACUC), Medical University of South Carolina.

Data availability. Microarray data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE94637. Source data for Figs 1e, 3e, 4a,h and 6b and Supplementary Figs 1b–3d,e and 4 have been provided as Supplementary Table 2. Unprocessed original scans of blots are shown in Supplementary Fig. 7. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

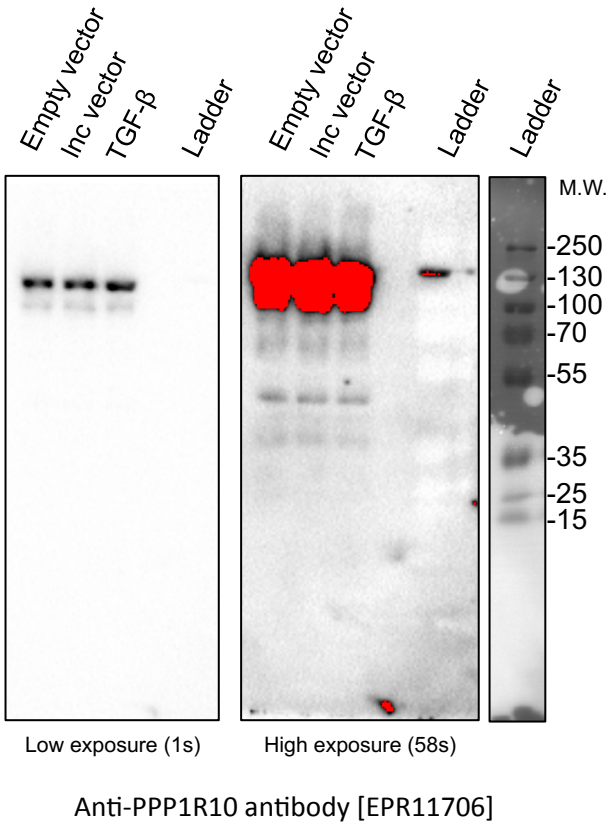
Addendum: A regulated *PNUTS* mRNA to lncRNA splice switch mediates EMT and tumour progression

Simon Grelet, Laura A. Link, Breege Howley, Clémence Obellianne, Viswanathan Palanisamy, Vamsi K. Gangaraju, J. Alan Diehl and Philip H. Howe

Nature Cell Biology **19,** 1105–1115 (2017); published online 19 July 2017; corrected after print 13 November 2017.

In this Addendum, the authors include western blot data using a C-terminal *PNUTS* antibody. This is important in that an annotation of the alternative spliced form of *PNUTS*, denoted in the UCSC genome browser (https://genome.ucsc.edu/), depicts it as a non-coding RNA. However, downstream of the alternative splice site is an alternative AUG located in frame in the *PNUTS* ORF at position 1039. The potential for a protein product of ~61 kDa being generated from this AUG was examined experimentally using a C-terminal raised antibody to *PNUTS* to exclude the possibility that the N-terminal deletion of the splice isoform was not the reason that the predicted 61-kDa protein was not detected in cells using an N-terminal generated antibody. The results presented here confirm our previous results using the N-terminal *PNUTS* antibody and originally presented in Supplementary Fig. 2b of the Article; namely, that this predicted ~61-kDa product is not detectable in cells under the conditions used, even under conditions of overexpression.

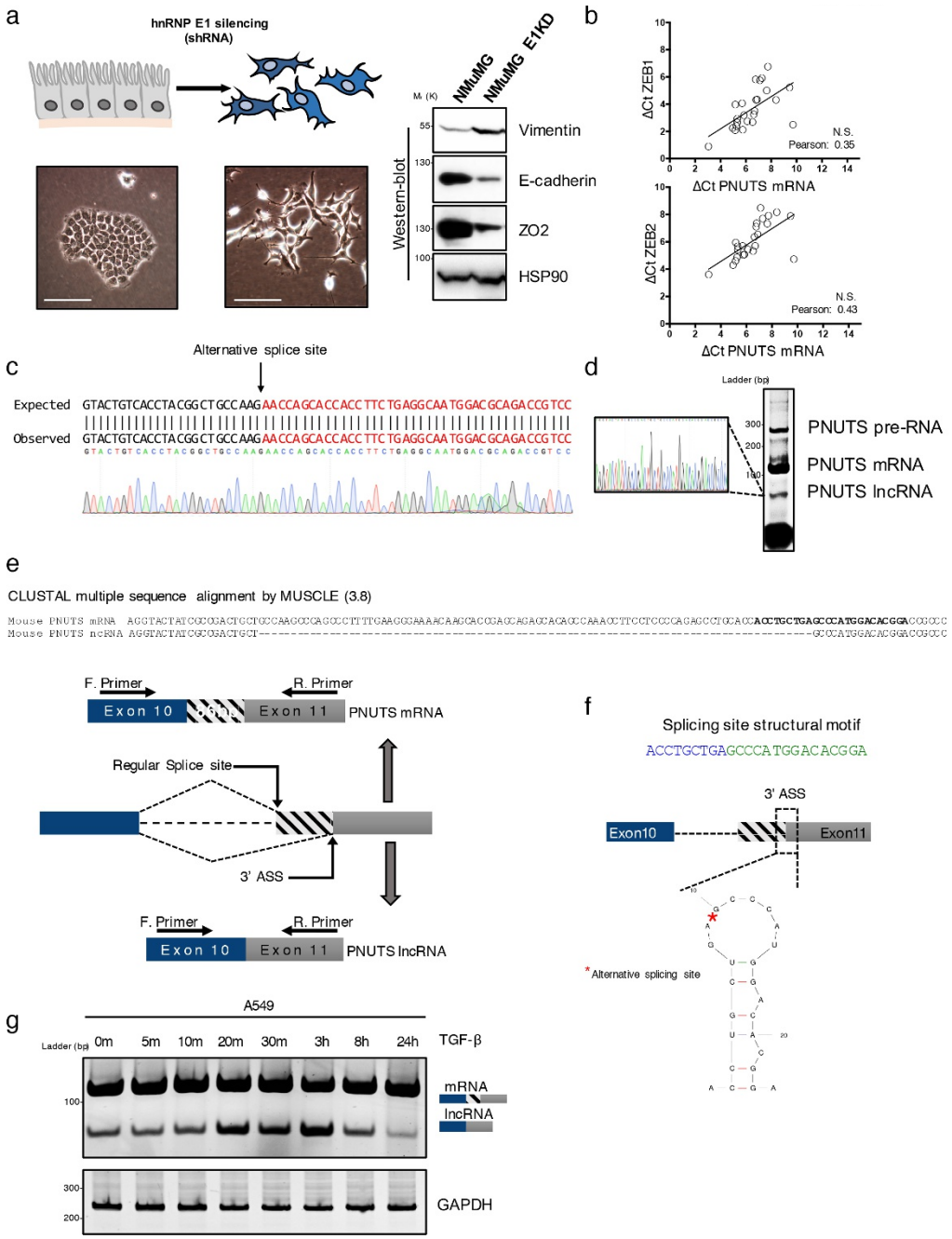
Figure: lncRNA-*PNUTS* does not encode for a N-terminal truncated-protein product. The result of a western blot analysis of PNUTS protein expression in CaCo-2 cells upon transient lncRNA-*PNUTS* expression (3 days) or TGFß treatment (1 day) is shown. The C-terminal antibody used was EPR11706 (Abcam: Ab173285; clone PPP1R10; 1/1000 dilution) raised against the C-terminal region of the PNUTS protein (amino acids 550–650). The western blot protocol and extracts used in this experiment were identical to those described in Supplementary Fig. 2 of the original Article.



nature cell biology

DOI: 10.1038/ncb3595

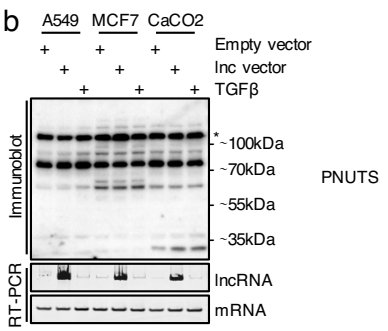
In the format provided by the authors and unedited.



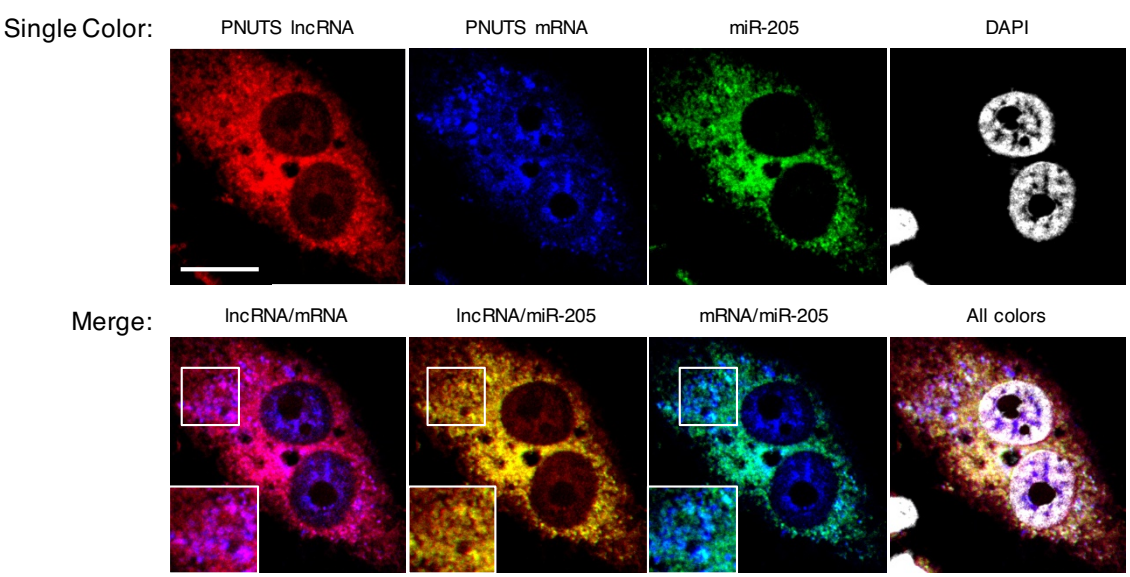
Supplementary Figure 1 Characterization of the EMT induced by hnRNP E1 knockdown and validation of the splicing model in mouse. (a) EMT induction following shRNA-mediated silencing of hnRNP E1 in NMuMG cells. Imaging of cells reveals a significant morphological change in cellular phenotype from an epithelial-like to a mesenchymal-like phenotype. EMT was validated by immunoblotting analysis using antibodies to the mesenchymal marker vimentin and the epithelial markers ZO2 and E-cadherin. Scale bar: 100μM. (b) Quantitative RT-PCR analysis of mRNA-PNUTS, ZEB-1 and ZEB-2 expression in 24 human breast tumor samples. Relative expression levels of transcripts were calculated using the ΔCt method normalizing to GAPDH. Correlations between transcript expression levels were then evaluated using Pearson correlation coefficient test. (Linear regression, df=24-2, a Pearson score > 0.515 and p<0.05 was considered as significant). Source data are available in Supplementary table 2. (c) Sanger sequencing result of the lower band obtained by end-point RT-PCR in Figure 1h. (d) Identification of an alternative splice product corresponding to the lncRNA-PNUTS in NMuMG cells. Sequence alignment of human and murine genomes indicates that exon-11 of the murine genome matches with exon 12 of the human genome. RT-PCR amplification of exon10-exon11 junction demonstrates a potential alternative splice product. The lower band was cloned and then sequenced using Sanger technology. (e) Sequencing results indicate a splicing pseudosite located 86 nucleotides downstream of the regular splicing site (Top). Schematic representation of the alternative splice region of the PNUTS variants based on Sanger sequencing results (Bottom). (f) Secondary structure of the BAT-like element located in the alternative splicing site of murine PNUTS RNA as predicted using the Mfold algorithm (ΔG= -2.10 kcal mol-1). (g) Extended time course experiment using RT-PCR analysis of *PNUTS* gene processing following addition of TGFβ in A549. GAPDH was used as a loading control.

a

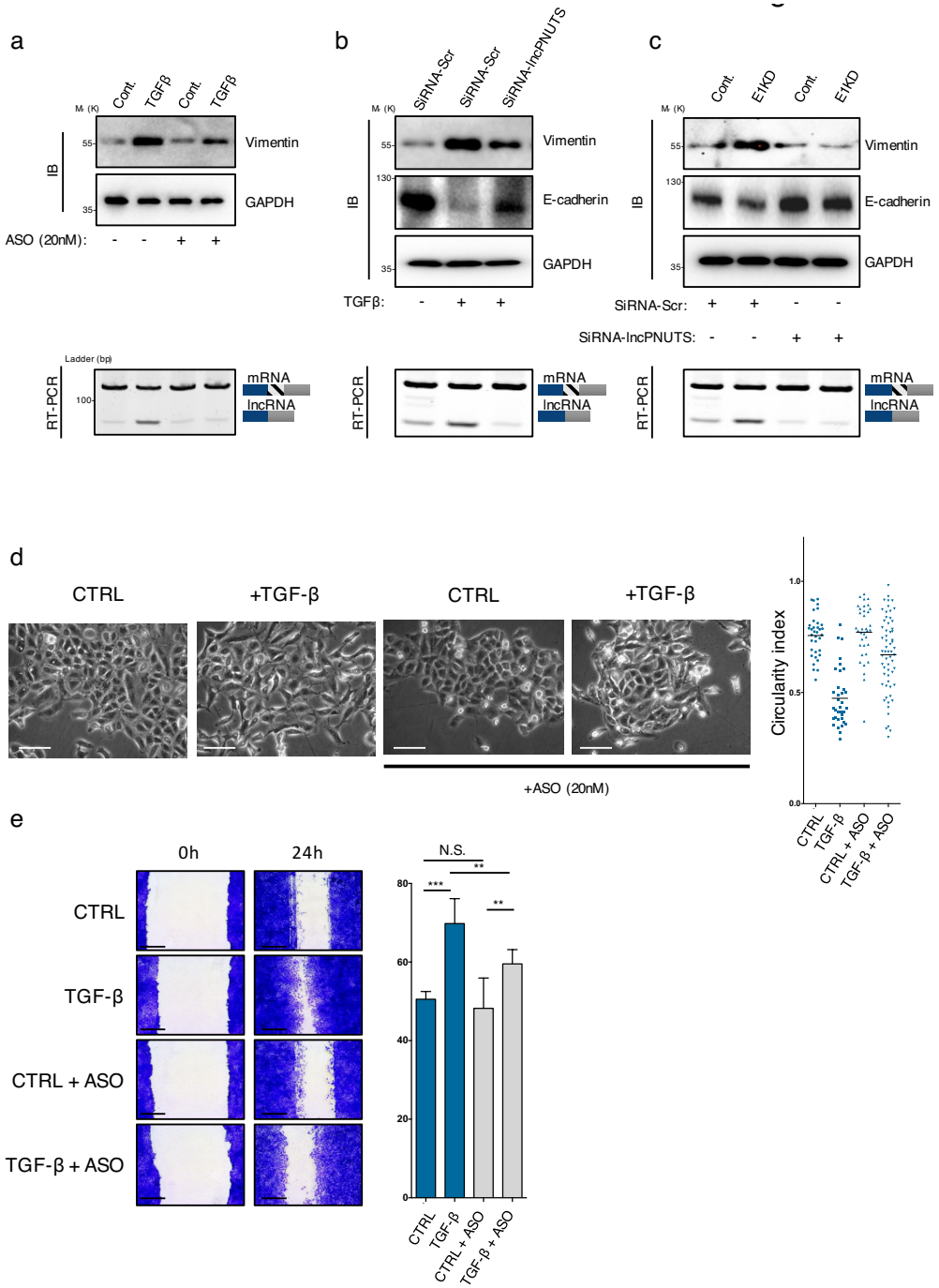
Position	Splice site type	Motif	New potential splice site	Consensus value
-12	**Acceptor**	**ctttcacaatagCC**	**ctttcacaatagCC**	79.38
-7	Acceptor	acaatagCCAAGCC	acaatagccaagCC	66.43
5	Acceptor	GCCCCTTTGAAGGG	gccccttgaagGG	81.41
20	Acceptor	AAACGAGCACAGAA	aaacgagcacagAA	74.41
27	Acceptor	CACAGAACCAAGCA	cacagaaccaagCA	67.16
32	Acceptor	AACCAAGCACAGCC	aaccaagcacagCC	75.73
50	***Acceptor***	***CTTCTTCCCCAGAA***	***cttcttccccagAA***	91.74
56	Acceptor	CCCCAGAACCAGCA	ccccagaaccagCA	77.88
70	Acceptor	CCACCTTCTGAGGC	ccaccttctgagGC	78.75
83	Acceptor	CAATGGACGCAGAC	caatggacgcagAC	77.57
92	Acceptor	CAGACCGTCCAGGC	cagaccgtccagGC	84.2
109	Donor	CCGGTTCCC	CCGgttccc	67.23
116	Acceptor	CCCCTGTTGAAGTC	cccctgttgaagTC	75.97
124	Acceptor	GAAGTCCCGGAGCT	gaagtcccggagCT	70.04
137	Acceptor	TCATGGATACAGgt	tcatggatacagGT	85.83
146	Acceptor	CAGgtaatctagaa	caggtaatctagAA	70.47
146	Donor	CAGgtaatc	CAGgtaatc	85.51 (WT)



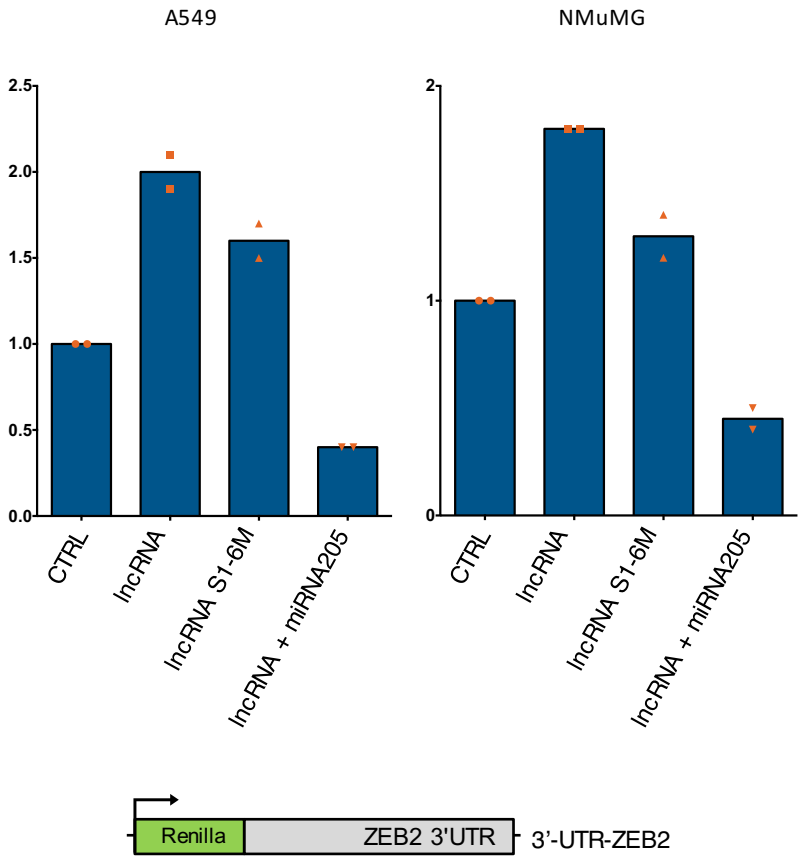
c



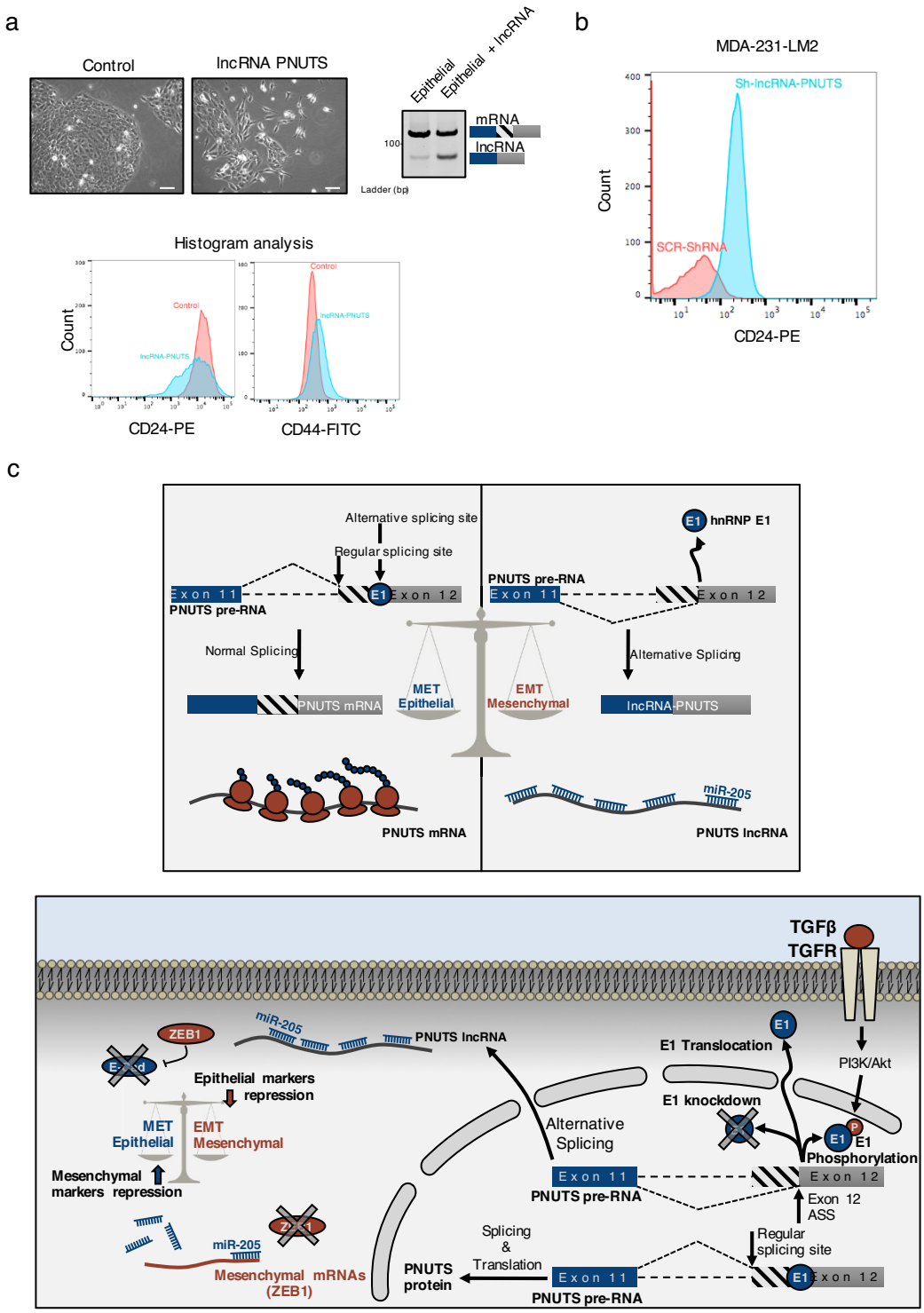
Supplementary Figure 2 The predicted lncRNA-PNUTS does not encode protein, is both nuclear and cytoplasmic and colocalizes with miRNA-205 (a) *In silico* prediction of PNUTS alternative splicing sites. Lower case: intronic sequences; Upper case: exonic sequences; Bold characters: natural acceptor splice site; Bold italicized characters: newly identified pseudosite. (b) Western-blot (top) analysis of PNUTS protein expression upon *PNUTS* predicted-lncRNA transient overexpression (3 days) or TGFβ treatment (1 day) in A549, MCF7 and CaCO2 cell lines. RT-PCR analysis (bottom) was used to monitor the PNUTS predicted-lncRNA overexpression.* indicates the PNUTS protein band. The fact that the lncRNA does not appear upregulated in this experiment is due to the time point assayed (24 h TGFβ treatment) (see Supplementary figure 1g). (c) Fluorescence *in situ* hybridization analysis (FISH) of the endogenous form of lncRNA-PNUTS, mRNA-PNUTS and miR-205 was performed using selective probes labeled with ATTO dyes. Images at 300x magnifications, obtained by using confocal microscopy. Scale bar: 10μM



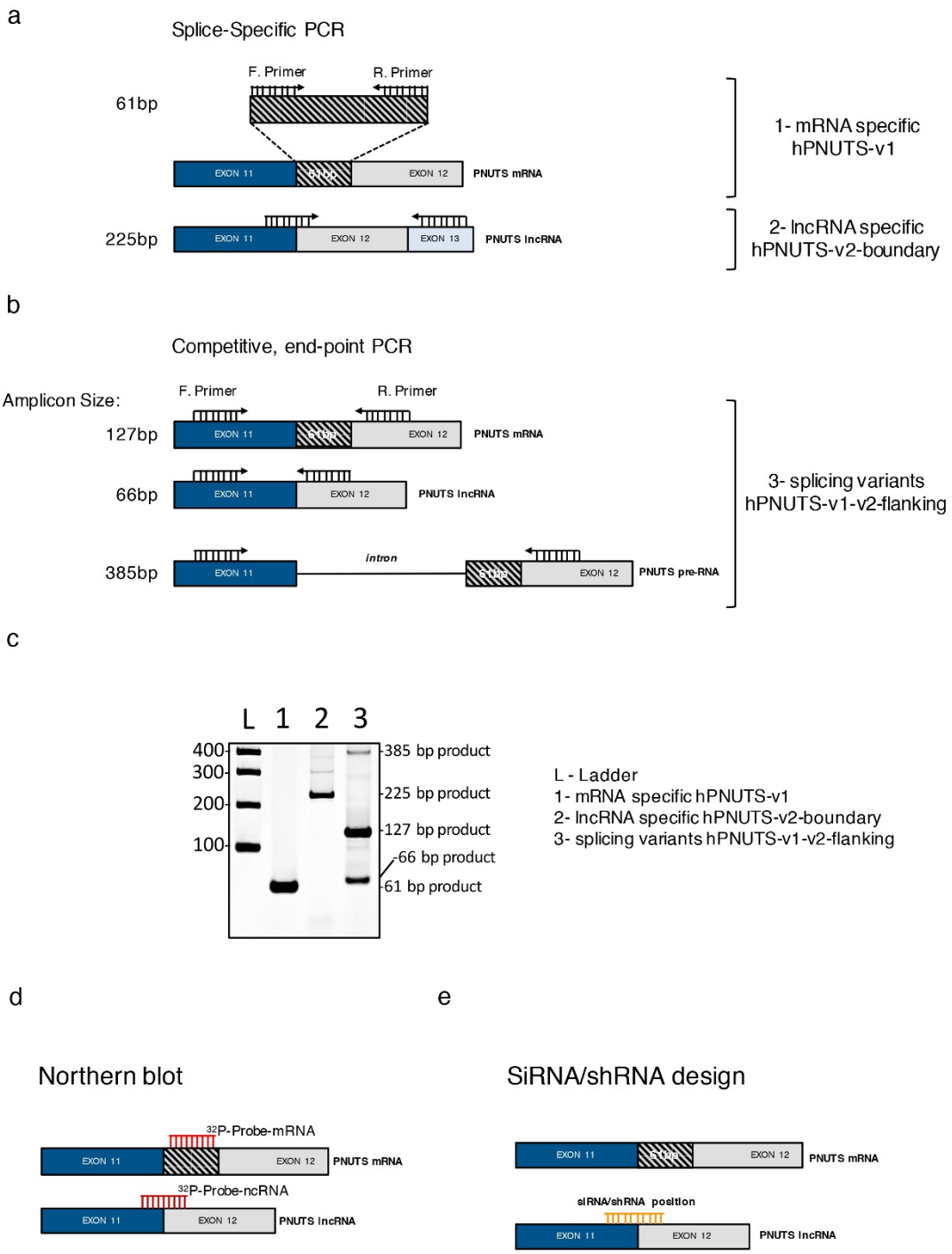
Supplementary Figure 3 The lncRNA-PNUTS silencing or splicing inhibition prevents both TGF-ß and E1KD-mediated EMT (a) The alternative splicing oligo (ASO) designed to block alternative splicing of PNUTS was used to inhibits TGFß-induced EMT in A549 cells (24 h TGFß treatment). The mesenchymal marker Vimentin was used to monitor the EMT induced by TGFß. (b) siRNA selectively targeting the lncRNA-PNUTS isoform was used to prevent TGFß-induced EMT in A549 cells or EMT occurring following transient hnRNP E1 knockdown in HMLE cells in (c). Vimentin (mesenchymal marker) and E-cadherin (epithelial marker) were used to monitor the EMT induced by TGFß in A549 cells or by E1KD in HMLE cells. (d) Changes in cells morphologies were assessed by bright-field microscopy and quantified by determining the cell circularity index of the cells with a decrease in cell circularity reflecting the acquisition of a more mesenchymal phenotype. Cellular circularity was measured using ImageJ software according to the following formula: "circularity = 4π (area/perimeter^2)". Data are from a single experiment, where 35, 35, 35 and 61 cells were scored per condition. Scale bar: 50µM. Source data are available in Supplementary table 2. **(e)** The impact of the ASO on TGF-ß induced EMT was monitored on A549 cells migration. We observed that ASO transfection significantly impaired the TGF-ß induced 2D migration of the cells. (mean ± s.d., n=5 fields quantified, pooled from 2 independent experiments, two-tailed Student t test, **p<0.01; ***p<0.001, NS, not significant). Scale bar: 200µM. Source data are available in Supplementary table 2. Unprocessed original scans of blots are shown in Supplementary Figure 7



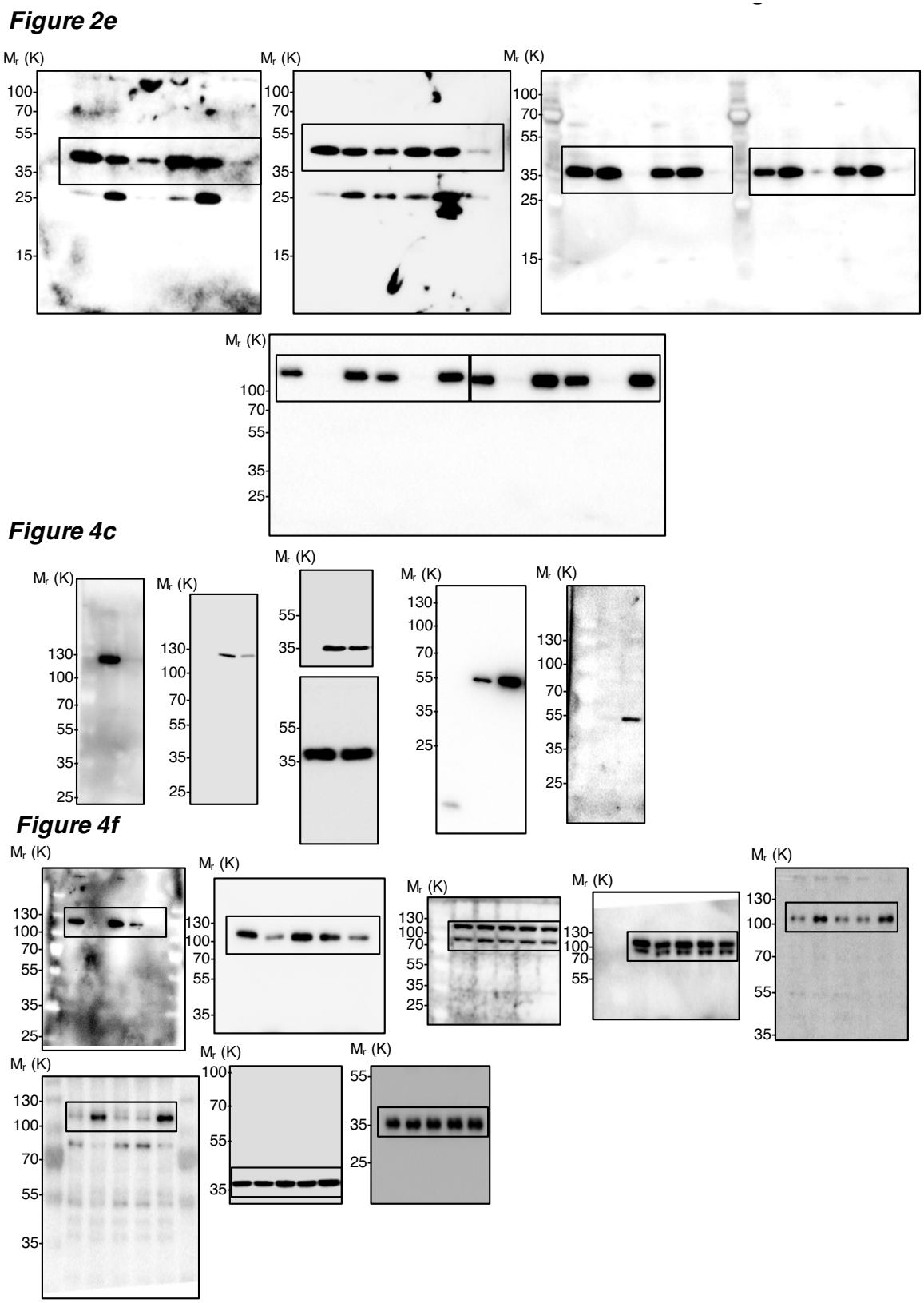
Supplementary Figure 4 The lncRNA-PNUTS regulates ZEB-1 3'UTR through its control of miRNA-205 A Renilla reporter containing 3'-UTR of ZEB2 cloned downstream of Renilla was transfected into A549 and NMuMG cells overexpressing wild-type (lncRNA) or mutated (lncRNAS1-6M) constructs of lncRNA-PNUTS and treated +/- synthetic miR-205 mimic. For each condition, Renilla luciferase activity was normalized to Firefly luciferase reporter. Data shown are from two experiments with a bar representing the mean.



Supplementary Figure 5 Impact and regulation of the lncRNA on stem cell properties (a) (Top) Cell morphology of control and lncRNA overexpressing HMLE cells were assessed by phase-contrast microscopy and lncRNA-PNUTS overexpression was validated by RT-PCR. Scale bar: 50μM. (Bottom) Histogram analysis related to FACS analysis presented in Figure 6g. (b) Flow cytometry analysis of CD24 cell surface expression levels in MDA231-LM2 control (Scr-ShRNA) and shRNA-lncRNA-PNUTS. (c) Graphical abstract of the study. hnRNP E1 inhibits the alternative splicing of PNUTS by directly binding to the alternative splicing site. Loss of binding following hnRNP E1 knockdown, phosphorylation or its translocation to the cytoplasm allows for alternative splicing of *PNUTS* and thus generates lncRNA-PNUTS. This lncRNA competes for miR-205 and thus contributes to the EMT by allowing expression of mesenchymal markers such as Zeb1, normally targeted by this miRNA.

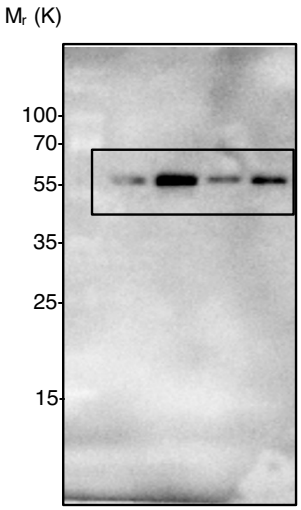


Supplementary Figure 6 Primers and probes design (a) Design of the splice-specific primers used either in end-point RT-PCR or quantitative RT-PCR. Primers were designed to overcome primer competition in order to analyze the specific expression of either mRNA-PNUTS or lncRNA-PNUTS isoforms. (b) Design of the primers used in competitive end-point RT-PCR to analyze the relative expression between mRNA-PNUTS, lncRNA-PNUTS and preRNA-PNUTS on the same PCR reaction. (c) Validation of the reliability of the primer sets presented in (a) and (b) by end-point RT-PCR (d) Design of the probes used for Northern-blot experiments to discriminate mRNA-PNUTS from lncRNA-PNUTS (e) Design of siRNA or shRNA used to selectively target the lncRNA isoform of PNUTS. Probes were designed to target the new exon11/exon12 splice junction specific to the lncRNA-PNUTS.



Supplementary Figure 7 Unprocessed original scans of blots

Figure S3 a



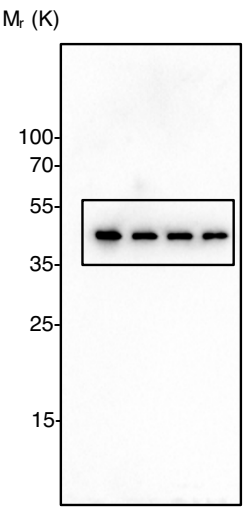
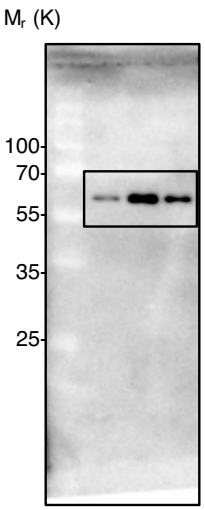
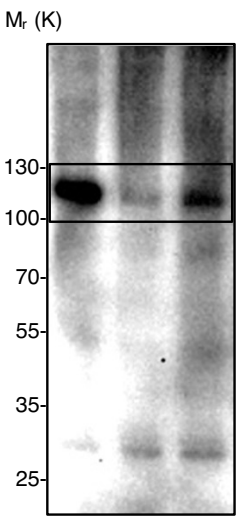


Figure S3 b





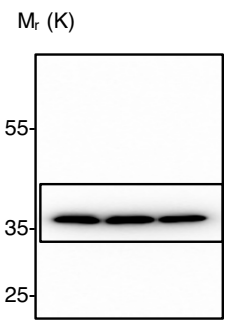
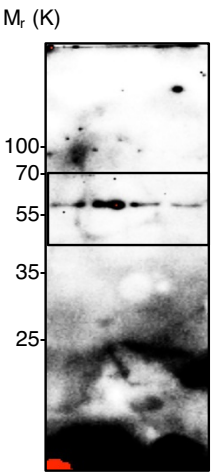
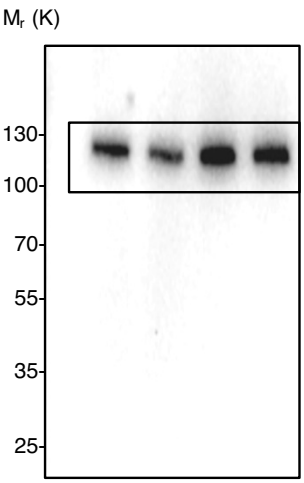
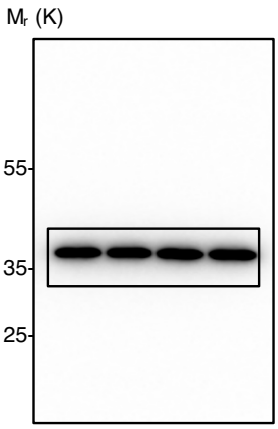


Figure S3 c







Supplementary Figure 7 Continued

Supplementary Tables Legends

Supplementary Table 1 List of predicted MiRNA binding sites to PNUTS lncRNA

Supplementary Table 2 Statistics Source Data

nature research

Corresponding Author: Philip HOWE

Date: 14/07/17

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   No statistical test was used to determine sample size.

2. Data exclusions

   Describe any data exclusions.

   No inclusion/exclusion criteria were used for samples or animals. However, 1 mouse of the MDA-231-SCR group was dead due to fast tumor progression of the MDA231-SCR cells (Fig6d). This does not impair our results and even supports it.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   Attempts at replication were succefull

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   No randomization was used in the study

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   No blinded experiments were conduced in this study.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	In silico prediction of MiR-205 binding sites was performed by using the DIANA-microT web server. Hybridization data were processed with Affymetrix Expression Console software. This data was imported into dChip software for hierarchical clustering and comparative analysis. Splicing sites scores were obtained by using the HSF finder software.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All the relevant materials of this study are available from the author.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Mouse monoclonal anti-E-cadherin (Clone [4A2], Cat. No. #14472, 1:2,000 dilution), Rabbit monoclonal anti-E-cadherin (Clone [24E10], Cat. No #3195, 1:2,000 dilution), Rabbit monoclonal anti-vimentin (Clone [D21H3], Cat. No. #5741, 1:2,000 dilution), Rabbit monoclonal anti-ZEB1 (Clone [D80D3], Cat. No. #3396, 1:1,000 dilution), Rabbit monoclonal anti-Ki67 (Clone [D2H10], Cat. No. #9027, 1:1,600 dilution) and Rabbit polyclonal anti-PARP (Cat. No. #9542, 1:2,000 dilution) were purchased from Cell signaling technology company. Mouse monoclonal anti-GAPDH (Clone [6C5], Cat. No. sc32233, 1:5,000 dilution) and Mouse monoclonal anti-HSP90 (Clone [F-8], Cat. No. sc-13119, 1:2,000 dilution) were purchased from Santa Cruz Biotechnology company. Rabbit monoclonal anti-vimentin (Clone [EPR3776], Cat. No. Ab92547, 1:2,000 dilution) was purchased from Abcam company. Rabbit polyclonal anti-PNUTS (Cat. No. 24450-1-AP, 1:1,000 dilution) was purchased from Proteintech company. Mouse monoclonal anti-hnRNPE1 (Clone [1G2] Cat. No. H00005093-M01, 1:1,000 dilution) was purchased from Abnova company. Mouse monoclonal anti-GFP (clones [7.1 & 13.1], Cat. No. 11 814 460 001, 1:500 dilution) was purchased from Roche company. Positive and/or negative controls such as protein silencing, overexpression or TGF-beta treatment of the cells were used to validate the relevant antibodies used in this work.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	NMuMG, A549, MCF7, CaCo-2, HMLE, MCF10a and MDA-MB-468 cells were obtained from the American Type Culture Collection (ATCC), and the MDA231 progression model was graciously provided by Dr. Joan Massagué.
b. Describe the method of cell line authentication used.	The cell lines were not authenticated
c. Report whether the cell lines were tested for mycoplasma contamination.	Our lab regularly checks the cell lines for mycoplasma contamination by using the "PCR Mycoplasma Test Kit" from Promokine manufacturer.
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No cell lines used in this study were found in the database of commonly misidentified cell lines that is maintained by ICLAC and NCBI Biosample.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

6-8 week-old female mice of the following strain were used : NOD.CB17-Prkdcscid/J (Jackson)
Animals were kept on a 12:12 h light–dark cycle and provided with food and water ad libitum. All experiments were performed according to approved protocols of the Institutional Animal Care and Use Committee (IACUC), Medical University of South Carolina.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The use of human breast tumor tissues and database were approved by the Institutional Review Board for Human Research of the Medical University of South Carolina. Written informed consent from the donors for research use of tissue in this study was obtained prior to acquisition of the specimen. Samples were confirmed to be tumor or normal based on pathological assessment.

Corresponding Author: Philip HOWE

Date: Jun 8, 2017

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Single-cell suspensions of HMLE cells were washed three times in PBS/1% BSA followed by incubation in 100 µl PBS/1% BSA containing anti-CD24 (PE) and anti-CD44 (FITC) antibody (BD Biosciences) for 2 hours at room temperature. Cells were then washed three times in PBS/1% BSA and resuspended in 500 µl PBS.
6. Identify the instrument used for data collection.	Samples were analyzed using the BD LSRFortessa Analytic Flow Cytometer.
7. Describe the software used to collect and analyze the flow cytometry data.	FACS sorting of CD44+/CD24- and CD44-/CD24+ HMLE populations was performed using the FACSAria II Cell Sorter and FACSDiva™ 6 software (BD Biosciences).
8. Describe the abundance of the relevant cell populations within post-sort fractions.	The Epithelial/Mesenchymal subpopulations ratio was aroung 95:5
9. Describe the gating strategy used.	Cells were first selected by using SSC-A/FSC-A gating followed by a subsequent FSC-H/FSC-A gating to remove aggregates and dying cells. Unstained cells (Antibodies omission), one stain controls and cells expressing a high level of either CD24 or CD44 were used as positive controls and negative controls to define the gates. (a figure exemplifying the gating strategy is provided in the "Flow Cytometry Gating strategy" document)

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒