Original Research Article

Identification of a novel human E-cadherin splice variant and assessment of its effects upon EMT-related events†

María Laura Matos[1,§], Lara Lapyckyj[1,§], Marina Rosso[1], María José Besso[1], María Victoria Mencucci[1], Clara Isabel Marín Briggiler[1], Silvina Giustina[1], Laura Inés Furlong[2], and Mónica Hebe Vazquez-Levin[1,*]

[§] both authors equally contributed to this work

Department affiliations
[1] Instituto de Biología y Medicina Experimental (IBYME).
National Research Council of Argentina (CONICET).
Vuelta de Obligado 2490. Zip Code C1428ADN, Buenos Aires, Argentina.
Telephone: 54-11-47832869. Fax: 54-11-47862564

[2] Research Programme on Biomedical Informatics (GRIB) Hospital del Mar Medical Research Institute (IMIM) DCEXS, Pompeu Fabra University (UPF)
C/Dr Aiguader 88. Zip Code 08003, Barcelona, Spain.
Telephone: 34-93-3160521.

[*] **Corresponding author:**
Mónica H. Vazquez-Levin, PhD.
Instituto de Biología y Medicina Experimental (IBYME). National Research Council of Argentina (CONICET)-Fundación IBYME
Vuelta de Obligado 2490. Zip Code C1428ADN. Buenos Aires, Argentina.
Phone: 541147832869 ext 248/271
Fax: 541147862564
email: mhvazquez@ibyme.conicet.gov.ar, mhvazl@gmail.com

Running Head: A novel human E-cadherin splice variant.

Keywords: Cancer, Epithelial to Mesenchymal Transition (EMT), Epithelial cadherin (Ecadherin), Alternative splicing, Nonsense Mediated mRNA Decay (NMD)

Additional Supporting Information may be found in the online version of this article.

Received 4 March 2016; Revised 25 September 2016; Accepted 27 September 2016
Journal of Cellular Physiology

DOI 10.1002/jcp.25622

Grant Information

(to MHVL)

Contract Grant Sponsor: Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT)**.**

Contract Grant Number: PICT-SU 1072

Contract Grant Sponsor: CONICET**. Contract Grant Number:** PIP-740

Contract Grant Sponsor: Fundación Roemmers. **Contract Grant Number:** 2011-2013

Contract Grant Sponsor: Instituto Nacional del Cáncer**. Contract Grant Number:** INC 2014-2015.

Note*:* Authors declare no financial interests or research funding received to compromise the integrity of the published work.

ABSTRACT

Epithelial Cadherin (E-cadherin) is involved in calcium-dependent cell-cell adhesion and signal transduction. The E-cadherin decrease/loss is a hallmark of Epithelial to Mesenchymal Transition (EMT), a key event in tumor progression. The underlying molecular mechanisms that trigger E-cadherin loss and consequent EMT have not been completely elucidated. This study reports the identification of a novel human E-cadherin variant mRNA produced by alternative splicing. A bioinformatics evaluation of the novel mRNA sequence and biochemical verifications suggest its regulation by Nonsense-Mediated mRNA Decay (NMD). The novel E-cadherin variant was detected in 29/42 (69%) of human tumor cell lines, expressed at variable levels (E-cadherin variant expression relative to the wild type mRNA = 0.05%-11.6%). Stable transfection of the novel E-cadherin variant in MCF-7 cells (MCF7Ecadvar) resulted in downregulation of wild type E-cadherin expression (transcript/protein) and EMT-related changes, among them cell acquisition of fibroblastic-like phenotype, increased expression of Twist, Snail, Zeb1 and Slug transcriptional repressors and decreased expression of ESRP1 and ESRP2 RNA binding proteins. Moreover, loss of cytokeratins and gain of vimentin, N-cadherin and Dysadherin/FXYD5 proteins was observed. Dramatic changes in cell behavior were found in MCF7Ecadvar, as judged by the decreased cell-cell adhesion (Hanging-drop assay), increased cell motility (Wound Healing) and increased cell migration (Transwell) and invasion (Transwell w/Matrigel). Some changes were found in MCF-7 cells incubated with culture medium supplemented with conditioned medium from HEK-293 cells transfected with the E-cadherin variant mRNA. Further characterization of the novel E-cadherin variant will help understanding molecular basis of tumor progression and improve cancer diagnosis. This article is protected by copyright. All rights reserved

Introduction

Epithelial Cadherin (CDH1, E-cadherin) is the founder member of the cadherin superfamily, a large group of cell surface molecules involved in calcium (Ca^{2+}) dependent adhesion events (Takeichi, 1995; Gumbiner, 1996; van Roy and Berx, 2008). E-cadherin is encoded by the *CDH1* gene located on chromosome 16q22.1, and encompasses 16 exons which span a region of around 100 kb (Berx et al., 1995). The *CDH1* gene is transcribed to a unique functional mRNA that encodes a protein precursor, which is processed to a 120 KDa mature glycoprotein. E-cadherin is organized in an ectodomain composed of five extracellular (EC) domains of around 110 amino acids each, a single pass-transmembrane domain and a highly conserved carboxy-terminal cytoplasmic domain (Angst et al., 2001). The EC domains mediate homophilic and heterophilic interactions between adjacent cells (Nose et al., 1990; Ozawa and Kemler, 1998). On the other hand, the cytoplasmic domain binds to β-catenin, linking the adhesion protein to a network of actin filaments that strengthen cell-cell adhesion (Nagafuchi et al., 1993), and is involved in cell signaling (Yap et al, 2003). Under physiological conditions, localization of E-cadherin is mainly restricted to cell-cell contact sites (Bryant and Stow, 2004; Ivanov and Naydenov, 2013).

Disruption of cell-cell adhesion is a key event during tumor progression and metastasis. Cancer cells gradually lose their epithelial characteristics and gain a mesenchymal-like phenotype, in a complex process called Epithelial-to-Mesenchymal Transition (EMT) (Thiery, 2002; Savagner, 2015; Serrano-Gomez et al., 2016). The decrease or loss of E-cadherin expression is a hallmark of the EMT and several events have been associated to alterations in the adhesion protein. Among them, the expression of transcription factors that negatively modulate E-cadherin is observed. Moreover, β-catenin is lost from the cell membrane and filamentous actin (F-actin) belts are replaced by a network of stress fibers. Another important event of the EMT is the increased expression of other classical cadherins, in particular neural cadherin (N-cadherin) and, in some cases, placental cadherin (P-

cadherin). The reduction or absence of E-cadherin expression sometimes is accompanied by a reciprocally increased expression of Dysadherin/FXDY5, a glycoprotein detected in many tumors and associated to metastasis (Ino et al., 2002; Lubarski Gotlib, 2016).

Despite its relevance, the precise underlying molecular mechanisms that trigger E-cadherin loss and the consequent EMT process have not been completely elucidated. E-cadherin can be functionally inactivated or silenced by different mechanisms, among them downregulation of gene expression through promoter hypermethylation, histone deacetylation and transcriptional repression, as well as post-translational modifications, i.e. protein degradation. In this regard, classical E-cadherin proteolytic processing mediated by matrix metalloproteinases involves cleavage of the adhesion protein near the plasma membrane, with the release of an 86 KDa E-cadherin soluble ectodomain, and the generation of a 38 KDa fragment (CTF1) (van Roy and Berx, 2008).

For other members of the cadherin superfamily, the Alternative Splicing (AS) mechanism has been reported as a regulator of gene expression and function(s). Examples are the pituitary gland and brain cadherin (PB-cadherin) (Sugimoto et al., 1996), cadherin-7 (Kawano et al., 2002), cadherin-11 or OB-Cadherin (Feltes et al., 2002), cadherin 24 (Katafiasz et al., 2003) and LI-Cadherin (Wang et al., 2005). In 2009, a variant form of E-cadherin, lacking exon 11 as a result of an AS event, was reported and was associated to the decrease of the wild type form (Sharma et al., 2009). This E-cadherin variant was detected in chronic lymphocyte leukemia (Jordaan et al., 2013) and in solid tumors (Sharma et al., 2011; Liao et al., 2013). The proposed role of some of these AS-isoforms is to interfere with the homotypic cadherin-cadherin coupling, reducing the adhesiveness of the cells involved. For cadherin-7, the variant protein would interact with the wild type form and inhibit it function in cellular adhesion (Kawano et al., 2002). Moreover, co-expression of wild type and variant cadherin-11 produces a seven times increases in cells invasive properties, compared with cells expressing only the wild type transcript (Feltes et al., 2002).

The present study reports the identification of a novel E-cadherin variant mRNA lacking a 34 nucleotide-segment in Exon 14, which would be generated by AS. The investigations included 1) A bioinformatics evaluation of the novel mRNA nucleotide sequence and a set of studies to assess its regulation by Nonsense-Mediated mRNA Decay (NMD), 2) The evaluation of the novel E-cadherin mRNA expression levels in a panel of 42 established tumor cell lines, 3) The implementation of transient and stable transfection models that express the novel E-cadherin transcript in mammalian cell lines, and 4) The assessment of cell morphology, expression of EMT markers and a set of functional assays in cells transfected with the novel variant. As result of these evaluations, the novel E-cadherin variant was detected in 69% of tumor cell lines analyzed. Evaluation of the nucleotide sequence led us to propose its generation by AS. Moreover, amino acidic sequence assessment of the novel variant and transient transfection studies revealed its secretory nature. Its expression in MCF-7 cells (and supplementation of cell cultures with conditioned medium (CM) from HEK-293 cells transfected with the variant mRNA) resulted in the downregulation of E-cadherin expression (transcript and protein) and triggered molecular and functional changes characteristic of the EMT. A schematic representation of the E-cadherin novel variant synthesis and its potential role in tumor progression has been elaborated and is shown in the last figure of this report.

Materials and Methods

All procedures reported in this study were done following protocols that complied with biosafety and ethical guidelines accepted worldwide.

Chemicals

Unless specified, chemicals were of analytical grade and purchased from Sigma Chemical Co. (St. Louis, MO). Electrophoresis reagents were products of BioRad (Richmond, CA), or specifically indicated throughout the article. Molecular biology reagents were of highest quality and purchased from Qiagen (Hilden, Germany) and Invitrogen (Carlsbad, CA), unless specified.

Antibodies

The following antibodies towards human E-cadherin (anti E-cadherin) were used throughout the study: a) polyclonal antibody H-108 (amino acids 600-707; cadherin 5 or membrane proximal extracellular domain; Santa Cruz Biotech., Santa Cruz, CA, USA), and monoclonal antibodies b) monoclonal antibody HECD-1 (amino acids 333-379; cadherin 2 domain; Zymed-Invitrogen Life Technologies; Carlsbad, CA, USA) and c) monoclonal antibody 610181 (amino acids 773-791; cytoplasmic domain; BD Biosciences, San Diego, CA, USA). In addition, antibodies towards β-catenin (610153; BD Biosciences), pan-cytokeratin (clone AE1/AE3, Dako; Carpinteria, CA, USA), N-cadherin (H-163, Sta. Cruz; clone 3-B9, Zymed-Invitrogen Life Technologies), P-cadherin (H-105, Sta. Cruz), Dysadherin/FDXY5 (NCC-M53) (Shimamura et al., 2003), vimentin (clone V9, Dako), actin (A2668 (Sigma) and Clone ACTN05 (Neomarkers, Basel, Switzerland)) and tubulin (clone D66; Sigma) were used in the study. For immunocytochemistry protocols, Cy3-labelled anti-mouse or anti-rabbit (Sigma) IgGs were used as secondary antibodies. Anti-mouse (Vector Lab. Inc., Burlingame, CA, USA) or anti-rabbit (Sigma) IgGs coupled to horseradish peroxidase were employed as secondary antibodies in Western immunoblotting assays.

Tissues and Cell Lines

Human epididymal tissues were obtained from adult patients undergoing orchiectomy as treatment for prostatic carcinoma, and not receiving any hormonal treatment prior to surgery. All human tissues

used in the study were obtained under donor's written consent, and protocols were approved by the Argentine Society of Clinical Investigation review board (Marín-Briggiler et al., 2008).

A total of 42 human cell lines were analyzed to assess the expression levels of wild type and novel variant E-cadherin mRNAs. Cell lines involved in the analyses were derived from melanoma (Mel 888, IIB-Mel Ian, IIB-Mel-J, IIB-Mel-Les, L1D, L2F6, LBlast, LCMV, SB-2, A375N), as well as from gastric (AGS, MKN45, SNU1, SNU5, SNU16, Kato III, N87), breast (MCF-7, T47D, HS578T, MDA-MB231), prostate (PC3, LNCap, 22RV1, C42, C42B, DU-145), ovarian (OV90, OAW-42, SKOV-3, TOV-112), pancreatic (Capan-1, MiaPaCa2, SW1990), colon (Caco-2, HT29), uterine (HeLa, Hec1A, Ishikawa), lung (A549), bladder (T24) and liver (Hep3B) cancer.

In addition, COS-7 (ATCC CRL-165™), MCF-7 (ATCC HTB-22™) and HEK-293 (ATCC CRL-1573 ™) established cell lines were used to develop working models of the novel E-cadherin variant mRNA. MCF-7 human breast cancer cells were also used as models to treat with conditioned medium obtained from HEK cells stably transfected with the novel E-cadherin variant form. All cell lines were obtained from the ATCC (Manassas, VA, USA) or kindly provided by colleagues.

In silico analysis

The nucleotide sequence of the identified clones was subjected to bioinformatics analysis. Initially the identified sequence was contrasted against the one reported for E-cadherin (Reference number: NM_004360.3; Homo sapiens cadherin 1, type 1, E-cadherin epithelial (*CDH1*), mRNA) and the analysis was extended to the evaluation of a set bioinformatics analysis programs, according to the specific type of analysis to be performed, as follows:

A) Analysis of multiple sequences: 1) Blast (Blast "two sequences") option: BLAST: Basic Local Alignment Search Tool; 2) Clustal X. In addition, PCR fragments and plasmids obtained in the

experiments were evaluated with Clustal X and Blast programs against the sequence corresponding to the total or the reference E-cadherin segment involved.

B) Analysis related to Alternative Splice events (Net2Gene)

C) Translation of the nucleotide sequence amino acid "Translate" (ExPASy tool Proteomics Server program.

D) Analysis of protein structure: 1) TopPred II: Software for Membrane Protein Structure Predictions; 2) HMMTOP: Transmembrane topology prediction server; 3) TMHMM: Prediction of transmembrane helices in proteins.

E) Analysis of cell fate: 1) Program Yloc: Interpretable Subcellular Localization Prediction, 2) ESLpred, 3) BaCelLo: Balanced Subcellular Localisation Predictor.

A survey on mutations in the *CDH1* gene was done using the COSMIC (Catalog of somatic mutations in cancer) (http://cancer.sanger.ac.uk/cosmic) and the CCLE (Cancer Cell Lines Encyclopedia) (https://portals.broadinstitute.org/ccle/home) databases.

Cell culture

Cell lines used throughout the study (parental and transfectants) were grown, harvested and processed following the recommendations provided by the manufacturer, as previously reported (Lapyckyj et al., 2010). Basically, cells were grown in the appropriate cell culture medium in the presence of 10% heat-inactivated fetal calf serum, 100 IU/mL penicillin, 100 mg/mL streptomycin, 2 mM glutamine and 2 mg/mL bovine insulin (cell culture media and supplements from Invitrogen Life Technologies) and 15 mM HEPES, and incubated at 37°C in a humidified atmosphere of 5% CO_2 in air. Low-passage cell lines were used in all cases; in addition, quality-control tests were done by monitoring cell morphology and growth rates, as well as mycoplasma detection with Hoechst 33258 or standard PCR procedures (data not shown).

In some experimental protocols, the Conditioned Medium (CM) was collected and analyzed. In those cases, cells were grown until 80% confluence was attained, the medium was changed to fresh one and supplemented or not with 10% heat-inactivated fetal calf serum; after 48 h, CM was collected and centrifuged 10 min at 2000 rpm, and stored at -20ºC until used.

Development and handling of an expression library

An expression cDNA library from human epididymis was constructed using the ZAP Express® vector (Stratagene, San Diego, CA) as previously reported (Marín-Briggiler et al., 2008). Purified phaghemid DNAs were subjected to PCR with T3 and T7 primers using a standard procedure.

Nucleotide sequence analysis of clones was performed in a set of selected clones at the Core Research Center from the University of Chicago (Chicago, IL, USA) with primers T3 and T7, as well as with internal primers designed from the partial sequencing results.

Subcloning of the novel E-cadherin variant transcript and transfection assays

The pcDNA3 plasmid containing the sequence of the E-cadherin wild type form was kindly provided by Dr. C. Niessen. To clone the novel E-cadherin variant mRNA, plasmids obtained from the expression library containing the novel sequence (and verified by nucleotide sequence analysis) were used as template, amplified by PCR and cloned in the wild type plasmid previously digested with *Bam HI* restriction enzyme to introduce the variant sequence. The pcDNA3 plasmid containing the novel sequence was purified with the Qiagen Plasmid Midi Kit (Qiagen). Nucleotide sequence analysis was done in the final construct (at least three recombinant clones).

For transient transfection protocols, COS-7 cells were used. PcDNA3 plasmids containing the sequences of wild type and novel variant E-cadherin were used for transient transfection protocols;

the pcDNA3 "empty" vector was used as control. Procedures were performed on a total of 100,000 cells at baseline and 1.5 µg of each plasmid with Lipofectamine™ (Thermo-Life, Carlsbad, CA, USA) / DNA ratio of 2.5 : 1, following the recommendations of the manufacturer. Briefly, a mixture of Lipofectamine™, culture medium and the indicated plasmid was made and added to the cell culture and incubated for 4 h 45 min. At the end of the incubation, the cell culture medium was removed and replaced by fresh medium with supplements and incubated for additional 24 h. CM and cells were recovered and processed for RNA and total cell protein analysis, as well as for fluorescent immunocytochemistry studies.

Stable transfection protocols were done using MCF-7 and HEK-293 cell lines and pcDNA3 plasmids containing the sequence of novel E-cadherin variant and control ("empty" plasmid), respectively. Transfection protocols were done with one µg of purified plasmid and Gene transfection reagent as Juice® (Novagen®-Merck-Millipore, USA) following the instructions of the manufacturer. To generate clonal transfectants, cells were seeded in 96-well plates in the presence of G418 (400 µg/mL, Invitrogen), and individual clones were obtained using the limiting dilution technique. The procedure of single cell cloning was repeated to obtain 100% clonal purity. Three clones were used in the studies presented in this report. Expression levels of the transfected gene were verified in all cases by means of quantitative RT-PCR procedures.

RNA extraction and RT-PCR procedures

Total RNA was isolated with the RNeasy kit (Qiagen) or the Trizol reagent (Invitrogen Life Technologies) according to the manufacturer instructions, and quantified (QUBIT, Invitrogen Life Technologies). Synthesis of complementary DNA (cDNA) was performed with oligo-dT and SuperScript™ III Reverse Transcriptase (Invitrogen Life Technologies). Negative controls omitting the RNA or the reverse transcriptase were included in all cases. Standard end point PCR amplification

protocols were carried out using *Taq*I DNA polymerase (Qiagen). Reverse transcription procedures were verified by amplification of a fragment for the Glyceraldehyde 3-Phosphate DeHydrogenase (GAPDH).

The quantitative assessment of genes analyzed along the study (E-cadherin wild type and variant forms, Twist, Snail, Zeb1 and Slug, ESRP1 and ESRP2, cytokeratin-19, vimentin and Dysadherin/FXDY5 mRNA levels) was done by real time PCR with the Applied Biosystems 7500 Real Time PCR unit, using the SYBR Green® PCR Master Mix (Applied Biosystems, Foster City, CA) as previously reported (Lapyckyj et al., 2010). All samples were run in triplicates and negative controls (no template) were included in all cases. To confirm specificity of the signal observed in real time protocols, melting curves were done in each run. Transcript relative expression was calculated using GAPDH as housekeeping control in all cases.

Transcripts expression levels were determined in established tumor cell lines by calculating $2^{-\Delta Ct}$, where ΔCt = [Ct transcript under study - Ct GAPDH housekeeping gene] (arbitrary units). In some cases, transcript expression levels were calculated using as reference the mRNA expression values detected in cells at the t0 condition, in parental cells or in cells transfected with pcDNA3, as specifically indicated in the Figure legend. The calculation describing these relations is $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct$ = [ΔCt test sample - ΔCt reference sample], and ΔCt= [Ct gene under study - Ct housekeeping gene]. The last calculation was also used to estimate E-cadherin variant mRNA expression levels relative to the E-cadherin wild type mRNA shown in ***Supplementary*** **Table 1**. In this last case the E-cadherin variant mRNA was the test sample, and the E-cadherin wild type mRNA was the reference.

In addition, nucleotide sequence analysis was performed on the PCR products of cDNA templates generated from the RNA of cell lines tested in this study. PCR products were purified using a commercial kit (Wizard SV PCR Clean Up Gel & System; Promega) following the manufacturer's

instructions. Purified products were quantified by agarose gel electrophoresis done in the presence of a molecular weight marker and mass (QuantiMarker, Biodynamics, Buenos Aires, Argentina) and analyzed at the Core Research Center from the University of Chicago (Chicago, IL).

Protein Extraction and Western Immunoblotting

For cell protein extraction, cell cultures were washed twice at 4°C with phosphate buffered saline solution (PBS, pH 7.4) supplemented with 2 mM $CaCl_2$, and placed at -20°C until processing. Cells were resupended in Laemmli sample buffer, boiled for 5 min and sonicated until the solution was no longer viscous (3 times for 30 sec at maximal power, Sonifier Cell Disruptor, model W 140; Heat Systems-Ultrasonics, Inc., Plainview, NY). Cell lysates were centrifuged at 4°C for 30 min at 10,000 xg to eliminate cellular debris and stored at -70°C. Protein concentration was determined by the Bradford assay (BioRad).

Protein mixtures (cell lysates and CM) were supplemented with β-mercaptoethanol to a 5% final concentration, boiled for 10 min, loaded onto the gel, and subjected to SDS-PAGE in 8 or 10% polyacrylamide gels, as specifically indicated. Molecular weight standards (BioRad protein broad range) were included in all runs.

A replica of protein patterns was obtained on nitrocellulose membranes (Hybond-ECL, Amersham/GE, Buckinghamshire, Great Britain) after Western immunoblotting using standard procedures. Membranes were placed for 1 h at room temperature with PBS containing 0.02% Tween-20 and 5% skimmed milk powder (blocking buffer), followed by overnight incubation at 4ºC with the specific primary antibody (see results) diluted in blocking buffer. Blots were washed, placed for 1 h at room temperature with anti-mouse or anti-rabbit horseradish peroxidase conjugated IgGs in blocking buffer, and developed with enhanced chemiluminiscence (ECL kit, Amersham/GE) following the

procedure suggested by the manufacturer. In all cases, immunodetection was specific, because no signal was detected when the primary antibody was replaced by the corresponding normal IgG at the same concentration (data not shown).

Immunofluorescence protein detection in whole cells

Cells grown on glass coverslips until 80% confluence were fixed with 4% (v/v) paraformaldehyde in PBS for 10 min. In some cases, cell fixation was followed by permeabilization in PBS supplemented with 0.2% (v/v) Triton X-100. For fluorescence immunocytochemistry, non-specific binding sites were blocked with 4% (w/v) BSA in PBS and overnight incubation at 4°C with anti E-cadherin (HECD-1, 2 μg/mL), anti-β-catenin (2.5 μg/mL), anti-cytokeratin (1 μg/mL), anti-vimentin (1 μg/mL) or anti-Dysadherin/FXYD5 (1 μg/mL) antibodies. After several washes, cells were incubated with purified secondary antibody coupled to Cy3-fluorophore for an additional hour. Nuclear cell staining was achieved by incubation with Hoechst 33342. Coverslips were mounted with Vectashield anti-fade solution (Vector).

To assess for the presence of F-actin, cells were incubated in the presence of Alexa Fluor 488 phalloidin (Invitrogen Life Technologies). Cells were evaluated with a confocal microscope (Nikon C1; excitation lines: 488nm and 544nm, emission filters: 515-530 nm and 570-LP nm); images were acquired using a 60x/1.40 oil objective.

Assessment of Nonsense-Mediated mRNA Decay

MCF-7 cells were cultured in DMEM medium F-12 and supplements until monolayers reached 75% confluence and treated with Actinomycin-D and Cycloheximide, basically as previously reported (Barbier et al., 2007). In the first case, cell treatment involved the addition 5 μg/mL of Actinomycin-

D/mL. In the second case, cell cultures were treated with medium supplemented with 10 µg/mL Cycloheximide. In both cases, aliquots were taken at 0, 2.5, 5 and 7.5 h and processed for mRNA isolation and analysis. Real time PCR analysis was done to quantify the expression levels of E-cadherin wild type and variant mRNAs.

Assessment of cell adhesiveness, motility, migration and invasion

Hanging Drop Assay

The procedure was done as previously described (Redfield et al., 1997). Basically, cells were grown, harvested, dissociated, counted and adjusted at a final concentration of $2.5x10^5$ cells/mL. Aliquots of 50,000 cells were placed on drops and cultured at 37°C in anchorage-independent conditions for up to 24 h to assess their ability to form cell aggregates. Photographs were taken at 0, 12 and 24 h to evaluate cell aggregates and their characteristics. Number of cells in each aggregate was determined and results grouped (1-5, 5-50 cells and more that 50 cells/aggregate) and plotted.

Wound Healing Assay

The procedure was done as previously described (Meunier et al., 2010). Basically, MCF7pcDNA3 and MCF7Ecadvar cells were grown at 37°C until confluence. The medium was removed and a small wound was done using a sterile 200-µL yellow plastic tip. Wounded cell monolayers were washed three times with PBS to remove loosely attached cells, after which fresh culture medium was added. Cells were cultured for up to 24 h. Photographs were taken at different times (0, 4, 8, 16, and 20 h) and analyzed using the Image J software. To score the results, measurement of the wound width to 2 points was done at initial time immediately after making the wound (t0) and at the last recording time (tf), in

this particular case 20 h after beginning the experiment. Taking the t0 value as 1, the remaining values (tf) were expressed in a relative fashion.

Migration and Invasion Assays

MCF7pcDNA3 and MCF7Ecadvar cells were grown in 6-well plates up to 70% confluence. Twenty four h prior to the assay, CM were removed and fresh media was added, cell monolayers were washed with sterile PBS, dissociated, harvested, counted and cell concentration was adjusted to 2 x 10^4 viable cells/200 µL of serum free medium. The cell suspension was seeded on top of a Transwell™ of 0.8 µm pore diameter device (Millipore) and placed on a 24 well-cell culture plate. At the bottom of the Transwell™ device, 10% fetal bovine serum was placed as chemo-attractant. Cell culture proceeded at 37°C for 8 h. Then, the Transwell™ device culture plate and the top of the culture medium were discarded. Three washes with PBS were performed to remove cells that did not migrate. Cells were fixed and stained with a 0.5% crystal violet solution in 20% methanol and quantification was done; photographs of at least 10 different fields were taken and cells were counted using Image J software. The same protocol was used for the invasion assay, with the exception that the device Transwell™ was previously covered with a layer of 100 µL of a commercial extracellular matrix (Matrigel™, BD). In this case the test was run for 16 h.

Note: In all four assays described above, no differences were found between MCF-7 and MCF7-pcDNA3 (working model control) cells.

Effect of CM from HEK293Ecadvar transfectants upon MCF-7 cell cultures

MCF-7 cells were maintained in Dulbecco-modified Eagle medium (DMEM) F12 supplemented with 10% of FBS, and HEK-293 were maintained in minimum essential media with non-essential amino acids (MEM/NEAA) containing 10% of FBS. Cells were grown at 37°C in a humidified atmosphere with 5% CO_2 in air.

For treatments with the different CM, 2 x 10^4 MCF-7 cells were seeded per well of a 6 culture plate in DMEM F12 supplemented with 10% of FBS. The day after seeding, cells were washed twice with 1x PBS and treated with a mixture of 70% CM (from HEK293Ecadvar or HEK293pcDNA3 (control) cells) + 30% DMEM F12 10% of FBS for 7 days, with medium changes every other day.

After one week, total RNA of MCF-7 cellular cultures treated with the different CM was extracted with TRIZOL (Invitrogen) and processed for cDNA synthesis and qRT-PCR analysis to assess wild type E-cadherin transcript expression levels.

Statistical Analysis

All experiments were run in triplicates. Data are presented as the mean ± Standard Error of the Mean (SEM). Evaluations done on MCF7Ecadvar and MCF7pcDNA3 control stable transfectants were done using Mann-Whitney test or Kruskall Wallis test followed by means of Dunn´s post-test analysis. A $P<0.05$ was considered significant for each comparison. Statistical analyses were performed using GraphPad Prism 5.01 (GraphPad Software, San Diego, CA).

RESULTS

1. Identification of a human epithelial cadherin novel variant mRNA

During the screening of an expression library prepared with RNA from human epididymis in the unidirectional lambda ZAP Express (Stratagene, EEUU) vector, a novel E-cadherin transcript was identified by serendipity. Briefly, a total of ten plates (35,000 pfu/plate) were analyzed and 19 positives were identified, of which six were randomly selected, cloned and sequenced. A graphical representation of the nucleotide sequence analysis done on the identified clones is shown in **Figure 1.A.** Basic bioinformatics tools were applied to the sequencing results (data obtained from NCBI database and use of programs to align sequences, BlastN and Clustal X). As result of this analysis, two clone families were identified. In one family, clones were found to carry the expected sequence of the mRNA encoding the functional E-cadherin, as provided over 99% homology with the reference sequence (NM_004360). In the other family, clones were found to have a nucleotide sequence highly homologous to the reference, but showed a 34 bp-sequence deletion (**Figure 1.B.**).

A PCR strategy, designed with primers flanking the region where differences between sequences of mRNAs had been found (**Figure 1.C.**, left), served to confirm the expression of both transcripts in mRNA samples of epididymal tissue from individuals other than those used for library construction, as well as in a commercially available testis cDNA library (data not shown). These findings ruled out the possible explanation for this variant form as being an artifact generated during construction of the expression library, or as a specific mRNA form produced solely in the epididymis.

Bioinformatics analyses were done to characterize changes in the novel nucleotide sequence identified. Firstly, alignment with the reference nucleotide sequence of E-cadherin (NM_004360) led us to identify the site of the 34-bp deletion, which was localized on the first 34 nt of Exon 14 and would introduce a premature stop codon. The sequence was then subjected to additional analyses, to evaluate whether this novel transcript could result from an alternative splicing event. For that, an analysis of putative splicing sites in the sequence of E-cadherin was performed using the NetGene 2 software. The analysis identifies constitutive splicing sites; in addition, it lists other potential sites that

display "lower level of confidence", indicating that site is used less frequently than a constitutive site (Hebsgaard et al., 1996). When the analysis was run with the human *CDH1* gene, the site used to produce the novel E-cadherin mRNA sequence was identified as a putative splicing site. Additional alignment analysis on the *CDH1* nucleotide sequences of other mammalian species showed conservation of key nucleotides reported to be relevant for recognition by the splicing machinery (data not shown). Thus, the novel E-cadherin variant mRNA would result from an alternative splicing event. The sequences involved in the E-cadherin constitutive and alternative splicing events between exon13, intron 13 and exon14 are shown in **Figure 1.D.**

At this point, it was highly relevant to define the strategy for detection of the novel E-cadherin variant form. Since difficulties were found to detect the variant transcript when the PCR was done with primers flanking the deleted region, a primer-specific sequence was designed as a new strategy (**Figure 1.C.,** right) using as template one plasmid of each family (see below). With this new approach, a quantitative analysis of both transcripts was optimized and found to be highly reproducible, specific, efficient and sensitive. Specificity was defined by the presence of a sole fragment (confirmed when melting curves were run and agarose gel-profiles were analyzed) and confirmation of the nucleotide sequence of each PCR product generated. Optimization also involved the construction of standard curves with different amounts of each plasmid, with the consequent establishment of the dynamic range that revealed a high efficiency of the assay (-3.32 slope) and a high sensitive detection (fg of plasmid) (data not shown). Using this methodology, a set of samples from different human tissues (epididymis, testis, breast, ovary, endometrium) were evaluated to assess the presence of the novel E-cadherin mRNA. Detectable levels of the novel variant form as well as of the wild type E-cadherin mRNA were found in all samples evaluated (data not shown), favoring the notion of the occurrence of an alternative splicing to generate this novel form rather than its occurrence as the result of a point mutation.

2. Expression of the novel E-cadherin variant mRNA in human tumor cells

Taking into account the association between the occurrence of AS and cancer (Shkreta et al., 2013; Zhang and Manley, 2013; Pagliarini et al, 2015) and the role of other cadherins' splice variants as negative regulators of the wild type transcript, it was of interest to evaluate the presence and expression levels of the novel E-cadherin variant transcript in a panel of 42 established human tumor cell lines of diverse tissue origin (urological: prostate, bladder; gynecological: breast, ovary, endometrium; digestive: gastric, liver, pancreas, colon; other: skin, lung; a list of cells lines studied is presented in Materials and Methods). Expression levels of the novel variant E-cadherin transcript were determined by applying the optimized highly sensitive quantitative Real Time PCR protocol (see Materials and Methods).

Results obtained in this analysis are shown in **Figure 2**. Specific amplification of segments of both transcript was observed. Moreover, detectable levels of the E-cadherin variant transcript were found in 29 of the 42 tumor cell lines evaluated (69%). The established cell lines in which the novel variant was detected were generated from tumors of several tissues/organs in which alterations in E-cadherin have been reported (breast, ovary, endometrium, stomach, colon, pancreas, liver, skin, lung, prostate, bladder). Transcript levels of the novel E-cadherin form varied among cells ($3.3\pm0.83 \times 10^{-4}$; mean ± SEM; median=$1.96\times10^{-4}$), being the highest found in SNU16 gastric cancer cell line (21×10^{-4}), and followed by OV-90 (12.5×10^{-4}), T47D (9.6×10^{-4}), OAW-42 (8.4×10^{-4}), Capan-1 (5.3×10^{-4}), and C42 (5.1×10^{-4}); the lowest levels were found in the T24 bladder cancer cell line (2.05×10^{-8}). No association was found between the levels of E-cadherin variant transcript and aggressiveness reported for the cell lines in the panel evaluated.

Of the 29 cell lines, 28 (97%) also had detectable expression levels of the wild type E-cadherin mRNA, although it was present at variable levels. In the Hep 3B liver cell line, the E-cadherin wild

type transcript was not detected. When mRNA levels of the E-cadherin variant transcript were calculated relative to the wild type mRNA, a 0.05 to 1.2% expression of the novel form was found in 23 cases (80%). The remaining cell lines showed higher relative levels of the E-cadherin variant transcript. Of those, the AGS and MKN45 gastric cancer cell lines were found to have 11.6% and 8.2% levels of the novel mRNA, respectively, and a 10.7% was observed for the Caco-2 colon cancer cell line. Other cell lines, such as Capan-1 (pancreas) and A549 (lung) depicted intermediate high mRNA levels (3.1 and 2%, respectively) (***Supplementary*** **Table 1**).

In order to search for the presence of mutations around the *CDH1* exon13-exon 14 constitutive splicing donor and/or acceptor sites, a survey was done on the COSMIC and the CCLE databases. Of the total 42 cell lines evaluated, 22 were found to be listed on these databases with information on the *CDH1* gene; all of them had detectable levels of the E-cadherin variant mRNA. Of those, 16 (73%) were found not to have any somatic mutations in *CDH1*. The remaining six (27%) were gastric cancer cell lines AGS, MKN45, Snu5 and Kato III, and prostate cancer cell lines DU-145 and LNCap. Specifically regarding cell lines with highest expression levels of the E-cadherin variant form, while no data was available on the *CDH1* sequence for SNU16 and C42, no mutations were reported in OV90, T47D, OAW42 and Capan 1 (***Supplementary*** **Table 2**). Altogether, presence of the E-cadherin variant form could not be associated to mutations in the sequence around exon13-exon14 donor and acceptor splicing sites.

3. Regulation of novel E-cadherin variant mRNA by Nonsense-Mediated mRNA Decay (NMD)

The Nonsense-Mediated mRNA decay (NMD) was initially identified as a post-transcriptional mRNA quality control mechanism operating in the cytoplasm associated to translation termination and capable to selectively degrade mutated mRNAs containing premature termination codons (PTCs) (Losson and Lacroute, 1979) Later it was shown that the NMD mechanism eliminates multiple non-

mutated mRNAs resulting from defects in pre-mRNA processing; it has been estimated that one third of the alternative splice transcripts are subjected to NMD (Lewis et al, 2003) . However, the efficiency of this mechanism may be altered; the NMD can be inhibited by wide variety of cellular stresses, some of which are reported associated to tumorigenesis (Gardner., 2010). In the last decade, the molecules and mechanisms underlying the NMD have been identified, as well as its role in the stabilization of alternatively spliced mRNA isoforms, and the role of NMD in tumorigenesis (Hug et al., 2016; Karousis et al., 2016).

In mammals, the NMD operates in mRNAs that have a PSC located 50-55 bp upstream to the next splicing site and in one exon different from the last one (Lejeune and Maquat, 2005). An *in silico* analysis was done on the novel E-cadherin variant mRNA sequence to evaluate the possible regulation of its cellular levels through the NMD degradation mechanism. As result of this analysis, the E-cadherin variant mRNA appears to have molecular characteristics to be detected and processed by the NMD mechanism.

The experimental verification of the results from the *in silico* analysis was done using MCF-7 breast cancer cells using an approach previously reported. Cells were subjected to treatment with Actinomycin-D or Cycloheximide. Regarding the first approach, Actinomycin-D inhibits RNA elongation, resulting in an impairment of cell transcription. Transcripts targeted to NMD are expected to undergo degradation at a higher rate than the wild type functional transcript, reflected in a differential decrease on its concentration with time, which depends on how the mRNA is being processed. As shown in **Figure 3.A.**, a significant decrease in the E-cadherin variant transcript was observed, becoming undetectable at 2.5 h post-treatment with Actinomycin-D (*; $P<0.05$). On the other hand, a reduction in E-cadherin wild type transcript levels was also found with time, although undetectable levels were observed 5 h after treatment. In agreement with our observations, a previous report estimated a half-life of 2 h for the wild type E-cadherin mRNA (Jacob and Udey, 1998).

Regarding the second approach, since Cycloheximide inhibits protein biosynthesis, the NMD is not activated, and accumulation of aberrant transcripts is expected. As shown in **Figure 3.B.**, higher levels of the E-cadherin variant mRNA were found at all times analyzed when compared to the wild type mRNA. Significant differences were found at 2.5 h (*; $P<0.05$), as well as at after 5 h and 7.5 h (*; $P<0.01$) after treatment. Contrasting, no significant changes in levels of the E-cadherin wild type transcript were detected at all times tested.

1. Expression of the novel E-cadherin variant mRNA in stable transfectants of MCF-7 cells

To study the effect of the novel E-cadherin variant mRNA expression upon the expression/functions of the E-cadherin wild type transcript and of related molecules, a working cell model was developed using the MCF-7 human breast cancer cell line. In addition to the vast information available in the literature on this cell line with regard to E-cadherin and other members of the adherent complex, MCF-7 cells have no reported somatic mutations in the *CDH1* gene (***Supplementary*** **Table 2**) and are known to express high levels of the wild type E-cadherin transcript and produce abundant quantities of the full-length functional cell-cell adhesion protein in comparison to other human breast cancer cells (Lapyckyj et al., 2010). In addition, the estimated expression of the E-cadherin variant mRNA in this cell line ($2^{-\Delta CT}=2.05 \times 10^{-4}$; arbitrary units) is similar to the median value of 1.96×10^{-4}, estimated for the 42 tumor-cell panel analyzed in this study (**Figure 2**). Moreover, MCF-7 cells were found to have low levels of the endogenous variant E-cadherin mRNA relative to the wild type transcript (0.24%; ***Supplementary*** **Table 2**).

Based on these findings, MCF-7 cells were stably transfected with the eukaryotic plasmid pcDNA3 containing the novel E-cadherin variant mRNA (MCF7Ecadvar cells); as control, stable transfectants were generated with the pcDNA3 "empty" plasmid (MCF7pcDNA3).

Three stable clones were obtained and the expression levels of the E-cadherin variant mRNA were verified using a quantitative real time PCR test previously optimized. MCF7Ecadvar cells were found to express ~30 times more of the transfected variant transcript than control cells (**Figure 4.A.**) (*; $P<0.01$). In association to the expression of the novel E-cadherin variant mRNA, a concomitant decrease in the wild type E-cadherin mRNA levels was observed, finding in MCF7Ecadvar cells 10,000 times less transcript than in MCF7pcDNA3 cells (**Figure 4.A.**). Such decrease was already noticed in MCF7Ecadvar polyclonal batch cultures (data not shown).

In agreement with the mRNA results, the 120 KDa-full length E-cadherin protein was not detected in total protein MCF7Ecadvar cells, even when loading double amount of protein extracts from MCF7Ecadvar transfectants. To rule out the possibility of decreased levels in the full length protein caused by proteolysis, presence of the 38 KDa CTF1 E-cadherin fragment was analyzed; however, no signal was detected for this fragment (**Figure 4.B.**), suggesting other mechanism involved in regulation of E-cadherin expression.

In association to the expression of the novel E-cadherin variant transcript, several changes were observed in MCF7Ecadvar cells. Firstly, cells depicted a fibroblast-like phenotype and a decrease in cell-cell contacts (**Figure 4.C.**). In agreement with these findings, MCF7Ecadvar cells were found to form only few colonies of a small cell number and did not generate a monolayer; meanwhile, a sustained growth and the development of several colonies that generated a monolayer was observed in MCF7pcDNA3 cells (**Figure 4.D.**).

By fluorescence immunocytochemistry, a very low signal, localized to the cell cytoplasm, was observed for E-cadherin in the MCF7Ecadvar cells. The adaptor protein β-catenin underwent re-localization from the cell membrane to the cytoplasm and around the nucleus in the MCF7Ecadvar cells. In line with these findings, F-actin was organized in non-junctional stress fibers oriented towards

the boundary between cells in these cells, a distribution that resembles F-actin reorganization found during cell invasion and metastasis (Yilmaz and Christofori, 2009). Contrasting with these observations, control cells showed localization of E-cadherin mainly in the cell borders and contact sites (**Figure 4.E., left panel**), the adaptor protein was associated to the plasma membrane, accompanying the adhesion protein distribution (**Figure 4.E., right panel**) and the F-actin belts co-localized with E-cadherin and β-catenin (**Figure 4.E**.).

2. ***Expression of the E-cadherin novel variant transcript and EMT***

As part of the changes associated to the EMT process, increased expression of transcription factors, among them Twist, Snail, Zeb1 and Slug has been reported (Tania et al., 2014).These factors are negative regulators of E-cadherin by their interaction with the *CDH1* promoter region (Lin et al., 2014). When evaluated in MCF7Ecadvar cells, expression levels of Twist, Snail, Zeb1 and Slug were significantly higher than in control cells (**Figure 5.A.**) (*; $P<0.01$).

The expression loss of ESRP1 and ESRP2 has been reported as an obliged and widespread event during EMT (Warzecha and Carstens, 2012). These RNA binding proteins are responsible for an AS regulated-program, which is essential for maintaining morphology and function of epithelial cells. Transcript expression analysis of ESRP1 and ESRP2 in our study model revealed significant lower levels of both transcripts in the MCF7Ecadvar cells when compared to MCF7pcDNA3 control cells (**Figure 5.B**.) (*; $P<0.05$).

An increased expression of E-cadherin transcriptional repressors has been related to the "*cadherin switch"* event characterized by an increased expression of N-cadherin and, in some cases, P-cadherin (van Roy and Berx, 2008). Protein analysis revealed a specific signal for the 135 KDa N-cadherin full length protein only in MCF7Ecadvar cell extracts. In agreement with these findings, a strong signal for N-cadherin was mainly detected in MCF7Ecadvar cells by fluorescence

immunocytochemistry (**Figure 5.C.**). A similar analysis done on P-cadherin showed no major changes in protein expression levels between MCF7Ecadvar and control cells (**Figure 5.D.**).

To further define the cell phenotype of MCF7Ecadvar transfectants, cells were evaluated for the expression of cytokeratins, typical epithelial markers, and vimentin, a classic mesenchymal marker (Serrano-Gomez et al., 2016). MCF7pcDNA3 control cells showed a typical epithelial cell profile, with expression of the cytokeratin 19 transcript and a strong protein signal for cytokeratins; in addition, lack of vimentin mRNA and protein was found in these cells. Contrasting, E-cadherin variant mRNA in the transfectants was associated to a loss of the cytokeratin epithelial marker and a gain of the vimentin mesenchymal marker (mRNA and protein) (**Figure 5.E. and Figure 5.F.**).

Finally, the expression of Dysadherin/FXYD5 was evaluated in MCF7Ecadvar and control cells. As indicated in the introduction, Dysadherin/FXYD5 is a cancer cell membrane-associated glycoprotein expressed in several cancer types that depict decreased E-cadherin expression (Ino et al., 2002). The evaluation of its expression revealed a significant increase in its mRNA levels in MCF7Ecadvar compared to MCF7pcDNA3 cells. Protein analysis confirmed the expression of Dysadherin/FXDY5 in MCF-7 cells transfected with E-cadherin variant, as shown by the detection of a specific signal localized mainly at the cell membrane and a Dysadherin/FXYD5 protein band of ~55KDa. No signal for Dysadherin/FXYD5 was observed in control cells (**Figure 5.G.**).

3. Expression of E-cadherin variant form and in vitro cell adhesiveness, motility, migration and invasion properties

Alterations found in MCF-7 cells transfected with the novel E-cadherin variant mRNA summarized in the previous sections prompted us to carry out a set of *in vitro* functional assays to evaluate their cell-cell adhesion, motility, migration and invasion properties.

Cell adhesiveness was assessed by means of the hanging drop method. Results, shown in **Figure 6.A.**, revealed striking differences between MCF7Ecadvar and control cells. Whereas 100% of MCF7pcDNA3 cells were grouped in >50 cell-aggregates after a 12 h-incubation, cells expressing the variant mRNA produced small aggregates (<5 cells/aggregate) and only in few cases groups of >50 cells were found after incubations of 24 h (*; $P<0.01$).

To assess cell motility characteristics, the Wound Healing assay was done. MCF7Ecadvar and control cell monolayers were monitored for up to 24 h after the wound was made. Again, differences in cell behavior were observed between both cell lines (**Figure 6.B**.). While a ~5 % reduction in the wound was observed in the control cells, a 43% reduction (*; $P<0.01$) was registered in MCF7Ecadvar cells, suggesting 8 times higher motility than control cells.

Finally, cell migration and invasion properties were assessed using the commercial Transwell™ (Corning) device. As shown in **Figure 6.C.** and **D.**, MCF7Ecadvar cells depicted a significantly (*; $P<0.01$) higher migration and invasion capacity than control cells.

7. Analysis of the protein encoded by the E-cadherin novel variant mRNA

The existence of isoforms containing PSC was initially inferred from EST (expressed sequence tag) data. These findings lead researchers to believe that a proportion of these mRNA manage to evade NMD degradation pathway. In line with this hypothesis, several studies have reported the ability of aberrant transcripts to escape NMD and translate to protein (Inácio et al., 2004; Stockklausner et al., 2006; Hamid et al., 2010).

Based on this background information, expression of the putative protein encoded by the novel E-cadherin variant mRNA was analyzed using a set of studies that combined bioinformatics tools and transfection assays. First, the amino acidic sequence of the protein encoded by the E-cadherin variant

mRNA was deduced using the Translate software (ExPASy Tools). The sequence identified showed 100% homology with the E-cadherin wild type form up to the residue in position 721, in which the sequence frame-shift gives rise to a new peptide sequence of 36 residues and a PSC. The novel protein spans 757 residues, contrasting with the 882 amino acids of the wild type form (**Figure 7.A.**). The predicted Mr for the novel polypeptide is 83.7 KDa. An amino acidic sequence analysis run through several software tools, among them TMHMM, YLoc, ESLpred, ESLpred2 and BaCelLo, anticipated the secretory nature of the novel protein (**Figure 7.B.**).

A transient expression protocol in COS-7 cells was designed to express the polypeptide encoded by the novel E-cadherin variant mRNA (COS7Ecadvar) as well as to assess its cell localization and estimate its apparent MW. COS-7 cells transfected with the empty plasmid served as negative control (COS7pcDNA3). In addition, a parallel transfection protocol was done with a plasmid containing the E-cadherin wild type transcript and used as an additional control of the transfection protocol (COS7Ecadwt). Fluorescence immunocytochemical analysis of E-cadherin in COS7Ecadvar cells revealed a low signal in the cell membrane and cytoplasm. COS7pcDNA3 cells showed no signal for E-cadherin, while COS7Ecadwt showed a strong signal in the cell borders, as expected (**Figure 7.C.**)

SDS-PAGE followed by Western immunobloting of cell lysates and CM from COS7Ecadvar cells revealed a specific positive signal for a secretory protein form of 94 KDa; an additional form of 104 KDa was detected in cell lysates. As expected, COS7Ecadwt cells were found to express the 120 KDa full length protein in cell lysates and the 86 KDa E-cadherin ectodomain in the CM (**Figure 7.D.**). No signal for E-cadherin was found in COS7pcDNA3 cell lysates and CM (data not shown).

To determine whether MCF7Ecadvar cells expressed the novel secretory E-cadherin protein, CM was collected and analyzed by SDS-PAGE and Western immunoblotting. In agreement with the results

obtained with COS7Ecadvar transfectants, a specific signal for a ~94 KDa protein was found in cell lysates. In MCF7pcDNA3 cells CM, the 86 KDa ectodomain was immunodetected (**Figure 7.E.**).

8- Effect of Conditioned Medium containing the novel E-cadherin variant protein upon MCF-7 cell cultures

As previously shown, MCF-7 cells stably transfected with the novel E-cadherin variant form underwent dramatic changes in cell morphology, lost their epithelial characteristics and acquired a fibroblast-like phenotype. Moreover, E-cadherin expression was significantly reduced and cells were found to secrete the 94 KDa protein encoded by the novel variant transcript. Based on these results, it was of interest to assess whether changes in cell morphology and E-cadherin expression could be attributed, at least in part, to an autocrine/paracrine effect of the secreted novel E-cadherin variant protein upon MCF-7 cells.

A strategy was designed to challenge MCF-7 cells with CM from cells producing the novel E-cadherin variant polypeptide (details on the experimental procedure are presented in the Materials and Methods). To generate the CM containing the E-cadherin novel form, the HEK-293 cell line was selected. HEK-293 is a commercially available established non tumor cell line of epithelial morphology, derived from human embryonic kidney. Compared to MCF-7 cells, HEK-293 cells were found to express less than 0.5% of E-cadherin wild type transcript (**Figure 8.A**.), ruling out the contribution of other E-cadherin protein forms in the CM besides the novel variant. HEK-293 cell stably transfected with E-cadherin variant (HEK293Ecadvar) and pcDNA3 empty vector (HEK293pcDNA3) were generated, and expression of the novel E-cadherin variant mRNA was verified (**Figure 8.B.**). HEK293Ecadvar cells were found to express more than 30 times E-cadherin variant mRNA than MCF7Ecadvar cells (**Figure 8.C.**). In addition, HEK293Ecadvar cells showed a

faint signal for E-cadherin by immunocytochemistry, but a strong specific signal for a 94 KDa E-cadherin protein in the CM; no immunoreactivity for E-cadherin was detected in control cells (**Figure 8.D**.).

Based on these findings, CM from stable transfectants of HEK293Ecadvar (VCM; variant CM) and control (TCM, transfection CM) was collected and used to challenge MCF-7 cells during 7 days. After 5 days of incubation, MCF-7 cells exposed to the CM from HEK293Ecadvar displayed a fibroblastic-like morphology (**Figure 8.E.**). Moreover, at the end of the study, these cells showed a significant (*; $P<0.01$) decrease in E-cadherin wild type transcript levels (**Figure 8.F.**). No changes were observed in cells incubated with media supplemented with CM from HEK293pcDNA3 cells.

Discussion

Alterations in the expression and function(s) of E-cadherin are a key event in deregulation of cell-cell adhesion during tumor progression and metastasis. Several mechanisms have been found to explain some of these abnormalities; among those, the literature has numerous examples of E-cadherin decrease caused by loss of heterozygosity, presence of mutations, promoter methylation, expression of transcriptional repressors and post-translational protein modifications (van Roy and Berx, 2008). In the present study, the identification and partial characterization of a novel human E-cadherin transcript is described. Screening of an expression library from a human reproductive tissue led us to the identification of clones with the sequence encoding the functional protein. Unexpectedly, other clones depicting a sequence with a 34-bp deletion were identified. Additional bioinformatics analyses led us to propose that the novel transcript would result from an alternative splicing event of the immature E-cadherin transcript. This novel transcript was called “E-cadherin variant” to distinguish it from the E-cadherin wild type form encoding the functional protein.

To characterize the novel E-cadherin variant mRNA expression, several studies were carried out. A quantitative PCR detection assay was developed and optimized to evaluate the expression levels of the wild type and novel variant transcripts, reducing artifacts associated to the reported low abundance of alternative-spliced variant mRNAs (Walton et al., 2007; Harvey and Cheng, 2016). Protocol optimization involved assessment of specificity, sensitivity, reproducibility and efficiency of the detection method. Two clones known to carry the sequences encoding the wild type and novel variant mRNA were used during assay optimization. Highly specific, sensitive, reproducible and efficient results were obtained using 1 fg of plasmid as template and detecting both transcripts even when a difference in concentration between them was five orders of magnitude. In recent years, implementation of Next Generation Sequencing (NGS) technology has enabled the identification of novel splice variants. However, in cases of alternative splice variants highly similar to constitutive transcripts, some difficulties have been highlighted due to limitations in computational approaches and bioinformatics tools (Bryant et al., 2012). The use of NGS to perform transcript quantification of splice variants has grown over the last 5 years (Beretta et al., 2014; Adamopoulus et al., 2016). In particular, recent reports have been released, describing new tools to improve data analysis (Xu et al., 2015; Song et al., 2016), which anticipates a promising future for the use of this technology in the identification and quantification of novel variants produced by AS.

The optimized quantitative PCR assay was used to evaluate the expression of the wild type and novel variant E-cadherin transcripts in a panel of human tumor cell lines. While over 69% of cell lines evaluated were found to simultaneously express both E-cadherin transcripts, relative low levels of the novel variant (less or equal than 1%) were detected, as previously seen for other transcripts (Vandenbroucke et al., 2001). Given the fundamental role reported for E-cadherin in the maintenance of homeostasis, it is expected that a variant transcript encoding a protein isoform different than the functional, should be tightly regulated due to the possible deleterious effects that it could bring to the

cell. In agreement with these observations, there are reports describing low levels of alternative isoforms relative to the constitutive transcripts subjected to AS (Madsen et al., 1995; Jung et al., 2011). Despite their low abundance, these transcripts are able to exert a modulatory effect upon the expression of the constitutive mRNA, and may be translated to proteins that, at low levels, are capable to produce a specific effect on the cells (Dreumont et al., 2005).

The NMD surveillance system was tested as a possible mechanism responsible for regulation of the novel E-cadherin variant levels. The strategies selected for these studies have been widely used for NMD analysis (Rajavel and Neufeld, 2001, Hilman et al., 2004; Jakobson et al., 2012). Moreover, biochemical verifications were done using MCF-7 cells; this established cell line is one, if not the most, *in vitro* model used worldwide to evaluate regulation of human E-cadherin gene expression. The high expression of the novel E-cadherin variant mRNA found accumulated in cells treated with Cycloheximide and its rapid degradation observed in cells incubated with Actinomycin-D strongly suggests an active transcription of the novel E-cadherin mRNA and its degradation by the NMD mechanism. These findings may explain the low levels of the variant transcript detected in a high proportion of the samples evaluated. Relative increased levels of the novel E-cadherin variant form may be explained by alterations in the NMD mechanism during cancer reported in several cells models (Gardner, 2010; Karam et al., 2013). Our findings are also in agreement with those observed for the variant de E-cadherin lacking exon 11 (Sharma et al., 2011).

Bioinformatics analyses done on the nucleotide sequence of the novel clones led authors to propose an AS event to explain the presence of this novel E-cadherin variant mRNA. It is now known that near 90% of protein-coding human genes produce multiple mRNA transcripts by AS (Pan et al., 2008; Gamazon and Stranger, 2014); this process represents a generalized mechanism to the expansion of proteomic diversity (Pan et al.; 2008; Nilsen and Graveley, 2010; Nellore et al., 2016). The AS is both spatially and temporally controlled, resulting in the expression of diverse splicing variants in the

same or different tissues, in different developmental stages, or in response to a pathologic process. Interestingly, not all alternative transcripts produce functional proteins. This mechanism also has errors, such as nucleotide sequences that would include a PSC and alteration of the protein structure, cellular localization, stability and changes in post-translational modifications, as the one reported in this study for the identified novel E-cadherin variant (Scholzova et al., 2007; Avenson and Barry, 2014).

Dysregulations in AS patterns have been associated with many human diseases including cancer (Chabot and Shkreta, 2016; Scotti and Swanson, 2016). In particular, in the last decade, alterations in the splicing of tumor suppressor genes or oncogenes have been identified (Bonomi et al., 2013). Aberrant AS events generate variants found to contribute to tumor development and progression. Changes in splicing have been attributed to several steps of this process, among them mutations in the splicing regulatory elements, altered expression of factors responsible for AS, and aberrant regulation of proteins and signaling pathways that control the AS. The AS- variants produced can frequently promote different or even opposite biological effects and have consequences on cellular functions (Zhang and Manley, 2013). These alterations affect a wide variety of cellular functions, including cell-cycle progression, signal transduction, proliferation and cell differentiation, avoidance of apoptosis angiogenesis, motility and invasion. Tumor cells quickly adapt to these extra- and intra-cellular stimuli, and survive in difficult environments and acquire resistance to therapy (Bechara et al., 2013; Pagliarini et al., 2015).

In association to the increased expression of the novel variant transcript, MCF7Ecadvar cells showed dramatic changes in cell morphology, as well as in the expression of EMT markers and in cell functions. The significant loss of functional E-cadherin (transcript and protein), the increased expression of transcriptional repressors, as well as the concomitant loss of cytokeratins and a gain of vimentin, and a switch towards the expression of the mesenchymal N-cadherin, all found in the E-

cadherin variant transfectants are well recognized EMT markers (Thiery, 2002; van Roy and Berx, 2008). Moreover, a fibroblast-like morphology, as well as reduced adhesiveness, and increased motility, migration and invasion properties was found in MCF7-Ecadvar cells, all indicative of a more aggressive cell behavior. In line with these findings, MCF7Ecadvar cells depicted significant low levels of ESRP1 and ESRP2. These RNA binding proteins are known to maintain the epithelial phenotype by AS-regulation of gene products involved in cell-cell adhesion, cell motility and cell-matrix adhesion (Warzecha et al., 2009; Ishii et al., 2014), and their loss was reported to induce phenotypic changes observed during EMT (Warzecha et al., 2010, 2011). Interestingly, changes described in this report on gene expression related to EMT observed in the MCF7Ecadvar cells could not be completely reproduced by the sole downregulation of E-cadherin wild type using a specific siRNA to block transcript expression (data not shown).

In addition to the studies done to assess the novel E-cadherin transcript, a set of bioinformatics and biochemical evaluations were done with the polypeptide sequence encoding the novel variant E-cadherin. These studies revealed the putative secretory nature of the protein expressed by translation of the novel transcript. The novel protein would have similar characteristics to the E-cadherin ectodomain, since both polypeptides share a large amount of the sequence, and both are soluble proteins. In addition, it would be localized in the extracellular medium, and would have, at least in part, intact homotypic cell-cell interaction domains. The use of models of transient and stable transfectants in cells that does express negligible amounts of wild type E-cadherin confirmed their localization mainly extracellular. In addition, they revealed the ability of the conditioned medium carrying the E-cadherin variant polypeptide to induce a decrease in the expression of E-cadherin wild type and to alter cell phenotype after 7 days of cell treatment.

The identification of transcripts encoding soluble E-cadherin forms may have great relevance with regard to diverse human pathologies. Several reports have described the presence of soluble E-

cadherin in tissue and fluids (plasma, serum, urine, conditioned medium) of bladder, breast, colorectal, gastric, liver, non-small cell lung, ovarian, prostate, esophageal and skin squamous cell carcinomas, as well as in multiple myeloma and Hodking´s disease (Grabowska and Day; 2014; Repetto et al., 2014). In addition, it has been found in patients with an inflammatory systemic response and syndrome of multiorgan dysfunction (Pittard et al., 1996) and with acute pancreatitis (Sewpaul et al., 2009), and associated to infections (Grabowska and Day; 2014). Detection of soluble E-cadherin has been done mostly by ELISA and RPPA techniques, evaluations that revealed a significant association was found between soluble E-cadherin levels and tumor aggressiveness and response to treatment (Repetto et al., 2014). The antibodies used in some of these studies recognize the E-cadherin variant form described in this report (reviewed in Repetto et al., 2014).

Specifically regarding to cancer, soluble E-cadherin has been proposed to act as a paracrine/autocrine signaling molecule, since it may diffuse into the extracellular environment and the blood, promoting cell migration, invasion and metastatic behavior in different tumors (Nawrocki-Raby et al., 2003; De Wever et al., 2007; Johnson et al., 2007; van Roy and Berx, 2008; Grabowska and Day; 2014; Repetto et al., 2014; Hu et al., 2016). The results presented in this report on the effect of supplementation of the CM expressing the novel E-cadherin variant form are in line with these findings.

In conclusion, the present study reports the identification of a novel human E-cadherin transcript that would result of an alternative splicing event and would be regulated by the NMD surveillance mechanism. The E-cadherin variant mRNA was detected in human tumor cells derived from different tissues. When transfected in MCF-7 human breast cancer cells, dramatic changes were found in cell morphology, as well as a significant decreased expression (transcript and protein) of E-cadherin, reduced cell-cell interaction, and increased motility, migration and invasion properties was observed. At molecular level, changes in gene expression resemble those characterized for cells

undergoing Epithelial to Mesenchymal Transition. Some of these changes were triggered by supplementation with conditioned medium of cell expressing the novel E-cadherin variant.

Based on the changes presented in this report, a schematic representation of the E-cadherin novel variant synthesis and its potential role in tumor progression has been elaborated and is shown in **Figure 9**. Future studies of this novel transcript will contribute to the understanding of the molecular mechanisms regulating gene expression of E-cadherin and specifically the impact of this novel mRNA in the deregulation of E-cadherin expression during tumor progression and metastasis. Understanding the expression of this novel E-cadherin variant mRNA may also help in the management of cancer diagnosis and treatment.

Acknowledgements

The pcDNA3 plasmid containing the sequence of the E-cadherin wild type form was kindly provided by Dr. C. Niessen from Germany. The NCC-M53 anti Dysadherin antibody was kindly provided by Dr. S. Hirohashi from Japan. Authors would like to thank Dr. J. Reventos from Vall d´Hebron Hospital (Spain), Dr. O. Podhajcer from Leloir Institute (Argentina) and colleagues at IBYME that provided us with human tumor cell lines and RNA preparations to address some of the screening studies presented in this report.

References

- Adamopoulos PG, Kontos CK, Tsiakanikas P, Scorilas A . 2016. Identification of novel alternative splice variants of the BCL2L12 gene in human cancer cells using next-generation sequencing methodology. Cancer Lett. 1;373:119-29.

- Angst BD, Marcozzi C, Magee AI. 2001. The cadherin superfamily. J Cell Sci 114:625-626.

- Avesson L, Barry G. 2014. The emerging role of RNA and DNA editing in cancer. Biochim Biophys Acta 1845:308-16.

- Barbier J, Dutertre M, Bittencourt D, Sanchez G, Gratadou L, de la Grange P, Auboeuf D. 2007. Regulation of H-ras splice variant expression by cross talk between the p53 and nonsense-mediated mRNA decay pathways. Mol.Cell Biol 27:7315-7333.

- Bechara EG, Sebestyén E, Bernardis I, Eyras E, Valcárcel J. 2013. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. Mol Cell. 52:720–33.

- Beretta S, Bonizzoni P, Vedova GD, Pirola Y, Rizzi R. 2014. Modeling alternative splicing variants from RNA-Seq data with isoform graphs. J Comput Biol. 21:16-40.

- Berx G, Cleton-Jansen AM, Nollet F, de Leeuw WJ, van de Vijver M, Cornelisse C, van Roy F. 1995. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. EMBO J 14:6107-6115.

- Bonomi S, Gallo S, Catillo M, Pignataro D, Biamonti G, Ghigna C. 2013. Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy. Int J Cell Biol. 2013:962038.

- Bryant DM, Stow JL. 2004. The ins and outs of E-cadherin trafficking. Trends Cell Biol 14:427-434.

- Bryant DW Jr, Priest HD, Mockler TC. 2012. Detection and quantification of alternative splicing variants using RNA-seq. Methods Mol Biol. 883:97-110.

- Chabot B and Shkreta L. 2016. Defective control of pre-messenger RNA splicing in human disease. J Cell Biol 212:13-27.

- De Wever O, Derycke L, Hendrix A, De Meerleer G, Godeau F, Depypere H, Bracke M. 2007. Soluble cadherins as cancer biomarkers. Clin. Exp. Metastasis 24:685-697.

- Dreumont N, Maresca A, Boisclair-Lachance JF, Bergeron A, Tanguay RM. 2005. A minor alternative transcript of the fumaryl acetoacetate hydrolase gene produces a protein despite being likely subjected to nonsense-mediated mRNA decay. BMC Mol Biol 6: 1.

- Feltes CM, Kudo A, Blaschuk O, Byers SW. 2002. An alternatively spliced cadherin-11 enhances human breast cancer cell invasion. Cancer Res 62:6688-6697.

- Gamazon ER, Stranger BE. 2014. Genomics of alternative splicing: evolution, development and pathophysiology. Hum Genet 133:679-87.

- Gardner LB. 2010. Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis. Mol Cancer Res. 8:295-308

- Grabowska MM, Mark L. Day ML. 2014. Soluble E-cadherin: More Than a Symptom of Disease. Front Biosci (Landmark Ed) 17:1948–1964.

- Gumbiner BM. 1996. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. Cell 84:345-357.

- Hamid R, Hedges LK, Austin E, Phillips JA 3rd, Loyd JE, Cogan J. 2010. Transcripts from a novel BMPR2 termination mutation escape nonsense mediated decay by downstream translation re-initiation: implications for treating pulmonary hypertension. Clin Genet 77:280-286.

- Harvey SE, Cheng C. 2016. Methods for Characterization of Alternative RNA Splicing. Methods Mol Biol 1402:229-241.

- Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. 1996. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res 24:3439-3452.

- Hillman RT, Green RE, Brenner SE. (2004). An unappreciated role for RNA surveillance. Genome Biol 5: R8.

- Hu QP, Kuang JY, Yang QK, Bian XW, Yu SC. 2016. Beyond a tumor suppressor: Soluble E-cadherin promotes the progression of cancer. Int J Cancer. 2016 138:2804-12.

- Hug N, Longman D, Cáceres JF. 2016. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Research 44:1483–1495.

- Inácio A, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, Faustino P, Lavinha J, Liebhaber SA, Romão L. 2004. Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. J Biol Chem 279:32170-32180.

- Ino Y, Gotoh M, Sakamoto M, Tsukagoshi K, Hirohashi S. 2002. Dysadherin, a cancer-associated cell membrane glycoprotein, down-regulates E-cadherin and promotes metastasis. Proc Natl Acad Sci USA 99:365-370.

- Ishii H, Saitoh M, Sakamoto K, Kondo T, Katoh R, Tanaka S, Miyazawa K. 2014. Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. J Biol Chem 289:27386–27399.

- Ivanov AI, Naydenov NG . 2013. Dynamics and Regulation of Epithelial Adherens Junctions. Recent Discoveries and Controversies. Int Rev Cell Mol Biol 303:27-99.

- Jakobson M, Lintulahti A, Arumae U. 2012. mRNA for N-Bak, a neuron-specific BH3-only splice isoform of Bak, escapes nonsense-mediated decay and is translationally repressed in the neurons. Cell Death. Dis 3, e269.

- Johnson SK, Ramani VC, Hennings L, Haun RS. 2007. Kallikrein 7 enhances pancreatic cancer cell invasion by shedding E-cadherin. Cancer 109:1811-1820.

- Jordaan G, Liao W, Sharma S. 2013. E-cadherin gene re-expression in chronic lymphocytic leukemia cells by HDAC inhibitors BMC Cancer 13: 88.

- Jung YD, Huh JW, Kim DS, Kim YJ, Ahn K, Ha HS, Lee JR, Yi JM, Moon JW, Kim TO, Song GA, Han K, Kim HS. 2011. Quantitative analysis of transcript variants of CHM gene containing LTR12C element in humans. Gene 489:1-5.

- Karousis E.D., Nasif S, Mühlemann O. 2016. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. WIREs RNA 2016. doi: 10.1002/wrna.1357

- Karam R, Wengrod J, Gardner LB, Wilkinson MF. 2013. Regulation of nonsense-mediated mRNA decay: implications for physiology and disease. Biochim Biophys Acta 1829:624-33.

- Katafiasz BJ, Nieman MT, Wheelock MJ, Johnson KR. 2003. Characterization of cadherin-24, a novel alternatively spliced type II cadherin. J Biol Chem 278:27513-27519.

- Kawano R, Matsuo N, Tanaka H, Nasu M, Yoshioka H, Shirabe K. 2002. Identification and characterization of a soluble cadherin-7 isoform produced by alternative splicing. J Biol Chem 277:47679-47685.

- Lapyckyj L, Castillo LF, Matos ML, Gabrielli NM, Lüthy IA, Vazquez-Levin MH. 2010. Expression analysis of epithelial cadherin and related proteins in IBH-6 and IBH-4 human breast cancer cell lines. J Cell Physiol 222:596-605.

- Lejeune F, Maquat LE. 2005. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. Curr Opin Cell Biol 17:309-315.

- Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A 100:189-192.

- Liao W, Jordaan G, Srivastava MK, Dubinett S, Sharma S, Sharma S. 2013. Effect of epigenetic histone modifications on E-cadherin splicing and expression in lung cancer. Am J Cancer Res 3:374-89

- Losson R, Lacroute F. 1979. Interference of nonsense mutations with eukaryotic messenger mRNA stability. Proc Natl Acad Sci U S A 76:5134-5137.

- Lubarski Gotliv I. 2016. FXYD5: Na(+)/K(+)-ATPase Regulator in Health and Disease. Front Cell Dev Biol 4:26.

- Madsen MW, Reiter BE, Lykkesfeldt AE. 1995. Differential expression of estrogen receptor mRNA splice variants in the tamoxifen resistant human breast cancer cell line, MCF-7/TAMR-1 compared to the parental MCF-7 cell line. Mol Cell Endocrinol. 109: 197-207.

- Marín-Briggiler CI, Veiga MF, Matos ML, Echeverría MF, Furlong LI, Vazquez-Levin MH. 2008. Expression of epithelial cadherin in the human male reproductive tract and gametes and evidence of its participation in fertilization. Mol Hum Reprod. 14:561-71.

- Meunier L, Puiffe ML, Le Page C, Filali-Mouhim A, Chevrette M, Tonin,P.N., Provencher DM and Mes-Masson AM. 2010. Effect of ovarian cancer ascites on cell migration and gene expression in an epithelial ovarian cancer in vitro model. Transl Oncol 3, 230-238.

- Nagafuchi A, Tsukita S, Takeichi M. 1993. Transmembrane control of cadherin-mediated cell-cell adhesion. Semin Cell Biol 4:175-181.

- Nawrocki-Raby B, Gilles C, Polette M, Martinella-Catusse C, Bonnet N, Puchelle E, Foidart JM, van Roy F, Birembaut P. 2003 E-Cadherin mediates MMP down-regulation in highly invasive bronchial tumor cells. Am J Pathol. 163:653-661.

- Nellore A, Jaffe AE, Fortin J-P, Alquicira-Hernández J, Collado-Torres L, Wang S, Phillips RA, Karbhari N, Hansen KD, Langmead B, Leek JT. 2016. Human splicing diversity across the Sequence Read Archive. doi: http://dx.doi.org/10.1101/038224.

- Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457-463.

- Nose A, Tsuji K, Takeichi M. 1990. Localization of specificity determining sites in cadherin cell adhesion molecules. Cell 61:147-155.

- Ozawa M, Kemler R. 1998. The membrane-proximal region of the E-cadherin cytoplasmic domain prevents dimerization and negatively regulates adhesion activity. J Cell Biol 142:1605-1613.

- Pagliarini V, Naro C, Sette C. 2015. Splicing Regulation: A Molecular Device to Enhance Cancer Cell Adaptation. Biomed Res Int. :543067.

- Pan Q, Shai O, Lee LJ, Frey B J, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat.Genet. 40:1413-1415.

- Pittard AJ, Banks RE, Galley H, Webster NR. 1996 Soluble E-cadherin concentrations in patients with systemic inflammatory response syndrome and multiorgan dysfunction syndrome. Br. J. Anaesth. 76:629-631.

- Rajavel KS, Neufeld EF. 2001. Nonsense-mediated decay of human HEXA mRNA. Mol Cell Biol 21:5512-5519.

- Redfield A, Nieman MT, Knudsen KA. 1997. Cadherins promote skeletal muscle differentiation in three-dimensional cultures. J Cell Biol 138: 1323-1331.

- Repetto O, De Paoli P, De Re V, Canzonieri V, Cannizzaro R. 2014. Levels of Soluble E-Cadherin in Breast, Gastric, and Colorectal Cancers. Biomed Res Int.: 408047

- Savagner P. 2015. Epithelial-mesenchymal transitions: From cell plasticity to concept elasticity. Curr Top Dev Biol.112:273-300

- Scholzova E, Malik R, Sevcik J, Kleibl Z. 2007. RNA regulation and cancer development. Cancer Lett. 246:12-23.

- Scotti MM , Swanson MS. 2016. RNA mis-splicing in disease. Nat Rev Genet. 17:19-32.

- Serrano-Gomez SJ, Maziveyi M, Alahari SK. 2016. Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. Molecular Cancer, 15:18.

- Sewpaul A, French JJ, Khoo TK, Kernohan M, Kirby JA, Charnley, RM. 2009. Soluble E-cadherin: an early marker of severity in acute pancreatitis. HPB Surg., 2009:397375.

- Sharma S, Liao W, Zhou X, Wong DT, Lichtenstein A. 2011. Exon 11 skipping of E-cadherin RNA downregulates its expression in head and neck cancer cells. Mol Cancer Ther 10:1751-1759.

- Sharma S, Lichtenstein A. 2009. Aberrant splicing of the E-cadherin transcript is a novel mechanism of gene silencing in chronic lymphocytic leukemia cells. Blood 114:4179-4185.

- Shimamura T, Sakamoto M, Ino Y, Sato Y, Shimada K, Kosuge T, Sekihara H, Hirohashi S. 2003. Dysadherin overexpression in pancreatic ductal adenocarcinoma reflects tumor aggressiveness: relationship to e-cadherin expression, J Clin Oncol 21:659-667.

- Song L , Sabunciyan S , Florea L . 2016. CLASS2: accurate and efficient splice variant annotation from RNA-seq reads. Nucleic Acids Research 44:e98.

- Shkreta L, Bell B, Revil T, Venables JP, Prinos P, Elela SA, Chabot B. 2013. Cancer-Associated Perturbations in Alternative Pre-messenger RNA Splicing. Cancer Treat Res 158:41-94.

- Stockklausner C, Breit S, Neu-Yilik G, Echner N, Hentze MW, Kulozik AE, Gehring NH. 2006. The uORF-containing thrombopoietin mRNA escapes nonsense-mediated decay (NMD). Nucleic Acids Res 34:2355-2363.

- Sugimoto K, Honda S, Yamamoto T, Ueki T, Monden M, Kaji A, Matsumoto K, Nakamura T. 1996. Molecular cloning and characterization of a newly identified member of the cadherin family, PB-cadherin. J Biol Chem 271:11548-11556.

- Takeichi M. 1995. Morphogenetic roles of classic cadherins. Curr Opin Cell Biol 7:619-627.

- Tania M, Khan MA, Fu J. 2014. Epithelial to mesenchymal transition inducing transcription factors and metastatic cancer. Tumour Biol. 35:7335-42.

- Thiery JP. 2002. Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer 2:442-454.

- Vandenbroucke II, Vandesompele J, Paepe AD, Messiaen L. 2001. Quantification of splice variants using real-time PCR. Nucleic Acids Res. 29, E68.

- van Roy F, Berx G. 2008. The cell-cell adhesion molecule E-cadherin. Cell Mol Life Sci 65:3756-3788.

- Walton HS, Gebhardt FM, Innes DJ, Dodd PR. 2007, Analysis of multiple exon-skipping mRNA splice variants using SYBR Green real-time RT-PCR. J Neurosci Methods 160:294-301.

- Wang XQ, Luk JM, Leung PP, Wong BW, Stanbridge EJ, Fan ST. 2005. Alternative mRNA splicing of liver intestine-cadherin in hepatocellular carcinoma. Clin Cancer Res 11:483-489.

- Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, Guo W, Xing Y, Carstens RP. 2010. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. EMBO J 29:3286-3300.

- Warzecha CC, Shen S, Xing Y, Carstens RP. 2009. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. RNA Biol 6: 546-562.

- Warzecha CC, Carstens RP. 2012. Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). Semin Cancer Biol. 22:417-427.

- Xu Z, Peters RJ, Weirather J, Luo H, Liao B, Zhang X, Zhu Y, Ji A, Zhang B, Hu S, Au KF, Song J, Chen S. Full-length transcriptome sequences and splice variants obtained by a combination ofsequencing platforms applied to different root tissues of Salvia miltiorrhiza andtanshinone biosynthesis. Plant J. 82:951-61.

- Yap AS, Kovacs EM. 2003. Direct cadherin-activated cell signaling: a view from the plasma membrane. J Cell Biol 160:11-16.

- Yilmaz M, Christofori G. 2009. EMT, the cytoskeleton, and cancer cell invasion. Cancer Metastasis Rev 28:15-33.

- Zhang J, Manley JL. 2013. Misregulation of pre-mRNA alternative splicing in cancer. Cancer Discov. 3:1228-37.

Figure Legends

Figure 1. Identification of the novel E-cadherin variant mRNA in a human epididymis RNA expression library. (A) Graphical representation of the nucleotide sequence analysis done on 6 of the 19 positive clones identified. **(B)** Nucleotide sequence analysis of the novel E-cadherin variant mRNA (variant). Sequence alignment of the novel transcript (variant) with the reference E-cadherin wild type transcript (NM_004360.3) shows a 34-bp deletion in the variant form. **(C)** Graphical representation of the PCR strategy used to detect the novel variant and the wild type E-cadherin mRNAs. Detection is based on the use of primers (arrows) flanking the region depicting differences between the variant and the wild type mRNA (Left) or the use of a primer-specific strategy (Right). In this last strategy, while the "forward" primer is common to both transcripts, the "reverse" primer is specific for each mRNA. **(D)** Graphical representation of the constitutive and the alternative splicing mechanisms involved in the generation of the E-cadherin wild type and the E-cadherin variant mRNAs, respectively. Nucleotide sequences around exon13, exon14 and intron 13 involved in both splicing events are shown.

Figure 2. Expression analysis of the novel E-cadherin variant mRNA in a panel of human tumor cell lines of diverse tissue origin. (**A**) Expression profile of E-cadherin transcripts (wild type and variant) and GAPDH housekeeping gene obtained in a total of 42 cell lines is reported. Images of 2% agarose gels stained with Ethidium Bromide following standard procedure are shown. Positive (plasmids containing the sequence of each E-cadherin mRNA form) and negative (no template) PCR control are included. **(B)** Quantitative expression of the E-cadherin variant mRNA in 29 established human tumor cell lines. Expression levels of the novel transcript were calculated according to the expression $2^{-\Delta Ct}$, where ΔCt= Ct variant E-cadherin - Ct housekeeping gene (GAPDH).

Figure 3. Novel E-cadherin variant transcript levels and the Nonsense-Mediated mRNA Decay (NMD) degradation mechanism. E-cadherin variant and E-cadherin wild type (wt) mRNAs expression levels were determined after 0, 2.5, 5 and 7.5 hours (h) incubation of MCF7 cells with **(A)** Actinomycin-D (5 µg/mL) or **(B)** Cycloheximide (10 µg/mL). Relative expression of E-cadherin variant and E-cadherin wild type (wt) was estimated according to the expression $2^{-\Delta\Delta Ct}$, where ΔΔCt = [ΔCt test sample - ΔCt reference sample], and ΔCt= [Ct gene under study - Ct housekeeping gene (GAPDH)], where test sample was the E-cadherin variant of wild type mRNA, and the reference was the corresponding sample at 0 hour (h)**.** (**P*<0.05).

Figure 4. Effects of the overexpression of the novel E-cadherin variant mRNA in MCF7 cells upon E-cadherin wild type (wt) levels, cellular morphology and adherent complex proteins localization. (A) Quantitative analysis of E-cadherin variant and E-cadherin wild type (wt) mRNA levels by real time PCR in non-transfected MCF7 cells (MCF-7), and in MCF7 cells stably transfected with a plasmid encoding the E-cadherin variant transcript (MCF7Ecadvar) or with the pcDNA3 empty plasmid (MCF7pcDNA3). Relative expression of E-cadherin variant and E-cadherin wild type (wt) mRNA levels was estimated according to the expression $2^{-\Delta\Delta Ct}$, where GAPDH was the housekeeping gene and MCF7 cells were used as reference sample. **(B)** Immunodetection of E-cadherin protein forms after SDS-PAGE and Western immunoblotting analysis. Total protein extracts of MCF7 (10 µg), MCF7pcDNA3 (10 µg) and MCF7Ecadvar (20 µg) cells were run and a commercial monoclonal antibody (BD) directed towards the intracellular domain of the adhesion protein was used. Immunodetection of Tubulin was used as loading control. **(C)** Representative images of MCF7pcDNA3 and MCF7Ecadvar cells registered by phase contrast microscopy. Bar: 20 µm. **(D)** Cell growth assessment in MCF7pcDNA3 and MCF7Ecadvar cells. The number of cells growing

individually or in colonies was measured during 12 days and plotted. **(E)** Fluorescence immunocytochemistry and confocal laser microscopy analysis of E-cadherin (HECD-1, Thermo-Invitrogen) and β-catenin (610153, BD) in MCF7pcDNA3 and MCF7Ecadvar cells. Presence of F-actin was evaluated by incubation with the Alexa Fluor 488 Phalloidin toxin. Negative controls are included; nuclei are visualized using HOESCHT 33342 (right panels). Bar: 20 µm.

Figure 5. Effects of the overexpression of the novel E-cadherin variant mRNA in MCF7 cells upon the expression levels of EMT typical markers. (A) Quantitative analysis of Twist, Snail, ZEB1 and Slug mRNAs in MCF7pcDNA3 and MCF7Ecadvar cells by real time PCR. Relative expression was estimated according to the expression $2^{-\Delta\Delta Ct}$, where GAPDH was the housekeeping gene and MCF7pcDNA3 cells were used as reference sample (the same for B, E, F and G). **(B)** Quantitative analysis of ESRP1 and ESRP2 transcripts in MCF7pcDNA3 and MCF7Ecadvar cells by real time PCR. **(C)** N-cadherin protein expression analyses in MCF7pcDNA3 and MCF7Ecadvar cells. Upper: Immunodetection of N-cadherin (3B9, Zymed) was done by SDS-PAGE and Western immunoblotting. Bottom: Fluorescence immunocytochemistry (H-163, Sta. Cruz Biotech.) and confocal laser microscopy analyses. Negative controls are included and nuclei are visualized using HOESCHT 33342 (right panels). **(D)** P-cadherin protein expression analyses in MCF7pcDNA3 and MCF7Ecadvar cells. Upper: Immunodetection of P-cadherin (H-105, Sta. Cruz) was done by SDS-PAGE and Western immunoblotting. Bottom: Fluorescence immunocytochemistry (H-105) and confocal laser microscopy analysis. Negative controls are included and nuclei are visualized using HOESCHT 33342 (right panels). (**E**) Cytokeratins expression analyses in MCF7pcDNA3 and MCF7Ecadvar cells. Right: Fluorescence immunocytochemistry (Clone AE1/AE3, Dako) and confocal laser microscopy analysis of cytokeratins. Left: Quantitative analysis of Cytokeratin 19 transcript by real time PCR. (**F**) Vimentin expression analyses in MCF7pcDNA3 and MCF7Ecadvar

cells. Right: Quantitative analysis of Vimentin mRNA by real time PCR. Left: Fluorescence immunocytochemistry (Clone V9, Dako) and confocal laser microscopy analysis of Vimentin . Negative controls are included and nuclei are visualized using HOESCHT 33342 (right panels). **(G)** Dysadherin mRNA (Left: real time PCR) and protein expression (Middle: SDS-PAGE/Western immunoblotting and Right: fluorescence immunocytochemistry; NCC-M53) analyses in MCF7pcDNA3 and MCF7Ecadvar cells. For fluorescence immunocytochemistry, negative controls are included and nuclei are visualized using HOESCHT 33342 (right panels).

Figure 6. Effects of the overexpression of the novel E-cadherin variant mRNA in MCF7 cells upon cellular adhesion, migration and invasion. (A) Evaluation of the cell-cell adhesion capacity of MCF7pcDNA3 and MCF7Ecadvar cells by the Hanging Drop method. Representative images of cells grown in suspension for 0, 12 and 24 h are shown. The number (arbitrary scale) and size (1 to 5, 5 to 50, more than 50 cells) of cellular aggregates were measured and plotted for each cell line. **(B)** Evaluation of the migratory behavior of MCF7pcDNA3 and MCF7Ecadvar cells by the Wound Healing assay. The relative distance was measured 12 hours after making the lesion and plotted for each cell line. **(C)** Evaluation of the migratory behavior of MCF7pcDNA3 and MCF7Ecadvar cells by the Transwell™ Migration test. The number of cells that were able to migrate through the transwells was measured and plotted. **(D)** Evaluation of the invasive capacity of MCF7pcDNA3 and MCF7Ecadvar cells by the Transwell™ Invasion test was done in the presence of an extracellular matrix. The number of cells that were able to invade was measured and plotted.

Figure 7. Characterization of the protein encoded by the novel E-cadherin variant mRNA. (A) Amino acidic sequence analysis of the protein encoded by the E-cadherin variant mRNA (Ecadvar). The sequence was deduced using the Translate software (ExPASy Tools) and was aligned with the

wild type sequence (Ecadwt). **(B)** Prediction of the secretory nature of the novel protein encoded by the E-cadherin variant mRNA. The analysis was done with several software tools (TMHMM, YLoc, ESLpred, ESLpred2 and BaCelLo). Results shown are from TMHMM tool. **(C)** Fluorescence immunocytochemical analysis of E-cadherin in COS7pcDNA3, COS7Ecadvar and COS7Ecadwt cells. Negative controls are included and nuclei are visualized using HOESCHT 33342. **(D)** E-cadherin protein expression analyses in COS7Ecadwt and COS7Ecadvar cells. SDS-PAGE followed by Western immunobloting analysis of cell lysates (Lys) and conditioned media (CM) of both cell lines was done. **(E)** E-cadherin protein expression analyses in MCF7pcDNA3 and MCF7Ecadvar cells. SDS-PAGE followed by Western immunobloting analysis of conditioned media of both cell lines was done.

Figure 8. Effect of the addition of conditioned medium from HEK293 cells transfected with the E-cadherin variant mRNA onto MCF7 cultures. (A) Quantitative analysis of E-cadherin mRNA in MCF-7 and HEK-293 cells by real time PCR. Relative expression was estimated according to the expression $2^{-\Delta\Delta Ct}$, where GAPDH was the housekeeping gene and MCF7 cells were used as reference sample. **(B)** Quantitative analysis of E-cadherin variant mRNA in HEK293pcDNA3 and HEK293Ecadvar cells by real time PCR. Relative expression was estimated according to the expression $2^{-\Delta Ct}$, where GAPDH was the housekeeping gene. **(C)** Quantitative analysis of E-cadherin variant mRNA in MCF7Ecadvar and HEK293Ecadvar cells by real time PCR. Relative expression was estimated according to the expression $2^{-\Delta\Delta Ct}$, where GAPDH was the housekeeping gene and MCF7Ecadvar cells were used as reference sample**. (D)** E-cadherin wild type and E-cadherin variant proteins analysis in HEK293pcDNA3 and HEK293Ecadvar cells. Left: Fluorescence immunocytochemistry and confocal laser microscopy analysis of E-cadherin wild type was done; negative controls are included and nuclei are visualized using HOESCHT 33342. Right: SDS-PAGE

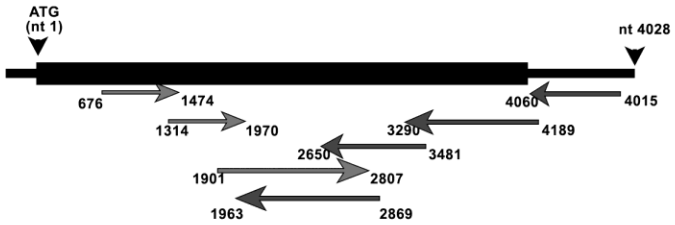
followed by Western immunoblotting study of the E-cadherin variant protein was done on conditioned media (CM). **(E)** Representative images of MCF7 cells incubated with fresh medium (MCF-7), CM from HEK293 cells stably transfected with the pcDNA3 empty plasmid (MCF-7 + TCM) and CM from HEK293 cells stably transfected with the E-cadherin variant mRNA (MCF-7 + VCM) during 7 days (d) (4x and 20x magnifications). **(F)** Quantitative analysis of E-cadherin wild type mRNA in MCF-7 + TCM and MCF-7 + VCM cells by real time PCR. Relative expression was estimated according to the expression $2^{-\Delta\Delta Ct}$, where GAPDH was the housekeeping gene and MCF7 + TCM cell condition was used as reference sample.

Figure 9. Schematic representation of the E-cadherin novel variant synthesis and its potential role in tumor progression. **(A)** In physiological conditions, both the wild type (wt) and the novel variant E-cadherin transcripts are generated from the *CDH1* gene by constitutive and alternative splicing, respectively. While the wild type mRNA is translated to a functional transmembrane protein responsible of the establishment of cellular adherent junctions, the novel variant transcript is recognized by NMD (Nonsense Mediated mRNA Decay), a translation-dependent mechanism involved in the degradation of aberrant transcripts. **(B)** Under pathological conditions in which the NMD mechanism is altered, the E-cadherin variant mRNA is translated and secreted into the extracellular space. Considering the high sequence homology between the E-cadherin variant protein and the extracellular domain of the wild type form, the novel protein may interfere with the correct establishment of the adherent junctions either by inhibiting the formation of E-cadherin *cis* and *trans* dimers. This phenomenon will contribute to destabilizing the adherent complex and triggering changes in gene expression, which will lead to alterations in the normal cellular morphology and behavior. Specifically, increased levels of the novel E-cadherin mRNA will cause loss of cellular epithelial markers (E-cadherin, ESRP1, ESRP2 and Cytokeratins) and expression of molecules related to the

acquisition of a mesenchymal phenotype and an invasive behavior (Vimentin, N-cadherin and Dysadherin). All these features have been associated to tumor progression.

Figure 1

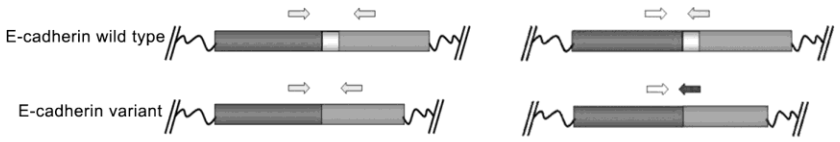
A.



B.

```
NM_004360.3     GCCTGTCGAAGCAGGATTGCAAATTCCTGCCATTCTGGGGATTCTTGGAGGAATTCTTGC
variant         GCCTGTCGAAGCAGGATTGCAAATTCCTGCCATTCTGGGGATTCTTGGAGGAATTCTTGC
                ************************************************************

NM_004360.3     TTTGCTAATTCTGATTCTGCTGCTCTTGCTGTTTCTTCGGAGGAGAGCGGTGGTCAAAGA
variant         TTTGCTAA--------------------------------GAGAGCGGTGGTCAAAGA
                ********                                ******************

NM_004360.3     GCCCTTACTGCCCCCAGAGGATGACACCCGGGACAACGTTTATTACTATGATGAAGAAGG
variant         GCCCTTACTGCCCCCAGAGGATGACACCCGGGACAACGTTTATTACTATGATGAAGAAGG
                ************************************************************
```

C.



D.

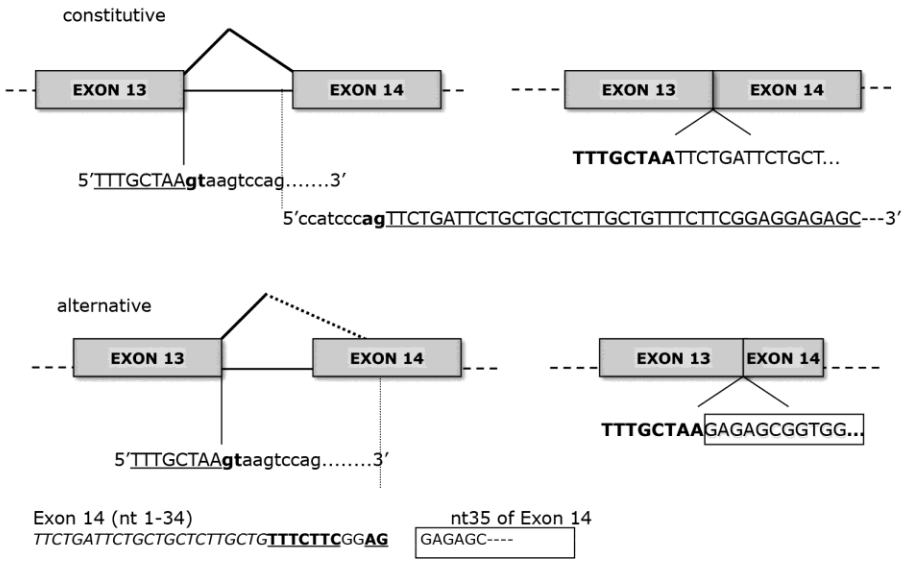


Figure 2

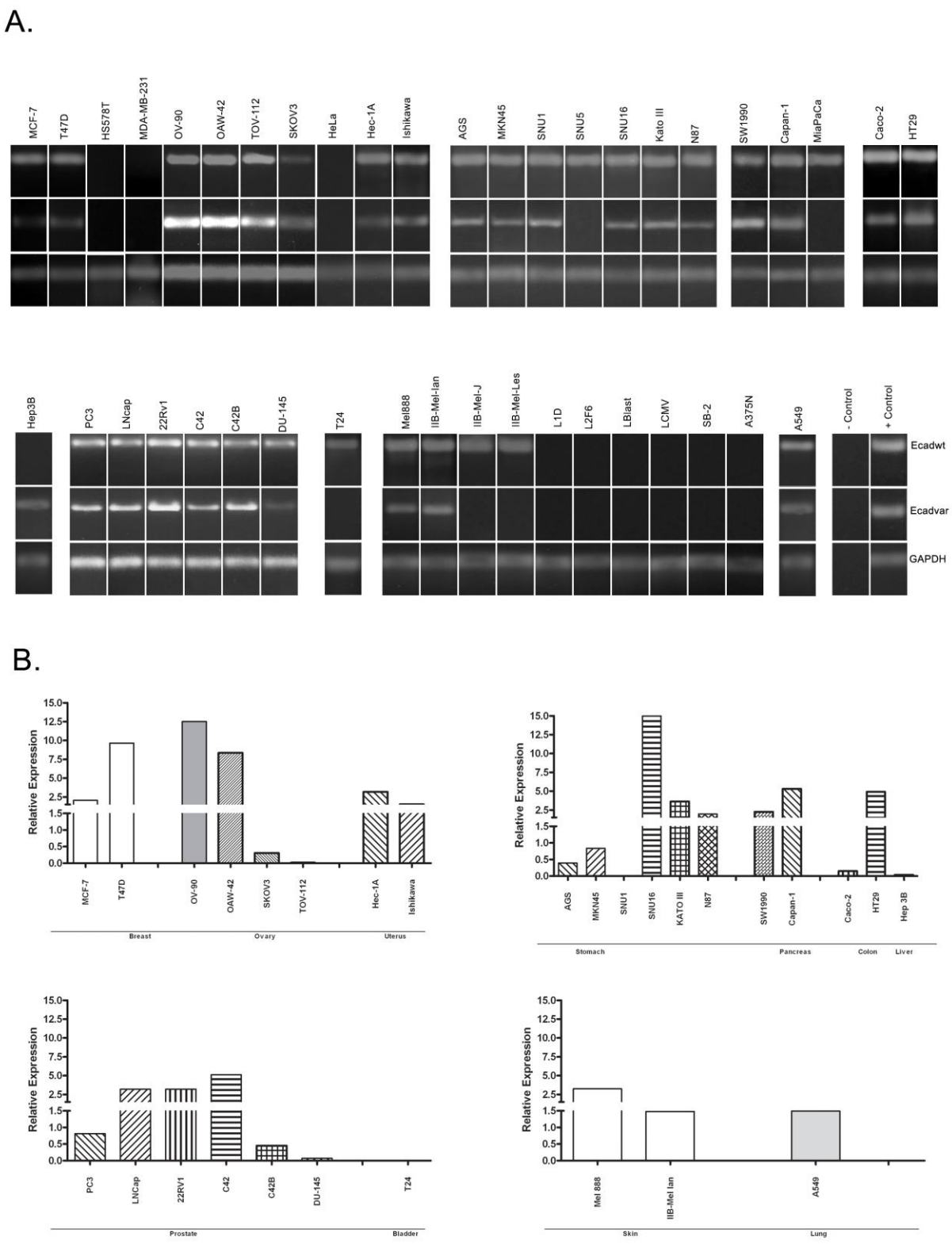
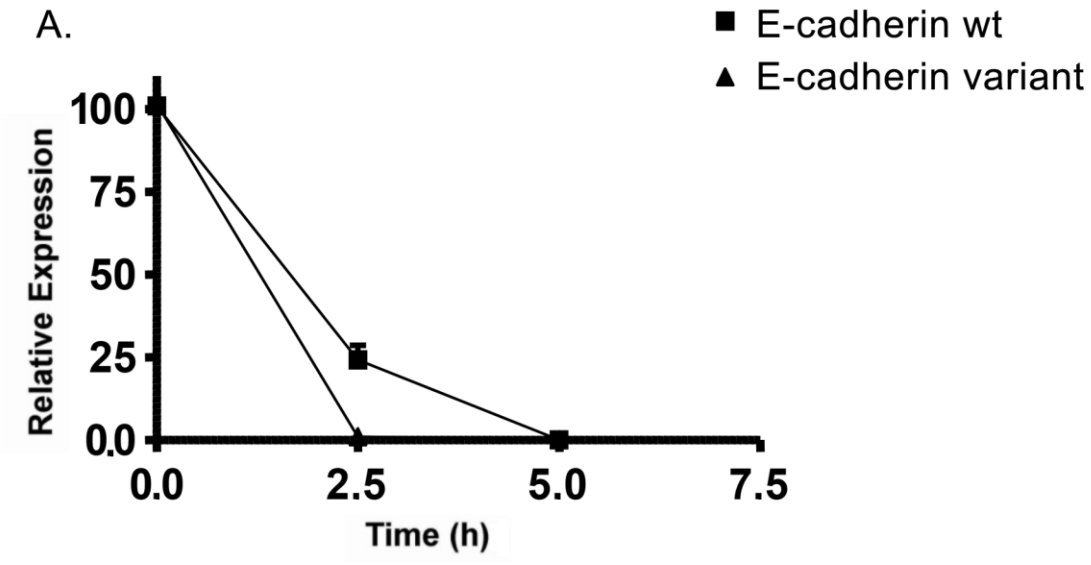
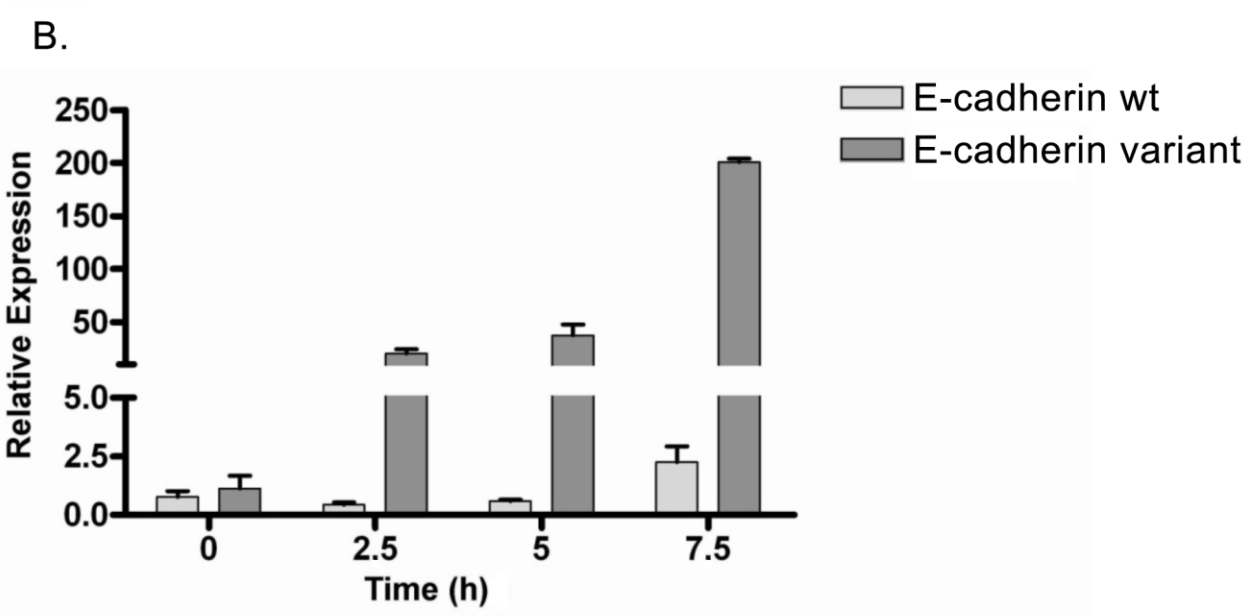


Figure 3





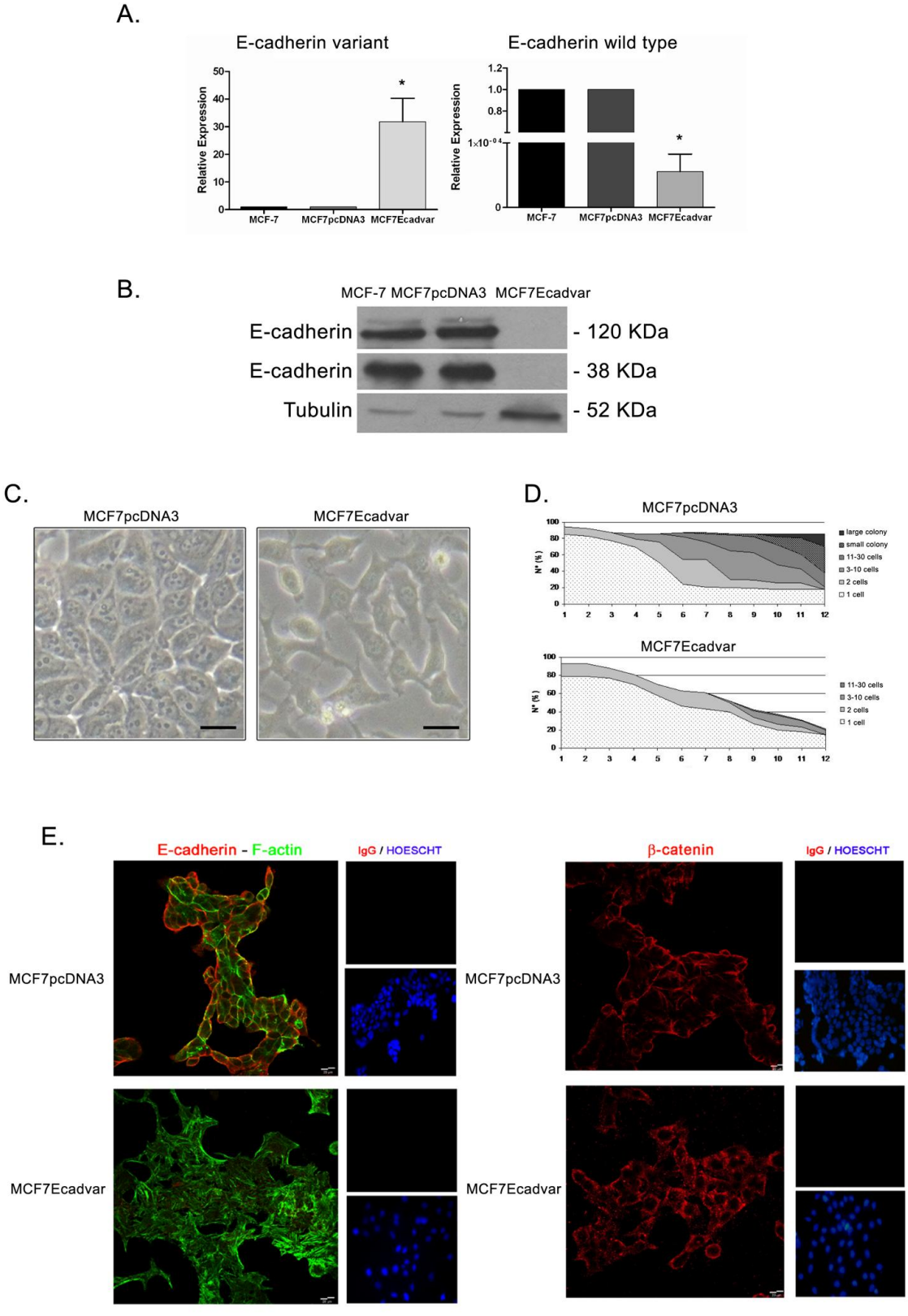
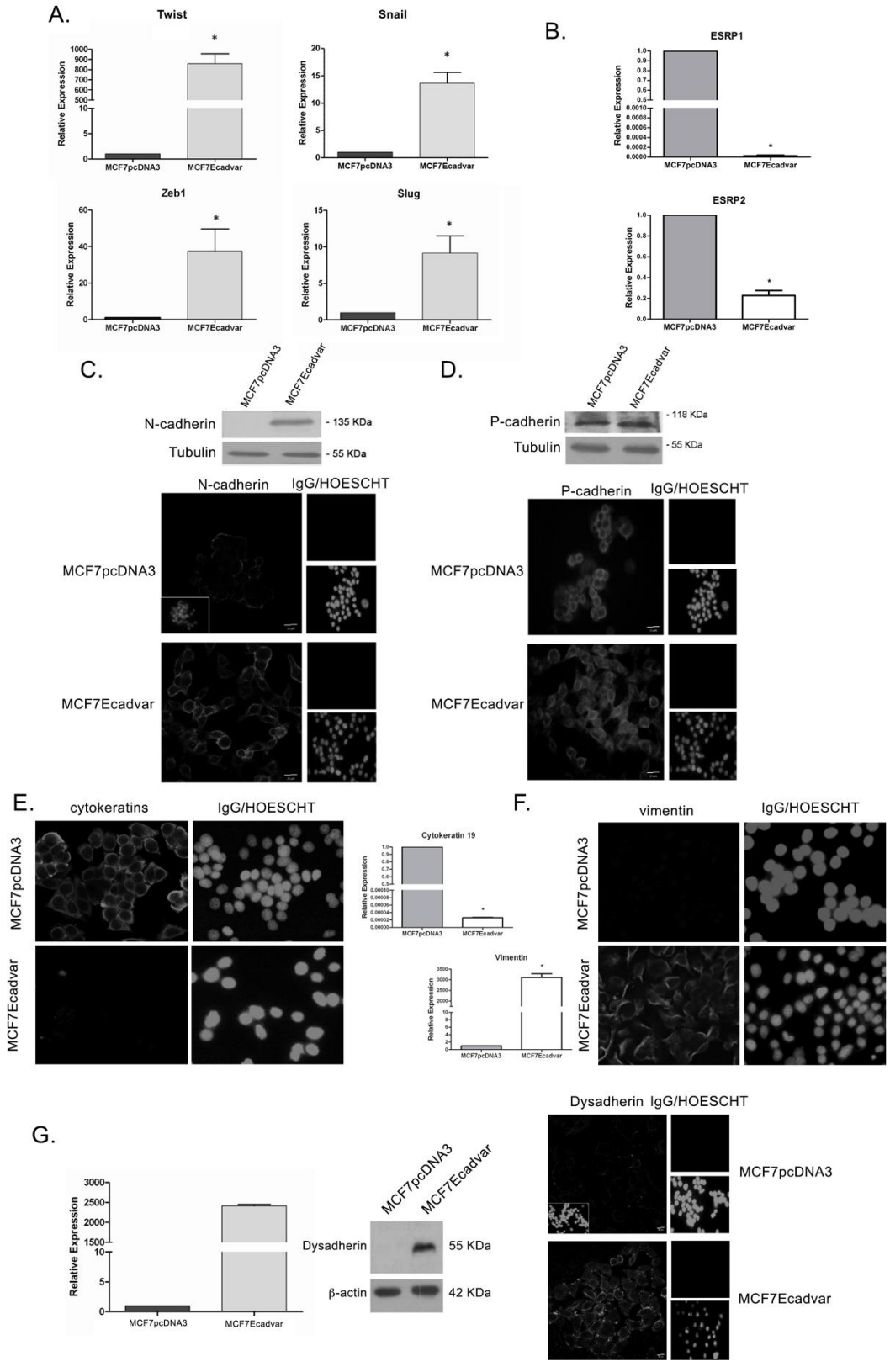


Figure 4

Figure 5



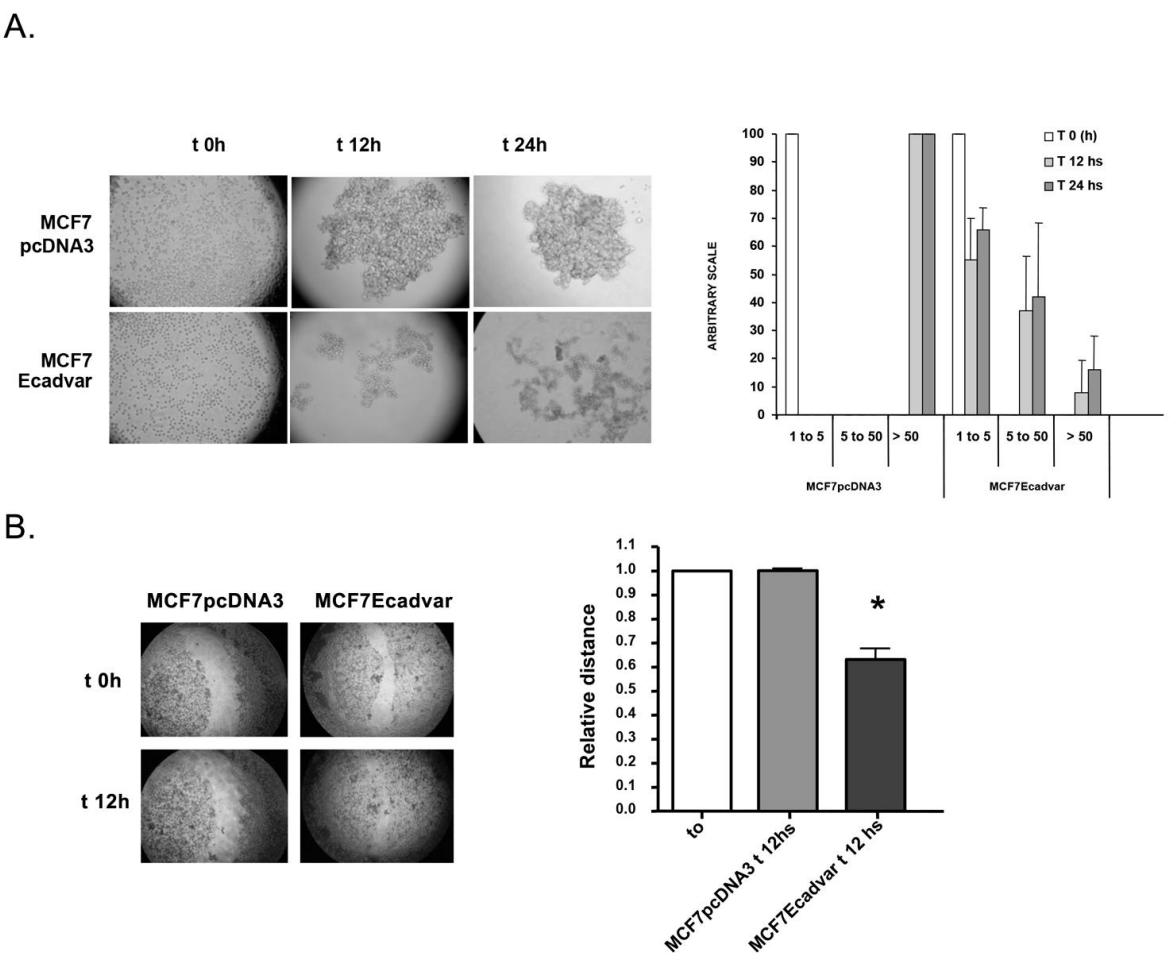


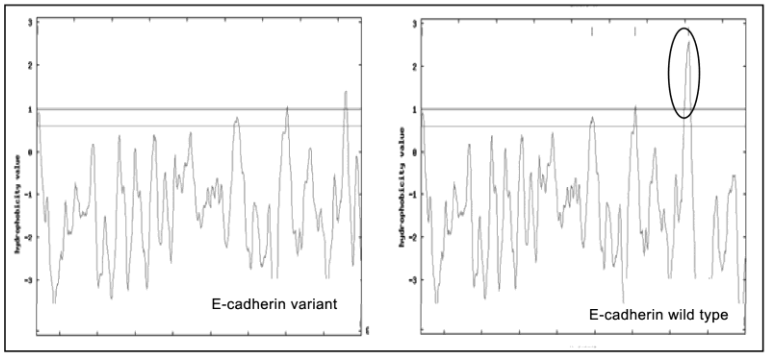
Figure 6

Figure 7

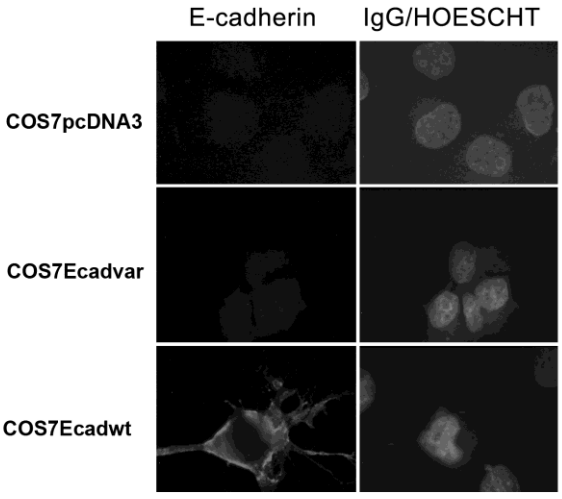
A.

```
Ecadwt     GDYKINLKLMDNQNKDQVTTLEVSVCDCEGAAGVCRKAQPVEAGLQIPAILGILGGILAL 720
Ecadvar    GDYKINLKLMDNQNKDQVTTLEVSVCDCEGAAGVCRKAQPVEAGLQIPAILGILGGILAL 720
           ************************************************************

Ecadwt     LILILLLLLFLRRRAVVKEPLLPPEDDTRDNVYYYDEEGGGEEDQDFDLSQLHRGLDARP 780
Ecadvar    LRERWSKSPYCPQRMTPGTTFITMMKKEAEKRTRTLT---------------------- 757
           *------------------------------------------------------------

Ecadwt     EVTRNDVAPTLMSVPRYLPRPANPDEIGNFIDENLKAADTDPTAPPYDSLLVFDYEGSGS 840
Ecad var   ------------------------------------------------------------

Ecadwt     EAASLSSLNSSESDKDQDYDYLNEWGNRFKKLADMYGGGEDD                   882
Ecadvar    ----------------------------------------
```

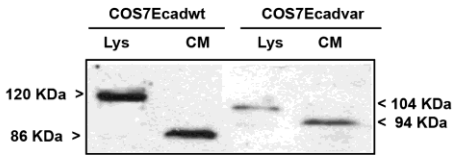
B.



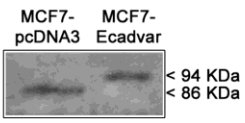
C.



D.



E.



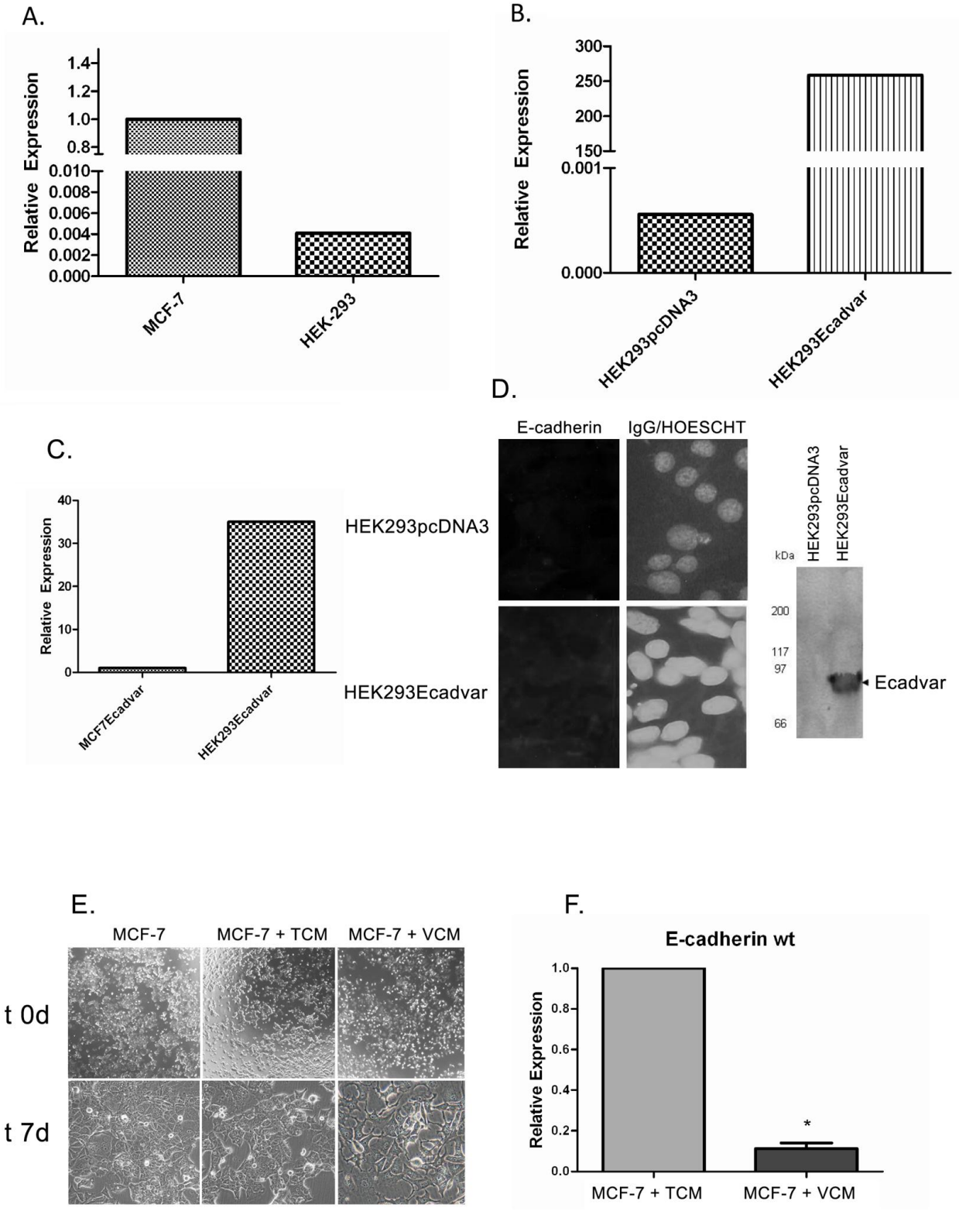


Figure 8

Figure 9