Linc00941 Is a Novel Transforming Growth Factor β Target That Primes Papillary Thyroid Cancer Metastatic Behavior by Regulating the Expression of Cadherin 6

Linc00941 is a novel TGFβ target that primes papillary thyroid cancer metastatic behavior by regulating the expression of Cadherin 6

Mila Gugnoni[1*], Veronica Manicardi[1,2+], Federica Torricelli[1*], Elisabetta Sauta[1,3*], Riccardo Bellazzi[3*], Gloria Manzotti[1*], Emanuele Vitale[1+], Dario de Biase[4*], Simonetta Piana[5†] and Alessia Ciarrocchi[1*]

* PhD + Master's degree † MD

[1] Laboratory of Translational Research, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy

[2] Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, Modena, Italy

[3] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[4] Department of Pharmacy and Biotechnology, Molecular Pathology Unit, University of Bologna, viale Ercolani 4/2, 40138, Bologna, Italy

[5] Pathology Unit, Department of Oncology and Advanced Technologies, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy

mila.gugnoni@ausl.re.it, veronica.manicardi@ausl.re.it, federica.torricelli@ausl.re.it, elisabetta.sauta@ausl.re.it, riccardo.bellazzi@unipv.it, gloria.manzotti@ausl.re.it, emanuele.vitale@ausl.re.it, dario.debiase@unibo.it, simonetta.piana@ausl.re.it, alessia.ciarrocchi@ausl.re.it

RUNNING TITLE

linc00941 in papillary thyroid carcinoma aggressiveness

KEYWORDS

Papillary Thyroid Cancer; Metastases; lncRNAs; CDH6; TGFβ; Autophagy

ABSTRACT

Background: Papillary thyroid cancers (PTCs) are common, usually indolent malignancies. Still, a small but significant percentage of patients have aggressive tumors and develop distant metastases leading to death. Currently, it is not possible to discriminate aggressive lesions due to lack of prognostic markers. LncRNAs which are selectively expressed in a context-dependent manner are expected to represent a new landscape to search for molecular discriminants. TGFβ is a multifunctional cytokine that fosters epithelial to mesenchymal transition and metastatic spreading. In PTCs, it triggers the expression of the metastatic marker Cadherin 6 (CDH6). Here, we investigated the TGFβ-dependent lncRNAs that may cooperate to potentiate PTC aggressiveness.

Methods: We used a genome wide approach to map enhancer-associated lncRNAs under TGFβ control. Linc00941 was selected and validated using functional *in vitro* assays. A combined approach using bioinformatic analyses of the thyroid cancer (THCA)-the cancer genome atlas (TCGA) dataset and RNA-seq analysis was used to identify the processes in which linc00941 was involved in and the genes under its regulation. Correlation with clinical data was performed to evaluate the potential of this lncRNA and its targets as prognostic markers in thyroid cancer.

Results: Linc00941 was identified as transcribed starting from one of the TGFβ-induced enhancers (ENHs). Linc00941 expression was significantly higher in aggressive cancer both in the TCGA dataset and in a separate validation cohort from our institution. Loss of function assays for linc00941 showed it promotes response to stimuli and invasiveness while restraining proliferation in PTC cells, a typical phenotype of metastatic cells. From the integration of TCGA data and linc00941 knockdown RNA-seq profiling, we identified 77 genes under the regulation of this lncRNA. Among these, we found the pro-metastatic gene *CDH6*. Linc00941 knockdown partially recapitulates the effects observed upon CDH6 silencing, promoting cell cytoskeleton and membrane adhesions rearrangements and autophagy. The combined expression of CDH6 and linc00941 is a distinctive features of highly aggressive PTC lesions.

Conclusions: Our data provide new insights into the biology driving metastasis in PTCs and highlight how lncRNAs cooperate with coding transcripts to sustain these processes.

INTRODUCTION

Papillary thyroid carcinoma (PTC) is the most common thyroid cancer subtype, usually characterized by slow proliferation and rare distant metastatic spread, resulting in a favorable prognosis for most patients (1). However, nearly 5% of PTCs behave aggressively and are associated with distant metastasis. PTCs are characterized by low mutational load. Although, low-risk PTC could be sufficiently treated with an ipsilateral lobectomy, some patients undergo aggressive treatment regimens, including total thyroidectomy and radioactive iodine (RAI) administration (2). The major clinical challenge in this setting, remains to find molecular features that correlate with PTC aggressiveness to better stratify and manage patients. In the last decade, some markers have been proposed. Re-expression of telomerase is a common feature in tumor, somatic mutations in telomerase reverse transcriptase (TERT) coding region are uncommon, but germline and somatic mutations in non-coding TERT promoter region were found in many cancers. The mutations occurring −124bp (C228T) and −146bp (C250T), enhancing promoter activity, were described as cancer hotspot mutations and proposed, alone or in combination with other markers, as prognostic markers to discriminate aggressive tumors (3–9). Dissecting the molecular bases underlying PTCs metastatic spreading is pivotal to implement predictive aggressiveness signatures and individualize patients' treatment.

The dual role of transforming growth factor β (TGFβ) in cancer is widely recognized. In the early phases of tumor development, TGFβ acts as repressor signal, restraining proliferation and favoring apoptosis. During tumor progression, on the contrary, TGFβ promotes metastatic spread by inducing epithelial-to-mesenchymal transition (EMT) (10–13). We previously demonstrated that TGFβ partakes in the induction of aggressive phenotype in PTCs by promoting EMT and inducing a transcriptional program that support cellular invasiveness and motility as well as changes in the metabolic condition of cancer cells (14).

A major mediator of TGFβ signaling in PTCs is Cadherin 6 (CDH6), a mesenchymal type-II cadherin, we showed to be a marker of metastatic spreading and reduced survival probability in PTCs (13, 15). Mechanistically, we demonstrated that CDH6 promotes EMT completion by inhibiting autophagy (15). This cadherin coordinates profound changes in cytoskeleton architecture and cell-cell interaction, which serve cancer cells to escape the

tight epithelial organization and overcome the mechanical stress that invasion into adjacent sites requires. CDH6 also blocks autophagy by sequestering GABARAP and restraining the formation of autophagosomes and the negative effect of this catabolic process on cell motility (15).

The improved ability of understanding the genome revolutionized the idea about its functioning and brought to light the existence of an impressive number of functionally relevant non-coding transcripts (16). Among these, lncRNAs are a class of non-coding RNAs longer than 200 bp sharing characteristics with coding-transcripts, including intron-exon organization and transcription by RNA Polymerase II (17). Enhancer (ENH)-associated lncRNAs are a peculiar class of lncRNAs transcribed from active ENHs. The current hypothesis is that they represent a further and finer layer of regulation of ENH activity in the gene expression processes (18). Indeed, the expression of ENH-associated lncRNAs correlates with transcription of specific protein-coding genes, implicating a role in their expression regulation (19, 20).

LncRNAs expression is strictly regulated in space and time and this specificity is thought to contribute to many physiological and pathological cellular processes (21).

The large number and high expression specificity of lncRNAs in cancer make these molecules promising candidates as diagnostic and prognostic markers. The role of lncRNAs in different tumor settings and the possibility to exploit these molecules, also in signatures composed by coding and non-coding transcripts, are currently under investigation. Despite the increasing number of studies on this topic, the role of lncRNAs in mediating TGFβ signaling during cancer progression is not completely known.

In this study, we explored the genome-wide landscape of lncRNAs under TGFβ regulation to define their role in mediating metastatic phenotype in thyroid cancer. Among the non-coding transcripts identified, we functionally characterized linc00941 showing that this molecule partakes in TGFβ signaling by regulating CDH6 expression, a crucial mediator of PTC invasiveness and metastatic spread.

MATERIALS AND METHODS

Cell culture and Transfection

B-CPAP, TPC1, Cal62 and 8505c cell lines were obtained from Dr. Santoro, University of Naples, Nthy-ori3-1 from Dr. Greco, Fondazione IRCCS-INT, Milan, FTC133

was purchased from Sigma-Aldrich (St. Louis, Missouri, USA) and SW579 was purchased from ATCC (Manassas, Virginia, USA). Cells were grown at 37°C and 5% CO_2, in DMEM, RPMI or DMEM Ham's F12 1:1(Gibco) with 10% FBS. 20 nM of either small interference RNA oligonucleotide against linc00941 (exon 5) (LOC40019 Silencer Pre-designed siRNA, Ambion, Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Silencer Select Negative Control (Thermo Fisher Scientific) were used to silence linc00941 and as a negative control, respectively. Lipofectamine RNAiMax and Lipofectamine 2000 (Thermo Fisher Scientific) were used for siRNAs and plasmids transfection, respectively. Plasmids used are listed in Supplementary Table 5. Plasmids used were pEGFP-C3_LC3B, p3XFlagCMV10_GABARAP and pCDNA-3.1.

TGFβ-induction and ChIP-seq

Nthy-ori3-1, TPC1, B-CPAP and Cal62 cells were starved with 1% FBS growth medium overnight, then treated with TGFβ (Peprotech, Rocky Hill, New Jersey, USA) for 24h. Chromatin immunoprecipitation (ChIP) on Nthy-ori3-1 was performed as previously described (22). Briefly, 10^7 Nthy-ori3-1 cells were crosslinked for 15 minutes with 1% formaldehyde, lysed and sonicated for 10 cycles (30 minutes ON, 30 minutes OFF) using Bioruptor Pico Sonicator (Diagenode, Liege, Belgium), to obtain 100-200 base pair chromatin fragments. Chromatin was precipitated overnight using Magna ChIP Protein G Magnetic Beads (Merck, Darmstadt, Germany) and 1µg of H3K27ac (Abcam, Trumpington, UK) or Rabbit IgG (Santa Cruz Biotechnology, Dallas, Texas, USA) antibodies. 1% of total chromatin was used as input. Samples were quantified at Qubit (Thermo Fisher Scientific) and the quality was evaluated by Bioanalyzer (Agilent, Santa Clara, California, USA). Library for sequencing were obtained following the TruSeq ChIP Sample Preparation protocol (Illumina Inc, San Diego, California, USA) using 5-10 ng ChIP DNA as starting material. Triplicates were sequenced on Illumina NextSeq500 high-output cartridge (single stranded, reads length 75 bp – 1x75).

RNA extraction, q-RT-PCR and RNA-seq

Matched normal and tumor fresh frozen patients' samples were obtained from the Research Tissue BioBank of the azienda unità sanitaria locale (AUSL) - istituto di ricovero e cura a carattere scientifico (IRCCS) of Reggio Emilia after written informed consent was

obtained. The project was approved by the local Ethical Committee (protocol n: 2014/0014425 of 06/05/2014).

RNA from patients' samples and from cell lines was extracted using TRIzol (Thermo Fisher Scientific) or by automatic extractor MaxwellRSC (simply RNA Cells kit, Promega, Madison, Wisconsin, USA). Bio-Rad (Hercules, California, USA) iScript cDNA kit was used for reverse transcription and Bio-Rad Sso Fast EvaGreen SuperMix was used to perform q-RT-PCR, read at Bio-Rad CFX96 Real Time PCR Detection System. Relative expression of target genes was calculated by the $2^{-\Delta\Delta Ct}$ method normalizing on the geometric mean of reference genes CyclophilinA (CYPA) or Beta-D-Glucuronidase (GUSB). Primers sequences are listed in Supplementary Table 6. RNA-seq libraries were obtained starting from 500ng of total RNA following Illumina TruSeq Stranded TotalRNA preparation protocol. Sequencing was performed using Illumina NextSeq high-output cartridge (double stranded, reads length 75 bp – 2x75).

Next-generation sequencing analysis

Next generation sequencing (NGS) was performed using a customized panel of genomic regions and sequenced by MiSeq Sequencer (Illumina Inc), as previously described (23).

Western Blot and Immunofluorescence

Cells ($5x10^5$ cells) were lysed in Passive Lysis Buffer (Promega). SDS-PAGE gels using 15-30 μg of lysates were used. Proteins were transferred by wet transfer method or by semi-dry method using Transblot Turbo (Bio-Rad). The membranes were blocked in 5% milk (Bio-Rad) and incubated with primary antibodies for 1 hour to overnight. Primary antibodies used for the study were: 1:5000 anti-β-actin (Sigma Aldrich), 1:1000 anti-phosphorylated polymerase II (Ser5), 1:1000 anti-p62/SQSTM1 (Abcam), 1:1000 anti-cJun, 1:1000 anti-LC3A/B, 1:1000 anti-GABARAP (Cell Signaling Technology, Danvers, Massachusetts, USA), 1:1000 anti-β-catenin (BD Bioscience) (For more information see Supplementary Table 7). Secondary antibodies were HRP-conjugated donkey anti-rabbit and sheep anti-mouse (NA934V and NXA931V GE Healthcare, Chicago, Illinois, USA. Used 1:5000). Membranes were incubated with secondary HRP conjugated antibodies for 1 hour. Bio-Rad Clarity ECL was used for detection using ChemiDoc Imager (Bio-Rad)

following the manufacturer's instructions. ImageJ software was used to normalize signals and quantify bands intensities.

For immunofluorescence, $7x10^4$ cells were seeded in Chamber slides (Nunc, Thermo Fisher Scientific) 24 hours upon linc00941 silencing. After 24 hours, cells were fixed with 4% PFA, permeabilized with 0.1% Triton (Sigma Aldrich), blocked with FBS-BSA (Gibco, Thermo Fisher Scientific), and incubated with primary antibodies for one hour. The following primary antibodies were used: 1:100 anti-β-catenin (BD Bioscience, San José, California, USA), 1:100 anti-Flag M2 (Sigma Aldrich). Antibody binding was revealed with secondary anti-mouse Alexa 555 or anti-rabbit Alexa 594 conjugated antibodies diluted 1:1000 (Invitrogen, Thermo Fisher Scientific) (For more information see Supplementary Table 7). Phalloidin conjugated with green fluorophore (Thermo Fisher Scientific) was used to mark actin filaments. DAPI (Thermo Fisher Scientific) was used to stain cell nuclei. ECLIPSE Ni fluorescent microscope was used for detection and analysis (Nikon Instruments, Melville, New York, USA).

All antibodies used are listed in Supplementary Table 7.

Cell Proliferation assay

Thirty hours after siRNA or control oligos transfection, $1.5x10^3$ TPC1 and $2x10^3$ B-CPAP were seeded in 96-well plates. IncuCyte NucLight Rapid Red diluted 1:500 for TPC1 and 1:250 for B-CPAP in complete medium was added to cells for staining. Cells were left to settle for 30 minutes at room temperature then placed in IncuCyte S3 Live-Cell Analysis System (Essen BioScience, Ann Arbor, Michigan, USA). Cell proliferation was monitored every 2 hours either using the phase contrast confluence metric or the labeling efficiency using the red fluorescence channel. To confirm IncuCyte data, Trypan-blue cell counting was performed. Thirty hours after siRNA transfection, $5x10^3$ cells were seeded in 96-well plates. Viable cells were counted after 24, 48 and 72 hours using trypan blue (Sigma-Aldrich) staining and the Countess Automated Cell Counter (Thermo Fisher Scientific).

Chemotaxis and Invasion assays

Chemotaxis assay was performed using IncuCyte S3 Live-Cell Analysis System (Essen BioScience) following the manufacturer's protocol. TPC1 cells were transfected with siCT or silinc00941, seeded on IncuCyte ClearView inserts under different serum concentrations

(no FBS, FBS 10%, FBS 20%) and counted on the top and on the bottom of invasion membrane overtime.

Thirty hours upon transfection, cells were plated on the IncuCyte ClearView inserts (Essen BioScience). Chemotaxis plate membranes were coated with 50 µg/ml matrigel diluted in DMEM without FBS. After 30 minutes of incubation at 37°C and 30 minutes at room temperature, excess matrigel was aspired. Cells ($1x10^3$ cells/well) were seeded in DMEM without FBS. Cells were left to settle at room temperature for 30 minutes then DMEM complemented with 10% or 20% FBS or without FBS was added. The ClearView Cell Migration plate was placed into the IncuCyte system and allowed to warm at 37°C for 15 minutes prior to scanning every 2 hours for 62 hours.

Invasion assay was performed as previously described (24) using BioCoat Matrigel Invasion Chambers (Corning, Tewksbury, Massachusetts, USA). Thirty hours after transfection, $3x10^4$ cells were seeded into invasion chambers inserts in DMEM without FBS. The wells in which the inserts were placed, were filled with DMEM without or with 10% FBS. After 24 hours, the excess matrigel and the non-invading cells were removed from the inserts. Invaded cells on the inserts' membranes were fixed with methanol and stained with crystal violet (Sigma-Aldrich). Invading cells were counted using ImageJ and normalized on control inserts following the manufacturer instructions.

Cell Fractionation

Cell fractionation was performed as previously described (25). Cells ($15x10^6$ cells) were harvested, 15% of the supernatant was used for total RNA extraction (TRIzol Reagent, Thermo Fisher Scientific) and 15% for total protein extraction (PLB, Promega), and 70% of the supernatant was used to isolate cellular fractions. To cellular isolate cytosolic fraction were incubated in 250 µl buffer A (10 mM HEPES ph 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5% NP40, RNAse and protease inhibitors) and then centrifuged. Incubation time was 8 minutes for TPC1, 4 minutes for B-CPAP, and 2 minutes for Nthy-ori 3-1. The supernatant was equally divided for protein and RNA extraction. Nuclei pellets were washed and resuspended in 250 µl buffer C (20mM HEPES pH 7.9, 25% glycerol, 0.42M NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, RNAse and protease inhibitors) and incubated 30 minutes on ice, vortexed then centrifuged. The supernatant was equally divided for protein and RNA extraction. Chromatin pellet was resuspended in TRIzol.

Flow cytometry assays

For cell cycle analysis, $5x10^5$ to $1x10^6$ cells were transfected for 24 hours then resuspended and incubated in Nicoletti solution (50µg/ml PI, Triton-X 0.1%, Sodium Citrate 0.1%), then read on the BD FACS Canto II (BD Bioscience). Blank cells were also read. TPC1 cells siCT or silinc00941 were incubated with 1µM 5-(and-6)-Carboxyfluorescein Diacetate, Succinimidyl Ester (CSFE), washed, suspended in culture medium and seeded. At different time points, cells were washed and suspended in PBS 1x, 0.5% BSA, 2mM EDTA for reading on the BD FACS CantoII (BD Bioscience).

Statistic and bioinformatic analyses

Statistical significance was determined by two-tailed *t* test in all experiments with a minimum of three biological replicates. For RNA-seq, raw paired-end reads were assessed for their sequencing quality using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) software, low-quality sequences were discarded. Filtered reads were aligned to the human reference transcriptome (GRCh38.p12) using STAR 2.7 (26). Mapped reads were determined and the expression at gene and transcript level was computed by RSEM (27). Differential expression analysis was performed with R package DESeq2 (28), using 0.05 FDR threshold. Genes significantly differentially expressed underwent enrichment analysis, performed on GO biological processes and KEGG pathways via enrichR package (29). Enriched pathways with Benjamini-Hochberg adjusted p value ≤ 0.05 were selected for network-based modeling using networkx package (30) within python 2.7 environment and Cytoscape software (https://cytoscape.org/).

For ChIPseq analysis, reads were mapped on the human reference genome hg38 and filtered using BWA and Samtools. Peak calling was performed using Macs2 after Picard duplicates removal.

NGS mutational analysis. The sequences obtained were analyzed using VariantStudio Software (Illumina Inc) and the Integrative Genomics Viewer 2.3 (IGV) tool (http://software.broadinstitute.org/software/igv/). Only mutations present in at least 5% of the total number of reads analyzed and observed in both strands were considered for mutational calls (23).

For the THCA-TCGA analysis, we used R TCGAbiolinks package to download and analyze RNAseq data (workflow.type "HTSeq - FPKM") and mi-RNA seq data (workflow.type "BCGSC miRNA Profiling") from THCA-TCGA project and expression differences were evaluated applying Kruskal Wallis test. For correlation analysis of linc00941 with *BRAF* and *TERT* promoter mutations we used TCGA Mutational data from TCGA, Cell 2014 (31).

Correlations and survival analysis were performed using R Corrplot and Survival package, respectively. For miRNAs correlation analysis, Linc00941-correlated miRNAs were analyzed with DIANA-microT-CDS algorithm to predict miRNA targets and perform enrichment analysis (32).

RESULTS

Linc00941 is a novel TGFβ target and it is highly expressed in PTCs

To identify ENH-associated lncRNAs under the direct control of TGFβ, we used a ChIP-seq approach to map TGFβ-response transcriptional elements. Changes in the enrichment profile of H3K27ac, histone marker of active transcription, was investigated in Nthy-ori 3-1 upon TGFβ stimulation. Based on our previous data, Nthy-ori 3-1 is the most suitable model to study TGFβ-mediated EMT in thyroid system (13). Analysis of treated and untreated cells identified 22775 peaks with significantly different levels of H3K27ac. Peak-to-target assignment identified 9677 transcriptional units potentially under the regulation of TGFβ. Of these, 1025 presented at least one peak enriched upon TGFβ treatment (Figure 1A). Categorization showed that the vast majority of TGFβ-associated transcriptional units (70.5%, 723/1025) were protein-coding genes, 28% (287/1025) were non-coding RNAs and 1.5% (15/1025) were pseudogenes (Figure 1B). Out of the non-coding transcripts, 38.7% (106/287) were non-coding or antisense RNAs, 28.8% (79/287) were long non-coding RNAs (lncRNAs), 28.1% (77/287) were microRNAs (miRNAs) and 4.4% (12/287) were small-nucleolar RNAs (snoRNAs) (Figure 1C).

To functionally prove that changes in the chromatin transcriptional status reflected different levels of expression of the identified targets, we selected a subset of the top scoring lncRNAs with enriched H3K27ac peaks in their locus upon TGFβ treatment (Supplementary Figure 1A). q-RT-PCR confirmed that all the analyzed transcripts underwent expression changes, consistent with chromatin status measured by ChIP-seq in response to TGFβ, confirming the validity of our approach (Figure 1D).

Among the lncRNAs analyzed, linc00941 has been associated with cancer progression and was chosen for further investigations (33, 34). First, we aimed to verify that linc00941 was overexpressed in thyroid cancer cells. Analysis of linc00941 levels in a panel of thyroid cancer cell lines of differentiated PTCs (TPC1 and B-CPAP), anaplastic thyroid cancer (8505c and Cal62, SW579) and follicular thyroid cancer (FTC133), the only cell line derived from metastasis, showed that this transcript was dramatically upregulated compared to normal thyrocytes (Figure 1E). Interestingly, its expression in PTC and anaplastic thyroid cancer cell lines was not further enhanced by TGFβ (Supplementary Figure 1B). To confirm these observations, we took advantage of the THCA gene expression dataset of TCGA project. Analysis of 58 matched PTCs and normal thyroid samples showed that linc00941 is overexpressed in tumor tissue as compared to normal (Figure 1F). We validated these data by q-RT-PCR analysis in a retrospective cohort of 12 matched PTCs and normal thyroid samples derived from the biobank at our institution (Figure 1G). Next, to evaluate the role of linc00941 in thyroid cancer pathobiology, we interrogated the THCA-TCGA dataset to explore changes in the expression levels of this transcript and disease aggressiveness. Expression data of 502 PTCs, 58 normal tissues and 8 metastasis samples were available in the THCA-TCGA dataset. Figure 1HI shows linc00941 expression trend in these samples. Linc00941 expression in tumor was higher compared to normal thyroid tissues, this analysis showed a significant upregulation of this transcript in metastasis as compared to primary lesions (Figure 1H,I). Analysis of matched primary and metastatic samples (n=8) showed a higher expression of linc00941 in metastatic tissue in 75% of patients, confirming the correlation of this lncRNA with metastatic behavior of PTC, although the number of cases was too small for statistical analysis (Figure 1I).

The expression of linc00941 correlates with aggressive features in PTC patients

We considered the association of linc00941 with clinical features of PTC (Figure 2, Table 1). Linc00941 expression was not associated with PTC stage (Figure 2A) and sex (Table 1), but associated with the follicular variant of PTC (0.083 ± 0.109) compared to classical (0.197 ± 0.236) and tall cell variants (0.167 ± 0.122) (Figure 2B). Mutation data were available for 396 samples, with 5 (1.3%) having *BRAF* mutations different from V600E, that were not considered in further analysis. Interestingly, linc00941 expression

was significantly higher in *BRAF*^{V600E} PTCs (0.210±0.200 vs 0.124± 0.254 in WT samples, p<0.0001), which was associated with extrathyroidal extension. We found no association between linc00941 expression and TERT promoter mutation status (Figure 2C, Supplementary Figure 1C). Samples with both mutations showed higher levels of linc00941 expression (Table 1) (Figure 2C). The association of linc00941 expression with *BRAF*^{V600E} mutation may account for the observed higher expression of linc00941 in tall cell variants as compared to follicular and classical variants of PTCs (Figure 2B). We also evaluated the correlation of linc00941 with *BRAF*^{V600E} and TERT promoter in 11 samples, 7 with *BRAF*^{V600E} mutation (63.6%) (Supplementary Table 1). In this samples set, we found significantly higher linc00941 expression in *BRAF*^{V600E} mutant samples (Supplementary Figure 1D).

Two years follow up were available for 121 of the 502 THCA-TCGA PTCs analyzed. Of these, 4 patients were dead while 117 alive. Linc00941 expression was higher in deceased (0.286 ± 0.156) compared to alive (0.187 ± 0.222) patients (Figure 2D), and patients with high expression of linc00941 (IV quartile) displayed worst overall survival than the low expressing ones (I quartile) (Figure 2E). Together, these data indicate a potential association of linc00941 with aggressive features of PTCs.

Linc00941 regulates cellular proliferation and invasion

To provide functional validation of the association between linc00941 and PTC aggressiveness in the THCA-TCGA dataset, we used cellular models of thyrocyte-derived cell line to perform *in vitro* experiments. First, since lncRNAs function is determined by their cellular localization, we performed fractionation assay on TPC1, B-CPAP (Figure 3A) and Nthy-ori 3-1 (Supplementary Figure 1E) cell lines. q-RT-PCR of linc00941 levels showed that it was ubiquitously present in different cellular compartments, implying its possible activity in both nuclear and cytoplasmic compartments. To define the biological relevance of linc00941 in PTC, we targeted it by siRNAs (Supplementary Figure 1F). Linc00941 silencing (silinc00941) led to a significant increase in cell proliferation compared to control (siCT) (Figure 3B), observed in a greater extent in TPC1 than in B-CPAP cells. Cell cycle analysis by PI staining showed that cell cycle phases were not affected by linc00941 knockdown (Figure 3C). CSFE labeling confirmed that linc00941 silencing leads to increased rate of cell division (Figure 3B and D), suggesting a restraining function of this lncRNA on cellular proliferation.

The ability to degrade extracellular matrix and invade surrounding tissues is a major feature of metastatic cells. Thus, we measured the effect linc00941 silencing on cellular chemotaxis and invasiveness. Chemotaxis analysis using live imaging in TPC1 cells, showed that silencing linc00941 resulted in the inability of cells to respond to chemical stimuli, independently from the concentration of chemoattractant used, compared to siCT cells (Figure 3E). Moreover, invasion assay showed that silencing linc00941 expression significantly reduced the invasive potential of TPC1 cells (Figure 3F). These features could not be evaluated in B-CPAP, which do not respond to chemical stimuli (Supplementary Figure 1G).

Linc00941 expression correlates with coding genes and miRNAs implicated in tumor aggressiveness in PTC

To get a deeper understanding of the molecular mechanisms linked to linc00941 function in PTCs, we used an integrated approach, combining gene expression data from human samples and functional readouts from cell lines. First, we searched the TCGA-THCA dataset for genes and miRNAs expression which correlated with linc00941 expression in human samples. We found 77 genes, 11 negatively and 66 positively correlated with linc00941 expression in PTC samples (Figure 4A, Supplementary Table 2). Next, we performed RNA-seq analysis on TPC1 cell line transfected with scramble or linc00941 siRNAs to define genes potentially affected by this lncRNA. Comparative analysis identified 2452 differentially expressed genes upon linc00941 knockdown. Among these, 1080 genes were upregulated, and 1372 genes were downregulated with silencing of linc00941 expression compared to siCT (Figure 4B). A subset of 37 genes (reported in Supplementary Figure 2A) was validated in an independent set of experiments. All the genes tested (30 down- and 7 up-regulated) were confirmed to be affected by linc00941 silencing (Supplementary Figure 2A).

GO and KEGG enrichment analysis was performed and the pathways affected by linc00941 silencing are displayed and ranked by p value in Supplementary Figure 2B. We identified several categories including signal transduction, cell growth and death, cell-cell adhesion and cytoskeleton organization, folding, sorting and degradation, transport and catabolism, besides tumor-specific pathways.

Downregulated genes were involved in the TNF, PI3K-Akt, MAPK, mTOR, NF-Kβ, cell adhesion, apoptosis and anoikis pathways. Interestingly, several members of the integrin family, playing key roles in TGFβ signaling transduction, and membrane-associated proteins regulating actin cytoskeleton, cell polarity, proliferation and chemotaxis were found. Many of the down-regulated targets upon linc00941 silencing are unfavorable prognostic markers in different cancer types (35).

Fewer enriched pathways resulted interrogating the databases with genes we found to be upregulated following linc00941 silencing. Among these were cell cycle, p53 signaling pathway and endocytosis. We also found CDC25B, required for entry into mitosis and several cyclins, which promote transition through G1/S and G2/M. We found as upregulated several genes responsible for vesicle trafficking, membrane shaping and remodeling and cytoskeletal organization.

The 77 genes we found significantly associated with linc00941 in the THCA-TCGA dataset were affected by linc00941 silencing in thyroid cancer cells in the RNA-seq analysis. Indeed, the 11 genes inversely correlated with linc00941 expression were upregulated on linc00941 silencing vs siCT condition, while the 66 genes positively correlated to linc00941 expression were downregulated upon linc00941 silencing (Figure 4A and C). To reconstruct the biological role of these genes, exploiting significant categories emerged by the enrichment pathways analysis of our RNA-seq data, we found that these 77 genes belong to pathways that fit well within the linc00941-associated phenotype (Figure 4D). Three main functional clusters emerged from this analysis: 1. Cell metabolism and vesicle trafficking, 2. Cell adhesion and extracellular matrix interaction, 3. Cancer related signal transduction. In addition, we identified 504 miRNAs that showed either a positive (n=251, 49.8%) or negative (n=253, 50.2%) correlation with linc00941 expression. Several studies reported that lncRNAs bind miRNAs sequestering them and blocking their activity (36–38). Thus, we mapped the functional pathways in which the top scoring linc00941-associated miRNAs were enriched. Setting an absolute correlation value of 0.3, a total of 58 miRNAs were used to interrogate DIANA-mirPath v.3 (Figure 4E). Thirty-five miRNAs had a positive correlation with linc00941 (group D- $R \geq 0.3$) and 23 were inversely correlated (group I- $R \leq -0.3$). The predicted pathways overlapped with the ones in Figure 4D, confirming what we observed with linc00941-associated coding genes. The

three major clusters identified were: 1. Cell adhesion and extracellular matrix interactions 2. Cancer related signal transduction 3. Metabolic and catabolic processes.

Linc00941 regulates CDH6 to modulate cytoskeleton architecture and autophagy

We recently reported that CDH6 mediates TGFβ signaling, promoting PTC spreading and modulating the balance between autophagy and cellular architecture (13,15). Interestingly, CDH6 was one of the 77 genes of the linc00941-associated PTCs signature. CDH6 expression was positively correlated with linc00941 in the THCA-TCGA dataset (Figure 4C) and significantly downregulated upon its silencing in PTC cells (Supplementary Figure 2A). Thus, we reasoned that modulation of this protein could explain part of the linc00941-associated phenotype. To test this hypothesis, we assessed whether linc00941 silencing recapitulated the effects we observed upon CDH6 loss of function.

We previously observed CDH6 silencing profoundly affects cytoskeleton structure and cell-cell junctions' organization in PTCs (15). Similarly, linc00941 silencing compared to siCT cells, cell membranes staining with β-catenin highlighted changes in cell-cell interactions, increased intimate connections and roof tile-like contacts among adjacent cells (Figure 5A). Recently, Yan et al. described the interplay of linc00941 with ANXA2 and miR-34a to allow β-catenin nuclear translocation and EMT promotion in hepatocellular carcinoma (33). β-catenin distribution analysis by immunofluorescence (Figure 5A) and cellular fractionation (Figure 5B) did not show significant alterations in β-catenin localization upon linc00941 silencing. However, ANXA2 was among the 77 linc00941-associated genes in PTC (Figure 4A).

Consistent with what is reported for CDH6 silencing, actin filaments staining demonstrated a peculiar reorganization of cytoskeleton in cells with linc00941 silencing compared to siCT cells. Linc00941 silencing induced actin assembly into thick fibers (stress fibers) spanning across the cytoplasm along the major cell axis, not visible in control cells (Figure 5A). These thick and stable actin fibers are distinctive features of non-motile cells, in line with the reduced invasion capability observed upon linc00941 silencing (39). Next, since we reported that CDH6 restrains autophagy, we evaluated linc00941's role in this process. At early time points (24-48h) after linc00941 silencing, autophagy markers LC3, p62/SQSTM1 and GABARAP levels increased compared to siCT. By contrast, at later time points (96h) LC3, p62/SQSTM1 showed a significant down-regulation (Figure 5C). This

modulation is compatible with autophagosomes formation and subsequent degradation of cargoes. Autophagic machinery activation upon linc00941 silencing was confirmed by immunofluorescence analysis of autophagosomes by EGFP-LC3 and Flag-GABARAP staining (Figure 5D). In siCT cells both these markers showed diffuse and homogenous distribution in the cytoplasm, but dramatically changed on linc00941 silencing, with both markers co-localizing into distinct dots highlighting autophagosome formation. The dots quantification is given in Figure 5E. Overall, these data indicate that linc00941 and CDH6 functionally cooperate along the same axis to potentiate PTC aggressive features.

Linc00941 and CDH6 expression are associated with PTC aggressiveness

A well-established mechanism by which lncRNAs post-transcriptionally control protein function is by interfering with miRNAs, acting as competing endogenous RNAs (ceRNAs).

We showed that linc00941 expression in PTCs is associated with a subset of cancer-related miRNAs, thus we reasoned that physical sequestration of CDH6-targeting miRNAs promote CDH6 expression and activity. We searched the THCA-TCGA for miRNAs correlated with both transcripts. CDH6 expression was associated with 616 miRNAs, 405 of which (66%) were significantly correlated also with linc00941 (Supplementary Figure 2C). Restricting the analysis to miRNAs with correlation coefficient $R \leq -0.3$ or $R \geq 0.3$, 47 miRNAs remained commonly associated to both transcripts (Supplementary Table 3, Figure 6A). 19 were inversely and 28 directly correlated to both linc00941 and CDH6 (Figure 6B). Among these, miR-223 has been reported to target CDH6 (40). These data support the hypothesis that linc00941 regulates CDH6 expression.

To translate these observations in a clinical setting, we explored whether the combined expression of CDH6 and linc00941 were associated with PTCs aggressiveness in the THCA-TCGA dataset. The expression of CDH6 and linc00941were subdivided into quartiles (I-II low, III-IV high expression). Then, we analyzed the percentage of patients for each subset distributed based on the expression of both (Supplementary Table 4, Figure 6C). In the normal samples, 93% were in the low expression quartiles for both markers, while 100% of metastases fell into the high expression quartile for at least one of the two markers. Tumor samples displayed a heterogeneous distribution, 41,3% of tumor samples expressed high levels and 31,4% low levels of both transcripts (Supplementary Table 4, Figure 6C).

We performed overall survival analysis considering both CDH6 and linc00941 expression level, the only 3 deaths in the THCA-TCGA cohort had high expression of both markers (Figure 6D).

Because our experiments indicated that CDH6 and linc00941 are both target of TGFβ in thyroid derived cells, we evaluated whether their expression correlated with TGFβ and receptors' expression in the THCA-TCGA dataset. The expression of both CDH6 and linc00941 was significantly correlated with the levels of TGFβ1, TGFβ2, TGFβ3 and TGFβR1 in PTC samples. In contrast, a poor correlation was observed with TGFβR2 (Supplementary Figure 2D). Together, these observations, in line with our previous data, suggest a role for CDH6 and linc00941 in mediating tumor aggressiveness and the potential utility of these markers in improving PTCs risk-based stratification.

DISCUSSION

The molecular mechanisms underlining PTC aggressiveness and metastatic potential remain poorly understood, limiting development of molecular-based risk-stratification that could inform optimal patient management. One of the reasons for this is the rarity of highly aggressive PTCs and thus the difficulty of collecting significant cohorts to profile. While aggressiveness of some cancer types relies on high mutational burden and specific mutations in coding-genes, this seems not true for PTCs (31).

We recently reported that highly metastatic PTCs are not characterized by a different overall mutational burden compared to non-metastatic tumors (9). No mutations in coding-genes were significantly associated with metastatic spreading and PTC patients' survival. By contrast, we and others reported mutations in non-coding elements like TERT promoter as main drivers of PTCs progression and dissemination (3–9). These observations indicate that non-coding genome, including non-coding transcripts, may represent a promising approach to identify molecular discriminants that explain differential aggressive behavior.

LncRNAs are interesting molecules to investigate. Unbiased genome-wide searches discovered thousands of lncRNAs, and their total number is estimated to be triple the number of protein-coding genes (41). Based on their large number and expression specificity, lncRNAs are thought to contribute to many cellular processes, including oncogenic signaling. More than 8,000 lncRNA genes were recently discovered to be highly

specific in cancer, representing a potential reservoir for biomarkers and therapeutic targets (42,43). The present work seems to support the idea that lncRNAs are major players in modulating PTCs biology. Here, we reported that linc00941 potentiates PTC aggressiveness enhancing responsiveness to chemical stimuli and invasiveness capability. We also showed that linc00941 expression is associated with disease aggressiveness.

The role of lncRNAs in PTC initiation and progression has been reported by many investigators. Among these, MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1) expression is higher in PTCs compared to normal tissues, but is significantly lower in ATCs, suggesting spatial-temporal expression regulation (44). LOC100507661 has been reported to promote thyroid cancer cells proliferation and to be associated with *BRAF*^{V600E} mutation (45). HOTAIR (HOX transcript antisense RNA) has also been reported to promote cancer cells colony formation and its expression was associated with worse survival in patients with PTC (46).

Many lncRNAs were also found to have tumor suppressive function in thyroid cancer. Low expression of MEG3 has been found to strongly correlate with lymph node metastases and increased invasion in TPC1 cells (36, 47). LINC00271 downregulation has been reported to be associated with extrathyroidal extension and advanced PTC stage (48). NONHSAT037832 expression is negatively correlated with lymph node metastases in PTC patients (49). In a recent genome-wide association study, single nucleotide polymorphisms (SNPs) leading to the transcriptional deregulation of the lncRNA PTCSC3 were found significantly associated with increased risk of developing PTC. Indeed, this lncRNA has been described to play tumor suppression function in thyroid cancer (50).

Several lncRNAs have been proposed to partake TGFβ signaling in cancer (51,52). Here, we provide evidence that linc00941, previously linked to EMT execution in other cancer types (33,34,53), is a TGFβ target and may mediate TGFβ-dependent phenotype. We also showed that linc00941 plays a dual role in this context. Its silencing led to a significant promotion of cell proliferation while completely abolishing cell response to chemical stimuli and invasive capability. This effect may seem paradoxical but it is established that during metastatic spreading inhibition of proliferation represents a necessary condition to allow cancer cells to reorganize cellular architecture to resist

mechanical stresses and allow cell movement (54–56). Thus, by slowing cell division and promoting invasiveness, linc00941 affects both, namely metastatic behavior.

Many functionally characterized lncRNAs act by conditioning the activity or stability of specific proteins, bridging the interplay between coding and non-coding genome. Our data indicate that lin00941 controls or modulates the expression of a set of coding-genes whose function is related to cellular organization and transduction of cancer-related signals. Among these, CDH6 is a linc00941 relevant target and a major mediator of its biological effect in PTC. At the interface between cellular structure and autophagy regulation, CDH6 candidates membrane adhesion molecules to be not just structural components but sensors of environmental stimuli and signaling centers for cellular pathways (15,57). We described, for the first time, that linc00941 controls CDH6 activity in PTCs and, through its modulation, restrains autophagy by promoting cytoskeleton rearrangements required for invasiveness. Their functional interplay and relevance for PTC progression is highlighted in the human PTC samples analyzed. We showed that the combined high expression of both these molecules is a distinctive feature of metastatic samples. We also showed that the expression of both CDH6 and linc00941 is positively associated with TGFβ1, 2 and 3 and TGFβR1 expression in PTC samples.

Many studies described the ability of cytoplasmic lncRNAs to sponge miRNAs. Sequestering these small transcripts, lncRNAs indirectly stabilize their target proteins, becoming an additional layer of post-transcriptional regulation (58). For instance, NEAT1 was described as upregulated in PTCs where it has established different functions, among which sponging miRNA-214 to aid thyroid cancer progression (37). Liu et al. showed how H19 expression in PTC cells promoted their capacity to proliferate and invade. Its mode of action includes miR-17-5p sponging to regulate YES1 oncogene levels (38). We did not provide experimental evidence that linc00941 effect on CDH6 regulation is mediated by its activity as miRNA sponge. However, we found that CDH6 and linc00941 expression in PTC patients is significantly associated with a specific subset of miRNAs. The biological pathways in which these miRNAs are involved overlapped with the ones predicted to be affected by linc00941 and are compatible with the biological function described for CDH6 in cancer. Experimental validation is required to support the idea that linc00941 may condition CDH6 transcript levels sequestering miRNAs. However, there are studies that

support this hypothesis as linc00941 has been shown to act as ceRNA (33) and CDH6 expression has been shown to be inhibited by miR-223, which is predicted to be involved in our model (40).

In summary, our data provide new evidence on the role of linc00941 in PTC and on the possibility of exploiting integrative signature to improve patients' risk-based stratification to optimize their management.

ACKNOWLEDGMENTS

We wish to thank Marina Grassi for the great technical support to this study, the Research Tissue BioBank of the Azienda USL-IRCCS of Reggio Emilia for providing patients samples and all the colleagues of our laboratory and WM for participating in stimulating discussion on this work.

AUTHORS DISCLOSURE STATEMENT

The authors declare that they have no competing financial interests.

FUNDING INFORMATION

This work was supported by the Italian Ministry of Health, GR-2011-02350937. Mila Gugnoni is supported by Associazione Italiana Ricerca sul Cancro (AIRC) fellowship for Italy, year of submission 2017, project code 20933. Alessia Ciarrocchi is supported by AIRC Investigator Grant number AIRC IG21772.

Correspondence should be addressed to:

Alessia Ciarrocchi, PhD

Laboratory of Translational Research

Azienda USL-IRCCS di Reggio Emilia

Viale Risorgimento 80

42123, Reggio Emilia, Italy

tel + 39 0522295668

fax + 39 0522295454

e-mail: Alessia.Ciarrocchi@ausl.re.it

REFERENCES

1. Cabanillas ME, McFadden DG, Durante C 2016 Thyroid cancer. Lancet **388**:2783–2795.

2. Filetti S, Durante C, Hartl D, Leboulleux S, Locati LD, Newbold K, Papotti MG, Berruti A, ESMO Guidelines Committee 2019 Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol.

3. Liu X, Bishop J, Shan Y, Pai S, Liu D, Murugan AK, Sun H, El-Naggar AK, Xing M 2013 Highly prevalent TERT promoter mutations in aggressive thyroid cancers. Endocr Relat Cancer **20**:603–610.

4. Landa I, Ganly I, Chan TA, Mitsutake N, Matsuse M, Ibrahimpasic T, Ghossein RA, Fagin JA 2013 Frequent somatic TERT promoter mutations in thyroid cancer: higher prevalence in advanced forms of the disease. J Clin Endocrinol Metab **98**:E1562-1566.

5. Melo M, da Rocha AG, Vinagre J, Batista R, Peixoto J, Tavares C, Celestino R, Almeida A, Salgado C, Eloy C, Castro P, Prazeres H, Lima J, Amaro T, Lobo C, Martins MJ, Moura M, Cavaco B, Leite V, Cameselle-Teijeiro JM, Carrilho F, Carvalheiro M, Máximo V, Sobrinho-Simões M, Soares P 2014 TERT promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas. J Clin Endocrinol Metab **99**:E754-765.

6. Liu X, Qu S, Liu R, Sheng C, Shi X, Zhu G, Murugan AK, Guan H, Yu H, Wang Y, Sun H, Shan Z, Teng W, Xing M 2014 TERT promoter mutations and their association with BRAF V600E mutation and aggressive clinicopathological characteristics of thyroid cancer. J Clin Endocrinol Metab **99**:E1130-1136.

7. Gandolfi G, Ragazzi M, Frasoldati A, Piana S, Ciarrocchi A, Sancisi V 2015 TERT promoter mutations are associated with distant metastases in papillary thyroid carcinoma. Eur J Endocrinol **172**:403–413.

8. de Biase D, Gandolfi G, Ragazzi M, Eszlinger M, Sancisi V, Gugnoni M, Visani M, Pession A, Casadei G, Durante C, Costante G, Bruno R, Torlontano M, Paschke R, Filetti S, Piana S, Frasoldati A, Tallini G, Ciarrocchi A 2015 TERT Promoter Mutations in Papillary Thyroid Microcarcinomas. Thyroid **25**:1013–1019.

9. Gandolfi G, Ragazzi M, de Biase D, Visani M, Zanetti E, Torricelli F, Sancisi V, Gugnoni M, Manzotti G, Braglia L, Cavuto S, Merlo DF, Tallini G, Frasoldati A, Piana S, Ciarrocchi A 2018 Genome-wide profiling identifies the THYT1 signature as a distinctive feature of widely metastatic Papillary Thyroid Carcinomas. Oncotarget **9**:1813–1825.

10. Dongre A, Weinberg RA 2019 New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol **20**:69–84.

11. Xue G, Restuccia DF, Lan Q, Hynx D, Dirnhofer S, Hess D, Rüegg C, Hemmings BA 2012 Akt/PKB-Mediated Phosphorylation of Twist1 Promotes Tumor Metastasis via Mediating Cross-Talk between PI3K/Akt and TGF-β Signaling Axes. Cancer Discov **2**:248–259.

12. David CJ, Huang Y-H, Chen M, Su J, Zou Y, Bardeesy N, Iacobuzio-Donahue CA, Massagué J 2016 TGF-β Tumor Suppression through a Lethal EMT. Cell **164**:1015–1030.

13. Sancisi V, Gandolfi G, Ragazzi M, Nicoli D, Tamagnini I, Piana S, Ciarrocchi A 2013 Cadherin 6 is a new RUNX2 target in TGF-β signalling pathway. PLoS ONE **8**:e75489.

14. Sancisi V, Borettini G, Maramotti S, Ragazzi M, Tamagnini I, Nicoli D, Piana S, Ciarrocchi A 2012 Runx2 isoform I controls a panel of proinvasive genes driving aggressiveness of papillary thyroid carcinomas. J Clin Endocrinol Metab **97**:E2006-2015.

15. Gugnoni M, Sancisi V, Gandolfi G, Manzotti G, Ragazzi M, Giordano D, Tamagnini I, Tigano M, Frasoldati A, Piana S, Ciarrocchi A 2017 Cadherin-6 promotes EMT and cancer metastasis by restraining autophagy. Oncogene **36**:667–677.

16. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR 2012 Landscape of transcription in human cells. Nature **489**:101–108.

17. Ransohoff JD, Wei Y, Khavari PA 2018 The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol **19**:143–157.

18. Ørom UA, Shiekhattar R 2013 Long noncoding RNAs usher in a new era in the biology of enhancers. Cell **154**:1190–1193.

19. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES 2016 Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature **539**:452–455.

20. Engreitz JM, Ollikainen N, Guttman M 2016 Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat Rev Mol Cell Biol **17**:756–770.

21. Kopp F, Mendell JT 2018 Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell **172**:393–407.

22. Sancisi V, Manzotti G, Gugnoni M, Rossi T, Gandolfi G, Gobbi G, Torricelli F, Catellani F, Faria do Valle I, Remondini D, Castellani G, Ragazzi M, Piana S, Ciarrocchi A 2017 RUNX2 expression in thyroid and breast cancer requires the cooperation of three non-redundant enhancers under the control of BRD4 and c-JUN. Nucleic Acids Res **45**:11249–11267.

23. de Biase D, Torricelli F, Ragazzi M, Donati B, Kuhn E, Visani M, Acquaviva G, Pession A, Tallini G, Piana S, Ciarrocchi A 2018 Not the same thing: metastatic PTCs have a different background than ATCs. Endocr Connect **7**:1370–1379.

24. Petrachi T, Romagnani A, Albini A, Longo C, Argenziano G, Grisendi G, Dominici M, Ciarrocchi A, Dallaglio K 2017 Therapeutic potential of the metabolic modulator phenformin in targeting the stem cell compartment in melanoma. Oncotarget **8**:6914–6928.

25. Rossi T, Pistoni M, Sancisi V, Gobbi G, Torricelli F, Donati B, Ribisi S, Gugnoni M, Ciarrocchi A 2019 RAIN Is a Novel Enhancer-Associated lncRNA That Controls RUNX2 Expression and Promotes Breast and Thyroid Cancer. Mol Cancer Res.

26. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**:15–21.

27. Li B, Dewey CN 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**:323.

28. Love MI, Huber W, Anders S 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**:550.

29. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A 2016 Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res **44**:W90-97.

30. Hagberg A, Schult D, Swart P 2008 Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman.

31. Cancer Genome Atlas Research Network 2014 Integrated genomic characterization of papillary thyroid carcinoma. Cell **159**:676–690.

32. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG 2015 DIANA-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic Acids Res **43**:W460-466.

33. Yan X, Zhang D, Wu W, Wu S, Qian J, Hao Y, Yan F, Zhu P, Wu J, Huang G, Huang Y, Luo J, Liu X, Liu B, Chen X, Du Y, Chen R, Fan Z 2017 Mesenchymal Stem Cells Promote Hepatocarcinogenesis via lncRNA-MUF Interaction with ANXA2 and miR-34a. Cancer Res **77**:6704–6716.

34. Liu H, Wu N, Zhang Z, Zhong X, Zhang H, Guo H, Nie Y, Liu Y 2019 Long Non-coding RNA LINC00941 as a Potential Biomarker Promotes the Proliferation and Metastasis of Gastric Cancer. Front Genet **10**:5.

35. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist P-H, Djureinovic D, Micke P, Lindskog C, Mardinoglu A, Ponten F 2017 A pathology atlas of the human cancer transcriptome. Science **357**.

36. Liu Y, Yue P, Zhou T, Zhang F, Wang H, Chen X 2018 LncRNA MEG3 enhances 131I sensitivity in thyroid carcinoma via sponging miR-182. Biomed Pharmacother **105**:1232–1239.

37. Li J-H, Zhang S-Q, Qiu X-G, Zhang S-J, Zheng S-H, Zhang D-H 2017 Long non-coding RNA NEAT1 promotes malignant progression of thyroid carcinoma by regulating miRNA-214. Int J Oncol **50**:708–716.

38. Liu L, Yang J, Zhu X, Li D, Lv Z, Zhang X 2016 Long noncoding RNA H19 competitively binds miR-17-5p to regulate YES1 expression in thyroid cancer. FEBS J **283**:2326–2339.

39. Tojkander S, Gateva G, Lappalainen P 2012 Actin stress fibers--assembly, dynamics and biological roles. J Cell Sci **125**:1855–1864.

40. Ji Q, Xu X, Song Q, Xu Y, Tai Y, Goodman SB, Bi W, Xu M, Jiao S, Maloney WJ, Wang Y 2018 miR-223-3p Inhibits Human Osteosarcoma Metastasis and Progression by Directly Targeting CDH6. Mol Ther **26**:1299–1312.

41. Hangauer MJ, Vaughn IW, McManus MT 2013 Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet **9**:e1003569.

42. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu Y-M, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM 2015 The landscape of long noncoding RNAs in the human transcriptome. Nat Genet **47**:199–208.

43. Schmitt AM, Chang HY 2016 Long Noncoding RNAs in Cancer Pathways. Cancer Cell **29**:452–463.

44. Zhang R, Hardin H, Huang W, Chen J, Asioli S, Righi A, Maletta F, Sapino A, Lloyd RV 2017 MALAT1 Long Non-coding RNA Expression in Thyroid Tissues: Analysis by In Situ Hybridization and Real-Time PCR. Endocr Pathol **28**:7–12.

45. Kim D, Lee WK, Jeong S, Seol M-Y, Kim H, Kim K-S, Lee EJ, Lee J, Jo YS 2016 Upregulation of long noncoding RNA LOC100507661 promotes tumor aggressiveness in thyroid cancer. Mol Cell Endocrinol **431**:36–45.

46. Zhu H, Lv Z, An C, Shi M, Pan W, Zhou L, Yang W, Yang M 2016 Onco-lncRNA HOTAIR and its functional genetic variants in papillary thyroid carcinoma. Sci Rep **6**:31969.

47. Wang C, Yan G, Zhang Y, Jia X, Bu P 2015 Long non-coding RNA MEG3 suppresses migration and invasion of thyroid carcinoma by targeting of Rac1. Neoplasma **62**:541–549.

48. Ma B, Liao T, Wen D, Dong C, Zhou L, Yang S, Wang Y, Ji Q 2016 Long intergenic non-coding RNA 271 is predictive of a poorer prognosis of papillary thyroid cancer. Sci Rep **6**:36973.

49. Lan X, Sun W, Zhang P, He L, Dong W, Wang Z, Liu S, Zhang H 2016 Downregulation of long noncoding RNA NONHSAT037832 in papillary thyroid carcinoma and its clinical significance. Tumour Biol **37**:6117–6123.

50. Jendrzejewski J, He H, Radomska HS, Li W, Tomsic J, Liyanarachchi S, Davuluri RV, Nagy R, de la Chapelle A 2012 The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. Proc Natl Acad Sci USA **109**:8646–8651.

51. Gugnoni M, Ciarrocchi A 2019 Long Noncoding RNA and Epithelial Mesenchymal Transition in Cancer. Int J Mol Sci **20**.

52. Farooqi AA, Attar R, Qureshi MZ, Fayyaz S, Sohail MI, Sabitaliyevich UY, Nurmurzayevich SB, Yelekenova A, Yaylim I, Alaaeddine N 2018 Interplay of long non-coding RNAs and TGF/SMAD signaling in different cancers. Cell Mol Biol (Noisy-le-grand) **64**:1–6.

53. Luo C, Tao Y, Zhang Y, Zhu Y, Minyao DN, Haleem M, Dong C, Zhang L, Zhang X, Zhao J, Liao Q 2018 Regulatory network analysis of high expressed long non-coding RNA LINC00941 in gastric cancer. Gene **662**:103–109.

54. Wang W, Goswami S, Lapidus K, Wells AL, Wyckoff JB, Sahai E, Singer RH, Segall JE, Condeelis JS 2004 Identification and testing of a gene expression signature of invasive carcinoma cells within primary mammary tumors. Cancer Res **64**:8585–8594.

55. Hoek KS, Eichhoff OM, Schlegel NC, Döbbeling U, Kobert N, Schaerer L, Hemmi S, Dummer R 2008 In vivo switching of human melanoma cells between proliferative and invasive states. Cancer Res **68**:650–656.

56. Schedin PJ, Eckel-Mahan KL, McDaniel SM, Prescott JD, Brodsky KS, Tentler JJ, Gutierrez-Hartmann A 2004 ESX induces transformation and functional epithelial to mesenchymal transition in MCF-12A mammary epithelial cells. Oncogene **23**:1766–1779.

57. Yu W, Yang L, Li T, Zhang Y 2019 Cadherin Signaling in Cancer: Its Functions and Role as a Therapeutic Target. Front Oncol **9**:989.

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

TABLE 1. Association between linc00941 expression and clinical features in PTC patients

	linc00941 expression (mean ± SD)		P-value	
Sample				
	Normal (58)	0.058 ± 0.028	Normal vs Primary Tumor	**< 0.0001**
	Primary Tumor (502)	0.171 ± 0.213	Normal vs Metastasis	**< 0.0001**
	Metastasis (8)	0.255 ± 0.12	Primary Tumor vs Metastasis	**0.0137**
Sex			0.8063	
	Male (135)	23.4 ± 26.018		
	Female (367)	22.686 ± 29.877		
Tumor Stage			0.09548	
	Stage I (281)	0.175 ± 0.22		
	Stage II (52)	0.152 ± 0.288		
	Stage III (112)	0.162 ± 0.158		
	Stage IV (55)	0.188 ± 0.203		
Status at 2 years of follow up			0.09548	
	Alive (117)	0.187 ± 0.222		
	Dead (4)	0.286 ± 0.156		
Histology				
	Papillary Classical	0.197 ± 0.236	Classical vs Follicular	**<**

(356)			**0.0001**
Papillary Follicular Variant (101)	0.083 ± 0.109	Classical vs Tall cell	0.349
Papillary Tall Cell Variant (36)	0.167 ± 0.122	Follicular vs Tall Cell	**< 0.0001**
Other (9)	0.146 ± 0.196		
***BRAF* and *TERT* promoter mutations**			
WT (144)	0.124 ± 0.254	*BRAF*^{V600E} vs *BRAF*^{V600E} + *TERT* promoter	0.358
BRAF^{V600E} (211)	0.210 ± 0.200	*BRAF*^{V600E} vs *TERT* promoter	**0.009**
TERT promoter (8)	0.075±0.085	*TERT* promoter vs *BRAF*^{V600E} + *TERT* promoter	0.545
BRAF^{V600E} +*TERT* promoter (28)	0.271±0.279	WT vs *BRAF*^{V600E}	**<0.0001**
		WT vs *BRAF*^{V600E} + *TERT* promoter	**<0.0001**
		WT vs *TERT* promoter	0.562

FIGURE LEGENDS

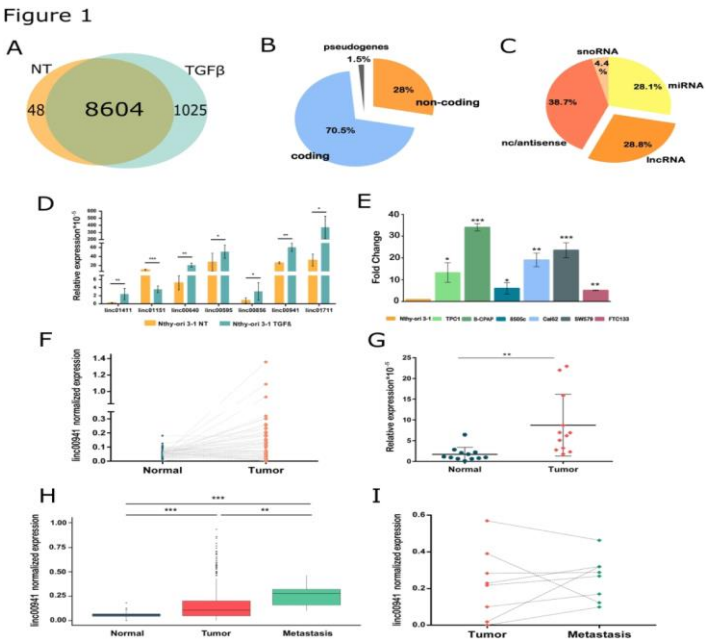


Figure 1 TGFβ-induced linc00941 expression in thyroid cell lines and PTC patients. A. Venn diagram showing all genes correlated to enriched H3K27ac peaks resulted from ChIP-seq analysis in Nthy-ori 3-1 not treated (NT), TGFβ treated or both conditions. **B.** Pie chart representing the different typologies of transcripts distribution among the genes correlated to enriched H3K27ac peaks in TGFβ condition. **C.** Pie chart displaying the distribution of non-coding RNAs typologies among the non-coding transcripts associated with H3K27ac enrichment in TGFβ condition. **D.** q-RT-PCR of 7 lncRNAs selected for validation in Nthy-ori 3-1 not treated (NT) or TGFβ treated **E.** qRT-PCR of linc00941 expression in normal thyrocyte-derived and thyroid carcinoma cell lines **F.** linc00941 expression in matched normal and tumor samples from PTC patients of THCA-TCGA dataset. **G.** linc00941 expression in matched normal and tumor patient tissues from our institutional biobank. **H.** Differential expression of linc00941 in normal thyroids, primary PTCs and metastases (THCA-TCGA). **I.** linc00941 expression in matched tumor and metastatic tissues (THCA-TCGA). *P<0.05, **P<0.01, ***P<0.001

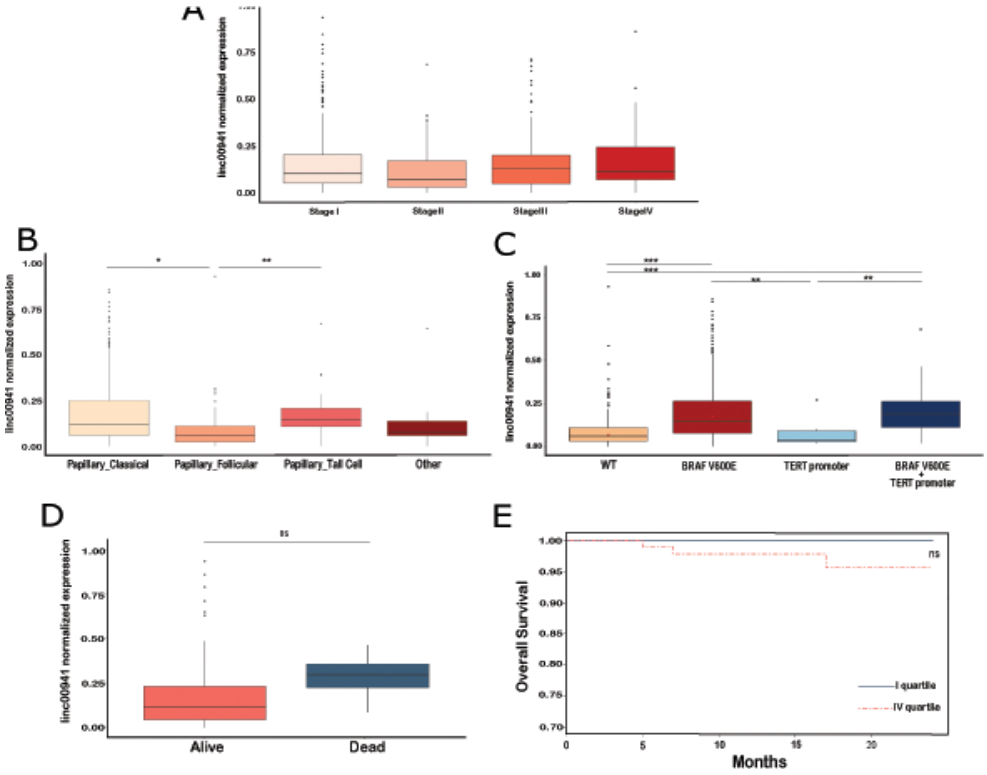


Figure 2 Linc00941 expression and clinical features. A. Distribution of linc00941 expression among different PTC stages. **B.** Distribution of linc00941 expression among different PTC histotypes**. C.** Association of linc00941 expression with *BRAF*^{V600E} and *TERT* promoter mutations. **D.** Linc00941 expression distribution in alive and dead patients. **E.** Overall survival in PTC patients from THCA-TCGA subdivided into quartiles based on linc00941 levels: I quartile: linc00941 expression <25%, IV quartile: linc00941 expression >75%. Follow-up 24 months. ns: not significant. *P<0.05, **P<0.01, ***P<0.001. Where not specified, the correlations are not significant.

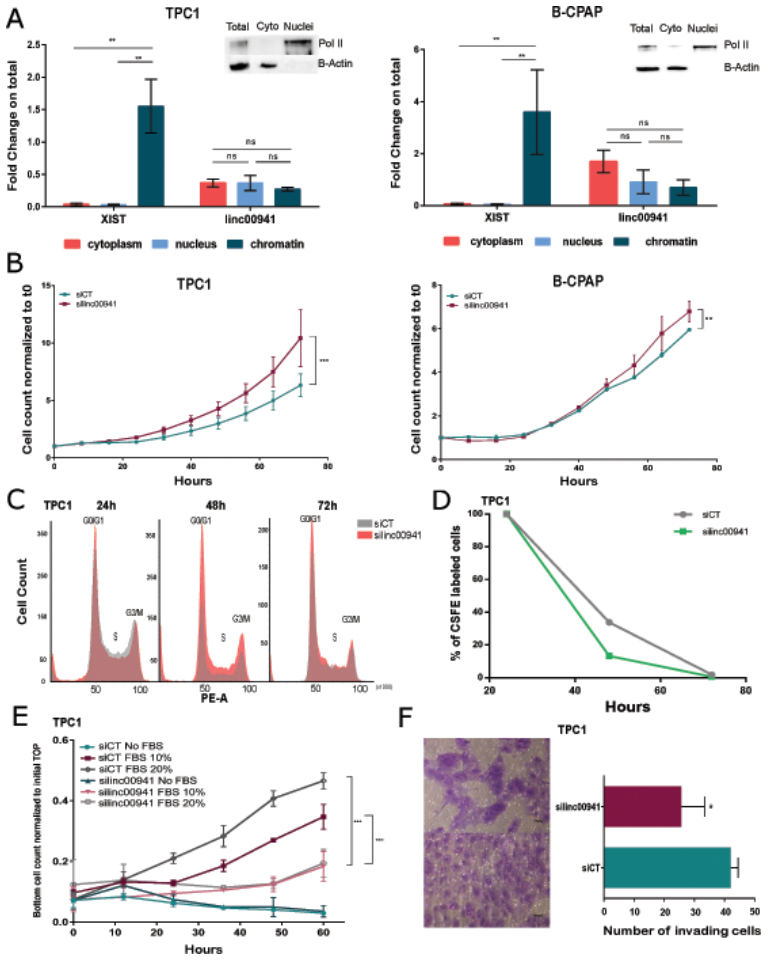


Figure 3 Linc00941 silencing effects on cellular proliferation and invasion of PTC cell lines. A. linc00941 localization in TPC1 and B-CPAP cell lines by fractionation. Transcription levels of linc00941 and Xist, used as positive control for chromatin fraction, were normalized to Glucuronidase Beta (GUSB). Data shown as fold change on total (+/- SD). qRT-PCR graphs show mean and standard deviation of three independent experiments. A representative Western blot shows the distribution of PolII and β-actin as control for fraction purity. **B.** Proliferation assay in TPC1 and B-CPAP cells. Cell count at each time point was normalized to initial cell number. **C.** Cell cycle distribution by PI of TPC1 siCT and silinc00941 cells. **D.** Representative FACS analysis of CSFE labeling of TPC1 siCT and silinc00941 cells. **E.** Chemotaxis assay on TPC1 siCT and silinc00941 with No, 10% or 20% FBS. Graph shows bottom cell count normalized to initial top cell count. **F.** Images and cell count of invading siCT and silinc00941 TPC1 cells. The assays were performed at least in three independent experiments. *P<0.05, **P<0.01, ***P<0.001.

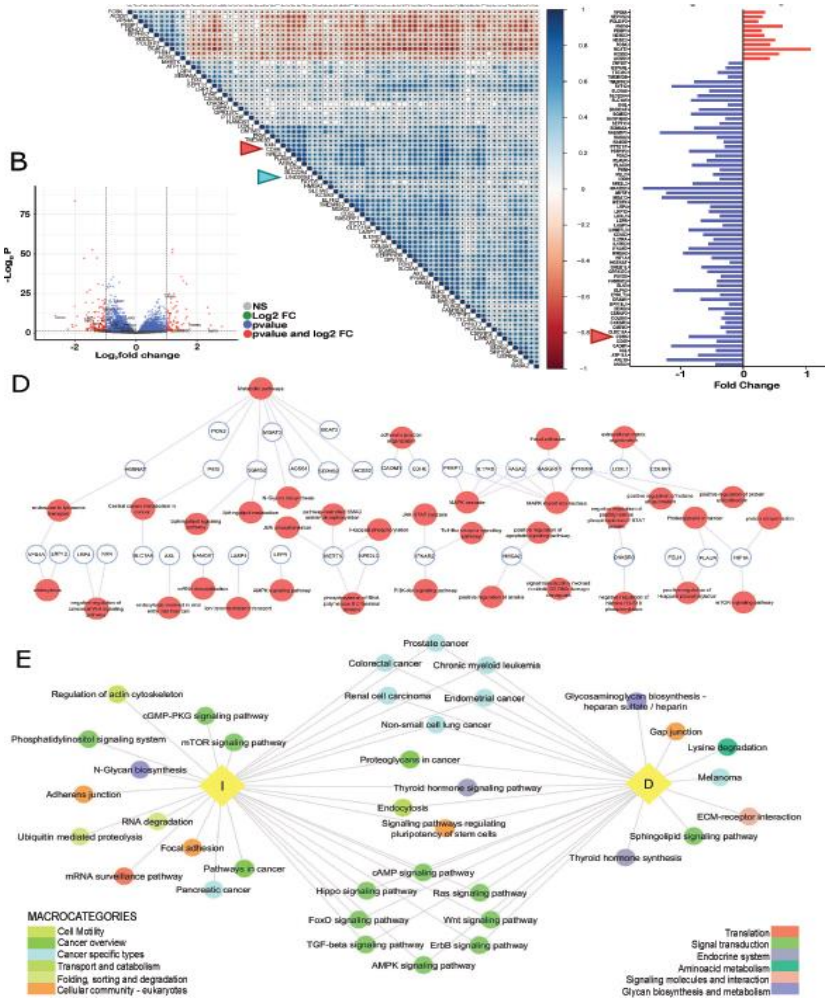


Figure 4 Linc00941 correlation patterns. A. Correlation matrix of the 77 genes correlated to linc00941 expression in PTC patients (THCA-TCGA). In red inversely correlated, in blue directly correlated genes. *indicates significance, dot size and color (legend on the right) are proportional to the correlation coefficient. **B.** Volcano plot of the deregulated genes in TPC1 silinc00941 compared to siCT ranked by p-value and FC. Red dots are genes with significant p value and Log_2 FC ≤ -1 or ≥ +1. **C.** Deregulation level of the 77 genes correlated to linc00941 expression in PTC patients (THCA-TCGA) in our RNA-seq analysis in TPC1 cells upon silinc00941. **D.** Visualization of linc00941-correlated genes (blue circles) and the pathways in which they are enriched (red circles). **E.** Visualization of linc00941-inversely (I) and directly (D) correlated miRNAs clusters (yellow diamonds) and the pathways in which they are enriched (colored circles). Pathways falling in the same macrocategory are highlighted with the same colors.

Arrows indicate CDH6 (red) and linc00941 (blue).

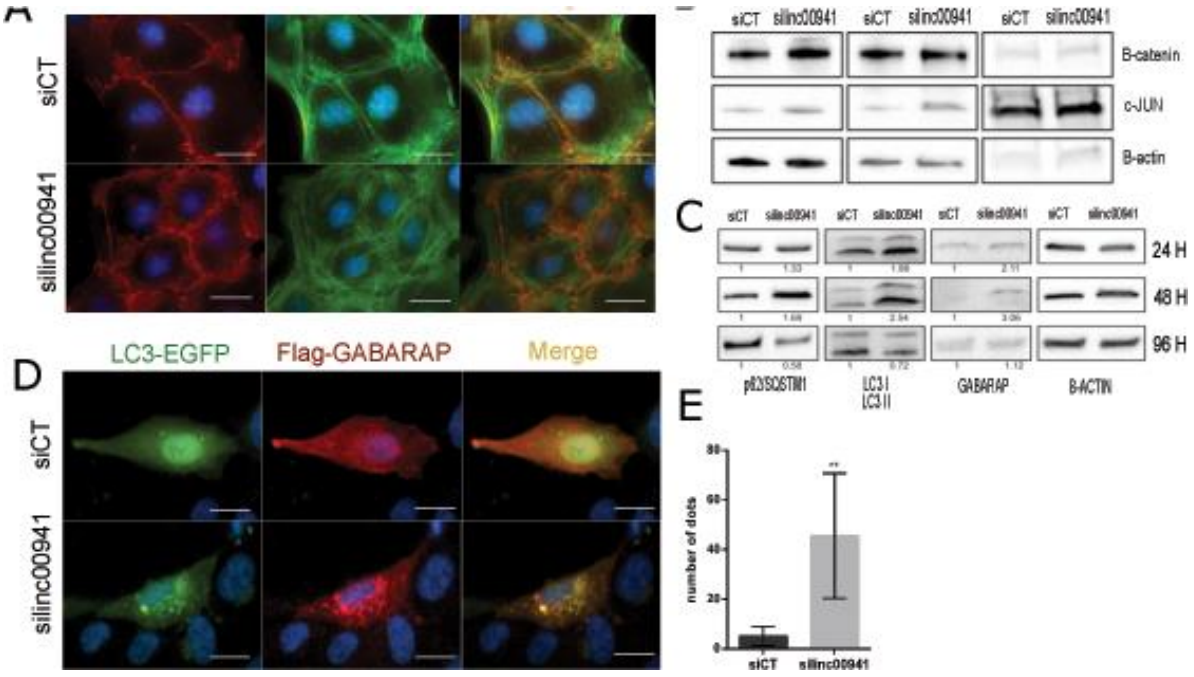


Figure 5 Linc00941 and CDH6 in autophagy. A. Immunofluorescence staining of β-catenin (red), actin filaments (green) and merged in TPC1 siCT and silinc00941 cells. **B.** Western blot analysis of β-catenin cellular localization in TPC1 siCT and silinc00941 cells. c-Jun and β-actin were used as positive control for nuclear and cytoplasmic fraction respectively **C.** Western blot of autophagy-related proteins in TPC1 at different time points upon linc00941 silencing. **D.** Immunofluorescence analysis showing co-localization of Flag-GABARAP (red) and EGFP-LC3 (green) in siCT and si-linc00941 transfected TPC1 cells. Nuclei were stained with DAPI (blue). **E.** Quantification of the number of autophagic dots per cell. ** P<0.01.

Magnification 400x. Scale bars IF: 10 µm. All the assays were performed at least in three independent experiments. ImageJ was used for WB quantification. **P<0.01.

Figure 6

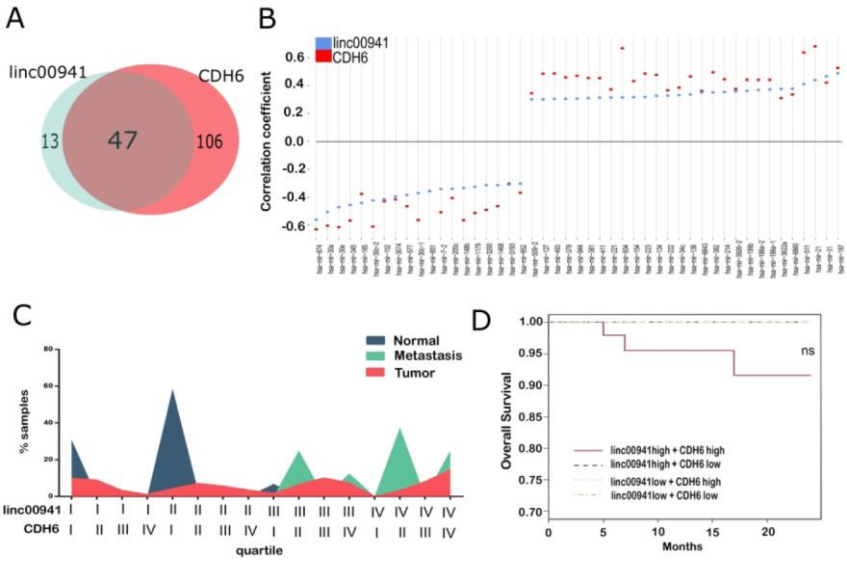
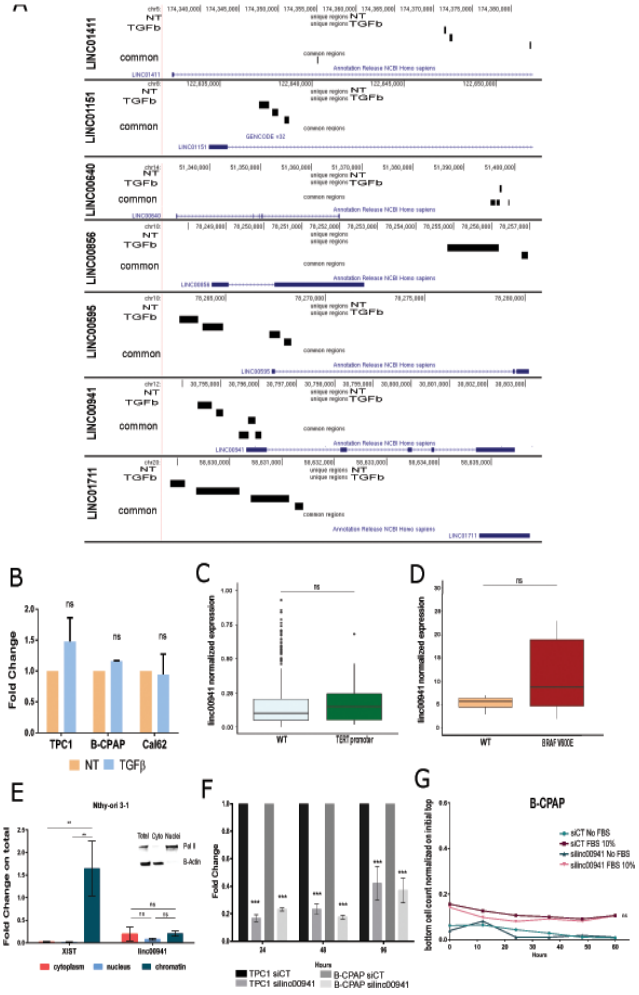


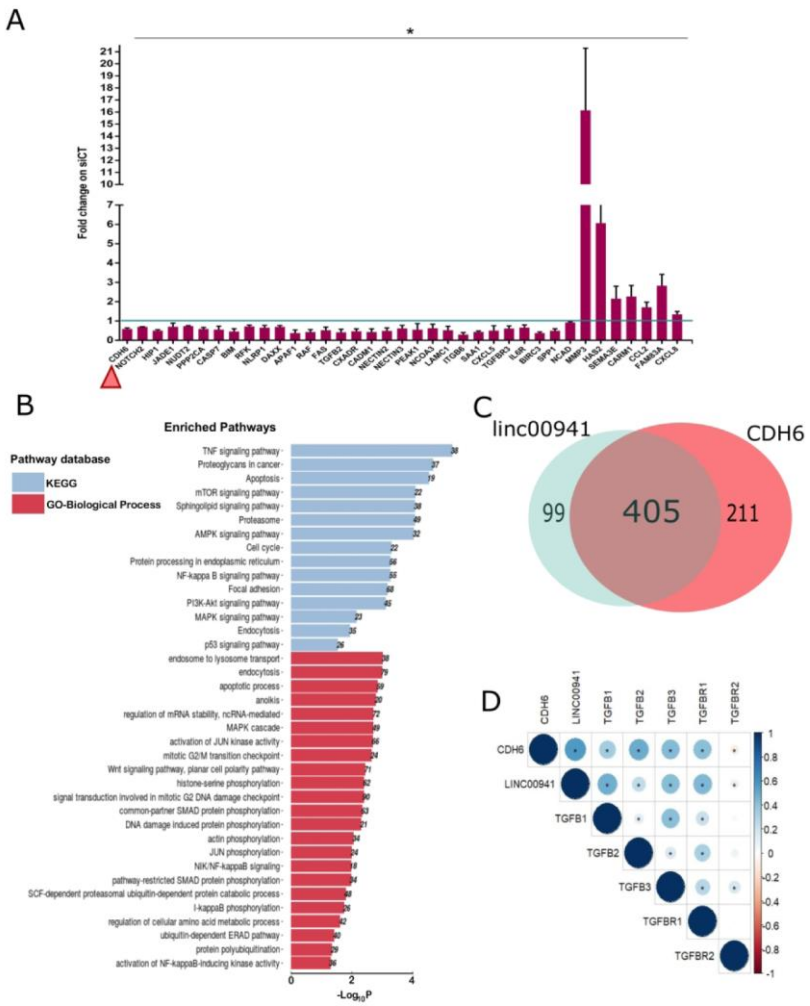
Figure 6 Linc00941 and CDH6 coexpression in PTCs A. Venn diagram showing the number of miRNAs correlated (R≤-0.3 and R≥0.3) to linc0041, CDH6 or both transcripts in the THCA-TCGA dataset. **B.** Correlation coefficient of linc00941 and CDH6 with the 47 commonly correlated miRNAs. **C.** Percentage of normal, metastasis and tumor samples from THCA-TCGA distributed for the combined expression of CDH6 and linc00941 in quartiles. **D**. Overall survival in PTC patients from THCA-TCGA subdivided in high or low linc00941 and CDH6 combined expression. Follow-up 24 months. ns: not significant.

SUPPLEMENTARY FIGURE LEGENDS



Supplementary Figure 1. A. Distribution of H3K27ac in Nthy-ori 3-1 not treated (NT), TGFβ treated or common peaks, in the locus of 7 lncRNAs selected for ChIP-seq validation. **B.** qRT-PCR of linc00941in TPC1, B-CPAP and Cal62 cells not treated (NT) or TGFβ treated. **C.** Association of linc00941 expression with *TERT* promoter mutations in THCA-TCGA dataset. **D.** Association of linc00941 expression with *BRAF*^{V600E} mutation in 11 samples of our cohort. **E.** Linc00941 localization in Nthy-ori 3-1 by fractionation. Linc00941 and Xist (chromatin fraction positive control) expressions, normalized on GUSB transcription levels, are reported as fold change on total. q-RT-PCR graphs show mean and standard deviation of three independent experiments. A representative western blot shows the distribution of PolII and β-actin to control fraction purity. **F.** linc00941 silencing efficiency in TPC1 and B-CPAP cells at 24, 48 and 96 hours. **G.** Chemotaxis assay on B-CPAP siCT and silinc00941 with No or 10% FBS. Graphs show bottom cell count normalized to initial topcell count. *P<0.05, **P<0.01, ***P<0.001. ns= not significant.

Supplementary Figure 2



Supplementary Figure 2. A. q-RT-PCR validation of a subset of RNA-seq deregulated genes upon linc00941 in TPC1. Fold change of normalized expression on siCT and SD calculated in at least three independent experiments is shown. Red arrow indicates CDH6. *P<0.05. **B.** GO and KEGG enrichment analysis on linc00941-deregulated genes from RNA-seq analysis. Pathways majorly affected by linc00941 silencing are displayed, ranked for pvalue in each group. Total number of genes enriched in each pathway are reported on the bars. **C.** Venn diagram displaying total number of miRNAs correlated to linc00941, CDH6 or both transcripts. **D.** Correlation matrix of CDH6 and linc0041 with TGFβ and TGFβ receptor family members in PTC patients (THCA-TCGA). In red inversely correlated, in blue directly correlated genes. *indicates significance, dot size and color (legend on the right) are proportional to the correlation coefficient.

SUPPLEMENTARY TABLES

Supplementary Table 1. Mutational analysis in PTC samples

Sample	Gene	AA Change	% Frequency	Coding
RE1	*BRAF*	p.Val600Glu	13	c.1799T>A
RE2	*BRAF*	p.Val600Glu	32	c.1799T>A
RE3	WT			
RE5	WT			
RE7	*HRAS*	p.Gln61Arg	37	c.182A>G
RE8	WT			
RE9	*BRAF*	p.Val600Glu	36	c.1799T>A
RE10	*BRAF*	p.Val600Glu	34	c.1799T>A
RE11	*BRAF*	p.Val600Glu	21	c.1799T>A
RE13	*BRAF*	p.Val600Glu	30	c.1799T>A
RE14	*BRAF*	p.Val600Glu	12	c.1799T>A

Supplementary Table 2. linc00941-correlated genes in THCA-TCGA

Gene	R coeff	p value
CADM1	0.44389966	0
SINHCAF	0.44309471	0
CD55	0.4417051	0
LEPR	0.43952682	0
FXYD5	0.43909947	0
HIF1A	0.43757896	0
SLC1A5	0.43378997	0
PTTG1IP	0.43278582	0
KCNS3	0.42749324	0
SKIL	0.42074673	0
RASA2	0.41951624	0
MYRF	0.40921423	0
PSTPIP2	0.4056704	0
RAB3B	0.3978094	0
USP6NL	0.3869673	0
LRP12	0.38158796	0
MERTK	0.37297047	0
CSRNP2	0.35983586	0
HGSNAT	0.35605547	0
NANOS1	0.34782961	4,44E-16

GPR137C	0.34765456	6,66E-16
PLAC8	0.34128356	2,22E-15
DYNLT3	0.30897896	9,61E-13
GREB1L	0.2804106	1,14E-10
CNKSR3	0.26741684	8,44E-10
FCSK	-0.361725	0
VPS4A	-0.4067712	0
PEBP1	-0.4224605	0
HDHD3	-0.4225454	0
HDDC3	-0.4354569	0
ACSS1	-0.4633077	0
SEPHS2	-0.4965019	0
ACSS2	-0.521941	0
BCAT2	-0.533493	0
POLDIP2	-0.5685086	0
PHYH	-0.6044332	0

Supplementary Table 3. miRNAs significantly correlated to linc00941 and CDH6 with -0.3 ≤R≥0.3

miRNA	R_linc00941	p value_linc00941	R_CDH6	p value_CDH6
hsa-mir-1179	-0.325027861	8.62E-14	-0.512419581	6.86E-35
hsa-mir-127	0.301738826	5.26E-12	0.484507097	0
hsa-mir-134	0.327196324	5.77E-14	0.477329448	7.14E-30
hsa-mir-136	0.336450948	1.00E-14	0.465217354	2.85E-28
hsa-mir-1468	-0.312989309	7.55E-13	-0.463841305	0
hsa-mir-148b	-0.332957637	1.96E-14	-0.564801775	0
hsa-mir-152	-0.413699144	3.91E-22	-0.427706277	0
hsa-mir-154	0.317340373	3.49E-13	0.431980408	3.42E-24
hsa-mir-185	-0.440817639	3.11E-25	-0.375969082	0
hsa-mir-187	0.487312516	3.05E-31	0.526706311	4.18E-37
hsa-mir-199a-1	0.372400456	6.31E-18	0.441897846	0
hsa-mir-199a-2	0.369293157	1.24E-17	0.441849464	0
hsa-mir-199b	0.362355983	5.44E-17	0.441998521	0
hsa-mir-200c	-0.339545738	5.52E-15	-0.406092325	0
hsa-mir-21	0.439474883	4.49E-25	0.68095942	0
hsa-mir-214	0.354683351	2.68E-16	0.445385595	0
hsa-mir-221	0.315057884	5.24E-13	0.372323592	0
hsa-mir-222	0.3282672	4.73E-14	0.366738173	7.95E-18

hsa-mir-223	0.318371497	2.90E-13	0.485375098	0
hsa-mir-3074	-0.394444421	4.24E-20	-0.415461221	0
hsa-mir-30a	-0.505775873	6.78E-34	-0.604397929	0
hsa-mir-30c-1	-0.36977535	1.12E-17	-0.563502254	0
hsa-mir-30c-2	-0.422391987	4.25E-23	-0.609448577	0
hsa-mir-30e	-0.470308931	6.16E-29	-0.614011117	0
hsa-mir-31	0.466369475	2.02E-28	0.421492001	5.37E-23
hsa-mir-3193	-0.309508313	1.39E-12	-0.303111005	4.17E-12
hsa-mir-3200	-0.313768185	6.58E-13	-0.49031591	0
hsa-mir-345	-0.45438328	6.82E-27	-0.566871102	0
hsa-mir-34c	0.331512392	2.57E-14	0.385575781	0
hsa-mir-3622a	0.376405392	2.61E-18	0.309929855	1.29E-12
hsa-mir-379	0.305196058	2.92E-12	0.458653148	0
hsa-mir-381	0.310436364	1.18E-12	0.454776966	6.09E-27
hsa-mir-382	0.351001827	5.69E-16	0.496129558	1.73E-32
hsa-mir-3926-2	0.358722753	1.16E-16	0.375314217	3.33E-18
hsa-mir-411	0.31385222	6.48E-13	0.453867185	7.91E-27
hsa-mir-493	0.305130623	2.96E-12	0.486421652	4.06E-31
hsa-mir-509-2	0.301033049	5.92E-12	0.345244429	1.80E-15
hsa-mir-511	0.410948551	7.78E-22	0.6372457	0
hsa-mir-577	-0.383686734	5.11E-19	-0.465030568	3.01E-28
hsa-mir-651	-0.3566906	1.78E-16	-0.356463617	1.86E-16

hsa-mir-652	-0.302529918	4.60E-12	-0.367962784	4.13E-18
hsa-mir-6843	0.349827392	7.21E-16	0.361735038	6.20E-17
hsa-mir-6860	0.377804858	1.92E-18	0.33596223	1.10E-14
hsa-mir-7-2	-0.341228556	3.98E-15	-0.507560268	3.68E-34
hsa-mir-874	-0.560829148	7.61E-43	-0.629511463	0
hsa-mir-934	0.315290871	5.03E-13	0.66774494	5.70E-66
hsa-mir-944	0.306285981	2.42E-12	0.46963588	7.55E-29

Supplementary Table 4. Expression of linc00941 and CDH6 in PTC samples

CDH6-linc00941	Metastasis%	Normal%	Tumor%
I - I quartile	0	**31**	10.2
II - I quartile	0	0	9.2
III - I quartile	0	0	3.6
IV - I quartile	0	0	1.4
I - II quartile	0	**58.6**	4.6
II - II quartile	0	**3.4**	7.4
III - II quartile	0	0	6
IV - II quartile	0	0	3.8
I - III quartile	0	**6.9**	2
II - III quartile	**25**	0	6.8
III - III quartile	0	0	10.4
IV - III quartile	**12.5**	0	7.6
I - IV quartile	0	0	0.4
II - IV quartile	**37.5**	0	3.6
III - IV quartile	0	0	8.4
IV - IV quartile	**25**	0	14.9

58 normal thyroids, 502 PTCs and 8 metastases comprised in this study, were subdivided into quartiles based on the expression of CDH6 or linc00941 (I-II low, III-IV high expression). The Table shows the percentage of patients for each subset distributed based on the expression of both.

Supplementary Table 5. Plasmids used

pEGFP-C3_LC3B
p3XFlagCMV10_GABARAP
pCDNA-3.1

Supplementary Table 6. Primers used

Gene	Sequence
CYPA Fw	GACCCAACACAAATGGTTCC
CYPA Rev	TTTCACTTTGCCAAACACCA
GUSB Fw	TTGAGCAAGACTGATACCACC TG
GUSB Rev	TCTGGTCTGCCGTGAACAGT
LINC00941 Fw	TTTTGTGTCCAAGCCCCAGA
LINC00941 Rev	GCCAAGAGTACAAGTCCAGC
LINC01711 Fw	CTTTCTTAACGGTCCGCTGG
LINC01711 Rev	GGACAGGCAGTAGTGTGAGT
LINC00595 Fw	TGGTCTCAAACTCCCGTGTT
LINC00595 Rev	GCTGATGACTCCCTCACTGT
LINC00856 Fw	AAACCACCTCCTTCCTCCTG
LINC00856 Rev	GGGCACTAGGTAGACTTCGG
LINC01151 Fw	TTCTTGGGCAGGGTGGATTT
LINC01151 Rev	GCCTCCCTCAAACCTTGTTG
LOC105375075 Fw	CACAAGGCCTGACACATAGC
LOC105375075 Rev	TTCCTGAGTTCAACCTGTGC
LINC00640 Fw	GCGTCCTTACTCCCAGTCTT
LINC00640 Rev	CCTAGGGACACAGCAAGACC
LINC01411 Fw	AGGTCATCAGCAACGTAAAG C

LINC01411 Rev	CAGCCTCAGAACCTCTTCGA
LOC100130111 Fw	GGAGTTCGAGACCTGCCTT
LOC100130111 Rev	GTCTAGCAACATCAGAGCGG
LOC101929705long Fw	TCCCGGCGATCTTGACTTAA
LOC101929705long Rev	CAAAGCAGGAAGACAGAGTGT
Xist Fw	GGCCAAGCTCCAGCTAATCT
Xist Rev	CGTCAAAGGGAATGGATCAC
APAF1 Fw	GGCTGTGGGAAGTCTGTATT
APAF1 Rev	CCCAACTGAAACCCAATGCA
BCL2L11 (BIM) Fw	GCCAGGCCTTCAACCACTAT
BCL2L11 (BIM) Rev	CAATACGCCGCAACTCTTGG
BIRC3 Fw	CATCCGTCAAGTTCAAGCCA
BIRC3 Rev	TGCATTTTCATCTCCTGGGC
CADM1 Fw	CGACCATCAGTTGCCAAGTC
CADM1 Rev	CAACTGAAACCTGCTGTCCT
CARM1 Fw	GAACGGCGAGATCCAGCG
CARM1 Rev	AGCACTTAAAGACACACACATCT
CASP7 Fw	AGATTCAGTGGATGCTAAGCC
CASP7 Rev	GGTCTTGATGGATCGCATGG
CCL2 Fw	CCCCAGTCACCTGCTGTTAT

CCL2 Rev	AGATCTCCTTGGCCACAATG
CDH2 Fw (NCAD)	GGTGGAGGAGAAGAAGACCAG
CDH2 Rev (NCAD)	GGCATCAGGCTCCACAGT
CDH6 Fw	TCACAGCCCAAGATCCAGA
CDH6 Rev	TCTGTCCATATCTGTGTGTCGAT
CXADR Fw	CTCCTGCTGTGCTTCGTG
CXADR Rev	GCAGTTTCCCCTTTGGCTTT
CXCL5 Fw	TCCTTCGAGCTCCTTGTGC
CXCL5 Rev	CAACGCAGCTCTCTCAACAC
CXCL8 Fw	AGAGTGATTGAGAGTGGACCA
CXCL8 Rev	ACTTCTCCACAACCCTCTGC
DAXX Fw	TGGGCATCAGGTTACAGGAG
DAXX Rev	GATAGTGCAGGGTCAACGC
FAM83A Fw	TTCGTTTGTGTGCTCCTGGA
FAM83A Rev	CTCTCCTTCCACACTCCGGA
FAS Fw	GACTCAGAACTTGGAAGGCC
FAS Rev	CATCCCCATTGACTGTGCAG
HAS2 Fw	CATGGTTGGAGGTGTTGGGG
HAS2 Rv	AAGACTGACAGGCCCTTTCT
HIP1 Fw	CTGTCAGCATCAATAAGGCCA

HIP1 Rev	GTCTGTGCCCCTTTCTCATG
IL6R Fw	CTGGGACTGTGCACTTGCT
IL6R Rev	ATTGCTGAGGGGGCTCTT
ITGB6 Fw	TCAGAAGGACTCCGGAAACA
ITGB6 Rev	CTTTTGGTGTTGGAAGAGGGT
JADE1 Fw	CCAGCAGCAGTGAGGATTCT
JADE1 Rev	TCGATCTTCATGTCTGGAGCA
LAMC1 Fw	AGTACCCCAGCTCCATCAAC
LAMC1 Rev	GCGCTTGTAAATGGCAAAGC
MAP3K1 (RAF) Fw	GCACGAATGGTTGGAAAGGA
MAP3K1 (RAF) Rev	TGGAGACTCAGCTGCTAAGT
MMP3 Fw	GACCTGGAAATGTTTTGGCCC
MMP3 Rev	GGCCAATTTCATGAGCAGCA
NCOA3 Fw	TGTGCGATTTTAAAGGAAACA GT
NCOA3 Rev	AACTCCCTGCCCTGTAGAAG
NECTIN2 Fw	GACGAGGGCAACTACACTTG
NECTIN2 Rev	CAGCTTGGTTCTTGGGCTTG
NECTIN3 Fw	AGCTGTTACATTCCCGCTTG
NECTIN3 Rev	TTATCAGGCTCACAGTGGGT
NLRP1 Fw	ATGGCCTCTGGATGAAACGT
NLRP1 Rev	GGCCTGCCTTTCTCTGATT

NOTCH2 Fw	CATGCTTTGTGTCTCGACCC
NOTCH2 Rev	CGTCCATTGGCACTCCTTAC
NUDT2 Fw	TCCGAAGATGCCTCATTCCC
NUDT2 Rev	TGCTTCTATGCCTGCTTCCT
PEAK1 Fw	GGAGAAGGGTAGAGCAGAGA
PEAK1 Rev	TCCTCTGTTACTTCATGGGTCA
PPP2CA Fw	ACGTTGGTGTCTAGAGCTCA
PPP2CA Rev	AGTTTGGAGCACTGAAAATCGT
RFK Fw	GGTGGTGAGCATAGGATGGA
RFK Rev	TGGCCACATTGAGGATTTCC
SAA1 Fw	GGGGAACTATGATGCTGCCA
SAA1 Rev	GCACCATGGCCAAAGAATCT
SEMA3E Fw	TGTGCCTTCATCAGAGTTGGA
SEMA3E Rev	ACATCTGCCCCTTCCTCTCT
SPP1 Fw	GACTCGTCTCAGGCCAGTTG
SPP1 Rev	AGGAGGCAAAAGCAAATCACTG
TGFB2 Fw	TGCCCTCCTACAGACTTGAG
TGFB2 Rev	TAAAGTGGACGTAGGCAGCA

Supplementary Table 7. Antibodies used

				Experimental Conditions	
Name	*Supplier*	*Reference*	*Species*	*WB*	*IF*
Anti-βactin	Sigma - Aldrich	A1978	Mouse (M)	1:5000	//
Anti-βcatenin	BD Bioscience	610153	Mouse (M)	1:1000	1:100
Anti-cJUN	Cell Signaling Technology	#9165	Rabbit (P)	1:1000	//
Anti-GABARAP	Cell Signaling Technology	#13733	Rabbit (P)	1:1000	//
Anti-LC3 A/B	Cell Signaling Technology	#4108	Rabbit (P)	1:1000	//
Anti-p62/SQSTM1	Abcam	ab91526	Rabbit (P)	1:1000	1:100
Anti-phosphorylated polymerase II (Ser5)	Abcam	ab5408	Mouse (M)	1:1000	//
Anti-Flag M2	Sigma Aldrich	F1804	Mouse (M)	//	1:100
Anti-mouse IgG-HRP	GE healthcare	NXA931	Sheep	1:5000	//
Anti-rabbit IgG-HRP	GE healthcare	NA934V	Donkey	1:5000	//
Anti-mouse Alexa 555	Invitrogen	A21422	Goat	//	1:1000
Anti-rabbit Alexa 594	Invitrogen	A11012	Goat	//	1:1000

M = Monoclonal P= Polyclonal