PhD degree in System Medicine (Curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore Disciplinare: Med/04

Investigation of the role of the endocytic protein Epsin3 in EMT and mammary stem cell regulation in breast cancer

Irene Schiano Lomoriello

IEO, Milan

Matricola n. R 11127

Supervisor: Prof. **Pier Paolo Di Fiore**

IEO, Milan and University of Milan

Added supervisor: **Sara Sigismund**

IEO, Milan and University of Milan

Anno accademico 2017-2018

TABLE OF CONTENTS

FIGURES AND TABLE INDEX……**8**

TABLE OF ABBREVIATIONS……...**12**

ABSTRACT……**15**

INTRODUCTION……..**17**

1. Endocytosis……..**17**

1.1 Endocytosis and different endocytic pathways……...**17**

1.2 Endocytosis as a master regulator of cell signaling……**21**

1.3 Endocytosis and cancer……...**24**

2. Epsin proteins……..**27**

2.1 The Epsin family……..**27**

2.1 Structure and domain organization of epsins.……**28**

2.3 Biological function of epsins……..**31**

2.4 Epsins and cancer……..**33**

3. Breast Cancer……..**36**

3.1 Normal mammary gland development……**36**

3.2 Histological and molecular characterization of breast cancer…..**39**

3.3 *In vitro* and *in vivo* study of breast cancer progression: non-tumorigenic cell lines and mouse models……..**43**

4. The epithelial-to-mesenchymal transition: mechanisms and pathways involved……..**46**

4.1 The basics of EMT……**46**

4.2 Epithelial cells and cell junctions……**49**

4.3 Mesenchymal cells, the Cadherin switch and EMT markers…….**52**

4.4 EMT master regulators and post-transcriptional mechanisms inducing EMT……..**53**

4.5 Signaling pathways involved in EMT……..**55**

4.6 The canonical TGFβ signaling pathway in EMT induction……..**57**

4.7 The WNT/β-catenin pathway……..**60**

4.8 Endocytosis and EMT……...**62**

5. Cancer stem cells……**65**

5.1 Mammary stem cells and breast cancer stem cells: origin and modality of division……...**65**

5.2 BCSCs: isolation and manipulation……..**68**

5.3 EMT and CSCs: two sides of the same coin……..**72**

5.4 CSCs: drug resistance and tumor relapse……...**73**

5.5 CSCs, EMT-MET and metastasis……**75**

5.6 Endocytosis, asymmetric cell division of SCs and CSCs ……**77**

PRELIMINARY DATA AND AIM OF THE PROJECT……..**80**

1. Preliminary data……**80**

1.1 The *EPN3* gene is amplified and overexpressed in breast cancer..**80**

1.2 Epn3 correlates with aggressive clinical/pathological parameters and with distant metastasis……..**82**

1.3 Epn3 knockdown reduces the *in vivo* tumorigenic potential of BT474 breast cancer cells……...**86**

2. Aim of the project……...**88**

MATERIALS AND METHODS……..**91**

1. Cell culture manupulations……..**91**

1.1 Cell lines……**91**

1.2 Retroviral and lentiviral infections……...**92**

1.3 RNA interference (RNAi)……...**93**

1.4 Stimulations and treatments……**94**

2. Cell lysis and protein and RNA extraction techniques……95

2.1 Western Blotting (WB)……..**95**

2.2 RNA extraction and Quantitative Real time PCR……..**96**

3. Biological assays……..**97**

3.1 Flow Cytometry……**97**

3.2 Mammosphere assay……**98**

3.3 ELISA assay……**99**

3.4 E-Cadherin internalization assay……**100**

3.5 Immunofluorescence studies……..**101**

4. *In vivo* xenografts……..**102**

5. Statistical analyses……**103**

6. Supplementary Tables……..**104**

RESULTS……..**109**

1. Analysis of Epn3-dependent phenotypes upon overexpression in non-tumorigenic MCF10A cells……**109**

1.1 Epn3 overexpression causes EMT in non-tumorigenic MCF10A cells……..**109**

1.2 Epn3-induced EMT expands the MCF10A mammary epithelial SC-like compartment……**113**

1.3 Epn3-induced EMT and expansion of the SC-like compartment in MCF10A cells are causally linked……..**115**

1.4 The epithelial population in Epn3-overexpressing MCF10A cells is highly plastic, in constrast to the meta-stable mensenchymal population……...**119**

2. Dissection of the Epn3-dependent molecular mechanisms in non-tumorigenic MCF10A-Epn3 cells……**122**

2.1 MCF10A-Epn3 cells secrete TGFβ ligands that cooperate with Epn3 and create a positive feedback-loop sustaining EMT…….**122**

2.2 MCF10A Epn3-dependent phenotypes are TGFβ dependent…..**125**

2.3 Epn3-overexpression in MCF10A cells induces E-Cadherin endocytosis and β-catenin translocation to the nucleus……..**129**

2.4 Epn3-induced EMT in MCF10A cells is TCF4 dependent…….**133**

2.5 The TCF/LEF and TGFβ signaling pathways crosstalk in Epn3-overexpressing cells……**135**

3. Investigation of the contribution of Epn3 to breast tumorigenesis *in vivo* ……**139**

3.1 Tumorigenic Epn3-overexpressing MCF10DCIS.com breast cancer cells partially recapitulate Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells……**139**

3.2 Epn3 overexpression in tumorigenic HCC1569 breast cancer cells partially recapitulates Epn3-dependent phenotypes observed in MCF10A cells and increases their *in vivo* tumorigenic potential……..**143**

DISCUSSION……**148**

1. Epn3 overexpression induces EMT and expansion of the SC-like compartment in non-tumorigenic mammary epithelial cells……**148**

1.1 Dissection of the molecular mechanisms: proposed model…….**148**

1.2 Partial EMT and its implications……**150**

1.3 EMT and the generation of cells with CSC properties……...**152**

1.4 The TGFβ signaling pathway in EMT and breast cancer……**154**

1.5 Feedback loops in EMT and breast cancer……...**155**

1.6 E-Cadherin in EMT and breast cancer……**157**

1.7 β-catenin/TCF-LEF axis in EMT and breast cancer……..**160**

1.8 Crosstalk between the TGFβ and the TCF-LEF axis in EMT and breast cancer……...**161**

1.9 Mechanisms inducing EMT that are active in Epn3-overexpressing MCF10A cells……**163**

2. Epn3 overexpression in breast cancer cells partially recapitulates phenotypes observed in non-tumorigenic MCF10A-Epn3 cells and increases their *in vivo* tumorigenic potential……..**165**

2.1 Epn3-overexpressing MCF10DCIS.com and HCC1569 breast cancer cells partially recapitulate Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells……**165**

2.2 High Epn3 levels increase the *in vivo* tumorigenic potential of HCC1569 breast cancer cells……**167**

2.3 Future directions: MCF10DCIS.com and human breast cancers: analysis of invasive phenotypes……**168**

3. Epn3 as a novel prognostic marker in breast cancer and possible therapies……...**170**

3.1 Epn3 as a putative novel oncogene in BC……**170**

3.2 Epn3 and HER2: two oncogenes in BC on the same amplicon…**171**

3.3 Anti-TGFβ signaling therapies……...**172**

3.4 Future directions: Epn3 and Epn1: chimeric approach……...**173**

REFERENCES……**177**

ACKNOWLEDGMENTS……**190**

FIGURES AND TABLES INDEX

INTRODUCTION

Fig. 1: Endocytic pathways: pathways of entry into the cell.....**18**

Fig. 2: NCE-RTN3-depdendent endocytic pathway of EGFR.....**21**

Fig. 3: Endocytosis regulates signaling at different levels.....**24**

Fig. 4: The Epsin family.....**29**

Fig. 5: Epsins mediate endocytosis of ubiquitinated cell surface receptors.....**31**

Fig. 6: Terminal lobular-ductal units and the surrounding stroma.....**37**

Fig. 7: Mammary gland development.....**38**

Fig. 8: Molecular characterization of breast cancer.....**42**

Fig. 9: The basics of epithelial-to-mesenchymal transition.....**46**

Fig. 10: Different types of EMT.....**48**

Fig. 11: Epithelial cell junctions.....**50**

Fig. 12: The Cadherin switch and EMT markers.....**53**

Fig. 13: Signaling pathways involved in EMT.....**56**

Fig. 14: TGFβ canonical signaling.....**59**

Fig. 15: The WNT/β-catenin signaling pathway.....**61**

Fig. 16: The origin of CSCs.....**66**

Fig. 17: MaSCs and PKH26 label retention……**71**

Fig. 18: The metastasis cascade……..**75**

PRELIMINARY DATA AND AIM OF THE PROJECT

Fig. 1: The *EPN3* gene is amplified and overexpressed in BC……**82**

Table 1: Epn3 overexpression correlates with clinical/pathological parameters indicative of aggressive disease……**84**

Fig. 2: High levels of Epn3 expression correlate with increased risk of distant metastasis……**85**

Fig. 3: Epn3 knockdown reduces *in vivo* tumorigenic potential of BT474 BC cells…..**87**

MATERIALS AND METHODS

Supplementary Table 1: RNAi oligo sequences……**104**

Supplementary Table 2: Solutions……**105**

Supplementary Table 3: List of antibodies……...**107**

Supplementary Table 4: TaqMan assays……..**108**

RESULTS

Fig. 1: Epn3 overexpression causes EMT in non-tumorigenic MCF10A cells……**112**

Fig. 2: Epn3-induced EMT expands MCF10A SC-like compartment……..**114**

Fig. 3: The $CD44^{high}/CD24^{low}$ population displays mesenchymal and SC-like traits in MCF10A-Epn3 and MCF10A-Twist cells……..**118**

Fig. 4: $CD44^{low}/CD24^{high}$ MCF10A-Epn3 or -Twist cells are able to convert into $CD44^{high}/CD24^{low}$ cells……...**121**

Fig. 5: A TGFβ-dependent feedback loop sustains EMT in MCF10-Epn3 overexpressing cells……...**12**

Fig. 6: MCF10A Epn3-dependent phenotypes are TGFβ dependent……**128**

Fig. 7: Epn3 overexpression induces E-Cadherin endocytosis and β-catenin translocation to the nucleus……...**132**

Fig. 8: Epn3-induced EMT in MCF10A is TCF4 dependent……**134**

Fig. 9: Crosstalk between the TCF/LEF and TGFβ signaling pathways……...**137**

Fig. 10: Epn3-overexpressing MCF10DCIS.com partially recapitulate the phenotypes observed in non-tumorigenic MCF10A……**142**

Fig. 11: Epn3-overexpression in HCC1569 breast cancer cells partially recapitulates Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells and increases their *in vivo* tumorigenic potential……**146**

DISCUSSION

Fig. 1: Proposed model for Epn3 overexpression-dependent phenotypes in non-tumorigenic MCF10A cells and breast cancer cells……...**150**

Fig. 2: Alignment of the conserved ENTH domain in Epn3 and Epn1 proteins.....**174**

Fig. 3: Alignment of Epn3 and Epn1 protein sequences.....**175**

LIST OF ABBREVIATIONS

ACD	Asymmetric cell division
AP2	Adaptor protein 2
AREG	Amphiregulin
BC	Breast cancer
BCSC	Breast cancer stem cell
CLIC	Clathrin- and dynamin-independent carriers
CCP	Clathrin-coated pit
CCV	Clathrin-coated vescicles
CK	Cytokeratin
CME	Clathrin-mediated endocytosis
CSC	Cancer stem cell
CSF1	Colony-stimulating factor 1
CTCs	Circulating tumor cells
DCIS	Ductal *in situ* carcinoma
DPW	Aspartate-Proline-Tryptophan
DSL	Delta/Serrate/Lag2
ECM	Extracellular matrix
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial-to-mesenchymal transition
ENTH	Epsin-N-terminal homology domain
EpCAM	Epithelial cell adhesion molecule
EPN1	Eps-15 interacting protein1

EPN2	Eps-15 interacting protein2
EPN3	Eps-15 interacting protein3
ER	Estrogen
ER-PM contact sites	Endoplasmic reticulum-plasma membrane contact sites
EV	Empty vector
E-Cadherin	Epithelial-cadherin
FACS	Fluorescence activated cell sorting
FISH	Fluorescent *in situ* hybridization
GEEC	Glycosil phosphatidyl-inositol-anchored protein enriched early endosomal compartments
GSK3β	Glycogen synthase 3 beta
HER2	Epidermal growth factor 2
HGF	Hepatocyte growth factor
HIF1α	Hypoxia inducible factor 1
HMECs	Human mammary epithelial cells
ICD	Invasive ductal carcinomas
IGF1	Insulin-like growth factor 1
IHC	Immunohistochemistry
ILV	Intraluminal vescicles
KD	Knockdown
KO	Knockout
LCIS	Lobular carcinoma *in situ*
LqF	Liquid facets
MaSC	Mammary stem cells
MET	Mesenchymal-to-epithelial transition
mRNA	messenger RNA

miRNA	micro RNA
MVB	Multivescicular bodies
NCE	Non-clathrin-mediated endocytosis
NPF	Asparagine-Proline-Phenylalanine
NSG mice	NOD (non-obese diabetic) - SCID (severe combined immunodeficiency) IL2RG null mice
N-Cadherin	Neural cadherin
PDAC	Pancreatic ductal adenocarcinoma
PM	Plasma membrane
PR	Progesterone
ROS	Reactive oxygen species
RTK	Receptor tyrosine kinase
RT-qPCR	Reverse transcriptase-quantitative polymerase chain reaction
SARA	SMAD-anchor for receptor activation
SC	Stem cell
SCD	Symmetric cell division
TICs	Tumor-initiating cells
TGFβ	Transforming growth factor β
TGFβR	Transforming growth factor βreceptor
TLDU	Terminal lobular-ductal unit
TNBC	Triple-negative breast cancer
UIM	Ubiquitin interacting motifs
VEGF	Vascular endothelium growth factor
VEGFR	Vascular endothelium growth factor receptor
WB	Western blotting

ABSTRACT

Endocytosis is a critical regulator of many cellular processes, including proliferation, migration, polarity, fate determination, epithelial-to-mesenchymal transition (EMT) and maintenance of the stem cell (SC) compartment. Subversion of endocytic routes plays an important role in cancer development, including breast cancer (BC). We recently uncovered an oncogenic role for the endocytic protein Epsin3 (Epn3) in BC. By screening a large cohort of BC patients, we found that the *EPN3* gene is amplified in ~10% of cases, while Epn3 protein is overexpressed in ~50% of BCs. Epn3 overexpression correlated with poor prognosis and distant metastasis and with increased *in vivo* tumorigenic potential in orthotopic transplantation experiments in NOD (non-obese diabetic) - SCID (severe combined immunodeficiency) IL2RG null mice (NSG mice).

The goal of the present Ph.D. thesis was to uncover the mechanisms through which Epn3 exerts its oncogenic properties in BC. To this aim, we overexpressed Epn3 in non-tumorigenic mammary epithelial cells, MCF10A, to mimic the overexpression occurring in BC. We showed that Epn3 overexpression causes a TGFβ-dependent EMT and expansion of the SC-like compartment, two events that are causally linked. Mechanistically, we found that Epn3 overexpression enhances E-Cadherin endocytosis, resulting in destabilization of cell-cell junctions and translocation of β-catenin in the nucleus. The β-catenin/TCF-LEF pathway leads to the activation of the EMT transcriptional program, consisting of the up-regulation of EMT master regulators (ZEB1, Snail, Twist), of mesenchymal markers (N-Cadherin, Vimentin, extracellular matrix (ECM) components, and proteases - Fibronectin and Serpine) and, importantly, of the transcription of TGFβ receptors and ligands. Consequent secretion of TGFβ ligands is a key event in the establishment of an autocrine

TGFβ positive feedback loop, which is able to increase Epn3-dependent E-Cadherin endocytosis and TGFβ ligand production and to self-sustain EMT. This loop could be reverted only by the abrogation of the TGFβ pathway itself or of the β-catenin/TCF-LEF pathway.

We also investigated the effects of Epn3 overexpression in two BC cell model systems, MCF10DCIS.com and HCC1569 and we observed that they partially recapitulate Epn3-dependent phenotypes observed in non-tumorigenic MCF10A-Epn3 cells. Indeed, although MCF10DCIS.com and HCC1569 are tumorigenic cells that have already undergone some form of EMT, they showed at least a partial cooperation between Epn3 overexpression and TGFβ stimulation in the induction of EMT. In addition, Epn3-overexpressing MCF10DCIS.com showed EMT reversion upon abrogation of the TGFβ signaling pathway (not tested for HCC1569). Importantly, Epn3 overexpression in tumorigenic HCC1569 cells was able to increase their *in vivo* tumorigenic potential upon orthotopic transplantation in NSG mice.

In conclusion, our results identify new oncogenic properties for the endocytic protein Epn3 and suggest an endocytic-based mechanism of how it contributes to EMT, cancer stem cell expansion and breast tumorigenesis.

INTRODUCTION

1. Endocytosis

1.1 Endocytosis and different endocytic pathways

Endocytosis is a process that allows eukaryotic cells to bring inside the cell (*endo*) nutrients and other types of molecules. It is an active energy-consuming process required by all cells to internalize large polar molecules that cannot pass through the hydrophobic plasma membrane (PM) by passive means. However, the fact that this process is not passive represents a huge advantage for the cell: indeed, through the finely regulated endocytosis of surface receptors, cells can control the composition of the PM. Moreover, cells also use endocytosis to internalize bacterial toxins, viruses and immunoglobulins, constituting therefore also a mechanism for protection. Finally, endocytosis, through the internalization of receptors and their bound ligands, is a master regulator of cell signaling and biological responses (see paragraph **1.2 Endocytosis as a master regulator of cell signaling**)[1].

Many endocytic pathways have been identified that can be broadly divided into phagocytosis, macropinocytosis and micropinocytic mechanisms. The latter, also known as small-scale endocytic processes, include clathrin-mediated endocytosis (CME), caveolin-dependent endocytosis, and clathrin- and caveolin-independent pathways. According to the specific entry route, different molecular players are

involved and the internalized molecules are subject to different fates[1,2]. These pathways are schematically depicted in in **Fig. 1**.

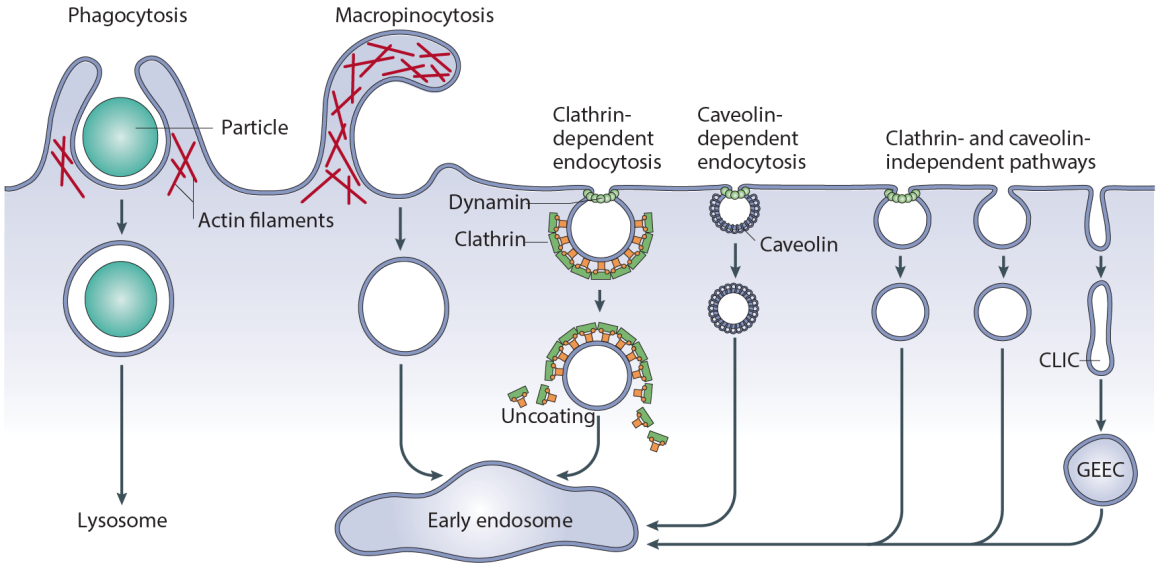


Fig. 1: Endocytic pathways: pathways of entry into the cell.

Schematic representation of the main endocytic pathways identified so far: phagocytosis, macropinocytosis and micropinocytosis. Micropinocytosis includes clathrin-dependent, caveolin-dependent and clathrin- and caveolin-independent pathways. The major molecular players in the pathways (actin filaments, dynamin, clathrin, caveolin) and the destination of the internalized molecules (lysosome, early endosomes, GEEC) are represented. See main text for a detailed description.
Fig. adapted from Mayor *et al.*[2].

Phagocytosis (cell eating) is the process by which cells internalize particulate matter larger than 0.75 µm in diameter, such as cell debris, bacteria and apoptotic cells. The internalization of these particles involves the uptake of larger membrane areas than clathrin- or caveolin- dependent or independent pathways and is characterized by large actin-based rearrangements of the PM. The destination of these molecules is lysosome degradation[3].

Macropinocytosis (cell drinking) usually occurs from highly ruffled regions of the PM and consists in the invagination of the cell membrane to form a pocket, which then forms a large vesicle filled with a large volume of extracellular fluid and molecules within it. The filling of the pocket occurs in a non-specific manner. These vesicles travel into the cytosol and fuse with other vesicles, such as early endosomes or lysosomes[4].

Micropinocytosis is instead characterized by smaller invaginations of the PM (<200 nm) and can be divided in CME and non-clathrin-mediated endocytosis (NCE). CME is the best characterized endocytic pathway. It is mediated by the production of small vesicles (~100-150 nm in diameter) that have a morphologically characteristic coat made up of the cytosolic protein, clathrin. Clathrin-coated pits (CCPs) assemble on the PM of all types of cells and pinch-off to release clathrin-coated vesicles (CCVs) in the cytosol. CME is responsible for receptor-mediated endocytosis of many types of ligands, including transferrin, growth factors and antibodies, among others[5].

Besides clathrin, the other major player of this pathway is the adaptor protein 2 (AP2), a multimeric protein that forms a heterotetrameric complex, known as the AP2 complex, consisting of two large adaptins (alpha and beta), a medium adaptin (mu), and a small adaptin (sigma)[6]. AP2 links the cargo to the clathrin machinery and plays a critical role in CCP formation at the PM[7]. Several other accessory proteins are involved in CME, for instance, epsins (see paragraph **2 Epsin proteins**) and Eps15 proteins[8]. Finally, the dynamin family of GTPases are responsible for the pinching step and scission of the CCV[7]. Internalized molecules are trafficked to early endosomes, which then mature into late endosomes or multivesicular bodies (MVBs). MVBs are characterized by their acidic pH and the formation of

intraluminal vesicles (ILVs), which then fuse to lysosomes, where cargo destined to degradation are entrapped and finally degraded[9].

The term NCE instead refers to a heterogeneous group of pathways that do not require the protein clathrin. Among these pathways we can distinguish caveolin-dependent or independent endocytosis. Caveolae are the most commonly reported non-clathrin-coated PM buds (~50 nm in diameter), which exist on the surface of many, but not all cell types. They consist of lipid rafts, cholesterol-rich PM domains, bound to the protein caveolin. Fission of the vesicle from the PM is mediated by dynamins and the released vesicle can fuse with early endosomes[10].

Finally, other pathways independent of both clathrin and caveolin have been identified. These entry routes can be dynamin-dependent or independent. For example, the CLIC (clathrin- and dynamin-independent carriers)/GEEC (glycosil phosphatidyl-inositol-anchored protein enriched early endosomal compartments) pathway is clathrin- caveolin- and dynamin-independent and is responsible for the internalization of PM glycoproteins and toxins[11]. In contrast, the recently characterized reticulon3 (RTN3)-dependent NCE pathway schematically depicted in **Fig. 2** requires dynamin function and is responsible for the internalization of the epidermal growth factor receptor (EGFR)[12].

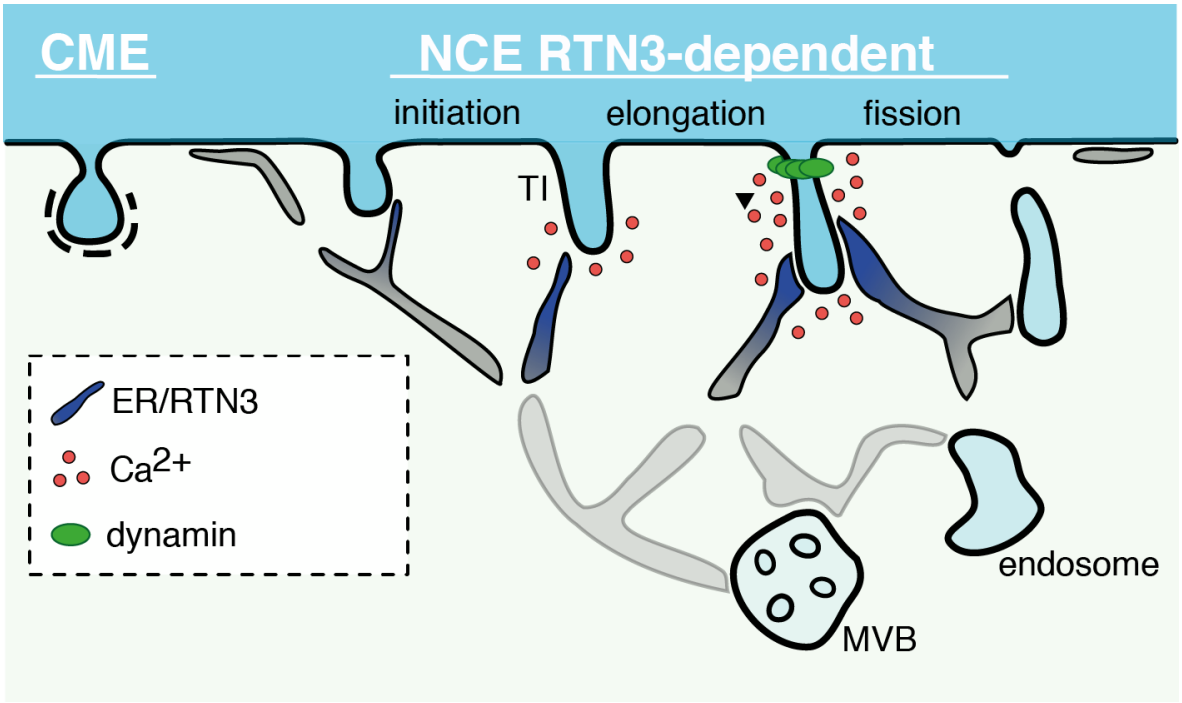


Fig. 2: NCE-RTN3-depdendent endocytic pathway of EGFR.

Schematic representation of the NCE-RTN3-depdendent endocytic pathway of EGFR. At high EGF doses, EGFR is internalized through both CME and through a specific NCE-RTN3-dependent endocytic pathway. This NCE consists of PM tubular invaginations (TI) in which the PM is in close contact with the endoplasmic reticulum (i.e. ER-PM contact sites). The formation and the fission of TI are dependent on the ER protein RTN3 at contact sites and on the local release of Ca^{2+} from the ER, respectively.
Fig. taken from Caldieri *et al.*[12].

1.2 Endocytosis as a master regulator of cell signaling

It is well established that endocytosis is much more than an internalization process of nutrients and molecules inside the cell; rather, it is now recognized as a master regulator of cell signaling. Indeed, internalization of receptors and their bound ligands is an efficient way of concentrating these molecules in signaling endosomes, regulated both in time and space.

Ligand-mediated endocytosis is essentially a mechanism for attenuating signaling from active receptors at the PM and targeting them to degradation. However, while in transit, the receptors are still bound to their ligands, and therefore, remain active and continue to signal. This is due to the fact that the small volume of endosomes could favor the receptor-ligand association and, consequently, sustain receptor activity. Hence, the name of "*signaling endosomes"* was coined, originally based on the findings of endosomes traveling retrogradely from the axon terminal to the soma[13]. Nowadays the expression signaling endosomes describes, more in general, vesicles that can sustain a signal originating from the PM.

In addition to sustaining signaling, signaling endosomes also have the potential to generate unique signals that could not be generated at the PM. In fact, sequestration of signaling receptors in endosomes might allow them to encounter substrates that were inaccessible from their PM location, acting as scaffolds for signaling components[14]. An example of this kind of signaling-scaffold endosome is the SARA (SMAD-anchor for receptor activation) endosome, involved in transforming growth factor β (TGFβ) signaling. Internalized TGFβ receptors are recruited to SARA endosomes where they encounter their targets SMAD2 and SMAD3 in close proximity, favoring their phosphorylation, which triggers the activation of pathway (see paragraph **4.6 The canonical TGFβ signaling pathway in EMT induction**)[15,16].

Vice versa, endosomes can isolate signaling components and prevent unwanted signaling interactions, for example, the sequestration of glycogen synthase 3 beta (GSK3β) in endosomes prevents the phosphorylation and subsequent degradation of β-catenin in the canonical WNT signaling pathway[17] (see paragraph **4.7 The**

WNT/β-catenin pathway). Thus, signaling endosomes contribute to signal diversification and specificity.

Endocytosis is also directly responsible for the activation of signaling pathways, for example, the Notch pathway, in which the endocytosis and recycling of Notch ligands delta/serrate/lag2 (DSL) in the signal-sending cell triggers the internalization and activation of Notch receptors in the signal-receiving cell[18,19] (see paragraph **2.3 Biological functions of epsins**).

Finally, endocytosis has the ability to respond to all the different signals coming from the extracellular environment and the final biological outcome depends on the integration of these signals and their entry routes. Indeed, receptors can be internalized through both CME and NCE routes, and the choice of the internalization pathway is one of the mechanisms used by endocytosis to control signaling outcome. For example, at low concentrations of the ligand, epidermal growth factor (EGF), EGFR is internalized through CME, which leads mainly to recycling of the receptor and sustained signaling. In contrast, in the presence of high EGF concentrations, EGFR is preferentially committed to NCE and targeted to degradation, thus leading to signal extinction[20]. Similarly, TGFβ receptors can be internalized via CME or NCE, leading to SMAD2/3 activation in the endosomes or to receptor ubiquitination and lysosomal degradation, respectively[15].

A summary of how endocytosis regulates signaling at different levels is depicted in **Fig. 3**.

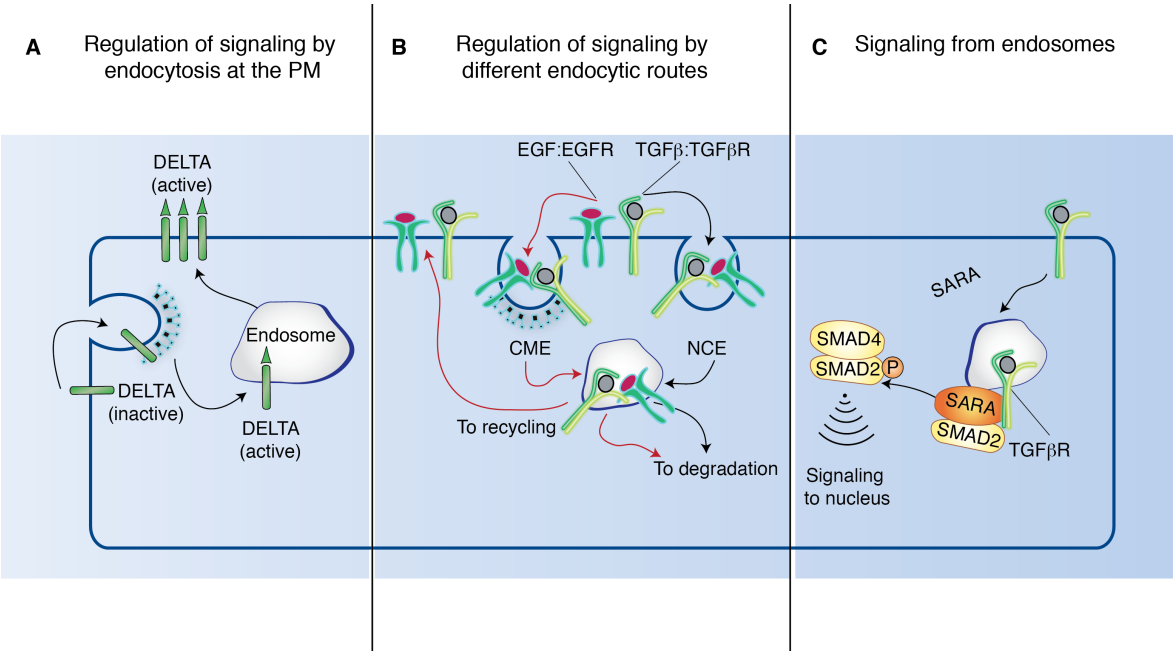


Fig. 3: Endocytosis regulates signaling at different levels.

Schematic representation of endocytic signaling regulation at different levels:

(A) Endocytosis regulates signaling at the PM, exemplified by the activation of the Notch pathway.

(B) Signaling regulation by choice of endocytic entry route (CME *vs.* NCE) in the case of the EGFR and TGFβ receptor, and the different fates associated with the specific entry route.

(C) Signaling regulation in the formation of signaling endosomes as in the case of SARA endosomes in TGFβ signaling.

Adapted from Sigismund *et al.*[1].

1.3 Endocytosis and cancer

Since endocytosis is a master regulator of cell signaling and biological responses, such as proliferation, differentiation and migration/invasion, the subversion of endocytic routes and aberrant signaling are likely to play an important role in cancer. In particular, since endocytosis regulates signaling both in a negative and a positive fashion, it could act either as a tumor suppressor or oncogenic mechanism. In the first case, endocytosis is an attenuator of signaling, inducing degradation of activated

growth factors receptors (e.g., EGFR), and thus is a potential tumor suppressor pathway. In the second case, endocytosis sustains signaling in endosomes or directs internalized receptors to recycling, preventing their degradation, thus, acting as a potential oncogenic pathway[21].

Indeed, alterations of several endocytic proteins have been reported in human tumors[22]. Among these, PM cargoes belonging to the receptor tyrosine kinase (RTK) family are frequently altered in human tumors, such as EGFR and Met (hepatocyte growth factor – HGF – receptor), both of which sustain proliferative growth. Genetic alterations consist in gene amplification/overexpression or point mutations that either inhibit proper internalization and degradation, or increase kinase activity of the receptor, resulting, in all cases, in a gain-of-function of sustained signaling[23]. Aberrant RTK signaling in cancer is also linked to mutations in proteins involved in receptor endocytosis and degradation, for example, the E3 ubiquitin ligase Cbl (casitas B-lineage lymphoma), responsible for ubiquitination of the receptor, a signal necessary for its degradation. Oncogenic forms of Cbl have been reported and are characterized by loss of the E3 ligase activity, again leading to sustained signaling[24]. In contrast, oncogenic RTKs have developed mechanisms to escape from Cbl-mediated downregulation[25].

Moreover, endocytosis is involved in the activation of pathways that have a key role in cancer, for example, the Notch pathway (see paragraph **2.3 Biological functions of epsins**), acting, also in this case, as a potential oncogenic pathway[26]. Furthermore, endocytic partitioning of ligand and receptor molecules in the Notch pathway or of the endocytic cell fate determinant Numb is crucial for asymmetric cell division (ACD) of stem cells (SCs) and consequently, endocytosis plays a major role in the maintenance of the SC compartment[27] (see paragraph **5.6 Endocytosis, asymmetric cell division of SCs and CSCs**).

Alteration of endocytic proteins with a causal role in tumors are now starting to be reported, for instance, in Rab GTPases, such as Rab25[28-30], as well as endocytic adaptors. Among these, epsin proteins were found to be altered in some human tumors[31-34] (see paragraph **2.4 Epsins and cancer**).

Finally, as we will discuss in the next paragraphs, endocytosis is emerging as a critical regulator of epithelial-to-mesenchymal transition (EMT), a hallmark of cancer progression in many human tumors (see paragraph **4.8 Endocytosis and EMT**).

2. Epsin proteins

2.1 The Epsin family

Epsins are a highly conserved family of endocytic adaptors, which in vertebrates (e.g. human, mice and rat species) is composed of Epsin 1 (Epn1), Epsin 2 (Epn2) and Epsin 3 (Epn3)[35]. In *Saccaromyces cerevisiae,* two epsin paralogs (Ent1 and Ent2) have been identified, whereas one epsin gene is present in *Caenorhabditis elegans* (epn1), and in *Drosophila melanogaster* (liquid facets: LqF)[36,37].

The first member of the family, Epn1, was discovered in rat as an interactor of the endocytic accessory protein, Eps15, thus the name of *epsin* was coined, which stands for Eps15-interacting protein[38]. Shortly after, the paralog Epn2 was discovered, a protein with a similar structural organization, high percentage of sequence similarity (see paragraph **2.2 Structure and domain organization of epsins** and **Fig. 4**), and similar cellular localization and profile of expression. Both Epn1 and Epn2 are ubiquitously expressed, and are particularly enriched in the brain[39]. In contrast, the third member of the family, Epn3, despite a similar structural domain organization and high percentage of sequence similarity, was shown to have a very restricted profile of expression. Indeed, under physiological conditions Epn3 is not expressed at all or expressed at very low levels, with the exception of high levels in gastric parietal cells[40]. Notably, Epn3 is overexpressed under pathological conditions, such as in migrating keratinocytes in cutaneous wounds[41] and some carcinomas[32,33].

While Epn1 is the best-characterized member of the Epsin family, its functions are largely phenocopied by Epn2. Indeed, both Epn1 and Epn2 were found to be localized at the PM and their role as endocytic adaptors in CME is well

characterized. Moreover, while single Epn1 or Epn2 knockout (KO) mice are healthy, the double Epn1/Epn2 KO is lethal and embryos do not survive beyond stage E10.5, due to impairment in the Notch signaling pathway[42] (see paragraph **2.3 Biological functions of epsins**). These data suggest redundancy between the two proteins, where one protein can compensate for the absence of the other. In contrast, very little is known about the physiological functions of Epn3 and its up-regulation in pathological conditions rather suggests an involvement in cell migration and in tumorigenesis.

Another protein with partial similarity to the protein structure and sequence of the epsin family has been discovered. This epsin-like protein, known as Epsin R or Epsin 4 (Epn4), is mainly involved in the formation of CCVs from the Golgi network and in retrograde transport, rather than in CME[43,44].

2.2 Structure and domain organization of epsins

The members of the Epsin family share a common structural domain organization, as shown in **Fig. 4.**

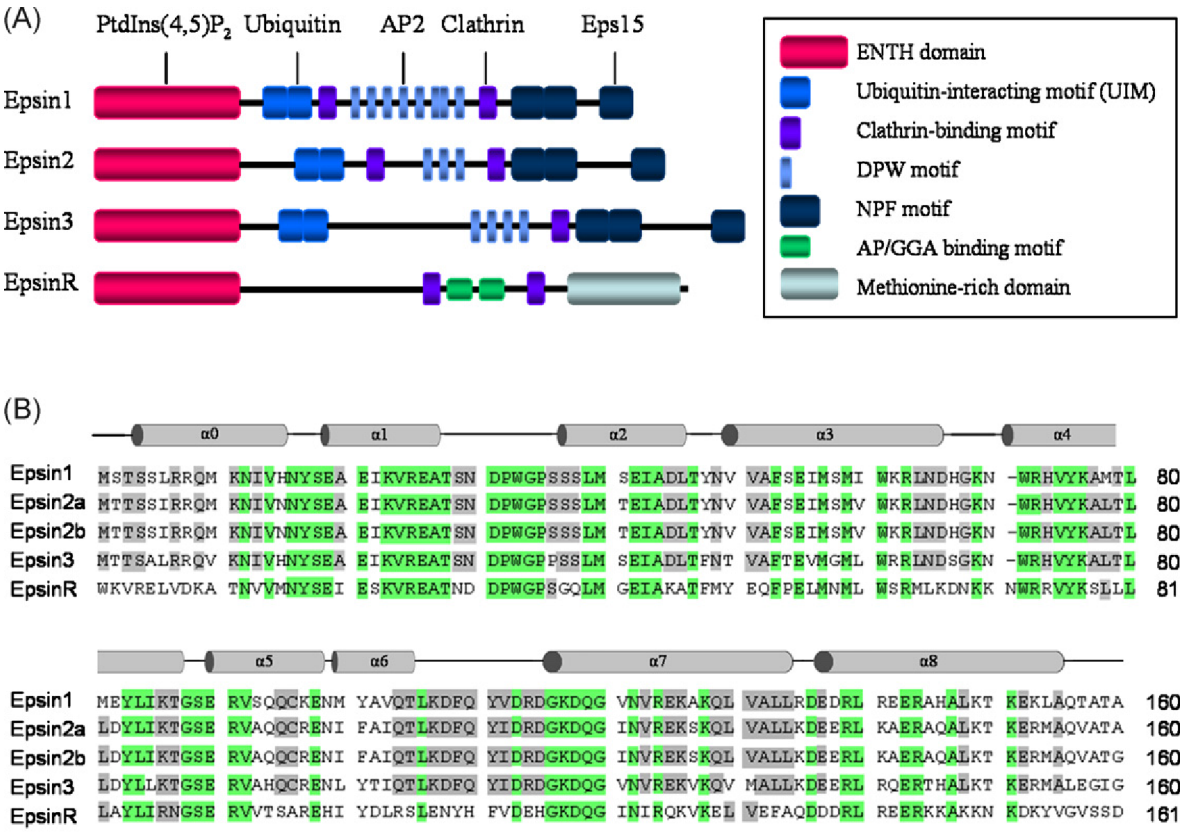


Fig. 4: The Epsin family.

(A) Schematic representation comparing the structural domain organization of human Epn1, Epn2, Epn3 and EpsinR. Epsin proteins share a similar modular organization: epsin-N-terminal homology domain (ENTH), ubiquitin interacting motifs (UIMs), clatrhin-binding, Asp-Pro-Trp (DPW), and Asn-Pro-Phe (NPF) motifs. EpsinR is a more distantly related protein and presents additional GGA motifs and a methionine-rich domain. See text for a detailed description.

(B) Sequence alignment of the ENTH domain in Epn1, Epn2, Epn3 and Epsin R. Identical amino acids are highlighted in green and similar residues in grey.

Fig. taken from Horvath *et al.*[45].

At the N-terminus, epsins contain their most conserved domain, approximately 150 amino acids long, known as the epsin-N-terminal homology domain (ENTH). This domain is responsible for binding to phosphatidylinositol 4,5-bisphosphate (PIP_2), a lipid enriched at endocytic sites of the PM. This interaction triggers a conformational change that induces membrane curvature required for the initiation of invagination and formation of CCPs[46].

Following the ENTH domain, epsins contain two or three ubiquitin interacting motifs (UIMs). These motifs interact with ubiquitinated cargoes: multiple monoubiquitination or limited polyubiquitination of cargoes is sufficient for the recruitment of epsin at the PM[47,48].

The region between the UIMs and the C-terminal of the protein is mostly unstructured and represents the region where the proteins differ the most. However, this region does contain short conserved signature motifs that allow epsin binding to different components of the clathrin endocytic machinery. In particular, one or two clathrin-binding and several (three to eight) DPW motifs (Aspartate-Proline-Tryptophan) that allow binding to the clathrin heavy chain and to the AP2 complex, respectively[38,49].

Finally, three NPF (Asparagine-Proline-Phenylalanine) motifs at the C-terminus are responsible for binding to the Eps15-homology (EH) domain of proteins, such as Eps15, intersectin and POB1[38,50,51].

Epsin R proteins share only the ENTH domain with epsin proteins. Indeed, since these proteins do not possess NPF motifs binding to Eps15, they are not classical epsins. Furthermore, they do not have UIMs and DPW motifs, while the clathrin-binding motif is conserved. Finally, in contrast to epsins, Epsin R proteins possess the Golgi-localized, Gamma-ear-containing, Arf-binding proteins (GGA) binding motif, which is involved in the formation of CCVs from the Golgi network, and a methionine-rich domain, the function of which is unclear[43,44].

2.3 Biological functions of epsins

Thanks to their structural domain organization, epsins can bind to both phospholipids and ubiquitinated cargoes at the PM, and interact with clathrin, AP2 and other endocytic accessory proteins, such as Eps15. Thus, epsins act as endocytic adaptors, linking surface receptors to the major components of the endocytic machinery, as depicted in **Fig. 5**.

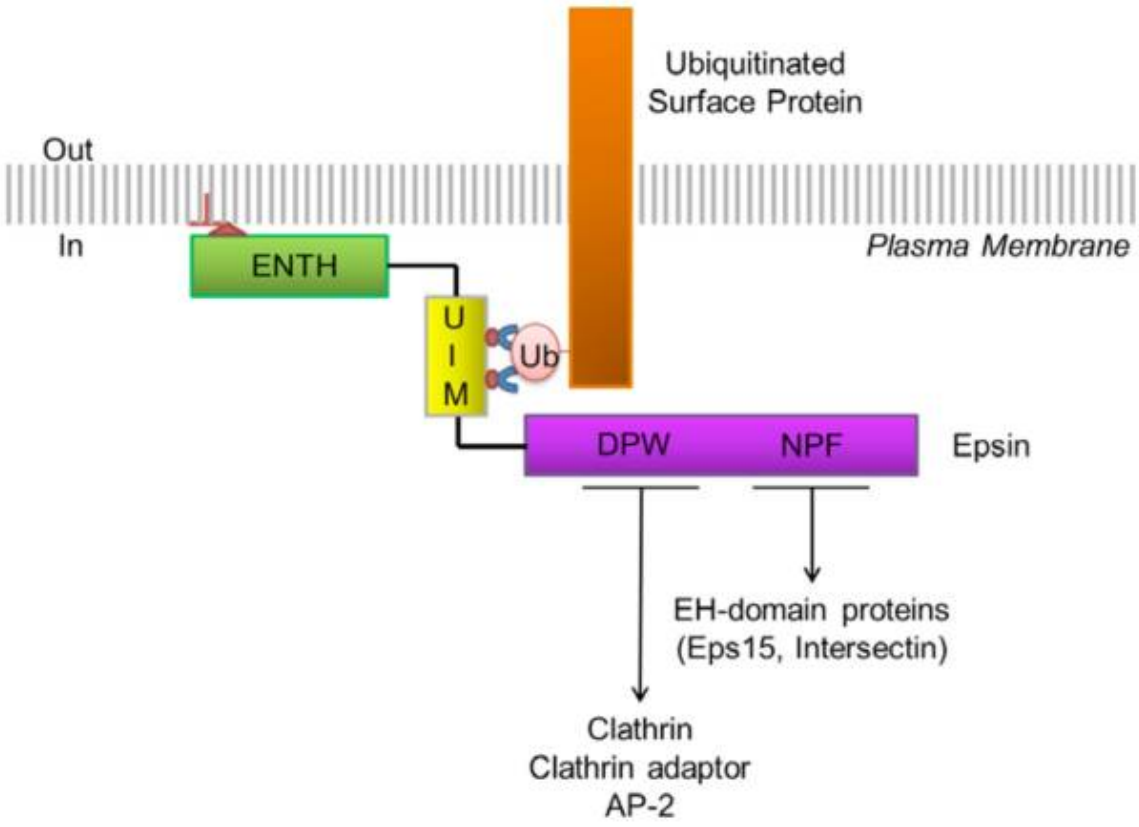


Fig. 5: Epsins mediate endocytosis of ubiquitinated cell surface receptors.

Schematic representation of a generic mammalian epsin involved in the endocytosis of a ubiquitinated surface receptor. The epsin domain organization is also represented: epsin-N-terminal homology domain (ENTH), ubiquitin interacting motifs (UIM), Asp-Pro-Trp (DPW), Asn-Pro-Phe (NPF) motifs.
Figure taken from Tessneer *et al.*[52].

For example, epsins are required for the endocytosis of the EGFR both through CME and NCE[48]. Similarly, they are also required for vascular endothelium growth factor receptor 2 (VEGFR2) endocytosis[53]. Both of these receptors have been extensively studied because of their role in cancer and studies have revealed that epsins, through the regulation of the endocytosis of these receptors, might play a role cancer progression (see paragraph **2.4 Epsins and cancer**).

As mentioned above, endocytosis has a dual role in signaling regulation: on one hand, it desensitizes cells to activated receptors (as in the case of the EGFR), on the other, it is required for the activation of signaling pathways (as in the case of Notch pathway)[1,21]. Epsins are important for both types of regulation. Indeed, studies in *Drosophila* have shown that the Epsin orthologue, LqF, is required for Notch activation. Notch signaling activation is dependent on the interaction of a signal-sending cell expressing Notch ligands and a signal-receiving cell, expressing Notch receptors. Upon proteolytic cleavage of Notch receptors in the signal-receiving cell, the notch intracellular domain (NICD) is released into the cytosol, from where it can translocate to the nucleus and activate gene expression. Epsins have been shown to regulate the endocytosis of Notch ligands, DSL, in the signal-sending cell[19]. It has been proposed that this endocytosis of DSL produces a mechanical pulling force that drives Notch receptor conformational change in the signal-receiving cell, necessary for its internalization and activation in the signal-receiving cell[54]. Of note, the LqF null mutation in *Drosophila* is embryonic lethal[55], similarly to the double Epn1/Epn2 KO in mammals, as described above[42]. Indeed, embryonic fibroblasts extracted from double Epn1/Epn2 KO mice embryos did not show general defects in CME. Lethality was rather due to the disruption of extraembryonic structures, defects in cardiovascular development, somitogenesis and neural tube differentiation[42]. All these phenotypes are similar to those produced by impairment of the Notch pathway

and show that epsins are fundamental for embryogenesis and development, possibly via their specific role in the endocytosis of DSL and the consequent impact on Notch signaling activation.

Besides their role in endocytosis, epsins are involved in other cellular processes, for example, in the regulation of Rho-GTPase signaling responsible for actin remodeling in motility and cell polarity. Indeed, studies in yeast have demonstrated that the ENTH domains of epsins binds to and inhibits Cdc42 GTPase-activating proteins (Cdc42 GAPs)[56]. This interaction is fundamental for the establishment of cell polarity and, consequently, for cell viability in yeast. Epsin involvement in Rho-GTPase signaling function is also conserved in mammals[57] (see paragraph **2.4 Epsins and cancer**).

Finally, the ability of epsins to induce membrane curvature is important for the formation of autophagosomes in autophagy[58] and for proper mitotic membrane organization[59]. It is still unclear whether these non-canonical functions of epsins are linked to their endocytic function or whether they go beyond endocytosis and signaling.

2.4 Epsins and cancer

Considering the pivotal role of epsins in the endocytosis and regulation of several signaling receptors (e.g., EGFR and Notch), and its non-canonical actin remodeling function in motility and cell polarity processes (see paragraph **2.3 Biological functions of epsins**), it is not surprising that epsins have been implicated in cell

processes, such as proliferation, differentiation, invasion, migration and angiogenesis, that is to say the major so-called *hallmarks* of cancer[60].

One of the most convincing evidences of epsins' role in cancer is its role in VEGFR2 endocytosis and, consequently, in tumor angiogenesis. Highly proliferating cells in growing tumor masses have an increased need of oxygen and nutrients. The secretion of growth factors, such as vascular endothelium growth factor (VEGF), by these cells promotes endothelial up-regulation of the VEGF receptor, VEGFR2, which allows the formation of new blood vessels from pre-existing ones, thus, ensuring adequate oxygen and nutrient delivery to the tumor cells (tumor angiogenesis)[61]. Conditional KO of Epn1/Epn2 in murine vascular endothelium showed no major defects under normal conditions, indicating that they are not required for normal blood vessel regulation; however, Epn1/Epn2 KO did result in increased vascular permeability and impaired tumor growth, highlighting a role in tumor angiogenesis. Mechanistically, this is due to impaired endocytosis and degradation of VEGFR2, resulting in excessive, yet non-productive, VEGF signaling[53]. In agreement, VEGFR2 reduction in primary endothelial epsin-depleted cells was sufficient to restore normal VEGF signaling, endothelial proliferation, migration and network formation[62].

Another hallmark of cancer development is tumor cell proliferation: cancer cells, because of the activation of oncogenes and/or loss-of-function of tumor suppressors, are able to escape cell cycle arrest, apoptosis and immune surveillance, and to proliferate at higher rates than normal cells. This gives tumor cells a proliferative advantage, allowing the formation of solid tumors. Although direct evidence of the role of epsins in tumor proliferation are currently lacking, epsins have been shown to regulate the signaling and endocytosis of several growth factor receptors involved in tumorigenesis, for example, the EGFR. Indeed, EGFR expression is frequently up-

regulated in many solid tumors, where it promotes cancer progression[63], suggesting that epsins might contribute to the regulation of tumor proliferation, acting through the endocytosis of EGFR.

Finally, the non-canonical role of epsins in actin remodeling via Rho-GTPase signaling might be relevant to the invasion and migration of tumor cells. Tumor cell invasion and migration depends highly on a functioning actin network. In agreement, overexpression of epsin proteins has been shown to promote fibrosarcoma migration and invasion. Mechanistically, this is achieved through the interaction of epsins, via their ENTH domain, with the signaling molecule, Ral-binding protein 1 (RalBP1), involved in migration and invasion[57].

Lastly, Epn1 and Epn2 have been reported to be up-regulated in prostate cancer and required for cancer progression[31], while Epn3 has been found to be up-regulated in a lung adenocarcinoma cancer cell line[32] and breast adenocarcinomas[33], and to be prognostic for both breast cancer (BC) and non-small cell lung cancer[34]. Nevertheless, even though these studies point to a link between epsin expression and some cancers, they were based on a lung adenocarcinoma cancer cell line (EKVX), a canine mammary tumor, and a bioinformatic analysis of publicly available databases, respectively. Therefore, whether epsins (especially Epn3) are indeed drivers in human cancer progression remains unclear.

3. Breast Cancer

3.1 Normal mammary gland development

The mammary gland is a glandular epithelium composed mainly of ducts that terminate in lobular structures, called terminal lobular-ductal units (TLDUs). These TLDUs are composed of two distinct lineages: luminal epithelial cells and basal/myoepithelial cells. As shown in **Fig. 6**, luminal cells constitute the inner lumen, whereas basal/myoepithelial cells form an outer layer in direct contact with the basement membrane. Immunohistochemically, these cells can be distinguished by the expression of different cytokeratin (CK) proteins: CK7/8 and CK18/19 identify luminal cells and CK5/6 and CK14/17 basal/myoepithelial cells. The ducts are surrounded by the stroma, commonly known as the *mammary fat pad*, which consists of adipose cells, fibroblasts, blood vessels, immune cells and nerve terminals[64].

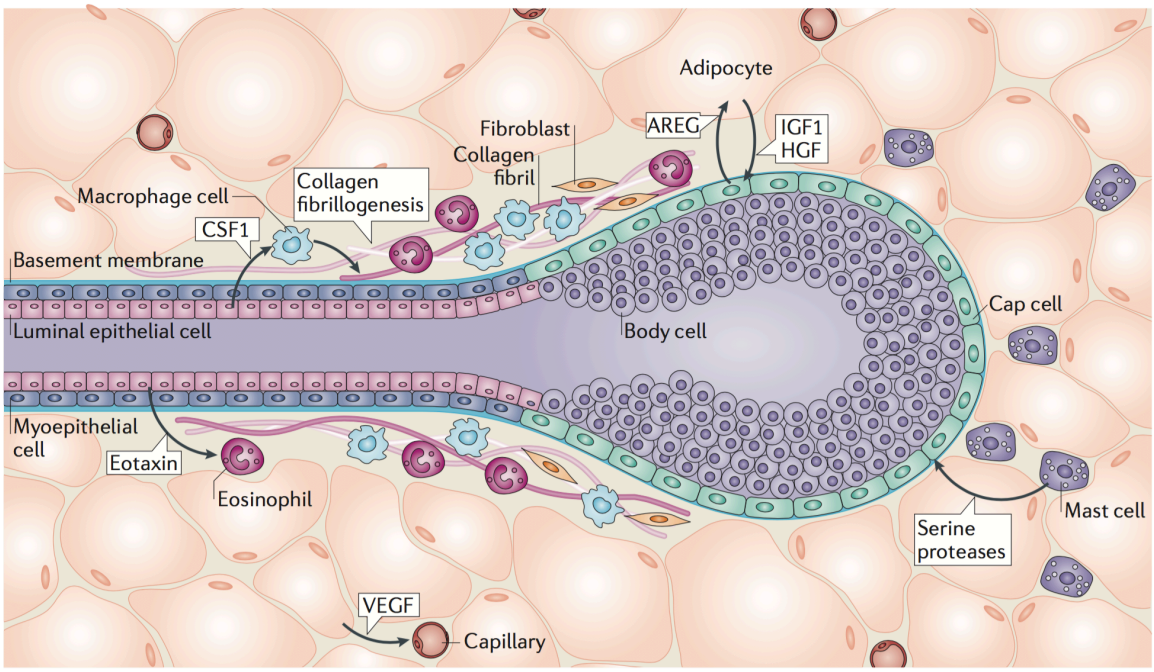


Fig. 6: Terminal lobular-ductal units and the surrounding stroma.

Schematic representation of a terminal lobular-ductal unit (TLDU) composed of an inner layer of luminal epithelial cells and an outer layer of basal/myoepithelial cells in contact with the basement membrane. The TLDU is surrounded by the stroma, composed of adipose cells, fibroblasts, blood vessels and nerve terminals. Interactions between epithelial and stromal cells through the production of growth factors (HGF, hepatocyte growth factor; IGF1, insulin-like growth factor 1; AREG, amphiregulin; VEGF, vascular endothelial growth factor; CSF1, macrophage colony-stimulating factor 1) coordinate mammary morphogenesis. Figure taken from Gjorevski *et al.*[64].

The mammary gland represents a unique organ because of its ability to undergo different cycles of remodeling during development and throughout the entire life of female mammals. As shown in **Fig. 7**, mammary morphogenesis starts at the embryo stage, in which an epithelial structure called the *primordium* or *mammary bud* is formed and which later develops rudimentary mammary epithelial trees that remain unchanged until puberty. At puberty, sex hormone signaling by estrogen and progesterone leads to duct expansion, side branching and the formation of mature TLDUs. At pregnancy, luminal epithelial cells of the TLDU undergo robust proliferation and differentiation to form lobular-alveolar buds, called *secretory*

alveoli, responsible for milk production during lactation, whereas basal/myoepithelial cells favor the lactation process through their contraction. Finally, at the end of pregnancy, the mammary gland undergoes an involution process in which massive apoptosis of the epithelium allows it to return to the virgin stage before pregnancy[65].

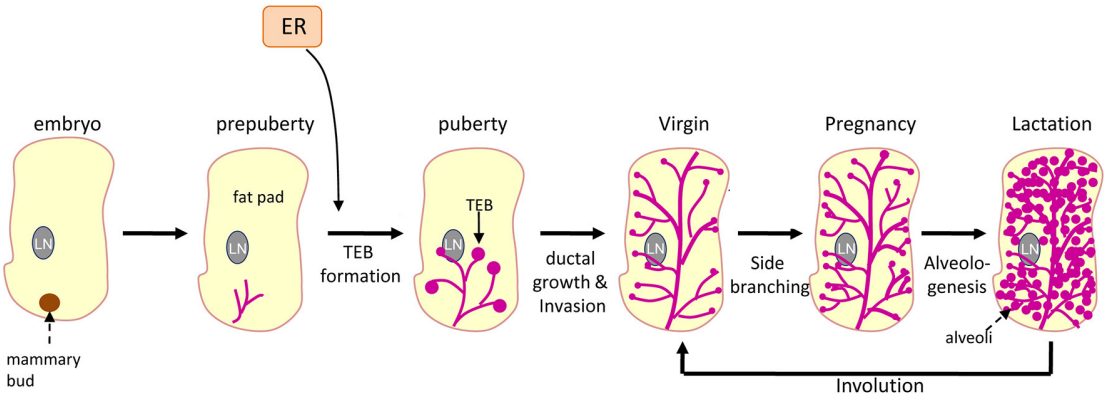


Fig. 7: Mammary gland development.

Schematic representation of the developmental stages of the mammary gland: embryo, pre-puberty, puberty, virgin, pregnancy and lactation.
Figure taken from Manavathi *et al.*[65].

The adult mammary gland is able to sustain multiple cycles of remodeling (pregnancy, lactation, involution) thanks to the existence of epithelial mammary stem cells (MaSCs) able to generate all the cell lineages of the mammary gland within the basal compartment[66] (see paragraph **5.1 Mammary stem cells and breast cancer stem cells: origin and modality of division**).

3.2 Histological and molecular characterization of breast cancer

BC is the most common cancer in women and a leading cause of cancer-related deaths among woman worldwide[67]. Indeed, despite the fact that prevention, early diagnosis and state-of-the-art therapies have led to significant improvements in the treatment of BC, tumor relapse and metastasis are common events and the major cause of cancer-related deaths.

The vast majority of BCs are carcinomas that originate in TLDUs from cells lining the milk-forming ducts of the mammary gland and, according to the epithelial origin, we can distinguish ductal or lobular carcinomas. Their transformation is dependent on genetic, environmental and dietary factors, which alter growth regulatory pathways, leading to uncontrolled proliferation, a hallmark of tumorigenesis.

Unlike other epithelial cancers, such as colon, different stages of BC are not yet clearly defined. However, benign stages, such as non-invasive BC (also known as carcinoma *in situ*), as opposed to invasive cancers have been identified. *In situ* carcinomas are characterized by the presence of cancer cells restricted to the basement membrane. The most common example of this type of BC is ductal *in situ* carcinoma (DCIS)[68]. Invasive BCs are instead characterized by cancer cells infiltrating the surrounding tissues and are associated with a higher risk of metastasis. Invasive ductal carcinomas (ICD) represent the most frequent type of invasive BC and constitutes 70-80% of all BCs[69].

Pathological classifications of BC also take into account clinical features of the tumor, such as tumor grade and tumor stage. Grading compares the morphological appearance of tumor cells with that of normal breast tissue cells. Cells are classified

as well differentiated (low grade or G1), moderately differentiated (intermediate grade or G2), and poorly differentiated (high grade or G3), as cancer cells progressively lose the features seen in normal mammary cells. Grading correlates with prognosis, with G3 tumors displaying the worst prognosis[70].

According to the international union against cancer (UICC), staging of the tumor is based on the TNM classification: the size of the tumor (**T**), whether or not the tumor has spread to the lymph nodes (**N**) in the armpits, and whether the tumor has metastasized (**M**). Larger size, nodal spread, and metastasis have a higher stage number and a worse prognosis. The main stages are: stage 0, a pre-cancerous condition, either DCIS or lobular carcinoma *in situ* (LCIS); stages 1–3, within the breast or regional lymph nodes; stage 4, metastatic cancer that has a less favorable prognosis since it has spread beyond the breast and regional lymph nodes.

Nevertheless, human breast tumors are very heterogeneous and histological or clinical classifications are not sufficient to predict prognosis or tumor cell response to therapy. Therefore, BCs are also classified according to their molecular features, and the different molecular subtypes better define breast tumor heterogeneity, patient outcome and therapeutic responses.

The major molecular classifications are based on the following features: i) hormone receptor status – estrogen (ER) and progesterone (PR) positive or negative BC; ii) overexpression of the human epidermal growth factor 2 (HER2) – HER2-positive or HER2-negative BC; iii) the proliferative index based on expression of the proliferation marker Ki67. Tumors expressing ER or PR have a better prognosis, whereas, tumors lacking all 3 receptors (i.e., ER^-, PR^-, $HER2^-$), are known as triple-negative BC (TNBC) and have a worse outcome.

More recent types of molecular classification, based on the microarray profiling of >8000 genes[71] or on the development of an efficient 50-gene classifier, named

PAM50 (Prediction Analysis of Microarray)[72], have subdivided BC into 5 main categories: luminal A, luminal B, HER2-amplified, basal-like and normal-like. After this initial classification, other categories have been added, such as the claudin-low subtype[73].

Luminal tumors express hormone receptors and their expression profiles are reminiscent of the luminal epithelial compartment. Luminal tumors respond well to hormone therapy, but poorly to conventional chemotherapy. In particular, Luminal A tumors are characterized by high level of ER and PR expression, no overexpression of HER2, low proliferation rate (Ki67), and low tumor grade. They constitute about 50-60% of all BC and are associated with the better clinical outcome. Luminal B tumors are a more molecularly heterogeneous group of tumors that express lower levels of ER and PR compared with Luminal A tumors, and can be either HER2-negative or positive. They represent about 10-20% of all BCs and are more aggressive than luminal A BCs, displaying a higher proliferative index (Ki67), higher histological grade (grade 2/3 as opposed to grade 1/2 for Luminal A) and worse prognosis[74].

The HER2-amplified BC subtype represents about 15-20% of all BCs and is characterized by high HER2 expression, high proliferative index, high histological grade (grade 2/3) and poor prognosis. Nevertheless, molecularly targeted therapies are available for this class, such as the HER2 monoclonal antibodies, trastuzumab[75] and pertuzamab[76], which have significantly improved the clinical outcome of this type of BC.

Finally, basal-like tumors represent about 10-20% of all BCs. They are characterized by their triple negative status for receptors and their expression profiles mimicking those of basal epithelial cells. They are high-grade (G3) tumors and have the worst clinical outcome (frequently associated with early relapse and metastasis

after surgery) especially because of the lack of molecularly targeted therapies for this BC subtype. Indeed, given the triple negative receptor status, basal tumors are not amenable to conventional targeted BC therapies, leaving chemotherapy the only option in advanced stages[77].

A summary of the histological and molecular classification of BC is shown in **Fig. 8**.

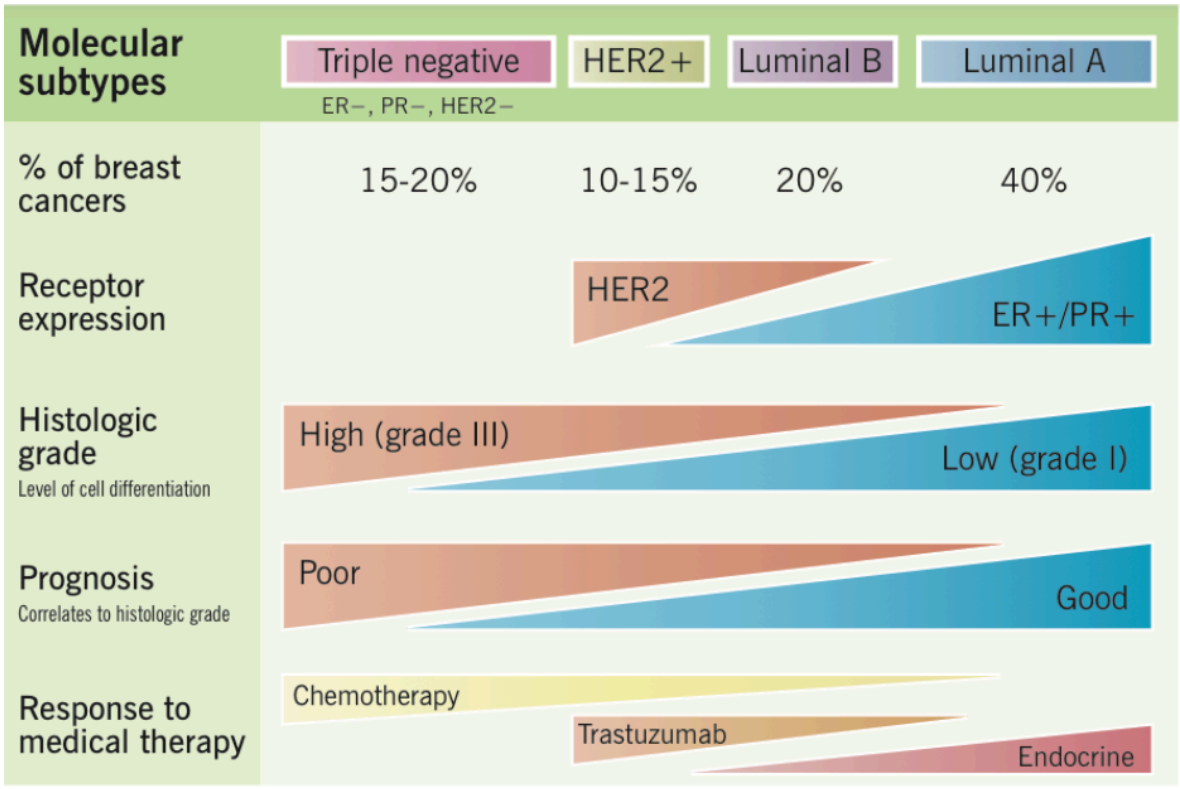


Fig. 8: Molecular characterization of breast cancer.

Table summarizing the major molecular subtypes of BC showing relative occurrence (% of all BCs), receptor expression (ER, PR, HER2), grade, prognosis and response to medical therapy. Figure taken from www.pathophys.org/breast-cancer).

3.3 *In vitro* and *in vivo* study of breast cancer progression: non-tumorigenic cell lines and mouse models

Various *in vitro* and *in vivo* models have been developed to unravel the mechanisms of breast tumorigenesis.

Human mammary epithelial cells (HMECs), which include luminal and basal/myoepithelial cells, obtained by reductive mammoplasties or mastectomies, can be grown *in vitro*. These cells have a limited life span in culture, eventually undergoing senescence. They have been invaluable for the characterization of the initial steps of tumorigenesis, involving the loss of senescence checkpoints and immortalization, both of which allow the cells to grow indefinitely in culture and to acquire additional mutations leading to their full transformation[78].

A fundamental gene involved in senescence is the enzyme telomerase, whose function is to maintain telomere length and protect chromosomes from losing important information at each cellular replication during embryonic development. The majority of adult somatic cells do not express telomerase, resulting in telomere shortening at each cell division, until their length is too short and cells senesce. However, many cancer cells are able to inhibit telomere shortening by expressing telomerase and, therefore, proliferate indefinitely[79]. Also DNA damage and overexpression of oncogenes are able to induce senescence and the tumor suppressors p53, p21 and $p16^{INK4a}$ play a crucial role in this process[80,81].

Consequently, *in vitro* immortalization is mediated through oncogenes that can overcome senescence, for example, viral oncogenes, such as the SV40 large T antigen[82] and the human papilloma virus E6-E7 oncoproteins[83], or through the overexpression of human proteins, such as hTERT (the catalytic subunit of human

telomerase)[79] or mutant p53[84]. Immortalized HMECs, however, are not able to form tumors when injected into the mammary fat pad of immunocompromised mice. Weinberg and colleagues showed that the combination of the expression of the SV40 small and large T antigens (inhibiting p53- and Rb-dependent senescence), hTERT (inhibiting shortening of telomeres), and the active Ras oncogene was able to generate transformed HMECs, named HMLERs (for the presence of active Ras in contraposition to non-tumorigenic HMLEs, non-tumorigenic cells derived by immortalization of HMECs)[85].

Besides primary HMECs, the most commonly used established normal human breast cell line is MCF10A. These cells were derived from benign proliferative breast tissue of a woman with fibrocystic disease by spontaneous immortalization, without viral oncogenes or chemical factors, and by switching culture conditions from decreased levels of calcium (0.04 mM) to normal calcium levels in the medium (1.05 mM). MCF10A are non-tumorigenic, mammary epithelial cells that are ER-, PR- and HER2- negative/normal low levels, and p53 wild-type. Cytogenetic analysis prior to immortalization showed normal diploid cells; later passages showed minimal rearrangement and near-diploidy, with deletion of the $p14^{ARF}/p16^{INK4a}$ locus[86].

In two-dimensional (2D) culture, MCF10A cell growth is dependent on hormones (hydrocortisone and insulin) and on the growth factor of EGF. In three-dimensional (3D) cultures, when cultured on top of Matrigel, an extracellular matrix (ECM) mixture derived from mouse sarcoma cells, MCF10A cells are capable of forming polarized organoids, acinus-like spheroids with a hollow lumen surrounded by basement membrane[87]. Indeed, like HMECs, MCF10A have proved to be a suitable model system to study BC tumorigenesis[88-90].

In addition to cell model systems, mouse models of transgenic overexpression of oncogenes or deletion of tumor suppressor genes that give rise to breast adenomas,

similar to those occurring in humans, have been critical to understanding the mechanisms underlying breast tumorigenesis. These include models expressing viral proteins (e.g., SV40), and also cellular proteins, such as c-Myc, ErbB2/neu, BRCA1/2, cyclin D1, mutant p53 and others[91].

4. The epithelial-to-mesenchymal transition: mechanisms and pathways involved

4.1 The basics of EMT

The epithelial-to-mesenchymal transition (EMT) is a complex biological process of transcriptional and post-transcriptional changes that define cell shape and function and reprogram gene expression. Through this program, epithelial cells lose their cell-cell adhesion and apico-basal polarity, and acquire a mesenchymal phenotype, as shown in **Fig. 9**. This consists of an elongated spindle-like cell shape, also known as a fibroblast-like morphology, enhanced migratory and invasiveness capacities, increased production of ECM components and matrix metalloproteases involved in ECM degradation, and expression of mesenchymal markers[92].

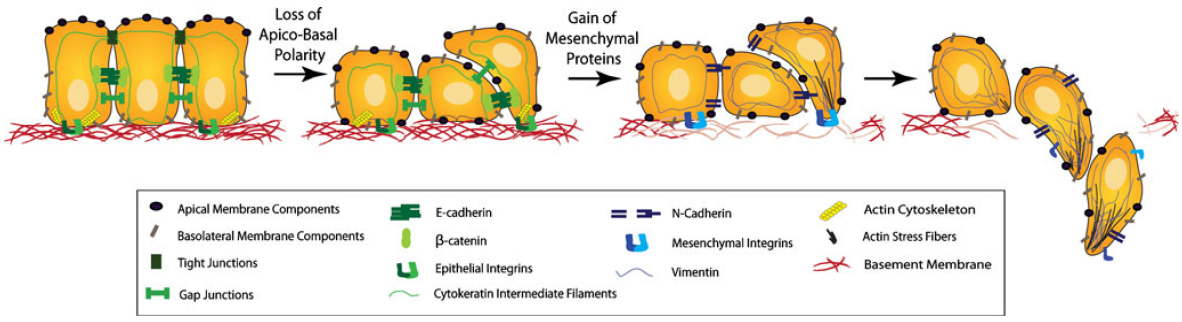


Fig. 9: The basics of epithelial-to-mesenchymal transition.

Illustration representing the basics of epithelial-to-mesenchymal transition (EMT): loss of apico-basal polarity and gain of mesenchymal features. The legend details the molecular players involved in this process. See main text for a detailed description of cell junctions, the Cadherin switch and the pathways involved in EMT.
Figure taken from Micalizzi *et al.*[93].

The result of EMT is, therefore, the formation of a mesenchymal cell able to invade the ECM and migrate far from the epithelial layer in which it originated. Conversely, mesenchymal cells can undergo the reverse process, known as mesenchymal-to-epithelial transition (MET). Consequently, EMT and MET have challenged the concept of terminal differentiation of cells and, indeed, the ability of epithelial cells to transition into mesenchymal cells and back again demonstrates the high degree of plasticity of the epithelial phenotype.

Nowadays, EMT is often associated with malignancy and is recognized as a hallmark of cancer[60]. However, EMT is, in the first place, a finely regulated physiological process that plays a crucial role in organogenesis during embryonic development or in wound healing and tissue regeneration. Cancer cells are able to hijack this process for metastasis formation (see paragraph **5.5 Cancer stem cells, EMT-MET and metastasis**). For this reason, EMT can be classified into three types according to the biological context in which it occurs: Type I (developmental), Type II (wound healing, tissue regeneration and fibrosis), Type III (cancer), as shown in **Fig. 10**.

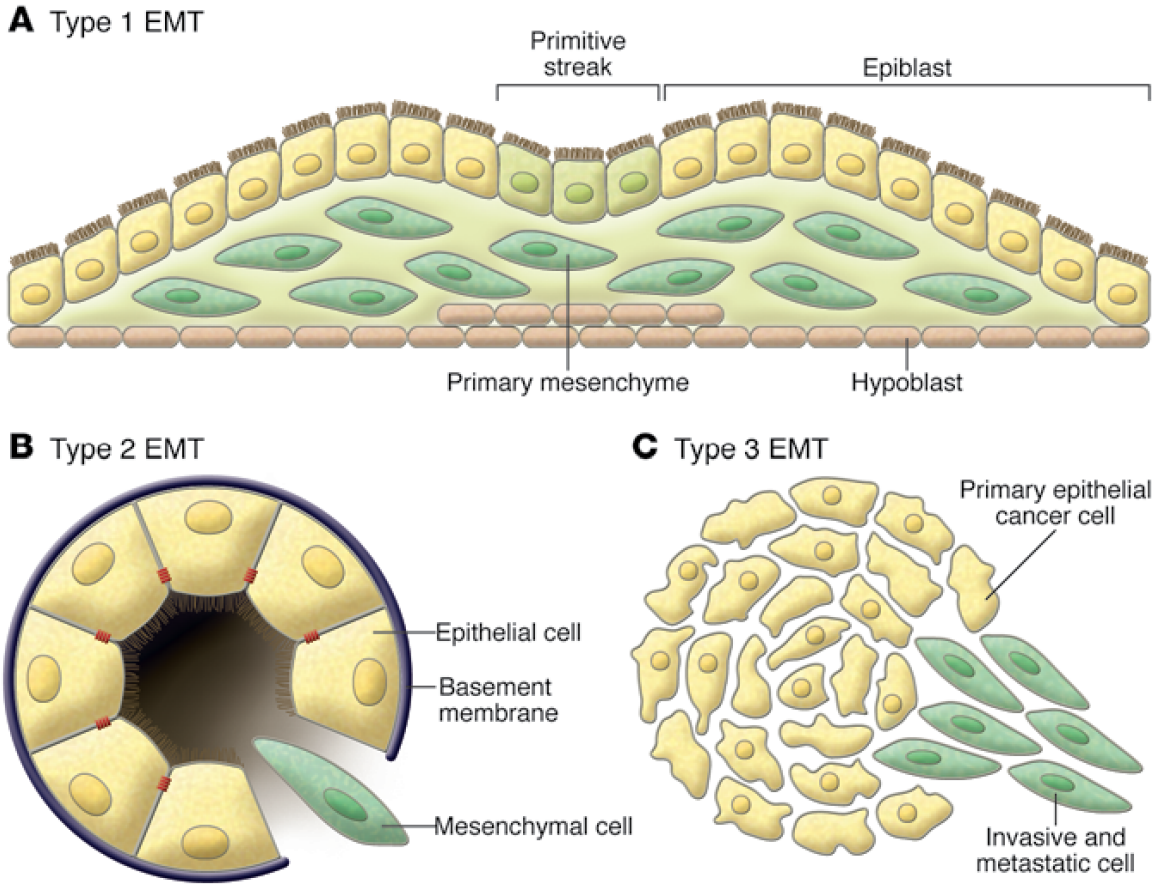


Fig. 10: Different types of EMT.

Illustration of the different types of EMT:
(A) Type I, occurring during embryonic development, for example, in the gastrulation stage from the primitive-streak.
(B) Type II, occurring during wound healing, tissue regeneration and fibrosis.
(C) Type III, occurring in cancer, leading to the generation of invasive metastatic cells.
Figure taken from Kalluri *et al.*[92].

In Type I, after the initial stage of embryogenesis, EMT is associated with the implantation of the embryo and the initiation of placenta formation. The trophoectoderm cells undergo EMT to facilitate the invasion of the endometrium and appropriate placenta placement, thus enabling nutrient and gas exchange with the embryo[94]. Later in embryogenesis, gastrulation generates three germ layers: the ectoderm, the endoderm and the mesoderm. In this process, the primitive epithelium, the epiblast, gives rise to the primary mesenchyme through EMT, which allows the

cells to ingress in a specific area of the embryo, called the primitive-streak, an invagination of the epiblast layer[95]. Finally, during embryonic development, EMT of the epithelial cells of the neuroectoderm gives rise to migratory neural crest cells[96].

Type II EMT, which occurs in wound healing, tissue regeneration and fibrosis, is similar to the physiological one active in development and acts to repair tissue upon injury. However, in contrast to Type I EMT, Type II EMT is associated with inflammation, and can be reactivated by it until inflammation is attenuated. Indeed, in the context of fibrosis, persistent inflammation leads to the persistent activation of EMT that can result eventually in organ destruction[97].

Lastly, epithelial cancer cells exploit the EMT program to acquire invasive and migratory properties, leading to tumor metastasis (Type III EMT)[98]. However, this type of EMT is far from similar to the other two types of EMT, affecting tumor suppressor genes and oncogenes, and the signals inducing this type of EMT are not only different from the physiological ones, but also heterogeneous and cancer specific.

4.2 Epithelial cells and cell junctions

Epithelial cells form a layer of cells that are closely adjoined by specialized membrane structures known as tight junctions, adherens junctions, desmosomes, hemidesmosomes and gap junctions[95]. Moreover, epithelial cells have an apico-basal polarity, in which specialized adhesion molecules are asymmetrically localized at the basal surface of the cell, in contact with the basal membrane and other adhesion molecules that form a lateral belt that allows cell-cell adhesion. Additionally, the

specialized apical surface on epithelial cells is in contact with the extracellular environment.

A schematic illustration of cell junctions and apico-basal polarity of epithelial cells is shown in **Fig. 11**.

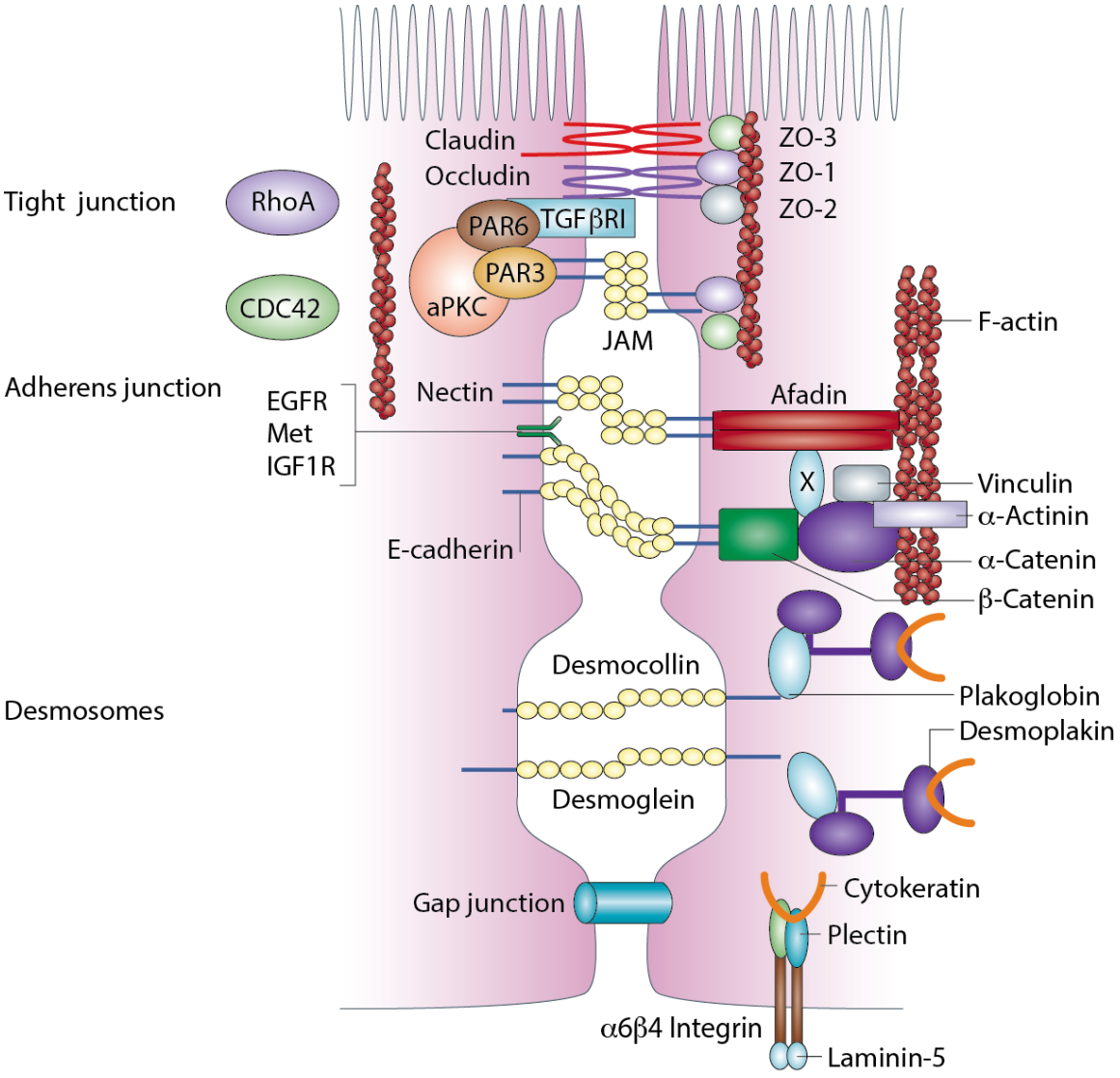


Fig. 11: Epithelial cell junctions.

Schematic illustration of the different epithelial cell junctions: tight junctions, adherens junctions, desmosomes, hemidesmosomes and gap junctions. The major molecular players are shown. See main text for a detailed description.
Figure taken from Thiery *et al.*[95].

At the molecular level, epithelial cells form adherens junctions between cells through homophilic binding of E-Cadherin (epithelial-cadherin), a transmembrane protein member of the cadherin superfamily. Its structure consists of an extracellular domain, which allows cell-cell junction formation, a transmembrane domain, and an intracellular domain, through which it binds the other components of the adherens junctions complex; i.e., members of the catenin superfamily (β-catenin, α-catenin, δ-catenin, also known as p120, and γ-catenin, also known as plakoglobin) and, indirectly, the actin cytoskeleton[99] (**Fig. 11**).

Tight junctions, instead, consist of claudins and the occludins, which associate with different peripheral membrane proteins, such as ZO-1, ZO-2, ZO-3, localized on the intracellular side of the PM, which in turn anchor the strands to the actin cytoskeleton (**Fig. 11**). As the name suggests, they form very tight junctions between cells, forming a protective barrier and preventing the passage of molecules between the cells[100].

In contrast, Gap junctions directly connect the cytoplasm of neighboring cells, allowing passage of molecules between cells. For this reason they are also known as communicating junctions[101] (**Fig. 11**).

Lastly, desmosomes and hemidesmosomes are composed of a cytoskeletal anchor, represented by intermediate filaments (IF) (made of cytokeratin in both cases), a transmembrane linker, which is made of cadherin proteins for desmosomes and integrins for hemidesmosomes, and other linker proteins (**Fig. 11**). Desmosomes and hemidesmosomes mediate cell-cell adhesion or cell-ECM adhesion, respectively[102].

Taken together, these specialized cell junctions coordinate the organization of the epithelial tissue.

4.3 Mesenchymal cells, the Cadherin switch and EMT markers

Mesenchymal cells, in contrast to epithelial cells, do not form an organized cell layer and do not have an apico-basal polarity; nor are they associated with the basal lamina. Moreover, while epithelial cells in culture grow as tightly packed clusters of cells (also known as islands or colonies) with cell-cell junctions, mesenchymal cells grow as individual cells with a fibroblast-like morphology.

Indeed, mesenchymal cells do not express E-Cadherin[103]. Therefore, a hallmark in EMT and cancer progression is loss of E-Cadherin expression through different mechanisms (transcriptional repression, protein degradation, inactivating somatic mutations, protein re-localization). This causes disruption of adherens and tight junctions, as well as, downregulation of claudins and occludins, favoring migration, invasion and extravasation steps[104].

Parallel to the loss of E-Cadherin, another hallmark of EMT is the so-called *cadherin switch*, i.e., the switch of expression from epithelial E-Cadherin to mesenchymal N-Cadherin[105]. N-Cadherin (neural cadherin) belongs to the same Cadherin family as E-Cadherin, sharing a similar structure and forming the same type of homophilic interactions between cells. However, N-Cadherin is expressed by mesenchymal cells, neural cells and many types of cancer cells and acts as a weaker adhesion molecule, facilitating cell migration and invasion[106].

Moreover, mesenchymal cells have a different type of IF, not composed of cytokeratins as in epithelial cells, but mainly of Vimentin, a type III IF protein, which constitutes the major cytoskeletal component of mesenchymal cells. Changes in IF composition enable cell motility and favor invasion[107].

Thus, the *cadherin switch* (loss of E-Cadherin and expression of N-Cadherin) and the switch in the composition of IF (from cytokeratins to Vimentin) are the most

commonly used EMT markers during both normal development and metastatic progression, as shown in **Fig. 12**.

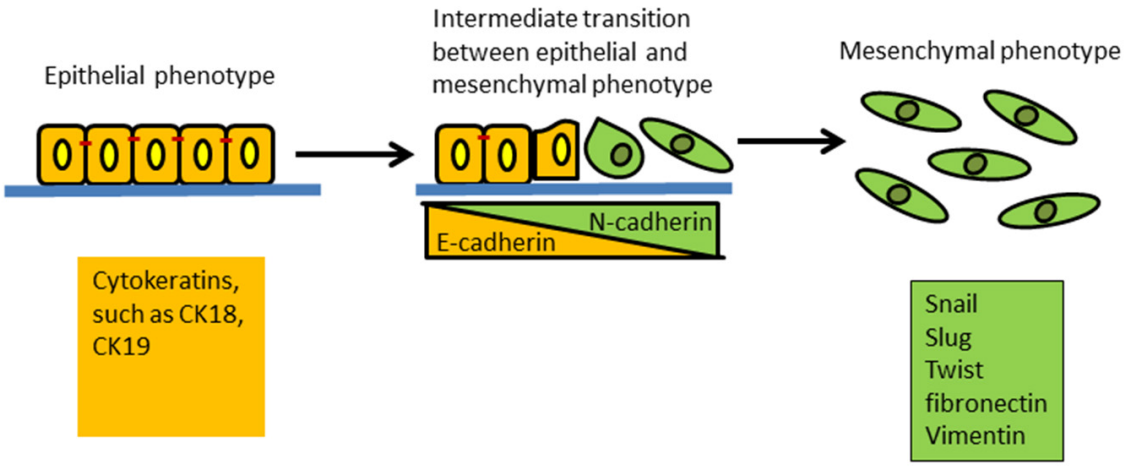


Fig. 12: The Cadherin switch and EMT markers.

Schematic illustration of the cadherin switch occurring during EMT: cells progressively lose E-Cadherin expression and acquire N-Cadherin expression. The process is accompanied by the loss of other epithelial specific proteins, such as cytokeratins, which compose the intermediate filaments (IF) of desmosomes, in favor of the expression of mesenchymal IF Vimentin. Other mesenchymal markers (fibronectin) and transcription factors (Snail, Slug, Twist) are also represented (see next paragraph for details). Figure taken from Wu *et al.*[108].

4.4 EMT master regulators and post-transcriptional mechanisms inducing EMT

EMT is a complex program that induces a drastic switch in cell differentiation and gene expression, down-regulating an epithelial gene expression signature, including those genes responsible for cell-cell junction complexes, and activating a mesenchymal gene signature. This process is mediated by key transcription factors, the so-called *EMT master regulators*, including SNAIL transcription factors (SNAI1,

also known as Snail, and SNAI2, also known as Slug), zinc-finger E-box-binding proteins (ZEB1 and ZEB2) and basic helix-loop-helix transcription factors, such as TWIST1 and TWIST2[109].

The function of these master regulators is finely regulated at the transcriptional, translational and post-translational levels. They functionally cooperate at target genes and often control the expression of each other. For example, ZEB expression often follows the activation of SNAIL, consistent with the fact that SNAI1, also in cooperation with TWIST1, directly activates ZEB1 transcription[110]. Together, the EMT transcription factors coordinate the repression of epithelial genes and the induction of mesenchymal genes, and frequently the same transcription factors direct both repression and activation[111].

An example of a major function of these EMT master regulators is repression of E-Cadherin expression, through recruitment of complexes that mediate epigenetic changes, specifically histone modifications at the promoter site. In particular, through the combination of repressive (mainly histone methylation) and active (mainly histone acetylation) marks, the E-Cadherin promoter is in a poised state as a '*bivalent domain'*, which typically occurs at many promoters in embryonic SCs in order to enable timely activation, while maintaining repression in the absence of further signals[112]. This bivalent control of the E-Cadherin promoter might contribute to the reversible nature of EMT[109].

Besides transcriptional factors driving EMT, post-transcriptional modifications also have a fundamental role in EMT progression. Alternative splicing indeed, allows the generation of proteins with structural and functional differences, for example, in the case of p120[113] or the adhesion cluster of differentiation 44 (CD44)[114], which is also used as a cancer stem cell (CSC) marker (see paragraph **5.2 Breast cancer stem cells: isolation and manipulation**). Moreover, non-coding

miRNAs selectively bind mRNAs encoding key proteins in EMT, thus inhibiting their translation or promoting their degradation. For example, miR-29b and miR-30a repress SNAI1 expression[115,116], whereas members of the miR-200 family repress the expression of ZEB proteins[117].

Finally, repression of the functions of epithelial proteins, mainly of E-Cadherin, is also achieved through protein re-localization. Indeed, several studies, have recently reported the so-called *partial EMT state* (also known as the 'P state' or 'hybrid epithelial/ mesenchymal', 'E/M' or 'intermediate EMT'), which is characterized by the concomitant presence of both epithelial and mesenchymal features[118,119]. In these cases, E-Cadherin is still expressed at normal levels, both at the transcriptional and protein level, however, it is re-localized in the cytosol and down-regulated at the PM, resulting in the same functional outcome as total loss of E-Cadherin[120]. Since, endocytosis has a major role in protein recycling and re-localization, finely regulating the composition of the PM in response to intra- and extra-cellular stimuli, it represents an alternative mechanism for the induction of EMT or partial EMT.

4.5 Signaling pathways involved in EMT

The reprogramming of gene expression during EMT are triggered and controlled by multiple signaling pathways that respond to extracellular cues, as shown in **Fig. 13**.

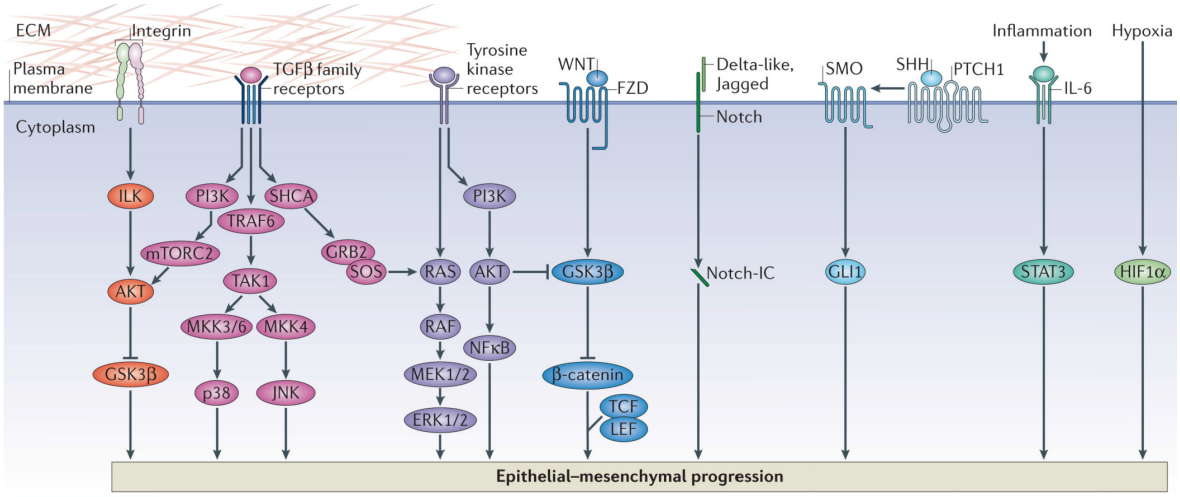


Fig. 13: Signaling pathways involved in EMT.

Schematic illustration of the major signaling pathways that can cooperate to induce EMT (see main text for more details). Figure taken from Lamouille *et al.*[109].

The major pathways involved in EMT are briefly described here, while a more detailed description of the canonical TGFβ signaling pathway and the WNT/β–catenin pathway is provided in the following two paragraphs (see **4.6 The canonical TGFβ signaling pathway in EMT induction**, and **4.7 The WNT/β–catenin pathway**). Note that the TGFβ pathway, besides inducing EMT through the canonical signaling pathway involving SMAD proteins, is also able to activate the PI3K-AKT, ERK MAPK, p38 MAPK and JNK pathways[121] (**Fig. 13**).

Growth factors that act through RTKs are able to induce EMT, such as EGF, fibroblast growth factor (FGF), HGF and VEGF. All these factors activate the RAS-RAF-MEK-ERK-MAPK signaling cascade[95] (**Fig. 13**).

The Hedgehog and Notch pathways are also able to induce EMT, activating the transcription factors SNAI1 through glioma1 (GLI1) or SNAI2 through the NCID, respectively[122,123] (**Fig. 13**).

Finally, the microenvironment also controls EMT: for instance, inflammation in cancer can promote interleukin-6 (IL-6) to induce EMT through the signal transducer and activator of transcription 3 (STAT3), which activates SNAI1[124]; instead, hypoxia induces the transcription factor hypoxia inducible factor 1 (HIF1α) which activates TWIST1[125] (**Fig. 13**).

4.6 The canonical TGFβ signaling pathway in EMT induction

Among the multiple signaling pathways, the TGFβ pathway has a predominant role in EMT, ECM remodeling, invasion and cell migration. Moreover, the TGFβ pathway has a dual action during tumorigenesis, characterized by tumor suppressor activity inhibiting cell proliferation at the early stages of tumorigenesis and tumor promoting activity at later stages.

Three homologous TGFβ genes exist in humans and their signaling is dependent on a heteromeric complex of two related transmembrane type I and type II receptors. The activation of the receptor complex leads to signaling through two different routes: the SMAD-dependent, or canonical signaling pathway, which plays a prominent role in EMT, and the SMAD-independent, or non-canonical pathways[126].

As shown in **Fig. 14**, in the canonical signaling pathway, TGFβ ligands bind TGFβ receptor type II (also known as TGFβR2) and TGFβ receptor type I (TGFβR1 or ALK5), which induces dimerization, cross-phosphorylation and activation. Activated TGFβ receptors are internalized and recruited to SARA endosomes[15,16]. This in turn leads to phosphorylation of receptor-specific SMAD proteins (R-

SMADs), such as SMAD2 and SMAD3, which bind the common mediator SMAD4 (co-SMAD) and then translocate to the nucleus, where they interact with other co-factors to regulate gene expression, inducing the transcription factors SNAI1, SNAI2 and ZEB1, thereby promoting EMT[127].

Inhibitory SMADS (I-SMADs), such as SMAD7, control R-SMADs and receptor ubiquitination through recruitment of the E3-ligase SMURF, targeting them to degradation and resulting in the inhibition of EMT activation[128,129].

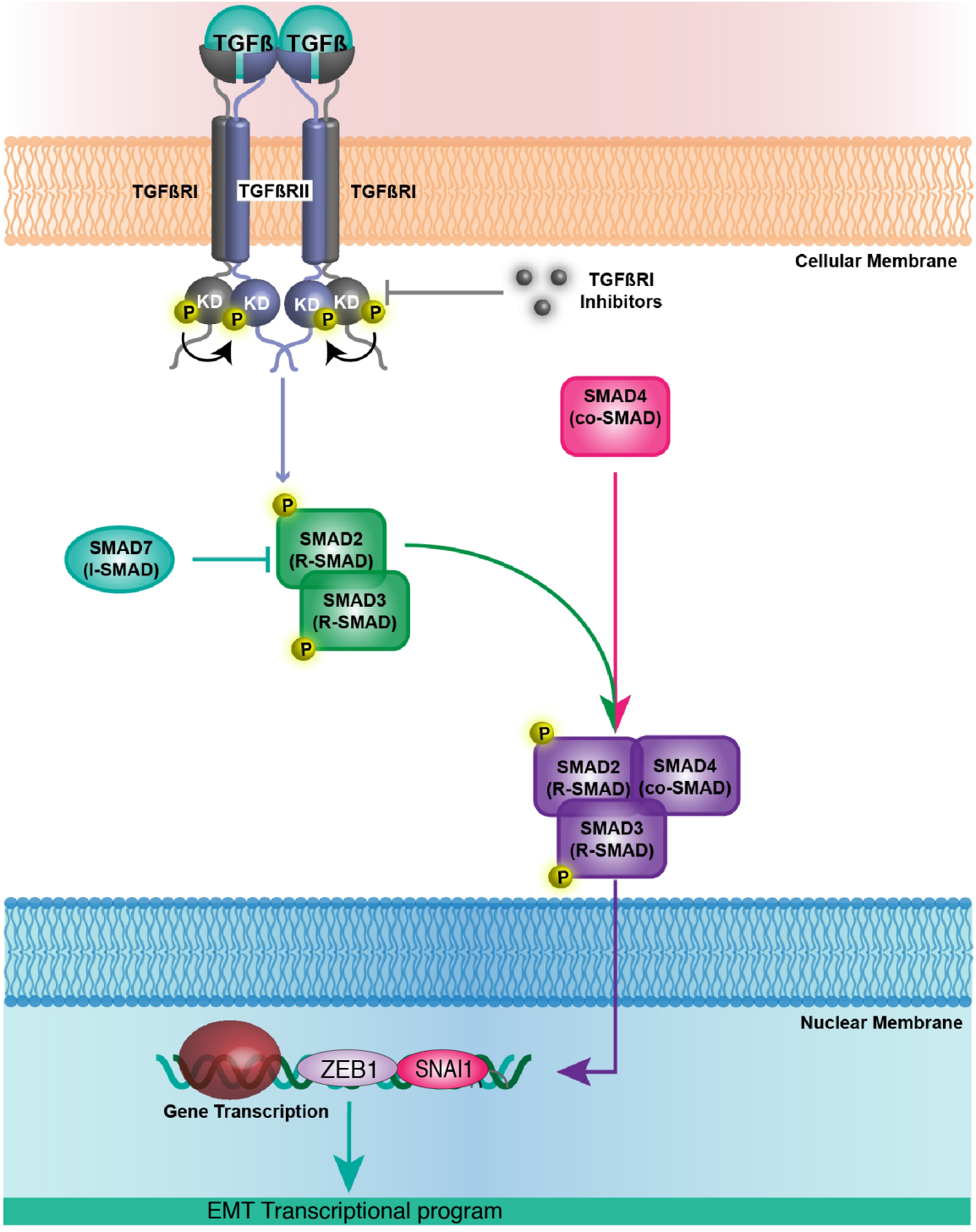


Fig. 14: TGFβ canonical signaling.

Schematic illustration of the TGFβ canonical signaling dependent on the phosphorylation and activity of SMAD proteins: depicted are R-SMADs, such as SMAD2 and SMAD3, co-SMAD (i.e., SMAD4), and the inhibitory SMAD (I-SMAD; i.e. SMAD7).
Figure adapted from www.mycancergenome.org/content/molecular-medicine/pathways/TGF-beta-signaling.

4.7 The WNT/β-catenin pathway

The WNT/β-catenin signaling pathway activates EMT during development or in cancer, through the regulation of gene expression mediated by β-catenin/TCF-LEF (T-cell factor/ lymphoid enhancer factor) transcription factors.

β-catenin is a protein with a dual function according to its subcellular localization. At the PM, it stabilizes cell adhesion by binding to the adherens junction proteins, i.e., E-Cadherin, α-catenin and δ-catenin (also known as p120). Through its binding with α-catenin, which in turn directly binds to actin filaments, β-catenin physically bridges the cadherins to the cytoskeleton[130]. In the absence of a stimulus, this is the main function of β-catenin inside the cell, at the junctions, while cytoplasmic β-catenin is kept at very low levels by the *destruction complex* through phosphorylation by CK1 (casein kinase 1) and GSK3β. Phosphorylated β-catenin is recruited by members of the destruction complex (APC and Axin), ubiquitinated by the E3-ligase β-TrCP and degraded via the proteasome[131].

However, in the presence of a stimulus, for example, in response to canonical WNT signaling, WNT binds to Frizzled receptors and the destruction complex is recruited at the PM, sequestering GSK3β in endosomes and preventing β-catenin phosphorylation. As a consequence, β-catenin accumulates in its dephosphorylated form and translocates to the nucleus, where it binds to the TCF-LEF transcription factors and behaves as a co-activator of target genes, including EMT related genes[95,132].

A schematic representation of the activation of β-catenin is shown in **Fig. 15**.

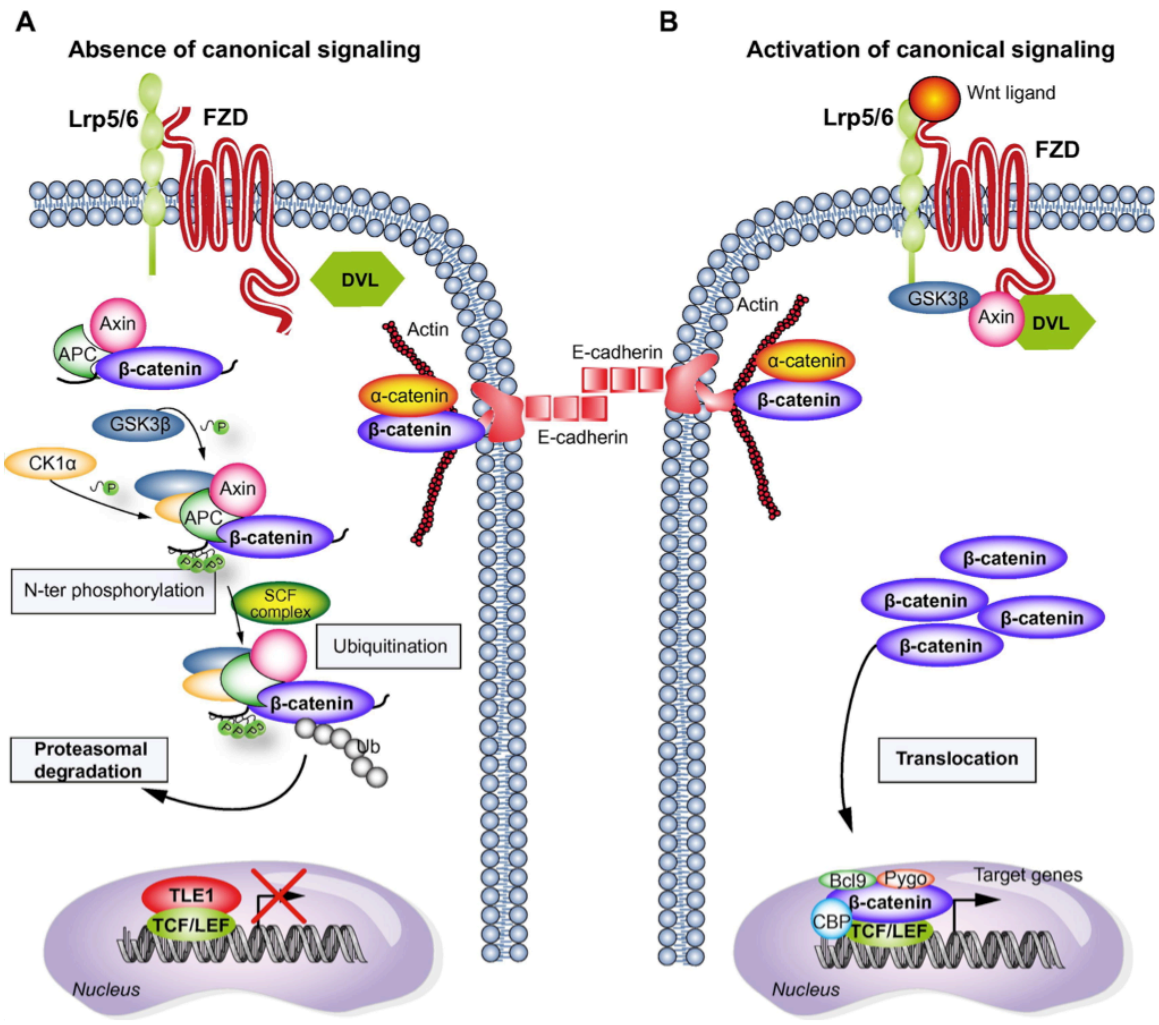


Fig. 15: The WNT/β-catenin signaling pathway.

Schematic illustration of the WNT/β-catenin signaling pathway and its different molecular players:
(A) In the absence of canonical signaling.
(B) In the presence of canonical signaling.
Subcellular localization of the proteins is also depicted (PM, cytosol, nucleus). Figure taken from pharmaceuticalintelligence.com/2015/04/10/targeting-the-wnt-pathway.

4.8 Endocytosis and EMT

Given the fundamental role of multiple signaling pathways in the activation of EMT, it is not surprising that endocytic circuitries are also involved in EMT regulation. The previously discussed TGFβ, WNT/β-catenin and Notch pathways involved in EMT activation and are indeed dependent on endocytic mechanisms. In particular, the final signaling outcome of these pathways is dependent on spatial constraints (i.e., the subcellular localization of the constituents of the signaling cascade) and on the specific endocytic trafficking route[133].

For example, the activation of TGFβ-dependent EMT is most likely dependent on the specific endocytic entry route used to internalize the receptor. Indeed, TGFβ receptors can be internalized via CME in SARA endosomes, a signaling-scaffold endosome in which internalized receptors encounter the targets SMAD2 and SMAD3, favoring their phosphorylation, thus triggering the activation of TGFβ canonical signaling pathway. In contrast, internalization of TGFβ receptors via caveolae, enriched with SMAD7 and SMURF2 proteins, results in rapid receptor turnover and attenuation of signaling[15]. Considering that in the canonical TGFβ signaling pathway phosphorylation of SMAD2 and SMAD3, binding to the common mediator SMAD4, translocation to the nucleus, leads to regulation of gene expression and induction of EMT transcription factors SNAI1, SNAI2, ZEB1[127], it is reasonable to assume that the specific endocytic entry route may have an impact on TGFβ-dependent EMT activation which indeed presumably requires the CME/SARA endosome entry route.

The activation of the WNT/β-catenin pathway, which controls EMT, also relies on endocytic regulation. Indeed, the sequestration in endosomal lysosomal

compartments of the Frizzled receptors, together with the destruction complex member, GSK3β, prevents its phosphorylation of β-catenin, which therefore can accumulate in its dephosphorylated form and translocate to the nucleus[17]. Direct evidence that endocytosis contributes to canonical WNT signaling comes from experiments in which pharmacological inhibition or genetic ablation of dynamin or CME prevents WNT-mediated increase of β-catenin levels[134,135,136]. Other lines of evidence suggest instead that CME acts in shutting off WNT signaling, while NCE via caveolae is responsible for the activation of the pathway[137]. Thus, while there are many indications of a role for endocytosis in WNT signaling, and therefore also in WNT-dependent EMT activation, the precise involvement of endocytosis is still unknown.

Finally, endocytosis is directly linked to one of the hallmarks of EMT: E-Cadherin loss (or down-regulation). Indeed, E-Cadherin endocytosis and trafficking control its turnover in mature epithelial tissues[138]. In agreement, *in vitro* studies on cultured epithelial cells showed that impairment of CME resulted in impairment of E-Cadherin recycling, demonstrated via impaired fluorescence recovery after photo-bleaching (FRAP)[139]. A key player in E-Cadherin endocytosis is p120 catenin (also known as δ-catenin), a member of the catenin superfamily, which binds to the cytoplasmic tail of E-Cadherin and inhibits its CME, stabilizing E-Cadherin at the PM. Mutations in the p120-binding site residues of E-Cadherin resulted in increased E-Cadherin endocytosis[140] and, accordingly, loss of p120 catenin also resulted in increased E-Cadherin endocytosis, cell-cell junction disruption, and is linked to EMT and invasiveness[141].

EMT-promoting factors, such as HGF, EGF, and SRC activation, lead to phosphorylation of E-Cadherin and association with the ubiquitin ligase, HAKAI

(*CBLL1*, Cbl proto-oncogene like1), which induces E-Cadherin ubiquitination and subsequent internalization and lysosomal degradation[142,143], thus, confirming that E-Cadherin endocytosis and degradation is directly involved in EMT activation.

E-Cadherin endocytosis and re-localization from the PM are relevant especially to the so-called *partial EMT state*, in which E-Cadherin is still expressed at normal levels, both at the transcriptional and protein level, however, through endocytosis it is re-localized from the PM to the cytosol, resulting in the same functional outcome as total loss of E-Cadherin. Moreover, a recent study has correlated the partial and complete EMT states with the different BC molecular subtypes using cell lines: luminal A, luminal B and normal-like subtypes correlated with a partial EMT state with persistent expression of E-Cadherin at the mRNA level and protein re-localization into the cytoplasm; in contrast, basal-like cells displayed loss of E-Cadherin expression, both at transcript and protein level. In the same study, the authors reported a similar scenario for pancreatic ductal adenocarcinoma (PDAC). In both cancer types these different EMT states were found to be associated with a different modality of cell migration: single cell invasion and migration in the case of complete EMT and collective migration in the case of partial EMT[120].

The partial EMT state has been proposed to be a more plastic condition compared with complete EMT, allowing cells to rapidly convert from an epithelial to a mesenchymal state (to allow cell invasion and migration) and *vice versa* (to allow the cell growth at mestastatic sites). Thus, it is tempting to speculate that this transition could occur via endocytosis, which, through the modulation of internalization and recycling rates, can rapidly change the E-Cadherin PM *vs.* intracellular levels, thereby acting as a primary mechanism for achieving cell plasticity.

5. Cancer stem cells

5.1 Mammary stem cells and breast cancer stem cells: origin and modality of division

MaSCs are adult tissue SCs that reside within the basal compartment of the mammary gland and, like all adult tissue SCs, are able to self-renew and to generate all the cell lineages of the mammary gland[66]. Indeed, experiments of *in vivo* transplantation in murine mammary fat pads cleared of endogenous epithelium, have demonstrated that MaSCs are capable of reconstituting the entire mammary gland[144].

The mechanism through which MaSCs are able to self-renew and differentiate is achieved by *asymmetric cell division* (ACD) of SCs, during which one daughter cell retains the SC identity (*self-renewal*), while the other becomes a proliferating progenitor cell that undergoes multiple rounds of cell division and generates differentiated progeny (*differentiation*). However, during development or under pathological conditions, SCs are able to promote rapid tissue generation through multiple rounds of *symmetric cell division* (SCD), where one SC gives rise to two daughter SCs, resulting in expansion of the SC pool[145].

Over the past decade, there has been much interest in the CSC hypothesis, which has challenged our traditional understanding of cancer development. According to this theory, tumors are generated from few tumor-initiating cells (TICs) able to generate all cells present in the tumor. These cells are also called CSCs because they share the two properties of normal SCs, self-renewal and differentiation potential[146].

However, despite sharing some functional characteristics with normal SCs, CSCs do not necessarily derive from normal SCs. Indeed, the origin of CSCs is still a debated issue and two main models have been proposed to explain their formation: the linear hierarchy model and the evolved CSC model, as shown in **Fig. 16**.

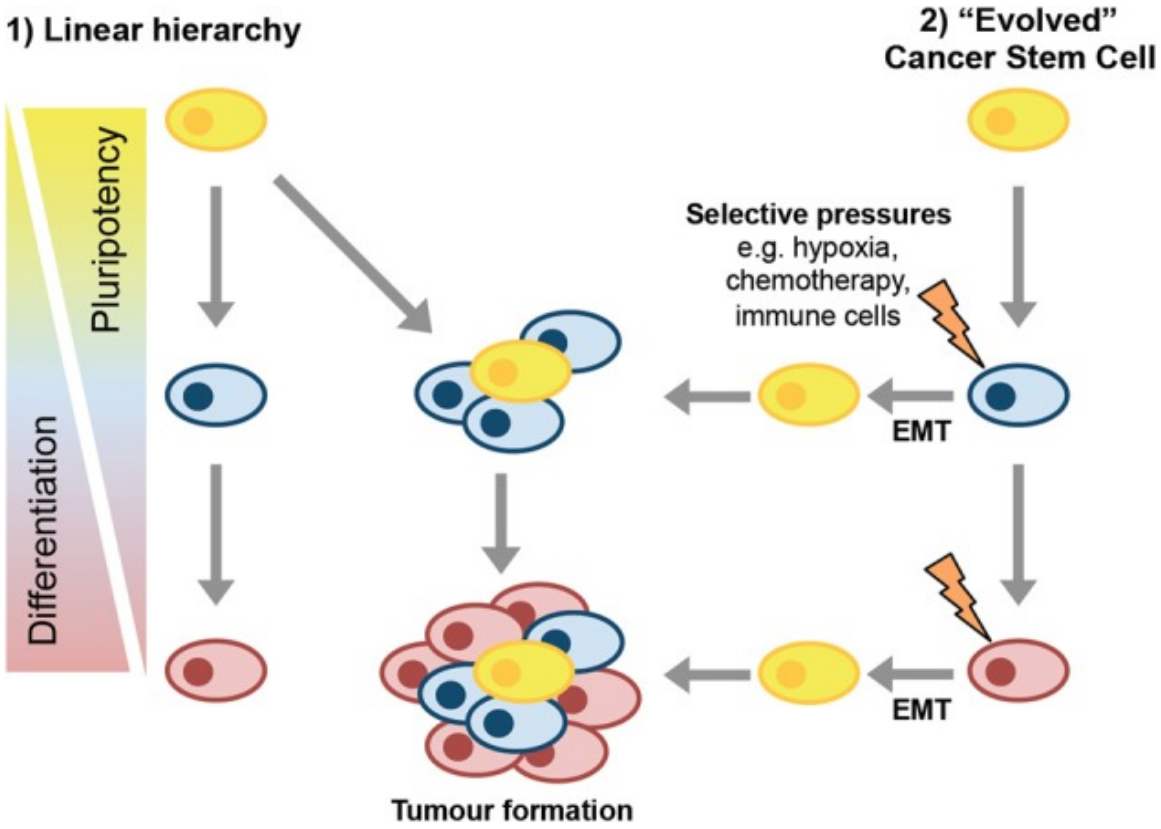


Fig. 16: The origin of CSCs.

Schematic illustration of the 2 main models that explain the origin of CSCs:
1) Linear hierarchy model: CSCs derive from normal SCs.
2) Evolved Cancer Stem Cell model: selective pressures generate CSCs from non-SCs, early progenitors or differentiated cells.
Figure taken from Owens *et al.*[147].

In the Linear Hierarchy Model, CSCs derive from normal SCs that acquire mutations and undergo oncogenic transformation, thus giving rise to differentiated progeny, i.e. all the cell types forming the tumor. This hypothesis is compatible with the fact that cells need to acquire many different mutations in multiple genetic loci in

order to reach full transformation. Thus, while the lifespan of differentiated cells would be too short for this to happen, long-lived SCs would be susceptible to the gradual accumulation of mutations. Instead, in the Evolved CSC Model, CSCs, rather than deriving from SCs, derive from differentiated cells as a response to genetic lesions and selective pressures from the tumor microenvironment (hypoxia, chemotherapy), which promote the acquisition of CSCs properties through de-differentiation and EMT mechanisms (see paragraph **5.3 EMT and CSCs: two sides of the same coin**). This second model is compatible with the fact that the degree of intratumoral heterogeneity cannot be easily explained by the simple Linear Hierarchy Model nor can the observation that relapsed tumors are more aggressive than the original tumor. Indeed, if tumors originated only through the Linear Hierarchy Model, even though chemotherapy would not be able to kill CSCs (see paragraph **5.4 CSCs: drug resistance and tumor relapse**), the remaining CSCs should give rise to a similarly-sensitive progeny. Since tumor relapse generally generates tumors that are more aggressive than the original one, it seems likely that tumors evolve mechanisms to adapt to the tumor microenvironment, self-sustain and generate CSCs from non-SCs. Therefore, the origin of CSCs still remains a debated issue and, most probably, both SCs and non-SCs contribute to their generation through different mechanisms which involve acquisition and accumulation of mutations.

In the case of BC, breast CSCs (BCSCs) were identified originally by Al-Hajj *et al*. by the use of cell surface markers, $CD44^{+}/CD24^{-/low}/Lin^{-}$. In particular, as few as 100 cells with this profile, isolated from the bulk tumor mass of BC patients, were able to form tumors in immunocompromised mice, whereas ten thousand counterpart cells failed to do so. Furthermore, the tumors formed could be serially passaged (meaning they had the ability to self-renew) and give rise to the entire heterogeneity of original tumor (ability to differentiate)[148].

Moreover, it has been demonstrated that BCSCs undergo SCD, rather than ACD, more frequently than their normal counterparts, leading to expansion of the CSC pool[149]. Therefore, elucidating the mechanisms that determine the mode of MaSC division (ACD *vs.* SCD) is crucial for understanding BC biology and for the development of therapeutic strategies able to restore normal cell division kinetics, thereby resulting in eradication of CSCs.

5.2 BCSCs: isolation and manipulation

Since their identification, many research groups have focused their efforts on the identification of "stem-like" profiles, based on cell surface protein expression, including BCSCs. However, the heterogeneous nature of BC makes the existence of a universal SC surface marker profile very unlikely. Indeed, different marker profiles are currently used to enrich for and analyze BCSCs, according to the molecular subtype.

As previously described, the first BCSC profile identified was the $CD44^{+}/CD24^{-/low}/Lin^{-}$ [148]. Subsequent studies have shown that this configuration is particularly suitable for enriching BCSCs from basal-like, HER2-negative and BRCA-null BCs[150] and from many breast tumorigenic cell lines[151], as well as normal MaSCs[152]. CD44 is the cell surface receptor for ECM hyaluronic acid[153]; CD24 (also known as heat stable antigen – HAS – or Nectadrin) is a cell surface glycoprotein involved in cell adhesion[154]; ESA (or EpCAM) is the epithelial specific antigen or epithelial cell adhesion molecule, a transmembrane glycoprotein mediating Ca^{2+}-independent homotypic cell-cell adhesion in epithelia[155]; Lin^{-} instead means not

expressing lineage markers that are normally expressed by mature hematopoietic cells[156].

Another CSC profile involving CD44, but not CD24, is that of $CD44^+/CD49f^{high}/CD133^+$. CD49f and CD133 are associated with normal MaSCs, but in combination with $CD44^+$ they have been shown to enrich for BCSCs[157]. CD49f, also known as integrin α6, is a member of the integrin family of transmembrane receptors that facilitate cell-ECM adhesion[158]. CD133, also known as prominin-1, is a glycoprotein involved in the organization of cell membrane topology[159].

Lastly, a widely adopted, non-cell surface marker of BCSCs is aldehyde dehydrogenase (ALDH). ALDH is an enzyme involved in detoxification of aldehydes and its activity has been shown to be elevated in normal MaSCs and BCSCs[160].Thus, the assay, known as the Aldefluor assay, measures ALDH activity.

Interestingly, the two cell populations isolated by the most commonly used configurations ($CD44^+/CD24^{-/low}$ and $ALDH^+$) show very little overlap. It has been proposed that these configurations identify distinct BCSC populations characterized by EMT and MET gene-expression profiles, respectively (see paragraph **5.5 CSCs, EMT-MET and metastasis**), reflecting the distinct profiles of basal and luminal SCs found in the normal breast[161].

Given the high level of heterogeneity in immunophenotypical markers, assays that exploit the functional properties of SCs, rather than their physical characteristics have been developed. In particular, Dontu *et al.* have developed the so-called mammosphere assay. This assay exploits the fact that MaSCs are normally quiescent or slowly proliferating, and, in contrast to progenitors, are able to survive *anoikis* and proliferate in anchorage-independent and serum-free conditions, to give rise to clonal spheroids (called mammospheres) that originate from single SCs[162]. Therefore, in

this assay, the number of spheres formed can be considered a proxy for the SC content of the initial cell population.

Building on the ability of MaSCs and BCSCs to form mammospheres, Cicalese *et al.* and Pece *et al*. developed a mammosphere label-retaining assay[149,163], using the lipophilic dye, PKH26, previously used for hematopoietic SC tracking[164]. This dye irreversibly binds to the lipid bilayer of cell membranes and is equally partitioned at mitosis into the two daughter cells. Thus, the PKH26 fluorescence intensity decreases in an exponential manner at each cycle of division. Therefore, MaSCs, which usually undergo one round of ACD, retain the dye, whereas progenitor cells, which continue to divide, progressively lose the dye. Consequently, MaSCs can be isolated by FACS-sorting of $PKH26^+$ cells and performing the mammosphere assay with $PKH26^+$ cells, coupled with time-lapse video-microscopy, allows the prospective determination of SC number. Instead, in SCD, both daughter cells retain the same amount of dye, which continues to be diluted at each division. A schematic representation of PKH26-labelled SCs undergoing ACD and SCD is shown in **Fig. 17**.

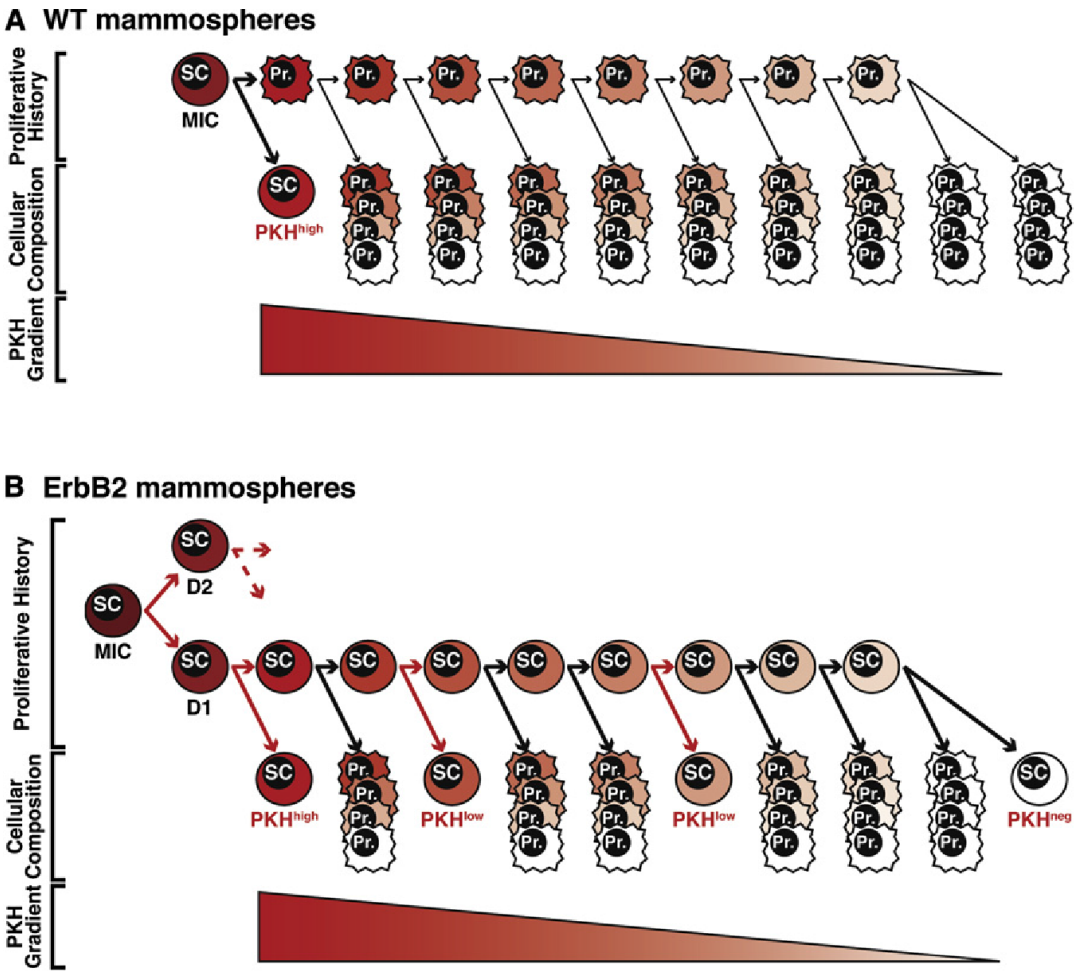


Fig. 17: MaSCs and PKH26 label retention.

Schematic illustration of SCs/CSCs and their modality of division and the ability to retain the PKH26 dye.

(A) In WT mammospheres normal SCs undergo ACD giving rise to a quiescent SC, which retains the highest level of PKH26 dye, and progenitors, which undergo rapid proliferation, diluting the dye at each cell division.

(B) In cancer, for example, in ErbB2-overexpressing mammospheres, CSCs can also undergo SCD, leading to expansion of the SC pool, each of which can undergo ACD or SDC, diluting the PKH26 dye, and generating a gradient of $PKH26^{high}$, $PKH26^{low}$, $PKH26^{neg}$ SCs and progenitors.

Figure taken from Cicalese *et al.*[149].

Finally, the gold standard assay to evaluate BCSC self-renewal and expansion is the *in vivo* transplantation assay at limiting dilutions, through injection of CSCs into mammary fat pad of immunocompromised mice and subsequent tumor formation or regeneration of the mammary gland in the case of injections into the cleared mammary fat pad[66,144].

5.3 EMT and CSCs: two sides of the same coin

As anticipated previously, EMT has a fundamental role in the de-differentiation of cells and acquisition of CSCs properties.

In particular, Mani *et al*. demonstrated that EMT *generates* cells with CSC properties. They used the non-tumorigenic HMLE cell line, which they previously derived by immortalization of primary human mammary epithelial cells (HMEC)[85]. These cells are widely used in EMT studies because they contain a relatively large subpopulation (~5%) of mesenchymal and SC-like cells, and show a high grade of plasticity upon external perturbations[165]. They showed that, epithelial cells that had undergone EMT by different mechanisms (i.e., upon overexpression of EMT transcription factors, Snail or Twist, or prolonged treatment with TGFβ) also expressed CSCs markers and were able to form mammospheres or colonies in soft agar and, importantly, had increased tumorigenic potential *in vivo*.

After this original discovery, several research groups confirmed the strict relationship between EMT and CSCs. Indeed, the same marker used to identify BCSCs, $CD44^+$, is actually a cell adhesion molecule, acting as a cell surface receptor for ECM hyaluronic acid. Alternative splicing generates several variant forms of CD44 (CD44v), which have been shown to be directly involved in EMT and invasion, and to correlate with increased malignancy and metastatic progression[114]. Moreover, CD44 is a β-catenin/TCF-4 target gene[166] and its splicing is dependent on a CD44/ZEB1 feedback loop[167], linking the EMT pathways of WNT and TGFβ to CSC maintenance. Finally, during physiological EMT (Type I) occurring in development, there is a close link with SCs, since the cells that undergo EMT in gastrulation, giving rise to the mesoderm, are pluripotent SCs of the epiblast[95].

The discovery that differentiated cells can generate CSCs, either, spontaneously, at a very low frequency rate, or through EMT or oncogenic transformation at a significantly higher rate[165], implies that de-differentiation is possible and that the Linear Hierarchical Model of CSC origin should take into consideration bidirectional interconversions.

Moreover, besides theoretical implications, the discovery that cells possessing mesenchymal properties, such as increased ability to invade and migrate, also have CSCs properties, such as tumor-initiating properties and the ability to self-renew and differentiate, has many important biological implications, including in the invasion-metastasis cascade (see paragraph **5.5 CSCs, EMT-MET and metastasis**).

5.4 CSCs: drug resistance and tumor relapse

CSCs are widely believed to be responsible for therapy failure and tumor recurrence, also known as tumor relapse.

Indeed, the majority of current chemotherapy strategies are mainly aimed at shrinking tumor mass by targeting the rapidly proliferating bulk tumor cells through the use of agents inhibiting the cell cycle, either by interfering with DNA replication (S-phase) or with mitosis (M-phase). For example, paclitaxel is a cytoskeletal drug that targets tubulin, therefore, paclitaxel-treated cells have defects in mitotic spindle assembly, chromosome segregation, and cell division. This blocks the progression of mitosis and prolongs activation of the mitotic division[168]. 5-Fluorouracil (5-FU), instead, triggers apoptosis or reversion to the G-phase of the cell cycle; it is a

thymidylate synthase (TS) inhibitor that blocks the synthesis of the pyrimidine thymidine, one of the four nucleosides required for DNA replication[169].

According to the CSC hypothesis, CSCs are able to escape and resist chemotherapy because of their slow proliferating nature. However, the mechanisms underlying CSC drug resistance are more complex and involve up-regulation of many specific proteins. Indeed, BCSCs, like adult tissue SCs have increased drug efflux activity[170] or enhanced ATM-dependent DNA-damage repair response[171]. This is not surprising since adult SCs are cells capable of living an entire lifetime, dividing when needed, and therefore cannot afford to be damaged. Moreover, BCSCs, similar to the normal MaSCs, have increased expression of free radical scavenging systems that maintain low levels of reactive oxygen species (ROS), which in turn results in less DNA damage and protection from radiotherapy[172]. Lastly, the BCSC marker, ALDH, is an enzyme involved in detoxifying aldehydes[173] and is directly involved in drug resistance[174]. In all cases, the resistance of CSCs to chemotherapy/radiotherapy is a possible mechanism to explain cancer recurrence.

In conclusion, the growing evidence that CSCs have a pivotal role in cancer initiation, progression, drug resistance and tumor recurrence imposes the need for the eradication of CSCs for efficient and long-term disease-free survival of cancer patients. Therefore, the development of anti-CSC therapeutic strategies is now a priority. In order to achieve this, it is crucial that we increase our understanding of pathways active in CSCs and identify CSC-specific molecular targets.

5.5 CSCs, EMT-MET and metastasis

Besides drug resistance and tumor relapse, CSCs are also involved in metastasis of different cancers, including BC[175,176]. This is particularly relevant given the fact that the majority of deaths caused by BC are due to secondary lesions that have disseminated from the original tumor, i.e., metastatic BC.

The metastatic process in cancer can be divided in 5 steps: invasion, intravasation, circulation, extravasation, and colonization, as shown in **Fig 18**.

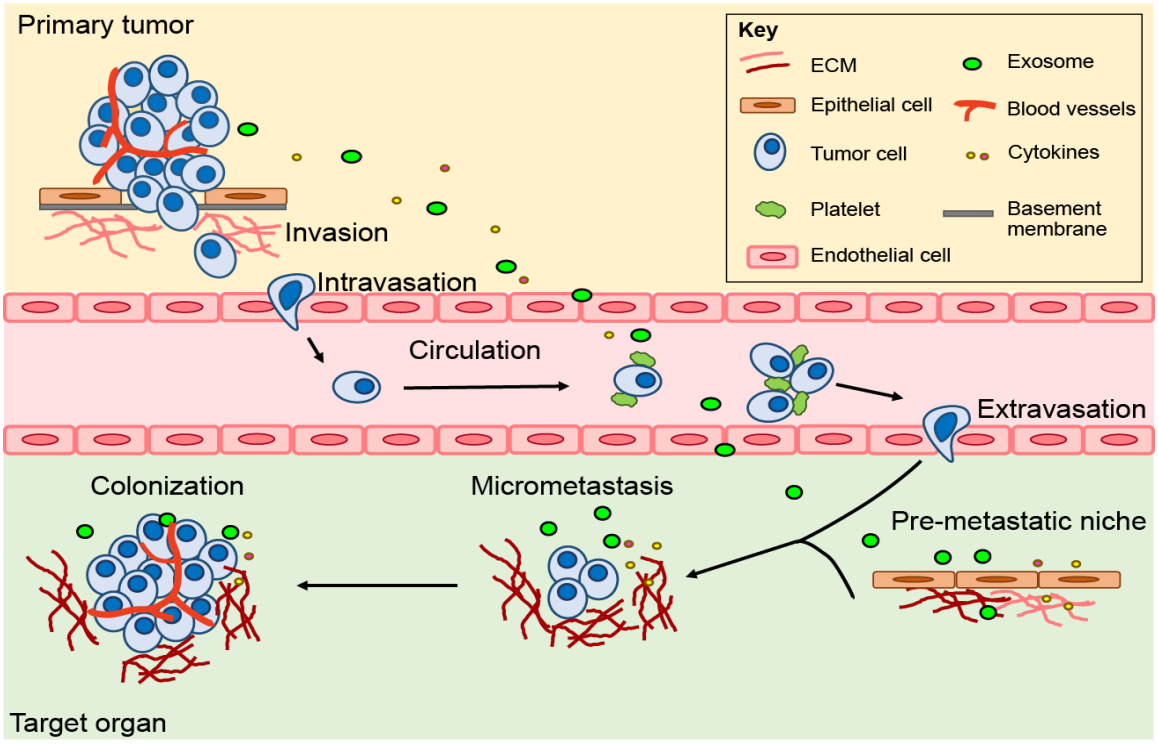


Fig. 18: The metastasis cascade.

Schematic illustration of the metastasis cascade and its steps of invasion of the extracellular matrix (ECM), intravasation, circulation, extravasation and colonization. Key components are color-coded and explained in the box in the top right corner.
See main text for a detailed explanation.
Figure taken from Gomez-Cuadrado *et al.*[177].

In the first step of the metastatic process, cancer cells detach from the original tumor site by acquiring the ability to invade the ECM that separates the tumor from adjoining tissues. Then, they migrate to other parts of the body via the bloodstream or the lymphatic system, both processes requiring an intravasation step. The final destination of the cells and the formation of the metastases is not random, with different types of cancer tending to spread to particular organs and tissues at a rate that is higher than expected by statistical chance alone. BC, for example, tends to metastasize to the bones and lungs. This specificity seems to be mediated by soluble signaling molecules, such as cytokines, chemokines and TGFβ[178]. Once the final destination is reached, the extravasation step is followed by the colonization step, leading to the formation of a distant tumor, similar to the one at the original site.

It is now well-established that the EMT program is one of the major mechanisms responsible for the first steps of invasion and migration, where epithelial tumor cells lose their cell-cell adhesion and apico-basal polarity and acquire the ability to invade the ECM and to migrate far from the original site of the tumor, through migration of single mesenchymal cells[92,109]. However, the last step of colonization requires MET, in order to regain epithelial characteristics and form the secondary tumor or macrometastasis. On note, EMT and MET enable solid tumor formations, 90% of which are of epithelial nature (carcinomas)[179].

The colonization step implies that metastasizing cells must have effective-tumor initiation and differentiation properties in order to fuel metastatic tumor growth and form highly heterogeneous tumors, similar to the original one. This means that migrating cells must have CSC properties, suggesting once again that EMT/MET and CSCs are strictly related. A recent study by Liu *et al.* has shown that BCSCs in an EMT-like state ($CD44^{+}/CD24^{-/low}$) are primarily quiescent and localized at the tumor invasive front, while those in a MET-like state ($ALDH^{+}$), are proliferating cells

located mainly at the center of the primary tumor and displaying high levels of cellular plasticity[161].

In conclusion, a complete metastasis cascade necessitates that tumor cells should have a high degree of plasticity switching between epithelial to mesenchymal states. According to the recent *partial EMT* theory, the E/M cells, which are characterized by the concomitant presence of both epithelial and mesenchymal features, are the ones with the highest plasticity able to switch between invasive and proliferative modes[180]. The result of partial EMT is, in fact, the mixture of epithelial (e.g., adhesion) and mesenchymal (e.g., migration) properties. Consequently, instead of a single-cell migration, cells move collectively as clusters. Indeed, experimental evidence shows that if these clusters reach the bloodstream intact, they can give rise to and be detected as clusters of circulating tumor cells (CTCs), which are enriched in CSCs and can resist *anoikis* in the bloodstream[119,181].

5.6 Endocytosis, asymmetric cell division of SCs and CSCs

Endocytosis is a critical regulator of many cellular processes, including proliferation, migration, polarity, fate determination, and maintenance of homeostasis of the SC compartment[1]. Indeed, several lines of evidence suggest an involvement of endocytosis in the regulation of ACD of SCs. During ACD, a single mitotic event gives rise to two different daughter cells: one daughter cell retains the SC identity (*self-renewal*), while the other becomes a proliferating progenitor cell that undergoes multiple rounds of cell division and generates differentiated progeny (*differentiation*). This is achieved through asymmetric partitioning of the so-called

cell fate determinants, molecules that instruct the cell that inherits them to adopt a certain identity[182].

Endocytosis plays an important role in asymmetric partitioning of cell fate determinants. For example, the differential activation of the Notch signaling pathway in ACD is achieved through an asymmetric regulation of the endocytic trafficking of both Notch DSL ligands and receptor molecules[26]. This is due to the asymmetric partitioning of the cell fate determinant and endocytic protein, Numb, which once asymmetrically partitioned is able to regulate cell fate by counteracting the Notch pathway in the SC in mammals (in contrast, in *Drosophila*, Numb inhibits neuroblast self-renewal and promotes differentiation)[27,183,184]. Errors in asymmetric partitioning and in cell fate determinant distribution can lead to expansion of the SC population and consequent over-proliferation in tumors: for example, mutation of cell fate determinants in *Drosophila* neuroblasts results in no daughter cell being driven into differentiation and, thus, two SCs are formed[182].

Moreover, in MaSCs, besides its role in counteracting Notch signaling pathway, the presence of Numb has been shown to be critical for maintaining normal ACD of MaSCs by regulating the modality of SC division through the formation of a trimeric complex with p53 and the E3-ubiquitin ligase MDM2. In this tricomplex, Numb prevents MDM2-mediated ubiquitination and consequent degradation of p53[185,186]. The tumor suppressor protein, p53, is a key determinant in the regulation of modality of division in MaSCs. Indeed, loss of p53 (or p53 mutation) favors the skewing of SC division from ACD to SCD, leading to the expansion of the CSC compartment and fueling tumor growth[149]. In agreement, loss of Numb results in an increase in p53 ubiquitination and degradation and therefore a similar biological outcome to p53 loss. Indeed, Numb expression is frequently attenuated in tumors: one third of BCs are Numb-deficient and characterized by poor prognosis[187]. Importantly, restoration

of Numb or p53 levels in Numb-deficient/p53-wildtype BCs, has been shown to be an effective SC-targeted strategy. In pre-clinical validation studies using patient-derived xenografts of BCs, the combination of the inhibitor Nutlin-3a that restores p53 levels and the chemotherapy drug, paclitaxel, successfully induced tumor growth inhibition or regression and prevented CSC-driven tumor relapse[188].

Finally, as described above (see paragraph **4.8 Endocytosis and EMT**), endocytosis has also been implicated in the activation of EMT through the TGFβ and WNT signaling pathways[95,127,132,189] and E-Cadherin trafficking[120,133]. Since EMT generates cells with CSC properties[165], alterations in endocytosis could represent a mechanism underlying the emergence of CSCs.

PRELIMINARY DATA AND AIM OF THE PROJECT

1. Preliminary data

In agreement with the emerging role of endocytosis in cancer, in our laboratory, we recently uncovered an oncogenic role for the endocytic protein Epn3 in BC. By screening a large cohort of BC patients, we found that the *EPN3* gene is amplified in ~10% of cases, while Epn3 protein is overexpressed in ~50% of BCs. Epn3 amplification/overexpression correlated with poor prognosis and distant metastasis and with increased *in vivo* tumorigenic potential in orthotopic transplantation experiments in NOD (non-obese diabetic) - SCID (severe combined immunodeficiency) IL2RG null mice, also known as NSG mice.

Here, I summarize the preliminary data representing the starting point of my thesis project.

1.1 The *EPN3* gene is amplified and overexpressed in breast cancer

The *EPN3* gene is located on chromosome 17, in region 17q21, about 11 Mb apart from the most frequently amplified BC oncogene *ERBB2* (also known as HER2) (**Fig. 1A**). A preliminary survey of a cohort of 1080 human BCs included in The

Cancer Genome Atlas (TCGA) database (http://cancergenome.nih.gov) revealed that the *EPN3* gene is potentially amplified in ~8% of BCs (**Fig. 1B left panel**), as opposed to only 2% for *EPN1*. To confirm this finding, we obtained direct evidence of *EPN3* gene amplification by fluorescent *in situ* hybridization (FISH) analysis on a cohort of 219 BC patients from the European Institute of Oncology (IEO), previously characterized in a clinical study of Veronesi *et al.*[190]. In particular, tissue microarrays (TMAs) from these BCs were available and were also previously used in other studies in the lab[185 ,191]. EPN3 was considered amplified when the *EPN3/CEP17* (centromere of chromosome 17) ratio was ≥ 2.5 and ERRB2 was considered amplified when the *ERRB2/CEP17* ratio was ≥ 2.0 according to Wolff *et al.*[192]. With this analysis, we found that the *EPN3* gene was amplified in ~10% of cases (21/219) and *ERBB2* was found to be amplified in 18% of the same cohort (39/219). Interestingly, despite the close proximity of the *EPN3* and the *ERBB2* genes, *EPN3* amplification can occur independently of *ERBB2* amplification: approximately half of the *EPN3*-amplified cases were negative for *ERBB2* amplification (11/21). Similarly, *ERBB2* amplification can occur independently of *EPN3* amplification, with co-amplification occurring in only 28% of *ERBB2*-amplified cases (11/39) (**Fig. 1B middle panel**).

Moreover, immunohistochemistry (IHC) analysis for Epn3 protein expression was performed on the same cohort of 219 BC patients and the results were compared with the FISH analysis (however, reliable results were obtained for 212 cases). Results revealed that: i) almost in all cases when the *EPN3* gene was amplified, it was also overexpressed (19/21); ii) Epn3 was overexpressed in 47% of the cases (100/212); iii) interestingly, in more than one third of all cases (81/212) Epn3 is overexpressed in the absence of gene amplification and accordingly, when Epn3 is overexpressed, in the high majority of the cases it is not amplified (81/100) (**Fig. 1B right panel**).

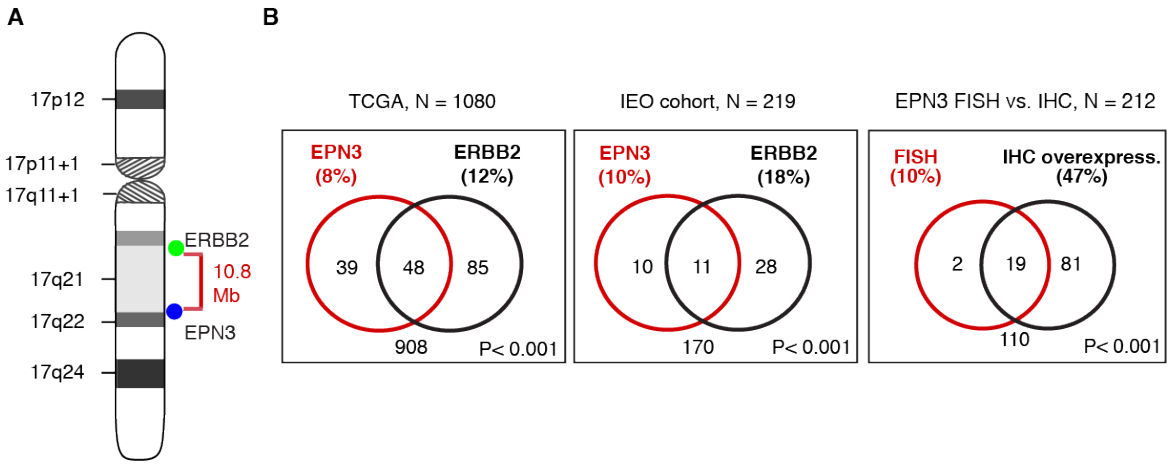


Fig. 1: The *EPN3* gene is amplified and overexpressed in BC.

(A) Scheme of human chromosome 17: the *EPN3* gene is located on chromosome 17q arm, ~10 Mb apart from the *ERBB2* gene (*left panel*).

(B) Venn diagrams showing *EPN3* and *ERBB2* gene amplification in the TCGA cohort of 1080 BCs (*left panel*) and in the IEO cohort of 219 BC samples analyzed by fluorescent in situ hybridization (FISH) (*middle panel*). Finally, EPN3 gene amplification *vs.* protein overexpression was evaluated by comparison of immunohistochemistry (IHC) and FISH analysis on the IEO cohort (N=212 evaluable cases) (*right panel*).

1.2 Epn3 correlates with aggressive clinical/pathological parameters and with distant metastasis

To investigate the potential prognostic significance of Epn3 overexpression, we extended our analysis to a large consecutive cohort of >2400 BCs (N=2453) of patients who underwent treatment at IEO between 1997-2000 (IEO BC 97-00 cohort). In this cohort, we found that Epn3 overexpression at the protein level was detectable in 37.6% of cases (**Table 1**), largely confirming results obtained with the 219-patient cohort (**Fig. 1**).

Epn3 overexpression was found to correlate with poor prognosis and clinical/pathological parameters indicative of aggressive disease. Indeed, high Epn3 levels correlated with increased tumor size (pT), positive nodal status (pN), high-grade tumors (grade), negative ER and PgR status, high proliferative index (Ki-67), and positive HER2 status. Moreover, Epn3 overexpression was significantly associated with the more aggressive molecular subtypes of BC, such as Luminal B, HER2-positive and triple-negative BCs (**Table 1**).

Finally, Epn3 overexpression correlated with increased risk of distant metastasis, rather than with loco-regional relapse, both at the protein and the mRNA level (**Fig. 2A** and **Fig. 2B** respectively) (Iavarone *et al.*, in preparation).

Variable	LOW N (%)	HIGH N (%)	Univariate analysis P	Multivariable analysis P
All	1133 (62.4)	684 (37.6)		
Histology				
Ductal	880 (58.7)	620 (**41.3**)	**<.01**	**<0.01**
No Ductal	253 (79.8)	64 (20.2)		
pT				
pT1a/b	123 (73.7)	44 (26.3)	0.01	0.09
pT1c	581 (60.7)	376 (39.3)	0.51	<.01
pT2	391 (62.4)	236 (37.6)		
pT3/pT4	38 (57.6)	28 (42.4)	0.45	0.33
pN				
pN0	555 (66.3)	282 (33.7)		
pN1-2-3	548 (58.0)	397 (**42.0**)	**<.01**	**<0.01**
pNX	30 (85.7)	5 (14.3)	0.02	0.10
Grade				
1	254 (84.1)	48 (15.9)		
2	543 (70.1)	232 (29.9)	<.01	0.05
3	310 (44.0)	394 (**56.0**)	**<.01**	**<.01**
NA	26 (72.2)	10 (27.8)	0.08	0.10
ER/PgR				
Non expressed (Both 0)	116 (42.8)	155 (**57.2**)	**<.01**	0.07
Expressed (ER>0 or PgR>0)	1017 (65.8)	529 (34.2)		
ERBB2 Status				
NEG	1028 (65.6)	539 (34.4)		
POS	85 (38.1)	138 (**61.9**)	**<.01**	**<.01**
NA	20 (74.1)	7 (25.9)	0.36	0.19
Ki-67				
<14%	404 (83.5)	80 (16.5)		
>=14%	728 (54.7)	603 (**45.3**)	**<.01**	**<.01**
NA	1 (50.0)	1 (50.0)	0.25	0.73
Subtype 2013				
LUM A	258 (83.5)	51 (16.5)		
LUM B ERBB2-NEG	677 (63.6)	388 (36.4)	**<.01**	Not included
LUM B ERBB2-POS	52 (41.3)	74 (58.7)	**<.01**	
ERBB2-POS NO LUM	31 (34.4)	59 (65.6)	**<.01**	
TN	83 (47.2)	93 (52.8)	**<.01**	
NA	32 (62.7)	19 (37.3)	**<.01**	

Table 1: Epn3 overexpression correlates with clinical/pathological parameters indicative of aggressive disease.

The correlation between Epn3 expression, evaluated by IHC, and clinical/pathological parameters was evaluated in the IEO BC 97-00 cohort, using univariate and multivariate logistic regression models. The p-values are shown (P) and marked in bold when significant (<0.01). The number of patients (N) and the relative percentages in parentheses are reported for each parameter according to Epn3 status: Low or High based on Epn3 expression levels (Epn3-Low IHC score ≤1.0, Epn3-High IHC score >1.0). pT: primary tumor size; pN: nodal status; ER: estrogen receptor; PgR: progesterone receptor; Ki-67: proliferation index; LUM: luminal; TN: triple negative; NA: not available.

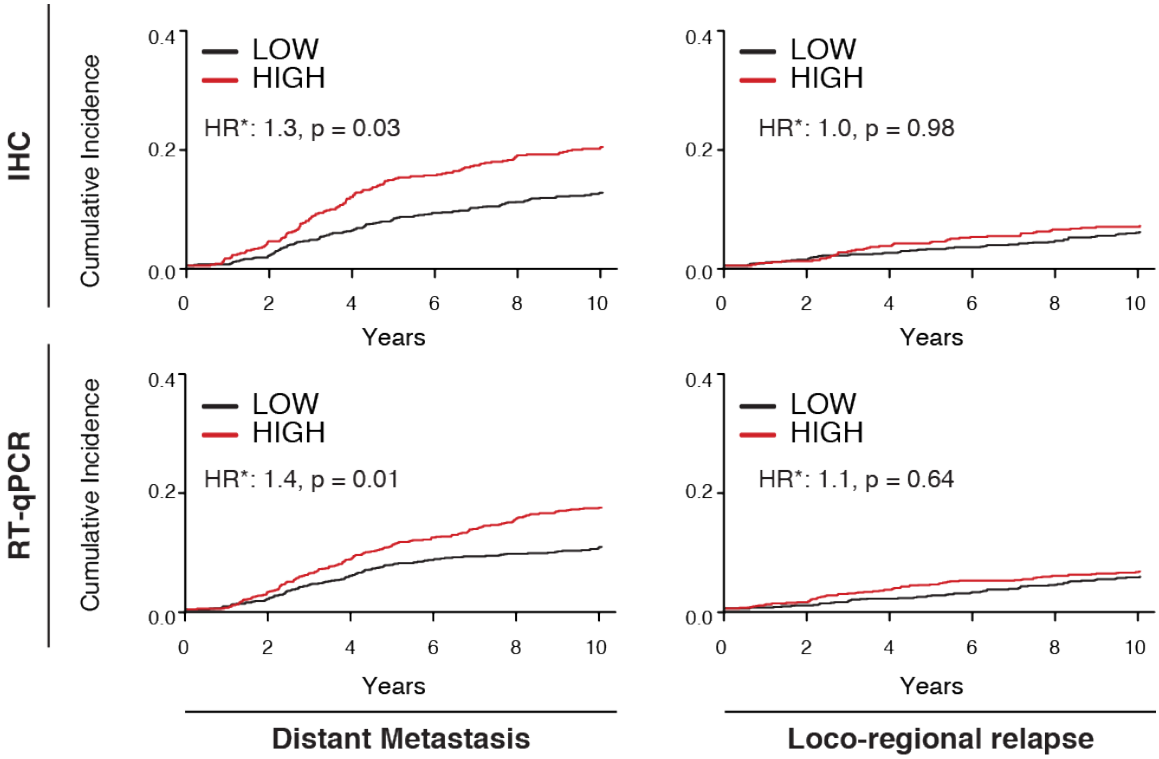


Fig. 2: High levels of Epn3 expression correlate with increased risk of distant metastasis.

Cumulative incidence probability of distant metastasis (*left graphs*) and of loco-regional relapse (*right graphs*) over a 10-year follow-up period in patients from the IEO BC 97-00 cohort. Patients were stratified according to low (*black curve*) *vs.* high (*red curve*) levels of Epn3 expression, both at protein level (*upper panels*) and at the mRNA level (*lower panels*).
At the protein level, Epn3 was defined Low or High based on IHC score (Epn3-Low IHC score ≤1.0, Epn3-High IHC score >1.0). At the mRNA level, Epn3 was defined Low or High according to qPCR threshold cycle (Ct) (Epn3-Low Ct ≥ 28.18, Epn3-High Ct < 28.18).
HR*, multivariable hazard ratio; p, Wald test P-value.

1.3 Epn3 knockdown reduces the *in vivo* tumorigenic potential of BT474 breast cancer cells

Considering the correlation between Epn3 overexpression and poor prognosis in BC patients (**Table 1**), we investigated whether Epn3 has a causal role in tumorigenesis by performing experiments of orthotopic transplantation in NSG mice of human BC cells and assessing the consequent *in vivo* tumorigenic potential. As a model system we chose BT474 $HER2^+$ BC cells, since they harbor *EPN3* gene amplification and protein overexpression. In this cell line we achieved stable knockdown (KD) of Epn3 expression through infection of cells with pSCIOR lentiviral constructs carrying an sh for *EPN3* gene (Epn3 KD), whereas control cells were transfected with empty vector (EV).

Compared with control cells, BT474 Epn3-KD cells formed tumors that were significantly smaller (**Fig. 3**), indicating that Epn3 overexpression contributes to tumor growth. Moreover, the fact that Epn3 KD in BT474 cells was able to cause a reduction of tumor size, despite the expression of the oncogene, HER2, indicates that this cell line is addicted to Epn3 overexpression and amplification for efficient tumor growth *in vivo*, independently of HER2, confirming that Epn3 has an oncogenic potential also independently of HER2.

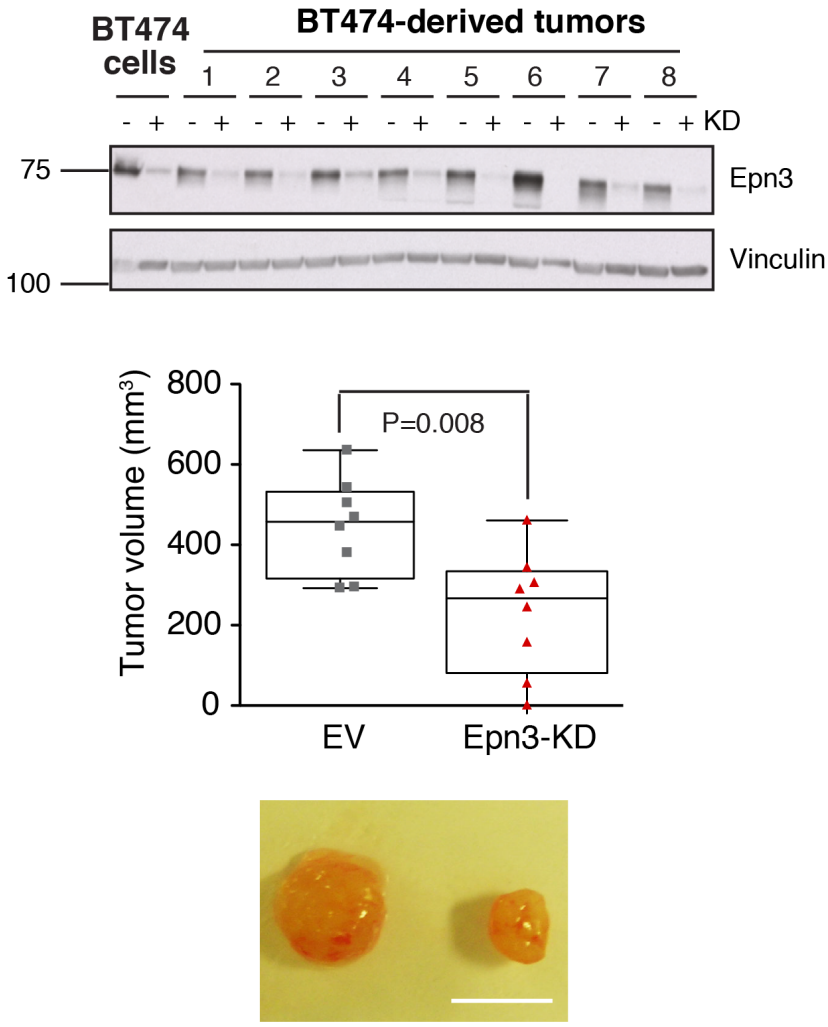


Fig. 3: Epn3 knockdown reduces *in vivo* tumorigenic potential of BT474 BC cells.

BT474 empty vector (indicated with '-' KD or EV) or Epn3-knockdown (indicated with '+' KD or Epn3 KD) cells ($1x10^5$) were injected into the mammary fat pads of NSG mice. Mice were sacrificed 13 weeks post-injection and tumors were measured for size and processed for protein expression analysis by western blotting (WB).
Upper panel: WB analysis of Epn3 expression in EV and Epn3-KD BT474 cells or BT474-derived tumors. Vinculin was used as a loading control. MW markers are shown on the left.
Middle panel: Tumor volume, calculated according to the standard formula (length x $width^2$)/2, is reported as a box-plot for tumors originating from control BT474 cells (EV) or BT474 Epn3-KD cells (Epn3 KD); p-value <0.05 *.
Bottom panel: Representative image of a BT474 EV-derived tumor (*left*) or BT474 Epn3 KD-derived tumor (*right*); bar, 5mm.

2. Aim of the project

Epn3 is defined as a member of the epsin family of endocytic adaptors, together with Epn1 and Epn2, since it shares with the other epsins a similar structural domain organization and a similar cellular localization, particularly at the PM and localization in clathrin-coated structures[40,41]. Nevertheless, the major difference and what makes Epn3 a unique member of the family is its tissue expression profile. In fact, while Epn1 and Epn2 are ubiquitously expressed and their role as adaptors in endocytosis is well characterized, under physiological conditions Epn3 is either not expressed or low expressed, with the exception of gastric parietal cells[40] and little is known about its function. However, Epn3 is overexpressed under pathological conditions, such as in migrating keratinocytes in cutaneous wounds[41] and some carcinomas[32,33].

Our preliminary data revealed Epn3 as a novel putative oncogene, prognostic marker and an independent predictor of distant metastasis in BC. Moreover, Epn3 conferred a proliferative advantage to BT474 BC cells in which it is amplified and overexpressed *in vivo*, suggesting it has a causal role in tumorigenesis. Based on these preliminary data, the overall aim of my project was to uncover the mechanisms through which Epn3 exerts its oncogenic properties in BC. To this aim, I adopted a dual approach:

1. *In vitro* experiments, in which Epn3 was overexpressed in both non-tumorigenic mammary epithelial cells, MCF10A, or in the tumorigenic BC cells, MCF10DCIS.com and HCC1569;

2. *In vivo* experiments, in which the tumorigenic potential of HCC1569 BC cells, overexpressing or not Epn3, was assessed by orthotopic transplantation in NSG mice;

In particular, we first aimed to characterize Epn3-dependent pro-tumorigenic phenotypes upon overexpression in non-tumorigenic MCF10A cells, mimicking the overexpression that occurs in human BC (see **Results, Section 1**), and to dissect the molecular mechanisms of the observed Epn3-dependent phenotypes (see **Results, Section 2**). We then overexpressed Epn3 in two BC cell lines, MCF10DCIS.com and HCC1569, and characterized its impact on the tumorigenic properties of these cells both *in vitro* and *in vivo* (see **Results, Section 3**).

Furthermore, given the predictive power of Epn3 overexpression on distant metastasis outcome, we also aim to define its role in the acquisition of invasive phenotypes. In particular, we plan to exploit Epn3 overexpression in MCF10DCIS.com cells that represent a great model to study *in vivo* invasion, the first step leading to the metastasis cascade. In fact, MCF10DCIS.com cells injected into immunodeficient mice form comedo type DCIS that spontaneously progress to IDC[193,194]. Therefore, the ongoing and the future plans of the described project are *in vivo* orthotopic transplantation of Epn3-overexpressing MCF10DCIS.com followed by the subsequent analysis of the derived tumors at different weeks from injection and *in vitro* migration/invasion assays (see **Discussion**, paragraph **2.3 Future directions: MCF10DCIS.com and human breast cancers: invasive phenotypes analysis**).

Finally, given the high percentage of sequence similarity and similar structural domain organization of Epn3 and the other members of the epsin family, such as Epn1, we plan to engineer chimeric molecules, swapping the domains of Epn3 and Epn1, overexpress them in MCF10A and investigate which combination is

sufficient to give similar phenotypes to the full length Epn3. This chimeric approach may help us identify the specific domain(s) responsible for Epn3-dependent phenotypes and consequently the function of that domain, therefore indicating a specific mechanism of action of Epn3 on its intercators (see **Discussion**, paragraph **3.4 Future directions: Epn3 and Epn1: chimeric approach**).

MATERIALS AND METHODS

1. Cell culture manipulations

1.1 Cell lines

MCF10A and HCC1569 cell lines were purchased from the American Type Culture Collection (ATTC), whereas MCF10ADCIS.com were kindly given to us by Giorgio Scita[195]. All cells were tested for STR profile and mycoplasma.

MCF10A cells were cultured in 1:1 mixture of DMEM and HAM'S-F12 (Lonza, Life Techonologies), supplemented with 5% horse serum (Invitrogen), 2 mM Glutamine (Euroclone), 50 ng/ml cholera toxin (Sigma), 10 μg/ml insulin (Sigma), 500 ng/ml hydrocortisone (Sigma), 1% Pen-Strep antibiotics (Life Techonologies) and fresh 20 ng/ml human EGF (Invitrogen).

MCF10DCIS.com cells were cultured in the same medium of MCF10A cells, but without cholera toxin.

HCC1569 cells were cultured in RPMI-1640 medium (Lonza), supplemented with 10% FBS, 1% Pen-Strep antibiotics (Sigma) and 2 mM L-Glutamine.

Phoenix and 293T cells, used for retroviral and lentiviral infections (see paragraph **1.2 Retroviral and lentiviral infections**), were cultured in DMEM (Lonza), supplemented with 10% fetal bovine serum (FBS, HyClone Euroclone) 2 mM Glutamine (Euroclone), 1% Pen-Strep antibiotics (Life Techonologies).

All cells were cultured at 37°C in a humidified atmosphere containing 5% CO_2.

1.2 Retroviral and lentiviral infections

MCF10A cells were infected with retroviral pBABE-FLAG puro vectors EV, or encoding for Epn1, Epn3 or Twist, whereas MCF10DCIS.com and HCC1569 with lentiviral pLVX-puro EV or Epn3. In all of these cases, puromycin selection was performed at 2 μg/ml for 7 days.
MCF10DCIS.com EV or Epn3 were superinfected with lentiviral pLKO.1 neo CMV-tGFP shEV or shTGFβR1.

Retrovirus infections were performed by overnight transient transfection with 10 μg of the above-described vectors of 70% confluent Phoenix cells, a second-generation retrovirus producer cell line for the generation of helper free ecotropic and amphotropic retroviruses. In particular, this cell line derives by manipulation of 293T cells in order to produce GAG-POL and ENV proteins necessary for viral packaging. 16 hours after transfection, the supernatant was discarded and replaced with half of the original volume (5 ml instead of 10 ml for a 100mm Petri dish) in order to concentrate viral supernatant. 24 hours later, supernatant containing viral particles was collected, filtered through a 45 μm filter and added to 40-50% confluent target cells for 3 hours at 37°C. After 3 hours, a second cycle of infection was performed with fresh viral supernatant (deriving from an independent transfection of Phoenix). Then the appropriate medium of target cells was added and cells were passaged and selected.

Lentivirus infections were performed by overnight transient transfection with 10 μg of the above-described vectors plus packaging GAG-POL, ENV, REV proteins of 70% confluent 293T cells. 16 hours after transfection, the supernatant was discarded and replaced with half of the original volume, as described for retroviruses, and, 24

hours later, supernatant containing viral particles was collected, filtered through a 0.45 μm filter and added to 40-50% confluent target cells for overnight infection at 37°C. The following day the appropriate medium of target cells was added and cells were passaged and selected.

1.3 RNA interference (RNAi)

RNAi was performed with Lipofectamine RNAimax reagent from Invitrogen, according to manufacturer's protocol: for TGFβR1 KD, cells were transfected with 2 cycles of 8 nM smart pool oligos (Dharmacon); for TCF4 KD, cells were transfected with 2 cycles of 20 nM smart pool oligos (Riboxx); for β-Catenin KD, cells were transfected with 2 cycles of 15 nM smart pool oligos (Riboxx).

Briefly, in all of the above-described conditions, 10^6 cells were transfected in reverse, meaning that liposomes complex carrying small-interfering RNAs (siRNAs) were allowed to form for 20 min at RT and then added to trypsinized cells in suspension. For the second cycle, cells were trypsinized and treated in the same way. All KD samples were analyzed after 7 days As a negative control cells were transfected with lipofectamine alone (MOCK) or not treated (NT).

See **Supplementary Table 1** for the nucleotide sequence of siRNAs employed.

1.4 Stimulations and treatments

MCF10A, MCF10DCIS.com and HCC1569 cells, 24 hours after plating, at 50% confluence, were treated with human recombinant TGFβ1 (from Peprotech). In particular, MCF10A cells were treated with both low dose and high dose (0.75 or 5 ng/ml), MCF10DCIS.com only with low dose and HCC1569 only with high dose of TGFβ1. TGFβ1 was added directly to the complete cell culture medium for the indicated time points (90 minutes and 24 hours for a short time-course experiment or 24 hours, 48 hours, 7 days and 14 days for a long time-course experiment). For long time-course experiments, cells were split every 2 days and, at each split, fresh TGFβ1 was added.

MCF10A cells, 24 hours after plating, at 50% confluence, were treated with 5μM TGFβ inhibitor LY2109761 (Selleckchem powder, which was dissolved in DMSO at 10mM) directly added to the complete cell culture medium for 72 hours, plus or minus 0.75 ng/ml TGFβ1 stimulation. As a control, cells were treated with DMSO vehicle alone for the same time point.

2. Cell lysis and protein and RNA extraction techniques

2.1 Western Blotting (WB)

Cells were lysed on ice by removing cell culture medium, washing twice with PBS and by adding RIPA buffer (see **Supplementary Table 2** for the composition of solutions) directly to the cell plates at 70% of confluence. Cell lysate was clarified by centrifugation at 16.000 g for 20 min at 4°C and supernatant was transferred in a new tube. Protein concentration was measured by Bradford Assay (Biorad). 15, 20 or 25 μg of protein lysate were prepared in Laemmli Buffer 1X (see **Supplementary Table 2** for the composition of solutions). Protein lysates were run on gradient 4-20% pre-cast Gels (Biorad) at 250 Volts and transferred by using Trans-Blot (Biorad) according to manufacturer's instructions. Filters were blocked with 5% non-fat dried milk (Applichem) diluted in TBS 0.1% Tween (TBS-T) (see **Supplementary Table 2** for the composition of solutions) and then incubated overnight (ON) with primary antibody according to data sheet. Following 3 washes with TBS-T, filters were then incubated with the appropriate secondary antibody conjugated with horseradish peroxidase. After 3 more washes, the signal was detected at Chemidoc (Biorad) using Amersham ECL Western Blotting Detection Reagent (GE Healthcare), for strong signals, or with Clarity Western ECL Blotting Substrates (Biorad) or with SuperSignal West Femto (Thermo Fisher) for low signals.

See **Supplementary Table 3** for the list of antibodies employed.

2.2 RNA extraction and Quantitative Real time PCR

Cells were lysed by removing cell culture medium, washing twice with PBS and adding Trizol (Life Tecnhologies) directly to the cell plates at 70% of confluence and lysate was clarified by centrifugation at 12.000 g for 15 min at 4°C. The aqueous phase was used to extract total RNA by using RNeasy kit (Quiagen) according to manufacturer's protocol. Single strand cDNA was obtained by SuperScript VILO reverse transcriptase (Invitrogen) according to manufacturer's protocol.

RT-qPCR experiments were performed according to Taqman chemistry (Thermo Fisher Scientific) in technical triplicates. Data were normalized on the average of expression of *18S-ACTB-GAPDH*, and results are reported as relative mRNA expression compared to control by the use of the $2^{-\Delta\Delta Ct}$ method.

See **Supplementary Table 4** for Taqman assays.

3. Biological assays

3.1 Flow Cytometry

For FACS assays, MCF10A cells at 70% confluence were stained with antibodies CD44-APC (BD) and CD24-PE (BD) diluted 1:5 in 1% PBS-BSA (100 µl/500.000 cells). Briefly, cells were trypsinized and, following cell counting, 2 x 10^6 cells were centrifuged at 1200 rpm for 5 minutes, resuspended and washed in 1% PBS-BSA, centrifuged once again in 100 µl of 10% PBS-BSA (blocking) for 1 hour at RT. Cells were centrifuged and resuspended in 1% PBS-BSA. The volume was equally divided in 4 eppendorf tubes, which were centrifuged and resuspended in 100 µl of 1% PBS-BSA with diluted antibody for 45 min at 4°C:

1. CD24-PE single staining
2. CD44-APC single staining
3. CD23-PE/CD44-APC double staining
4. Not stained

Cells were centrigued and washed for 3 times with 1% PBS-BSA and then fixed with 1% formaldehyde for 20 minutes on ice. Cells were finally pelleted, resuspended in PBS and acquired by by BD FACSCantoII. Analysis was performed using BD FACSDiva Software. Not stained and single stained cells were used as control and to set the gates.

For FACS-sorting assays, cells were stained as described previously, however not fixed. Live $CD44^{high}/CD24^{low}$ and $CD44^{low}/CD24^{high}$ cells were sorted by BD

FACSAria and immediately lysed in Trizol for qPCR analysis or re-plated for WB analysis and FACS time course (2-5-7 days) or subjected to mammosphere assay.

3.2 Mammosphere assay

The mammosphere assay was performed with some slight technincal modifications compared to the assay described by Dontu *et al.*[162] and routinely used in our laboratory[196]. In particular, MCF10A cells were trypsinized and centrifuged at 1200 rpm for 5 minutes and resuspended in PBS. Cells were centrifuged once again and resuspended in stem cell medium (MEBM, Lonza, supplemented with 1% Glutamine, 1% Pen-Strep, 5 μg/ml insulin, 500 ng/ml hydrocortisone, 1U/ml heparine) and filtered at 20 μm in order to obtain a single cell suspension. Cells were then counted and 250.000 cells were resuspended in 25 ml complete stem cell medium (stem cell medium plus fresh growth factors 20 ng/ml EGF, 20 ng/ml FGF and 1xB27, Gibco) and embedded in 1% methylcellulose (Sigma powder at 4000 cp, which was previously dissolved at 2% in stem cell medium). Cells were finally plated on ultralow attachment 150mm plates coated with Poly-HEMA and grown as mammospheres in serum-free and anchorage-independent conditions for 10 days.

After 10 days, the medium containing mammospheres was collected into falcon tubes, with the help of PBS to was the plates, then centrifuged at 1200 rpm for 5 minutes and resuspended in stem cell medium. The medium was then filtered at 40μM in order to discard all the single cells that did not grow in the mammosphere assay, whereas the spheres are retained on cell filter. The filter was then turned upside down on a six-well plate and, by washing it with stem cell medium, the

spheres were transferred into the well. Finally, spheres were transferred into wells of a 24 well, and pictures of mammospheres were acquired through confocal microscopy (Leica). Spheres with diameter larger than 50 μm were counted and used to calculate %SFE (number of spheres/number of cells plated x100).

The same protocol was applied for sorted populations $CD44^{high}/CD24^{low}$ and $CD44^{low}/CD24^{high}$, with the exception that 10.000 cells/ml were resuspended in 1 ml of complete stem cell medium, embedded in 1% methylcellulose and plated on ultralow attachment 12 wells coated with Poly-HEMA. After 10 days, pictures of mammospheres were directly acquired from the 12 well through confocal microscopy and spheres with diameter larger than 50 μm were counted and used to calculate %SFE, as described previously.

3.3 ELISA assay

ELISA assay was performed to measure TGFβ1 concentrations in MCF10A-EV and Epn3-overexpressing cell culture medium, according to manufacturer's instructions (Human TGFβ1 Quantikine ELISA Kit, R&D). Briefly, cells were plated in 150 mm dishes with 20 ml complete medium; at confluence, after 2 days from plating, cells were serum starved for 24 hours with 10 ml of the following medium: 1:1 mixture of DMEM and HAM'S-F12 (Lonza, Life Techonologies), supplemented with 2 mM Glutamine (Euroclone), 50 ng/ml cholera toxin (Sigma), 10 μg/ml insulin (Sigma), 500 ng/ml hydrocortisone (Sigma), 1% Pen-Strep antibiotics (Life Techonologies), 1% BSA (Life Technologies), 5 μg/ml Transferrin (Life Technologies) and fresh 20 ng/ml human EGF (Invitrogen). The following day, cell culture media were

concentrated using Amicon Ultra 15 10 KDa cutoff tubes (Merck) up to ~750 μl. Supernatants were then activated to detect latent TGFβ1 in the medium according to protocol with 1N HCl for 10 min, neutralized with 1.2 N NaOH/0.5 M Hepes and immediately spotted onto microplate coated with monoclonal antibody α-TGFβ1 for 2 hours at RT. Wells were washed 3 times with manufacturer's wash buffer and then hybridized with a polyclonal antibody specific for TGFβ1 conjugate with horseradish peroxidase for 2 hours at RT. After 3 more washes substrate solution (hydrogen peroxide and stabilized chromogen tetramethylbenzidine) was added for 30 minutes at RT. Finally, stop solution (hydrochloric acid) was added and plate absorbance was immediately read at Glomax (Promega) at the specific wavelength of 450 nm and at aspecific wavelength of 750 nm.

Cell culture supernatants were assayed in triplicate, together with duplicates of known TGFβ concentrations (standards), which were used to create a standard curve and to calculate the equation of the curve and hence derive the concentration of TGFβ in cell culture supernatants. The concentration of TGFβ was divided by the number of cells at the day of the experiment and normalized for 10 million cells and therefore data are reported as pg/ml / x 10^6 cells.

3.4 E-Cadherin internalization assay

MCF10A cells were trypsinized and, following cell counting, 150.000 cells were plated on coverslips (4 coverslips into a well of a six-well plate). After 24 hours cells were EGF starved ON. The following day, cells were incubated with an E-Cadherin antibody, recognizing extracellular domain (HECD-1, Abcam) diluted 1:25 in cell

culture medium (see **Supplementary Table 3**) for 1 hour at 4°C. Time zero was immediately fixed with 4% paraformaldehyde, whereas the others were stimulated with recombinant TGFβ1 for 90 and 180 minutes at 37°C and then fixed. Immunofluorescence was performed by permeabilizing fixed cells with PBS 0.1% Triton X-100 (Sigma) for 10 min at RT, 3 washes with PBS, blocking with 5% PBS-BSA to prevent non-specific binding, 3 more washes with PBS and incubation for 1 hour with secondary antibody Alexa- 488-conjugated (Molecular Probes). Coverslips were finally mounted in 90% glycerol solution (Sigma) and images were acquired at AOBS confocal microscope (Leica) at 63X oil objective and processed using ImageJ software.

3.5 Immunofluorescence studies

MCF10A cells were trypsinized and, following cell counting, 150.000 cells were plated on coverslips, as described previously, and, after 2 days, immunofluorescence was performed. Briefly, cells were fixed in methanol for 10 minutes, washed with PBS and blocked with 2% BSA to prevent non-specific binding; then cells were stained with primary antibody α-active β-Catenin (Millipore 8E7) (see **Supplementary Table 3**) diluted in PBS 0.2% BSA for 1 hour; after 3 washes with PBS, cells were incubated for 1 hour with secondary antibody (mouse Alexa 488-conjugated, Molecular Probes); after 3 more washes with PBS, cells were stained with nuclear marker DAPI (Sigma) for 5 minutes; finally after 3 more washes with PBS, coverslips were mounted on a 90% glycerol solution (Sigma) and examined under a confocal microscope (AOBS SP2, Leica). Images were processed with Image J software. For each image (.tiff), a DAPI mask was used to identify nuclei and then

to measure the mean fluorescence intensity of active β-catenin in the nuclei per each cell. The mean fluorescence intensity of total cells was instead calculated dividing the mean fluorescence intensity of the entire field of view by the number of cells in that field. Finally, the results are reported as the ratio of active β-catenin staining in nuclei/total.

4. *In vivo* xenografts

HCC1569 EV or Epn3-overexpressing cells were trypsinized and resuspended in cell culture medium. Following cell count, 10^6 or 10^5 cells were centrifuged at 1200 rpm for 5 minutes and resuspended in 40 μl of 1:1 Matrigel (BD at 9-12 mg/ml protein concentration)-PBS solution and injected into mammary fat pad of NOD-SCID IL2RG null female mice (6-8 weeks old). Mice were monitored by hand-palpation for tumor development and sacrificed before reaching the diameter of 1.2 cm, according to Ministerial Project details. Tumor volume was measured with a caliper and according to the standard formula (length x width^2)/2. Tumors were explanted and processed for cell lysis by the use of RIPA buffer (see **Supplementary Table 2** for the composition of solutions), followed by homogenization with the help of magnetic beads and Tissue Lyser magnetic rack (Quiagen). Briefly, 2 magnetic beads were added to each tumor sample in eppendorf tube; tubes were inserted into the Tissue Lyser and subjected to 30 Hz for 2 minutes, until the tumor was completely dissociated. The homogenized tissues were clarified at 13.000 g for 1h at 4°C; the resulting supernatant was ultracentrifuged at 45.000 rpm for 1h at 4°C and this final supernatant was transferred in a new tube. Protein lysates were quantified and loaded as described previously.

5. Statistical analyses

Two-sided student's T-test on JMP software was employed to compare multiple conditions and calculate statistical significance of experimental differences in RT-qPCR experiments. For all the other experiments Two-sided student's T-test type 3 was calculated on Excel to compare two conditions.

In all cases p-value <0.05 *, p-value <0.01 **, p value <0.001 ***.

6. Supplementary Tables

Supplementary Table 1: RNAi oligo sequences

TGFBR1 ON-TARGERTplus SMART pool	GAGAAGAACGUUCGUGGUU
	UGCGAGAACUAUUGUGUUA
	GACCACAGACAAAGUUAUA
	CGAGAUAGGCCGUUUGUAU
TCF7L2 iBONI siRNA pool	AUAAUACAGAACCAACUCCCCC
	UUCUUCCAAACUUUCCCGGCCCCC
	UUAAAGUCUGCUGCCUACCCCC
CTNNB1 siRNA one	AUGAAAUAGAUCCACCUGCCCCC

Supplementary Table 2: Solutions

1X RIPA buffer

Tris HCl, pH 7.6	50 mM
NaCl	150 mM
NP-40	1%
SDS	0.1%
Deoxyxholic acid	0.5%
EGTA	5 mM
Protease Inhibitor cocktail (Calbiochem)	500X
Sodium pyrophosphate pH 7.5	20 mM
Sodium fluoride	250 mM
PMSF	2 mM
Sodium orthovanadate	10 mM

4X Laemmli buffer

SDS	8%
Tris-HCl pH6.8	200 mM
Glycerol	40%
Bromophenol blue	0.1%
β.Mercaptoethanol	2.4%

For 1X Laemmli buffer, the 4X stock was diluted 1:4 with 1X RIPA buffer.

10X SDS-PAGE running buffer

Glycine	192 mM
Tris HCl, pH 8.3	250 mM
SDS	1%

For 1X SDS-PAGE running buffer, the 10X stock was diluted 1:10 with ddH_2O.

10X Tris-buffered saline (TBS)

NaCl	137 mM
KCl	2.7 mM
Tris HCl, pH 7.4	25 mM

The pH was adjusted to 7.4 with HCl and distilled H_2O was added to bring the volume up to 1 litre.

1X TBS-Tween (TBS-T)

TBS	1X
Tween	0.1%

Supplementary Table 3: List of antibodies

Antibody	Producers	Clone or Epitope	Catalog number	Dilution
Epn3	homemade	VI31, epitope 464-483		1:1000
Epn1/Epn2	homemade	ZZ3, epitope 249-401		1:100
FLAG	Sigma	Clone M2	F1804	1:1000
Twist	Santa Cruz	Twist2C1a	sc-811417	1:500
GAPDH	Santa Cruz	6C5	sc-32233	1:1000
E-Cadherin for WB	BD	Clone 36	610181	1:1000
E-Cadherin for IF	Abcam	HECD-1	ab1416	1:25
Vimentin	BD	RV202	550513	1:1000
N-Cadherin	BD	Clone 32	610920	1:1000
Snail	Cell Signaling	L70G2	#3895	1:1000
pSMAD2	Cell Signaling	138D4	#3108	1:1000
ZEB1	Santa Cruz	H102	sc-25388	1:500
TCF4	Cell Signaling	C48H11	#2569	1:1000
β-Catenin for WB	BD	14/β-Catenin	610153	1:1000
β-Catenin for IF	Millipore	Clone 8E7	05-665	1:100
Fibronectin1	Abcam	IST-9	6328	1:1000
HER2	Cell Signaling	44E7	2248	1:1000
CD24-PE for FACS	BD	ML5	555428	1:5
CD44-APC for FACS	BD	G44-26	559942	1:5

Supplementary Table 4: TaqMan assays

Gene symbol	Assay ID	Ref Seq
TGFβ1	Hs00998133_m1	NM_000660.5
TGFβ2	Hs00234244_m1	NM_003238.3
TGFβR1	Hs00610320_m1	NM_004612.3
TGFβR2	Hs00234253_m1	NM_003242.5
18S	Hs99999901_s1	NR_003286.2
GAPDH	Hs99999905_m1	NM_002046.5
ACTB	Hs99999903_m1	NM_001101.3
VIM	Hs00185584_m1	NM_003380.3
CDH2	Hs00169953_m1	NM_001792.4
CDH1	Hs00170423_m1	NM_004360.4
SNAI1	Hs00195591_m1	NM_005985.3
ZEB1	Hs00232783_m1	NM_01128128.2
TWIST1	Hs00361186_m1	NM_000474.3
FN1	Hs00365052_m1	NM_001306129.1
SERPINE1	Hs01126606_m1	NM_000602.4

RESULTS

1. Analysis of Epn3-dependent phenotypes upon overexpression in non-tumorigenic MCF10A cells

1.1 Epn3 overexpression causes EMT in non-tumorigenic MCF10A cells.

To uncover the oncogenic properties of Epn3 in BC, we initially employed an *in vitro* cell model system. We used MCF10A, which are non-tumorigenic mammary epithelial cells that express low levels of Epn3 and display no alterations of the *EPN3* locus[86] (for a detailed description of this cell line see also **Introduction,** paragraph **3.3 *In vitro* and *in vivo* study of breast cancer progression: non-tumorigenic cell lines and mouse models**). In this cell line, we overexpressed Epn3 to mimic Epn3 overexpression observed in human BC and to investigate the mechanisms through which Epn3 overexpression can contribute to tumorigenesis. In particular, we used a retroviral pBABE vector to overexpress FLAG-tagged Epn3. As a control, we overexpressed FLAG-tagged Epn1, member of the same endocytic adaptor family, which is not amplified in BC (see **Preliminary Data and Aim of the Project,** paragraph **1.1 The *EPN3* gene is amplified and overexpressed in breast cancer)**. Cells infected with EV were used as a negative control.

We obtained high levels of Epn3 overexpression, comparable to those observed in the BT474 BC cell line that harbors *EPN3* gene amplification and protein

overexpression, as evidenced by WB analysis using an homemade anti-Epn3 antibody (**Fig. 1A;** see also **Materials and Methods**, **Supplementary Table 3**). Moreover, we assessed whether the levels of Epn3 and Epn1 overexpression were similar, exploiting the fact that the two proteins are fused with a FLAG-tag at the N-terminus. Indeed, the hybridization with an anti-FLAG antibody reveled similar overexpression levels (**Fig. 1A**).

Interestingly, by close observation of cell morphology under the microscope, we observed that Epn3 overexpression induced a clear change in cell morphology, promoting an elongated and spindle-like morphology, as opposed with the more compact and rounded epithelial morphology of MCF10A cells infected with EV or Epn1. Moreover, MCF10A-EV or -Epn1 cells formed large islands of highly packed cells in culture, indicative of strong cell-cell junctions; in contrast, Epn3 cells grew more sparsely, appearing scattered, suggestive of looser junctions (**Fig. 1B,** compare EV and EPN1 *vs.* EPN3). These morphological changes resembled those of an EMT. For this reason, we decided to compare the changes induced by Epn3 overexpression with those induced by the well-known EMT-inducing transcription factor, Twist[197]. We observed that the morphological changes induced by Twist- and Epn3-overexpression were indeed similar, although Twist overexpression induced a stronger mesenchymal phenotype, characterized by the presence of mainly fibroblast-like cells and very few islands of epithelial cells (**Fig. 1B,** compare EPN3 *vs.* TWIST).

In agreement with the observed morphological changes, WB analysis revealed that Epn3- (but not Epn1-) overexpression induced an increase in the levels of mesenchymal markers, N-Cadherin and Vimentin, and a partial decrease in the epithelial marker, E-Cadherin, and this analysis also confirmed a stronger mesenchymal phenotype upon Twist-overexpression (**Fig. 1C**). The “cadherin

switch", i.e., the switch of expression from epithelial E-Cadherin to mesenchymal N-Cadherin, is a hallmark of EMT and is usually accompanied by the production of Vimentin intermediate filaments (see **Introduction** paragraph **4.3 Mesenchymal cells, the Cadherin switch and EMT markers**).

Accordingly, as shown in **Fig. 1D**, EMT master regulators were up-regulated at the mRNA level upon Epn3 overexpression, namely, the zinc-finger E-box-binding ZEB1, SNAI1 (also known as Snail) and TWIST1. Also transcripts encoding for proteins involved in binding to ECM components or in ECM degradation were up-regulated, specifically: fibronectin (FN1), involved in binding with integrins and ECM components, such as collagen and fibrin; SERPINE1 (also known as PAI1) involved in ECM degradation. Finally, the increased protein levels of Vimentin and N-Cadherin were confirmed at the mRNA level (*VIM* and *CDH2* genes, respectively). Interestingly, the reduction in E-Cadherin protein levels upon Epn3 overexpression was not accompanied by a reduction in its mRNA levels (gene name *CDH1*) compared to EV cells, suggesting that post-transcriptional mechanisms are responsible for its down-regulation at the protein level.

Taken together, these data indicate that Epn3 overexpression in the non-tumorigenic MCF10A cell line is able to activate major reprogramming of cells, consisting in the activation of the transcriptional EMT program, through increased expression of transcription factors and their targets, inducing in turn, the "cadherin switch", ECM remodeling and morphology changes associated with EMT. Since we did not observe the complete transcriptional loss of E-Cadherin, the phenotype induced by Epn3 is consistent with the 'partial EMT' state, where mesenchymal markers are expressed concomitant with epithelial ones (see **Introduction** paragraph **4.4 EMT master regulators and post-transcriptional mechanisms inducing EMT**).

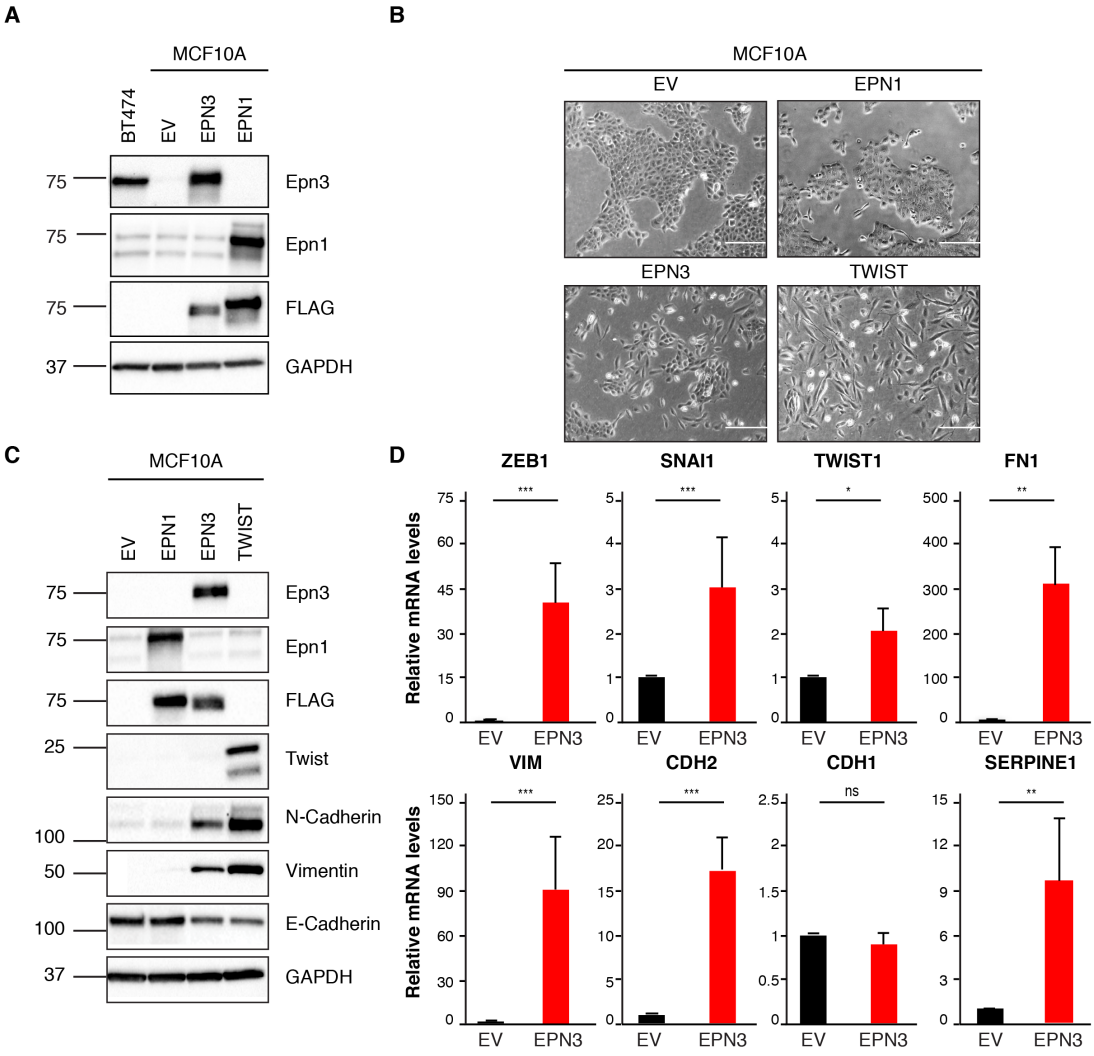


Fig. 1: Epn3 overexpression causes EMT in non-tumorigenic MCF10A cells.

(A) MCF10A cells were infected with retroviral pBABE empty vector (EV) or expressing FLAG-Epn3 (EPN3) or FLAG-Epn1 (EPN1). Epn3, Epn1 and FLAG expression levels were analyzed by WB. The level of Epn3 overexpression was compared to that observed in the BT474 BC cell line. Relative Epn3 *vs.* Epn1 overexpression in MCF10A was compared using the FLAG tag. GAPDH was used as a loading control. MW markers are shown on the left.

(B) Phase-contrast microscopy of MCF10A-EV, -Epn3 or -Twist cells; bar, 250 µm.

(C) Lysates of the MCF10A transfectants described in (A) and of MCF10A-Twist cells were analyzed by WB for the expression of mesenchymal markers (N-Cadherin and Vimentin) and of epithelial marker E-Cadherin. GAPDH was used as a loading control. MW markers are shown on the left.

(D) Activation of the EMT transcriptional program was assessed by RT-qPCR analysis of EMT transcription factors (*ZEB1*, *SNAI1*, *TWIST1*), of extracellular matrix protein components, fibronectin (*FN1*) and protease *SERPINE1*, and of EMT markers (*VIM*, *CDH2* and *CDH1*) in MCF10A-EV and -Epn3 cells. Data were normalized on the average expression of *18S-ACTB-GAPDH* and results are represented as relative mRNA expression compared to EV ± standard deviation of biological replicates, each performed in technical triplicates (n=9 *ZEB1*; *SNAI1*; *TWIST1*; *VIM*; *CDH2*; *CDH1*; n=3 *FN1 SERPINE1*); p-value <0.05 *; p-value <0.01**; p-value <0.001 ***.

1.2 Epn3-induced EMT expands the MCF10A mammary epithelial SC-like compartment.

Based on the correlation between EMT and the generation of cells with CSC properties (see **Introduction**, paragraph **5.3 EMT and CSCs: two sides of the same coin**), we investigated whether Epn3 overexpression was able to alter the SC-like compartment in MCF10A by analyzing the expression of MaSC surface markers, CD44 and CD24, by FACS analysis. In particular, we exploited the $CD44^{high}/CD24^{low}$ configuration, previously identified by Al-Hajj *et al*. for the isolation of CSCs from the bulk tumor mass[148]. This configuration is now extensively used for the identification of normal MaSCs, breast CSCs and cells undergoing EMT[152,165]. While MCF10A-EV or -Epn1 control cells displayed a small percentage of $CD44^{high}/CD24^{low}$ cells (~0.1 and ~0.9%, respectively), this population was significantly increased in Epn3- or Twist-overexpressing MCF10A cells (~10 and ~30%, respectively) suggesting an expansion of the SC-like compartment (**Fig. 2A** and **Fig. 2B**).

To confirm this hypothesis, we performed the mammosphere assay developed by Dontu *et al.*[162] and routinely used in our laboratory[196] (see **Materials and Methods**, paragraph **3.2 Mammosphere assay** for details). This assay exploits the functional properties of MaSCs, which are normally quiescent or slowly proliferating; however, in contrast to progenitors, they are able to survive *anoikis* and to proliferate in anchorage-independent and serum-free conditions, giving rise to clonal spheroids (called mammospheres) originating from single SCs. Therefore, through this assay, the number of spheres formed can be considered a proxy of the SC content of the initial population. MCF10A-Epn3 cells displayed a 0.006% sphere-forming

efficiency (SFE), i.e. a 10-fold higher SFE than controls MCF10A-EV and -Epn1 cells (0.0006 and 0.0005 %SFE, respectively) and accordingly, our positive control MCF10A-Twist cells, displayed a 0.021% SFE, i.e. a 35-fold increase in SFE (**Fig. 2C**). Interestingly, the extent of the expansion of the $CD44^{high}/CD24^{low}$ cell population mirrored the increase in SFE measured in the mammosphere assay (compare **Fig. 2B** and **Fig. 2C**).

Collectively, these data suggest that Epn3-overexpression, besides inducing a partial EMT state, is able to induce differentiated cells to acquire CSCs properties, consistent with previous reports showing an association between these states.

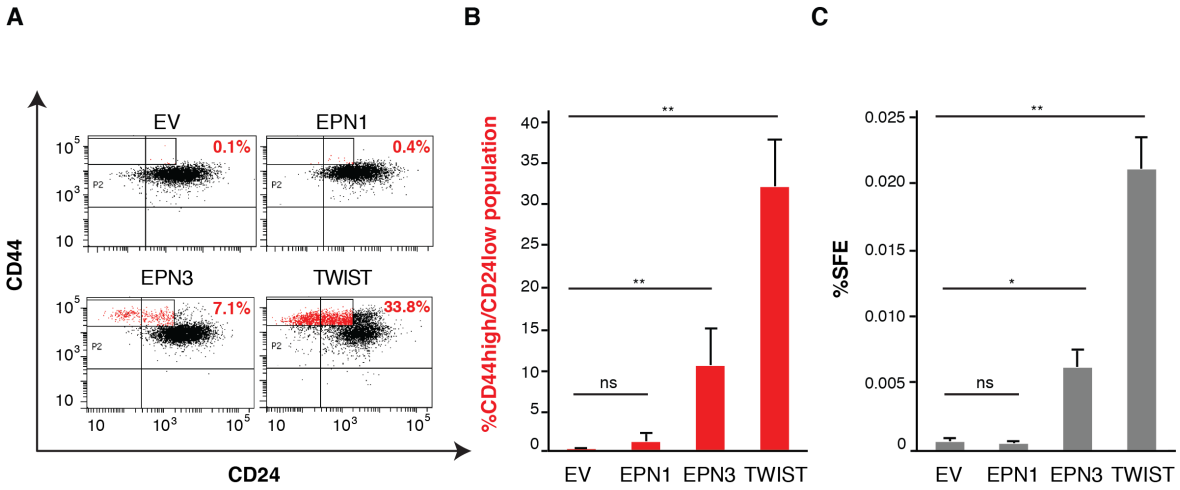


Fig. 2: Epn3-induced EMT expands MCF10A SC-like compartment.

(A) Representative bi-parametric FACS analysis of the cell surface markers CD44/CD24 of MCF10A-EV, -Epn1, -Epn3 or -Twist cells. The $CD44^{high}/CD24^{low}$ population is boxed and represented in red with the relative percentage.

(B) Quantification of the $CD44^{high}/CD24^{low}$ population, reported as the mean ± standard deviation from biological replicates of the FACS analysis (n=4 EV; n=2 Epn1; n=8 Epn3; n=6 Twist); p-value <0.05 *, <0.01 **, ns: not significant.

(C) Mammosphere assay on the indicated MCF10A transfectants described in (A). Results are reported as mean sphere forming efficiency (SFE) ± standard deviation of two biological replicates, each performed in technical triplicates; p-value <0.05 *; p-value <0.01 **.

1.3 Epn3-induced EMT and expansion of the SC-like compartment in MCF10A cells are causally linked.

We hypothesized that the above-mentioned induction of partial EMT and expansion of the SC-like compartment following Epn3 overexpression in MCF10A cells are causally linked. To test this hypothesis, we further characterized the $CD44^{high}/CD24^{low}$ population derived from MCF10A-Epn3 cells, to verify if this population is indeed enriched in cells with EMT and SC-like characteristics. For this reason, we FACS-sorted the $CD44^{high}/CD24^{low}$ and the $CD44^{low}/CD24^{high}$ cell populations from MCF10A-Epn3 and MCF10A-Twist cells, and we cultured and analyzed them separately. The two populations showed dramatically different morphologies, with the former displaying a more mesenchymal morphology and the latter, an epithelial morphology (**Fig. 3A**).

Furthermore, we decided to couple FACS-sorting with RT-qPCR and WB analysis of EMT markers, and with the mammosphere assay to assess SC properties. For the RT-qPCR analysis, we extracted RNA immediately after FACS-sorting (day 0) to control and avoid eventual artifacts induced by cell re-plating. As a control, we used the total unsorted population of Epn3- and Twist-overexpressing cells that had been processed equally to the sorted cells, including the passage in the FACS-sorting beam and the exposure to lasers (which represent two stressful events for cells), but were left unstained. Data are expressed as relative fold change to the internal control of the total population of Epn3-overexpressing cells (represented in grey). Results showed a clear difference in EMT markers expression in the sorted cell populations similarly in MCF10A-Epn3 and -Twist cells: the epithelial marker E-Cadherin

(*CDH1*) was exclusively expressed in the $CD44^{low}/CD24^{high}$ populations (represented in black), whereas the mesenchymal markers, N-Cadherin (*CDH2*) and *ZEB1* were almost exclusively expressed in the $CD44^{high}/CD24^{low}$ populations (represented in red). Other mesenchymal markers, Vimentin or Snai1, were instead enriched, but not exclusively expressed, in the $CD44^{high}/CD24^{low}$ populations (**Fig. 3B**).

Consistently, a time-course analysis of the expression of EMT markers at the protein level by WB analysis of cell lysates harvested on days 2, 5 and 7 after FACS-sorting, confirmed that the epithelial marker, E-Cadherin, is almost exclusively expressed in the $CD44^{low}/CD24^{high}$ population, whereas mesenchymal markers, N-Cadherin and Vimentin, are enriched in the $CD44^{high}/CD24^{low}$ population (**Fig. 3C**). Therefore, we concluded that the $CD44^{high}/CD24^{low}$ population represents the mesenchymal population, whereas the $CD44^{low}/CD24^{high}$ population represents the epithelial one in both Epn3- and Twist-overexpressing MCF10A cells.

Of note, the FACS-sorted $CD44^{high}/CD24^{low}$ Epn3-overexpressing cells displayed a substantial decrease of E-Cadherin not only at the protein but also at the mRNA level, as opposed to the unsorted population (compare RTqPCR for *CDH1* in **Fig. 3D** and **Fig. 1D**), suggesting therefore that only a minor fraction of Epn3-overexpressing cells have transcriptionally lost E-Cadherin expression and that post-transcriptional mechanisms are responsible for its down-regulation at the protein level in the unsorted population (see **Fig. 1C**).

Lastly, coupled analysis of the FACS-sorted populations with the mammosphere assay revealed that in Epn3-overexpressing cells only the $CD44^{high}/CD24^{low}$ population (i.e., the mesenchymal population) was able to form mammospheres, indicating that this population is also the population with SC-like traits (**Fig. 3D**). Interestingly, the SFE of MCF10A-Epn3 $CD44^{high}/CD24^{low}$ cells was notably increased upon sorting, compared to the total population (0.15% SFE and 0.006%

SFE, respectively; ~25-fold increase), and was similar to the SFE of Twist $CD44^{high}/CD24^{low}$ sorted cells, which instead was only slightly further increased compared to the unsorted population (0.1% SFE and 0.021 %SFE, respectively; ~5-fold increase) (compare **Fig. 2C** for unsorted populations and **Fig. 3D** for sorted populations). Indeed, although the SFE of unsorted MCF10A-Twist cells was greater than the SFE of unsorted MCF10A-Epn3 cells, the $CD44^{high}/CD24^{low}$ populations from the two cell lines had a similar SFE. This result suggests that MCF10A-Twist cells have a higher SFE in mammosphere assay since they have a larger $CD44^{high}/CD24^{low}$ population, compared with MCF10A-Epn3 cells (see **Fig. 2A** and **Fig. 2B**); however, once purified, the $CD44^{high}/CD24^{low}$ populations from the two cell lines display similar SC self-renewal properties, independently of the oncogenic stimulus.

In conclusion, the data deriving from the combination of FACS-sorting with various assays (RT-qPCR, WB and mammosphere assay) indicate that EMT and expansion of SC-like compartment induced by Epn3 overexpression in MCF10A cells are two events causally linked, with the $CD44^{high}/CD24^{low}$ population displaying indeed both mesenchymal and SC-like traits.

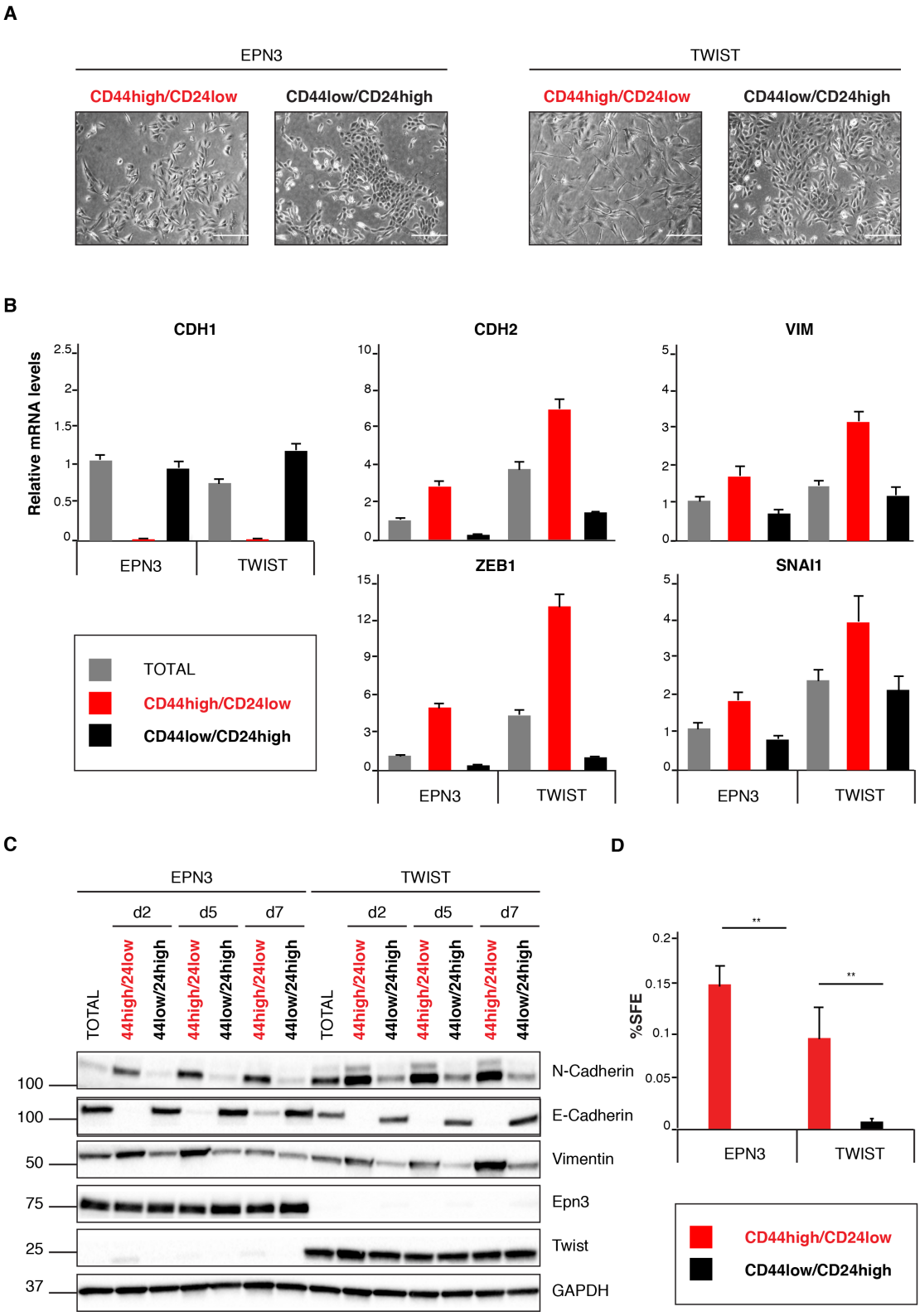


Fig. 3: The CD44^{high}/CD24^{low} population displays mesenchymal and SC-like traits in MCF10A-Epn3 and MCF10A-Twist cells.

Continued on the next page

Fig. 3: The CD44^{high}/CD24^{low} population displays mesenchymal and SC-like traits in MCF10A-Epn3 and MCF10A-Twist cells.

MCF10A-Epn3 and MCF10A-Twist cell lines were FACS-sorted into CD44^{high}/CD24^{low} and CD44^{low}/CD24^{high} populations. Sorted populations were analyzed by:

(A) Phase-contrast microscopy to assess mesenchymal (elongated spindle-like) *vs.* epithelial (compact and rounded) morphology; bar, 250 μm.

(B) RT-qPCR analysis to assess the expression of the epithelial marker, *CDH1* (E-Cadherin), and of the mesenchymal markers, *CDH2* (N-Cadherin), *VIM* (Vimentin), *ZEB1* and *SNAI1* at day 0 (immediately after FACS-sorting). Data were normalized on the average expression of *18S-ACTB-GAPDH* within each sample and results are expressed as relative fold-change to the internal control of the total population of MCF10A-Epn3 unsorted cells ± standard deviation of technical triplicates.

(C) WB analysis to assess the expression of EMT markers (N-Cadherin, Vimentin, E-Cadherin) in FACS-sorted CD44^{high}/CD24^{low} and CD44^{low}/CD24^{high} populations from MCF10A-Epn3 and MCF10A-Twist cells at 2, 5 and 7 days after FACS-sorting (indicated as d2, d5, d7). Unsorted (Total) cells were also assessed. GAPDH was used as a loading control. MW markers are shown on the left.

(D) Mammosphere assay to evaluate the SC properties. Results are represented as sphere forming efficiency (SFE) ± standard deviation of two biological replicates, each performed in technical triplicates; p-value <0.01 **.

1.4 The epithelial population in Epn3-overexpressing MCF10A cells is highly plastic, in constrast to the meta-stable mensenchymal population

To conclude the characterization of the CD44^{high}/CD24^{low} as opposed to the CD44^{low}/CD24^{high} population in MCF10A-Epn3 cells, we investigated their degree of plasticity in culture, i.e., whether these populations are stable or whether they are able to switch from a mesenchymal to an epithelial state (in the case of the mesenchymal population) or *vice versa* from an epithelial to a mesenchymal state (in the case of the counterpart epithelial population). For this purpose, we coupled FACS-sorting of the two different populations with a time-course FACS analysis of CD44/CD24 cell surface markers. As shown in **Fig. 4A**, once sorted and purified

from the bulk population, the $CD44^{high}/CD24^{low}$ is relatively a stable population: indeed, there was only a small increase in the appearance of the epithelial population (maximum 3.7% after 7 days in culture for MCF10A-Epn3 cells, **upper panel,** and 0.8% after 7 days in culture for MCF10A-Twist cells, **lower panel**). However, this small increase is indeed very far from the initial situation, where the epithelial population is the predominant population in both Epn3- and Twist-overexpressing cells (compare **Fig. 2A** and **Fig 4A, upper and lower panels**).

In contrast, the $CD44^{low}/CD24^{high}$ population appeared to be highly plastic and after 7 days it reached the initial equilibrium of 8% of mesenchymal cells for MCF10A-Epn3 cells (compare **Fig. 2A** and **Fig 4B, upper panel**). Instead, MCF10A-Twist cells, showed ~13% of mesenchymal population after 7 days, suggesting that this time point is not yet sufficient to allow them to reach the initial equilibrium of ~30% of $CD44^{high}/CD24^{low}$ cells (compare **Fig. 2A** and **Fig. 4B, lower panel)**.

In theory, the $CD44^{high}/CD24^{low}$ population carrying SC-like properties should be able to self-renew and to differentiate, and should therefore give rise to $CD44^{low}/CD24^{high}$ cells (also referred to as progenitors or differentiated cells), whereas differentiated cells under physiological conditions should not be able to give rise to SCs. However, many lines of evidence indicate now that oncogenic stimuli can alter the equilibrium between the two cell populations, causing reversion of progenitors to SC-like cells[198]. Therefore, our data are in agreement with the possibility that Epn3 (or Twist)-overexpression, via EMT, can drive differentiated cells to a more naïve, stem-like state with CSC properties. Together, these data suggest that Epn3 overexpression could contribute to breast tumorigenesis by driving the emergence of CSCs.

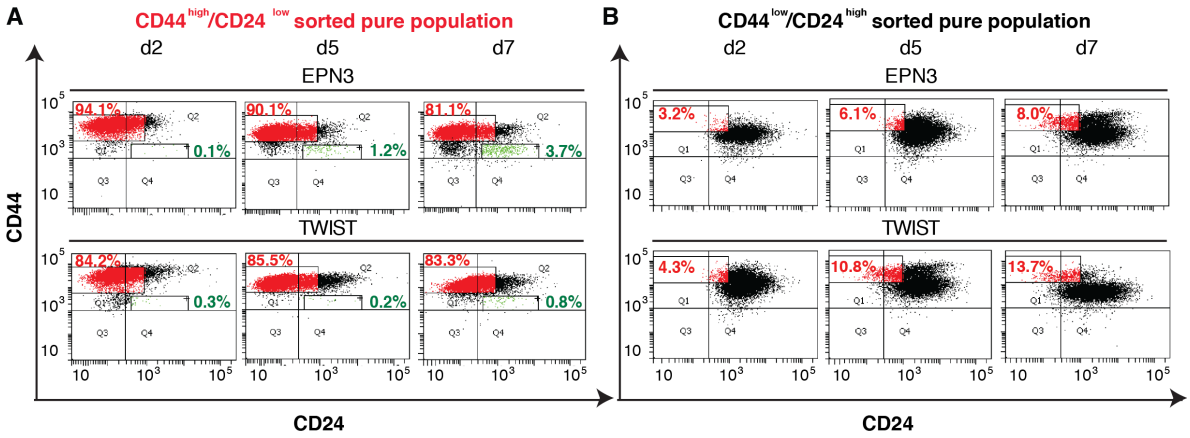


Fig. 4: CD44^{low}/CD24^{high} MCF10A-Epn3 and -Twist cells are able to convert into CD44^{high}/CD24^{low} cells.

(A) CD44^{high}/CD24^{low} populations derived from MCF10A-Epn3 or -Twist cells were FACS-sorted, re-plated in culture and analyzed by FACS for cell surface markers CD44/CD24 after 2, 5 and 7 days after sorting (indicated as d2, d5, d7). The CD44^{high}/CD24^{low} and the CD44^{low}/CD24^{high} populations are boxed and represented in red and green, respectively. The size of each population is indicated as a percentage of total cells.

(B) FACS analysis time-course of the counterpart CD44^{low}/CD24^{high} populations from the experiment described in (A). The CD44^{high}/CD24^{low} population is boxed and represented in red with the relative percentage.

2. Dissection of the Epn3-dependent molecular mechanisms in non-tumorigenic MCF10A-Epn3 cells

2.1 MCF10A-Epn3 cells secrete TGFβ ligands that cooperate with Epn3 and create a positive feedback-loop sustaining EMT.

To unravel the mechanisms responsible for Epn3-dependent phenotypes of EMT and SC-like expansion, we initially investigated the TGFβ pathway. This pathway is, in fact, involved in inducing and sustaining EMT, in metastasis and in SC regulation[189,199]. Hence, we measured the transcript levels of TGFβ ligands (TGFβ1, TGFβ2, TGFβ3) and receptors (TGFβR1 and TGFβR2) in MFC10A-Epn3 cells and control EV cells. The ligands TGFβ1 and TGFβ2 were significantly and particularly up-regulated in Epn3-overexpressing cells (~15 folds TGFβ1 and ~4 folds TGFβ2, **Fig. 5A**; TGFβ3 was unchanged, data not shown), whereas the receptors TGFβR1 and TGFβR2 showed a more moderate up-regulation (~2 folds both receptors), with only TGFβR2 reaching statistical significance (**Fig. 5A**).

Consequently, we hypothesized that endogenous TGFβ ligands produced by Epn3-overexpressing cells, might be secreted in the medium and create a positive feedback loop of TGFβ signaling that cooperates with Epn3 to establish and sustain the EMT phenotype. To test this hypothesis, we exploited commercially available ELISA assays for TGFβ1 and TGFβ2 to analyze conditioned media from MCF10A-EV and -Epn3 cell cultures (see **Materials and Methods**, paragraph **3.3 ELISA assay** for details). This assay confirmed that cell culture media harvested from Epn3-

overexpressing cells contained more TGFβ1 than that of EV cells (~3,5 folds). Moreover, this assay allowed the quantification of TGFβ1 ligand protein levels in the cell medium: MCF10A-Epn3 media had ~400 pg/ml / 10^6 cells of TGFβ1, compared with ~110 pg/ml / 10^6 cells in the control (**Fig. 5B**). Also TGFβ2 was up-regulated in Epn3 cell culture supernatants (data not shown).

Building on the hypothesis that secreted ligands bind to their receptors (whose levels are increased), thereby activating the TGFβ signaling pathway and transcription of EMT genes, we initially checked whether TGFβ stimulation cooperates with Epn3 overexpression to potentiate EMT. To this aim, we treated cells with recombinant TGFβ1 at two different doses, a low dose 0.75 ng/ml (closer to the amount that was secreted in the medium) and a high dose of 5 ng/ml, in a time-course experiment. We found that TGFβ stimulation was able to enhance Epn3-induced EMT, with increased levels of late transcriptional targets, such as N-Cadherin, Vimentin and ZEB1, and decreased levels of E-Cadherin compared to control cells (**Fig. 5C**). Moreover, while control MCF10A cells responded only to chronic stimulation with the high dose of TGFβ, Epn3-overexpressing cells responded also to the low TGFβ dose (**Fig. 5C**).

Collectively, these data suggest that TGFβ ligands, which are up-regulated and secreted by Epn3-overexpressing cells, could synergize with Epn3 overexpression in EMT induction, and that cooperation occurs already at low TGFβ concentrations. Finally, further up-regulation of endogenous TGFβ1 ligand upon stimulation with recombinant TGFβ1 in Epn3-overexpressing cells demonstrates the existence of a transcriptional positive feedback loop, which further sustains the TGFβ signaling pathway (**Fig. 5D**).

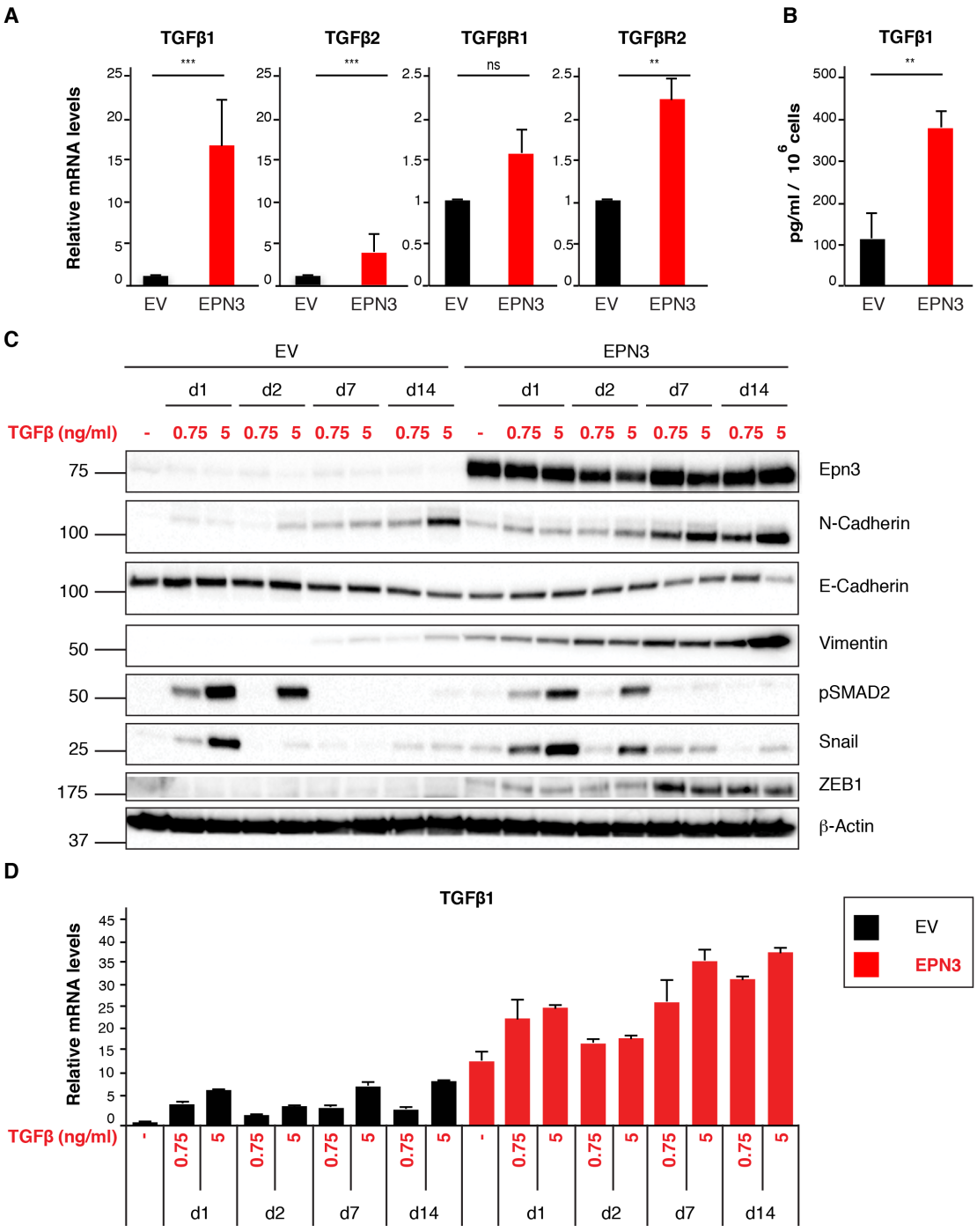


Fig. 5: A TGFβ-dependent feedback loop sustains EMT in MCF10-Epn3 overexpressing cells.

Continued on the next page

Fig. 5: A TGFβ-dependent feedback loop sustains EMT in MCF10-Epn3 overexpressing cells.

(A) RT-qPCR analysis of the expression of TGFβ ligands (TGFβ1, TGFβ2) and receptors (TGFβR1 TGFβR2) in MCF10A-EV and -Epn3 cells. Data were normalized on the average expression of *18S-ACTB-GAPDH* and results are reported as relative mRNA expression compared to EV control ± standard deviation of biological replicates, each performed in technical triplicates (n=9); p-value <0.05 *; p-value <0.01 **; ns, not significant.

(B) ELISA assay of secreted TGFβ1 in conditioned media from MCF10A-EV and -Epn3 cell cultures. Results are reported as pg/ml / 10^6 cells of two biological replicates, each performed in technical triplicates.

(C) WB analysis from a representative experiment of the expression of EMT markers (N-Cadherin, Vimentin, E-Cadherin) and of TGFβ signaling pathway activation (pSMAD2, Snail, ZEB1) in MCF10A-EV and -Epn3 cells, not treated (-) or treated for the indicated times with 0.75 or 5 ng/ml recombinant TGFβ1. β-Actin was used as a loading control. MW markers are shown on the left.

(D) RT-qPCR analysis of TGFβ1 in MCF10A-EV and -Epn3 cells. Data were normalized on the average expression of *18S-ACTB-GAPDH* and results are reported as relative mRNA expression compared to EV control ± standard deviation of technical triplicates.

2.2 MCF10A Epn3-dependent phenotypes are TGFβ dependent.

The fact that Epn3-overexpressing MCF10A cells displayed up-regulation of the TGFβ pathway and cooperation between this pathway and Epn3 in inducing EMT, led us to hypothesize that Epn3-induced EMT is dependent on TGFβ signaling. To test this hypothesis, we abrogated the TGFβ pathway by KD of TGFβR1, achieved using the small interfering RNA (siRNA), in MCF10A-EV, -Epn3, or -Twist cells. We used mock transfection or not treated (NT) samples as negative controls.

Cell morphology was dramatically different between mock transfected cells and TGFβR1 KD cells, the latter assuming a more epithelial phenotype in both Epn3- or Twist-overexpressing cells (**Fig. 6A**). Accordingly, WB analysis of EMT markers

showed an almost complete reversion of the induction of mesenchymal markers N-Cadherin and Vimentin in Epn3-overexpressing cells and a partial reversion in the case of Twist-overexpressing cells (**Fig. 6B**, **first 9 lanes**). To verify that TGFβ signaling was efficiently abrogated following TGFβR1 KD, we stimulated KD cells with 0.75 ng/ml of TGFβ1 for 90 minutes and analyzed by WB the activation of early signaling targets, i.e. SMAD2 phosphorylation and Snail protein expression levels. These targets were both reduced, however not completely, in KD samples, suggesting that silencing was not complete (**Fig. 6B**, **last 6 lanes**). Nevertheless, the level of KD was sufficient to revert Epn3-induced EMT phenotype (see **Fig. 6C** for TGFβR1 transcript levels).

To test whether Epn3-induced expansion of the SC-like compartment was also reverted upon TGFβR1 KD, we performed the CD24/CD44 FACS analysis. As shown in **Fig. 6D**, the percentage of $CD44^{high}/CD24^{low}$ cells in MCF10A-Epn3 TGFβR1 KD cells (**middle panel**), was significantly reduced, almost to control levels (**upper panel**), whereas in Twist TGFβR1 KD cells it was only partially reduced (**lower panel**), consistent with partial EMT reversion as assessed by WB analysis. Furthermore, RT-qPCR analysis showed that also the up-regulation of TGFβ1 in Epn3- and Twist-overexpressing cells was reduced upon TGFβR1 KD (data not shown, for a complete analysis of transcript levels upon TGFβR1 KD see **Fig. 9)**.

These results indicate that also the expansion of the SC-like compartment in Epn3-overexpressing cells is dependent on TGFβ signaling, as observed for the EMT phenotype. Instead, Twist-overexpressing MCF10A cells showed only a partial reversion of both EMT and SC-like expansion phenotypes upon TGFβR1 KD, suggesting that these cells are not completely dependent on the TGFβ signaling

pathway. Moreover, these data strongly show, once again, that the two phenotypes are causally linked, since a complete reversion of one of them (i.e. EMT) causes the loss also of the other phenotype (SC-like expansion), whereas a partial reversion of the first causes only a partial reversion of the second.

Lastly, to exclude potential off-target effects of transient silencing and to confirm our data, we decided to use a chemical approach, besides the previous genetic approach. In particular, we used a commercially available chemical inhibitor of TGFβ signaling: LY2109761 at 5 μM for 72 hours. This inhibitor blocks TGFβ signaling by blocking the kinase activity of TGFβRs, thereby inhibiting SMADs phosphorylation and abrogating their signaling[200]. Control samples were treated with vehicle, DMSO, alone. As shown in **Fig. 6E**, LY2109761 treatment was able to decrease the expression of mesenchymal markers, most notably Vimentin and to a lesser extent N-Cadherin, and to up-regulate the epithelial marker, E-Cadherin, in both Epn3- and Twist-overexpressing cells. The inhibitor also efficiently abrogated SMAD2 phosphorylation upon TGFβ1 stimulation at 0.75 ng/ml for 90 minutes, as expected.

Taken together, these data indicate that Epn3-dependent phenotypes are almost completely dependent on TGFβ signaling, and that abrogation of the TGFβ signaling pathway is able to revert activation of late EMT targets and expansion of SC-like compartment in Epn3-overexpressing MCF10A cells. In contrast, the Twist-dependent phenotype in these cells appears to be only partially dependent on TGFβ signaling.

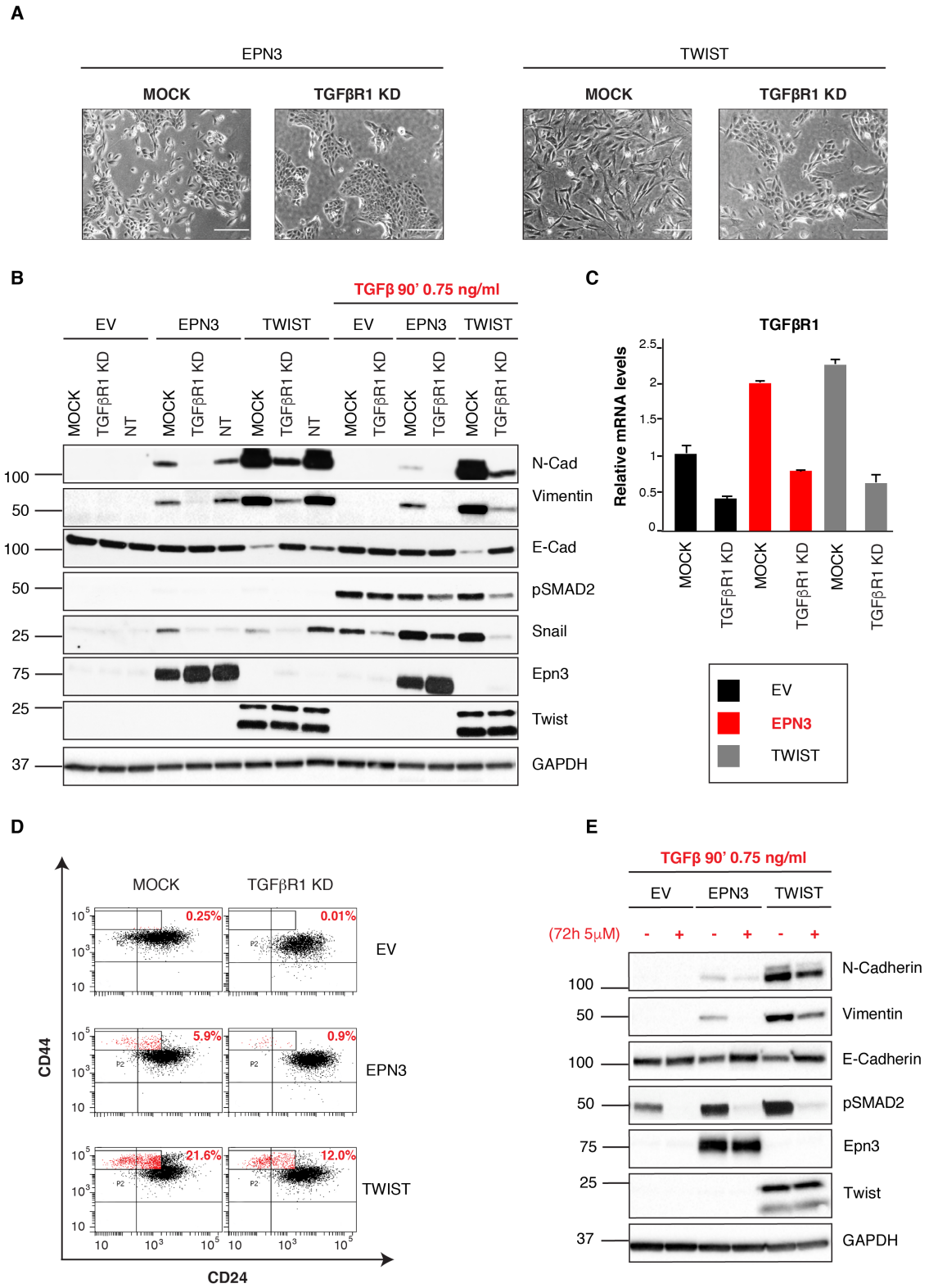


Fig. 6: MCF10A Epn3-dependent phenotypes are TGFβ dependent.

Continued on the next page

Fig. 6: MCF10A Epn3-dependent phenotypes are TGFβ dependent.

(A) Phase-contrast microscopy of MCF10A-Epn3 and -Twist cells transfected with only lipofectamine (MOCK) or with a pool of siRNAs to knockdown TGFβR1 (TGFβR1 KD); bar, 250 μm.

(B) WB analysis of the expression of controls (Epn3, Twist), EMT markers (N-Cadherin, Vimentin, E-Cadherin) and the activation of TGFβ signaling (pSMAD2, Snail) in MCF10A-EV, -Epn3 or Twist cells, under control conditions (NT; not treated), following transfection described in (A). Cells were lysed 7 days after transfection and following stimulation (last 6 lanes) for 90 minutes with 0.75 ng/ml recombinant TGFβ1, or not (first 9 lanes). GAPDH was used as a loading control. MW markers are shown on the left.

(C) RT-qPCR analysis of TGFβR1 expression in mock or TGFβR1 KD samples of MCF10A-EV, -Epn3 or -Twist cells. Data were normalized on the average expression of *18S-ACTB-GAPDH* and results are reported as relative mRNA expression compared to EV mock sample ± standard deviation of technical triplicates from a representative experiment.

(D) Representative bi-parametric FACS analysis of the cell surface markers CD44/CD24 in mock control or TGFβR1 KD samples of MCF10A-EV, -Epn3 or -Twist cells of three independent biological replicates. The $CD44^{high}/CD24^{low}$ population is boxed and represented in red with the relative percentage of total cells. p-value <0.05 * for the difference between MOCK and TGFβR1 KD in MCF10A-Epn3 , p-value <0.01 ** for MCF10A-Twist, p-value not significant for MCF10A-EV.

(E) WB analysis of the expression of controls, EMT markers and TGFβ signaling activation, as described in (B), in MCF10A-EV, -Epn3- or -Twist cells, treated with 5 μM LY2109761 (+), or DMSO vehicle alone (-), for 72 h. Cells were lysed after 90 minutes of stimulation with 0.75 ng/ml recombinant TGFβ1. GAPDH was used as a loading control. MW markers are shown on the left.

2.3 Epn3-overexpression in MCF10A cells induces E-Cadherin endocytosis and β-catenin translocation to the nucleus.

A hallmark in EMT and cancer progression is loss of E-Cadherin through different mechanisms, such as transcriptional repression, protein degradation and inactivating somatic mutations, which allow the disruption of cell-to-cell junctions, favoring migration, invasion and extravasation steps[104]. An alternative mechanism of E-Cadherin down-regulation and destabilization of adherent junctions is E-Cadherin

endocytosis[133,138]. Given that Epn3 is a member of the epsin endocytic adaptor family, we hypothesized that Epn3 overexpression could be directly affecting E-Cadherin endocytosis. In agreement with this possibility, E-Cadherin it is unchanged at the mRNA level upon Epn3-overexpression in not sorted MCF10A cells (see **Fig. 1D)**, suggesting that post-transcriptional mechanisms are responsible for its partial down-regulation observed at the protein level (see **Fig. 1A**). Of note, E-Cadherin down-regulation is particularly evident upon TGFβ stimulation, especially at longer time points and, importantly, TGFβ is a known inducer of E-Cadherin endocytosis[201].

Based on these observations, we decided to investigate E-Cadherin endocytosis in MCF10A-EV or -Epn3 cells, either not stimulated or at early time points of TGFβ1 stimulation, to investigate whether Epn3 overexpression increases E-Cadherin endocytosis induced by TGFβ. Using an *in vivo* E-Cadherin internalization assay (see **Materials and Methods**, paragraph **3.4 E-Cadherin internalization assay** for details), we showed that in not stimulated cells (time 0) Epn3-overexpression correlated with an increase in the number of cells that have already lost E-Cadherin expression, especially those at the leading edge (indicated with white asterisks in **Fig. 7A**). After 90 minutes of TGFβ stimulation, most control EV cells still exhibited a PM E-Cadherin staining (~80%), whereas the majority of Epn3-overexpressing cells had already internalized E-Cadherin with less than 20% displaying PM staining. After 180 minutes, also control EV cells showed a remarkable decrease of PM E-Cadherin staining (~50%), while in Epn3-overexpressing cells E-Cadherin was completely internalized and was no longer detectable at the PM (see representative images of E-Cadherin staining and internalization in **Fig. 7A** and relative quantification of cells with remaining E-Cadherin at the PM in **Fig. 7B**).

Taken together, our results indicate an accelerated endocytosis and turnover of E-Cadherin upon Epn3 overexpression in MFC10A cells. Of note, down-regulation of E-Cadherin expression and the EMT phenotype are frequently associated with active, i.e. dephosphorylated, β-catenin translocation and accumulation in the nucleus, where it binds to the TCF/LEF transcription factors and behaves as a co-activator of target genes, including EMT related genes[95,132] (see **Introduction** paragraph **4.7 The WNT/β-catenin pathway**). To test whether such a scenario is occurring in Epn3-overexpressing cells, we performed an immunofluorescence staining of active β-catenin in MCF10-EV and -Epn3 cells (see **Materials and Methods**, paragraph **3.5 Immunofluorescence studies** for details). As shown in **Fig. 7C**, the levels of active β-catenin in the nucleus of EV control cells were very low, while they were particularly up-regulated in Epn3-overexpressing cells. A quantification of the data from biological replicates, confirmed a statistically significant increase in the ratio of nuclear/total β-catenin per cell in Epn3-overexpressing cells (**Fig. 7D**).

Collectively these data suggest that the mechanism through which Epn3 induces a major reprogramming of cells, consisting in the activation of the transcriptional EMT program, is the acceleration of E-Cadherin endocytosis at the PM that leads to destabilization of adherent junctions, nuclear translocation of β-catenin, which in turn can bind to the TCF/LEF transcription factors, altering gene expression.

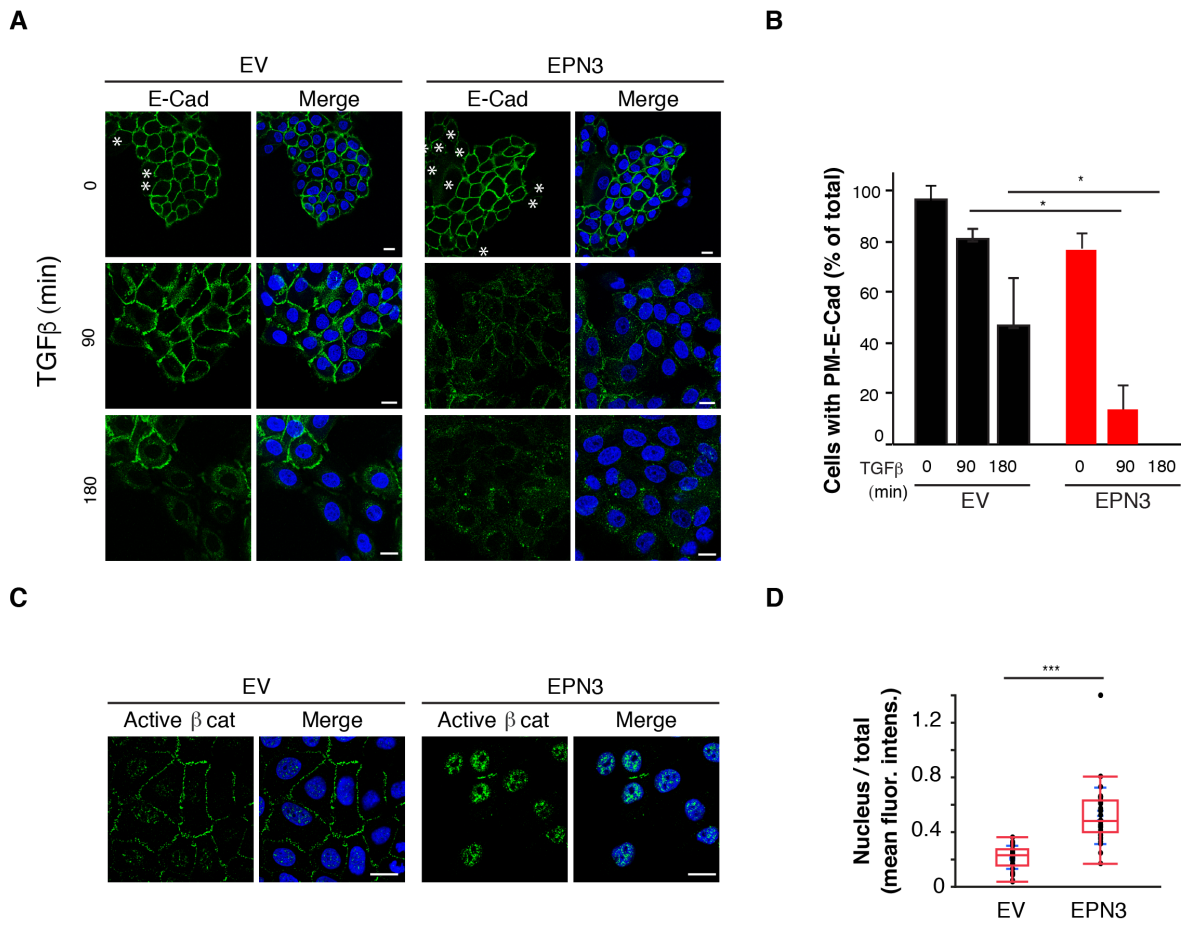


Fig. 7: Epn3 overexpression induces E-Cadherin endocytosis and β-catenin translocation to the nucleus.

Experiments performed by Claudia Iavarone and Giovanni Giangreco.

(A) E-Cadherin *in vivo* internalization assay of MCF10A EV and Epn3-overexpressing cells. Cells were EGF starved overnight then incubated with an E-Cadherin antibody recognizing the extracellular domain. E-Cadherin internalization was followed in untreated cells (time 0, immediately fixed) or in cells stimulated for 90 or 180 min with 5 ng/ml recombinant TGFβ prior to fixation. Representative images are shown: E-Cadherin (green) and Merge of E-Cadherin and DAPI (blue); asterisks indicate cells lacking E-Cadherin before TGFβ addition; bar, 10 μm.

(B) Quantification of cells with plasma membrane E-Cadherin (PM-E-Cad) expressed as mean % of total ± standard deviation; >100 cells counted per condition in two independent experiments performed as described in (A); p-value <0.05 *.

(C) Representative images of active β-catenin immunofluorescence in MCF10A-EV and -Epn3 cells: active β-catenin (green) and Merge of active β-catenin and DAPI (blue); bar, 10 μm.

(D) Quantification of active β-catenin staining from three biological replicates in cells stained as described in (C). The intensity of staining was determined on the total area of cells and on the area of nuclei only; >30 cells were counted per condition in each experiment. Results are reported as the relative fold change compared to EV of the ratio of active β-catenin staining in nuclei/total area per cell ± standard deviation; p-value <0.05 *.

2.4 Epn3-induced EMT in MCF10A cells is TCF4 dependent.

To test whether the E-Cadherin/β-catenin/TCF-LEF axis could be responsible for the induction of the EMT transcriptional program in MCF10A Epn3-overexpressing cells, we silenced the β-catenin gene, or the most studied member of the TCF-LEF family, *TCF7L2* (also known as TCF4)[202], using siRNAs. We used as negative controls, mock transfected or not treated cells, as previously described, and as a positive control, TGFβR1 KD cells. Cells were analyzed by WB to assess the levels of EMT markers. While β-catenin KD did not affect EMT markers, TCF4 KD, was able to revert the levels of N-Cadherin and Vimentin almost to the levels of control MCF10A-EV cells, similarly to our positive control of TGFβR1 KD (**Fig. 8A**).

Once again, the morphology of cells was in agreement with the WB analysis, with the negative control and β-catenin KD cells displaying a mesenchymal morphology, while TGFβR1 KD and TCF4 KD cells displayed an epithelial phenotype (**Fig. 8B**). This analysis also gave us a possible explanation as why β-catenin KD was not able to revert EMT. In fact, β-catenin KD resulted in the destabilization of cell-to-cell junctions and therefore cells grew very sparsely. This is probably due to the fact that the majority of β-catenin is present not as the active β-catenin in the nucleus, but as the phosphorylated β-catenin at the adherent junctions, which is consequently destabilized upon KD. Instead, TCF4 KD acts specifically on the active β-catenin/TCF complex in the nucleus that is responsible for activation of the EMT transcriptional program. Accordingly, MCF10A-Epn3 TCF4 KD cells showed an epithelial phenotype.

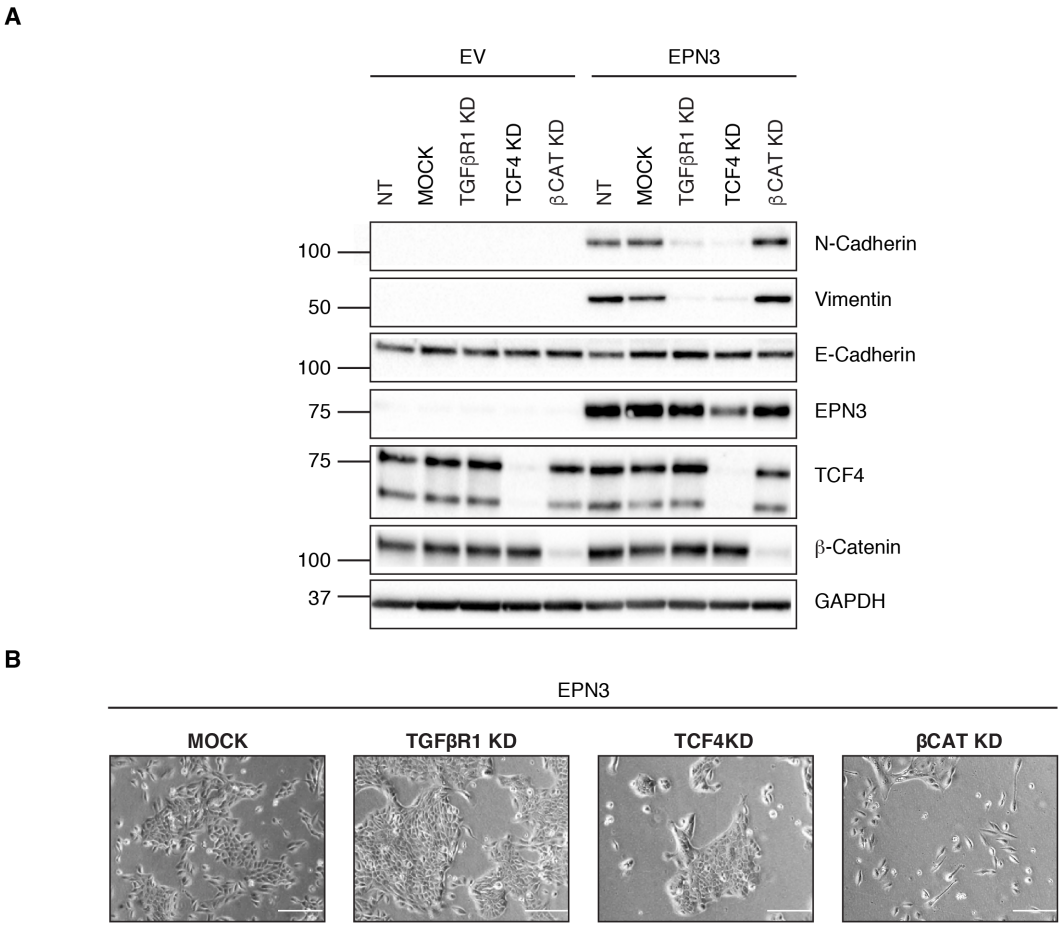


Fig. 8: Epn3-induced EMT in MCF10A is TCF4 dependent.

(A) WB analysis from a representative experiment of the expression of EMT markers (N-Cadherin, Vimentin, E-Cadherin) and controls (Epn3, TCF4, β-catenin) in MCF10A EV and Epn3-overexpressing cells; not treated (NT), transfected with only lipofectamine (MOCK) or transfected with a pool of siRNAs to knockdown TGFβR1, TCF4 and β-catenin (TGFβR1 KD, TCF4 KD and βCAT KD, respectively). Cells were lysed 7 days after transfection. GAPDH was used as a loading control. MW markers are shown on the left.

(B) Phase-contrast microscopy of representative images of mock, TGFβR1 KD, TCF4 KD and βCAT KD MCF10A Epn3-overexpressing cells; bar, 250 μm.

2.5 The TCF/LEF and TGFβ signaling pathways crosstalk in Epn3-overexpressing cells.

The fact that TCF4 or TGFβR1 silencing was able to revert Epn3-induced EMT to the same extent suggests that the two pathways are converging on EMT activation and behaving as non-redundant pathways, i.e., abrogation of one has the same effect as abrogation of the other, and that both are needed to sustain the mesenchymal phenotype. Therefore, we decided to investigate the relationship between the TCF4 and TGFβ pathways and the eventual crosstalk between them. For this reason, we performed a TCF4 KD followed by TGFβ stimulation at high dose of 5 ng/ml for 48 h in EV control and Epn3-overexpressing MCF10A cells. As negative controls we used mock transfected and not treated cells, as in previous experiments. In particular, we were interested in investigating whether TGFβ-signaling requires TCF4 to activate EMT in MCF10A cells, both in control EV or in Epn3-overexpressing cells. Thus, if TCF4 KD cells are not impaired in TGFβ-dependent EMT activation, it would suggest that the two pathways are acting independently of each other, while a reduced response to TGFβ in TCF4 KD cells would suggest that the TGFβ pathway is acting through the TCF4 pathway in the activation of EMT.

WB analysis of the expression of EMT markers in MCF10A-EV and -Epn3 cells upon TCF4 or TGFβR1 KD followed by 48 h of TGFβ stimulation, showed that Epn3-overexpressing cells were impaired in TGFβ-dependent EMT activation, with only a slight up-regulation of N-Cadherin, Vimentin, Snail, Fibronectin and ZEB1 expression, compared with negative controls (mock, NT), mirroring the behavior of TGFβR1 KD cells (**Fig. 9A**). However, no effects on MCF10A-EV cells were

observed at the protein level. Thus, we performed RT-qPCR experiments, which allowed us also to investigate the effect of TCF4 or TGFβR1 KD on TGFβ ligands and receptors and to better quantify the differences in MCF10A-EV cells. This analysis showed that N-Cadherin, Vimentin and importantly also TGFβ1 expression was significantly reduced in TGFβR1 KD and TCF4 KD MCF10A-Epn3 cells at the mRNA level, both at steady state and after TGFβ stimulation, compared to not treated Epn3 and to stimulated Epn3, respectively. ZEB1 and Serpine1 expression in TGFβR1 KD and TCF4 KD MCF10A-Epn3 cells, instead, was only significantly down-regulated following TGFβ stimulation compared to stimulated Epn3 control cells (**Fig. 9B**).

The fact that both MCF10A-EV and -Epn3 cells following TCF4 KD were strongly impaired in the up-regulation of EMT targets upon TGFβ stimulation, to a similar level to that of TGFβ–stimulated TGFβR1 KD cells, demonstrates that the TGFβ and TCF4 pathways are converging in the activation of EMT and that both pathways need to be functional to activate EMT.

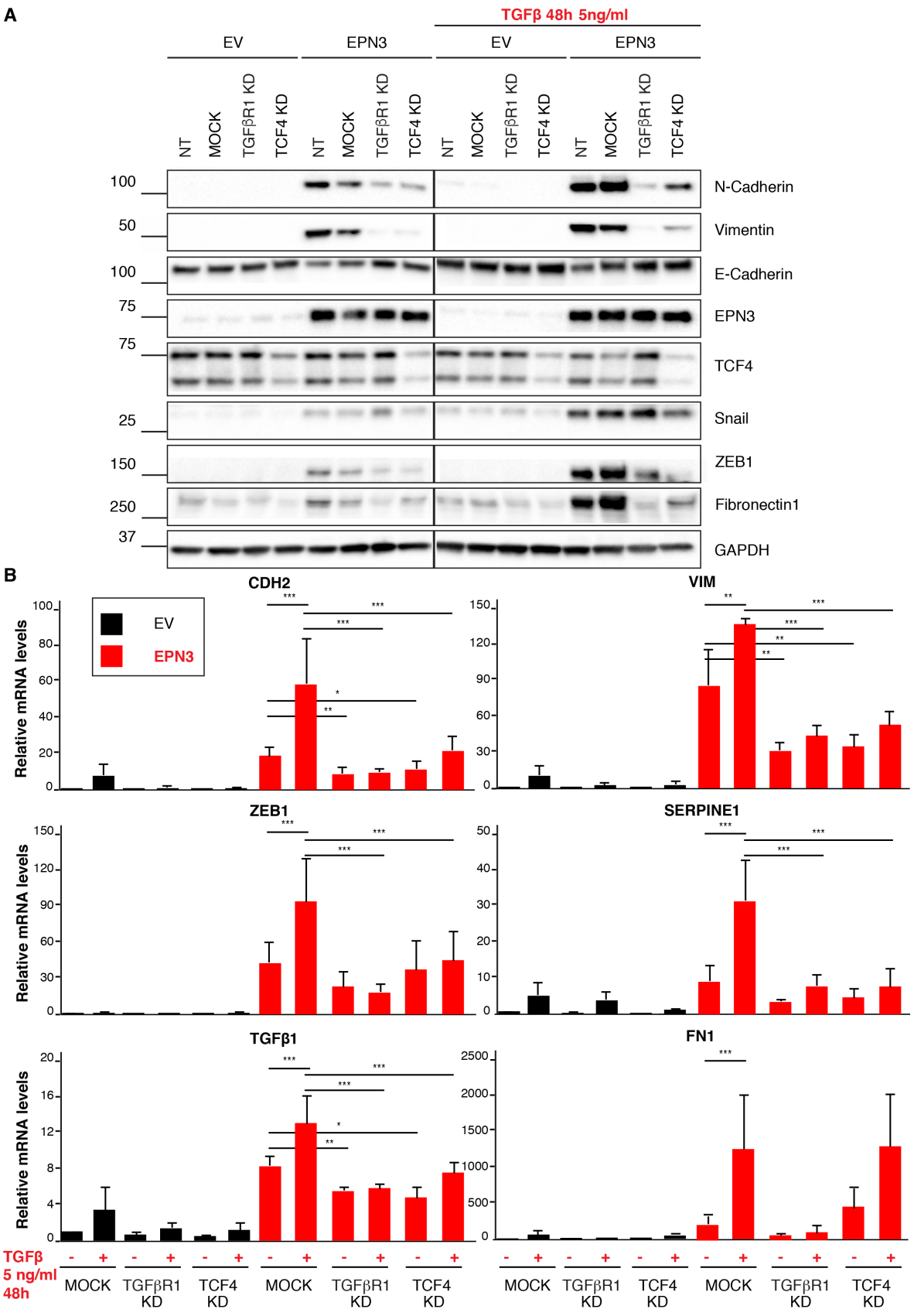


Fig. 9: Crosstalk between the TCF/LEF and TGFβ signaling pathways.

Continued on the next page

Fig. 9: Crosstalk between the TCF/LEF and TGFβ signaling pathways.

(A) WB analysis from a representative experiment of the expression of EMT markers (N-Cadherin, Vimentin, E-Cadherin) and of TGFβ signaling activation (Snail, ZEB1, Fibronectin) in MCF10A-EV and Epn3-overexpressing cells. Controls of TCF4 expression and Epn3 overexpression are also shown. Cells were lysed 7 days after transfection with siRNAs to KD TCF4 or TGFβR1, or with lipofectamine alone (MOCK), and following stimulation with TGFβ for 48 h at 5 ng/ml, or not (NT, not treated). GAPDH was used as a loading control. MW markers are shown on the left.

(B) RT-qPCR analysis of mesenchymal markers *CDH2, VIM, ZEB1, SERPINE1, TGFβ1* and *FN1* in MOCK, TGFβR1 KD and TCF4 KD samples in MCF10A EV and Epn3-overexpressing cells. Data were normalized on the average expression of *18S-ACTB-GAPDH* and results are reported as relative mRNA expression compared to EV mock samples ± standard deviation of three biological replicates, each performed in technical triplicates; p-value <0.05 *; p-value <0.01**; p-value <0.001 ***.

3. Investigation of the contribution of Epn3 to breast tumorigenesis *in vivo*

3.1 Tumorigenic Epn3-overexpressing MCF10DCIS.com breast cancer cells partially recapitulate Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells

In order to study the contribution of Epn3 to breast tumorigenesis *in vivo*, we first asked if the mechanisms identified in non-tumorigenic MCF10A cells were recapitulated in a tumorigenic cell model system. For this reason, we initially used the MCF10DCIS.com cell line, since it belongs to a unique progression model of MCF10 isogenic cell lines, therefore representing the most similar cell line to that of MCF10A, however tumorigenic and commonly used in the study of BC progression[88-90] (see **Discussion** for details on this cell line, paragraph **2.1 Tumorigenic Epn3-overexpressing MCF10DCIS.com and HCC1569 breast cancer cells partially recapitulate Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells**).

We overexpressed Epn3 (and EV control) through lentiviral infection in MCF10DCIS.com (from now on indicated as DCIS) and compared them to MCF10A EV- or Epn3 in WB analysis, as shown in **Fig. 10A**. MCF10DCIS.com showed a partial up-regulation of EMT markers N-Cadherin and Vimentin already in basal condition. Nevertheless, these markers (particularly Vimentin) were further up-regulated in Epn3 overexpressing cells compared to EV cells, however not reaching the levels observed for MCF10A-Epn3 cells. This difference may be attributed to the

difference in overexpression levels of Epn3, which are much higher in MCF10A cells compared to DCIS cells and/or to the fact that DCIS start form an already partial EMT state. Moreover, the increase in the levels of N-Cadherin (*CDH2*) and Vimentin (*VIM*) in Epn3-overexpressing DCIS cells was also visible at the transcriptional level, together with EMT master regulator *ZEB1* (**Fig. 10B**), although again not reaching the levels of the up-regulation of MCF10A-Epn3 (compare **Fig. 1D** and **Fig. 10B**).

To investigate if also in this model system there is cooperation of Epn3 with TGFβ, we stimulated EV and Epn3-overexpressing DCIS cells with recombinant TGFβ1 at low dose of 0.75 ng/ml (as employed previously for MCF10A cells) and we performed WB analysis at short time points (90 minutes and 24 h) and RT-qPCR analysis at a short and long time point (24 h and 7 days, respectively). As shown in **Fig. 10C** TGFβ stimulation was able to enhance Epn3-dependent EMT, as shown by the progressive increase in Vimentin protein levels (N-Cadherin, instead, does not change much between EV and Epn3 stimulated cells). Moreover, as shown in RT-qPCR in **Fig. 10D**, TGFβ stimulation was able to increase the transcription of TGFβ1, suggesting that also in this cell line there might be an establishment of a positive autocrine TGFβ loop.

To investigate on this hypothesis, we performed a TGFβR1 KD by infecting DCIS EV and Epn3-overexpressing cells with a lentiviral construct encoding for an shRNA targeting TGFβR1 (shTGFβR1) or shEV and carrying a GFP reporter gene. WB analysis, shown in **Fig. 10E**, revealed a similar decrease in the levels of N-Cadherin in EV and Epn3-overexpressing cells and a reversion in the Vimentin increase exclusively in Epn3-overexpressing cells, suggesting that also the partial EMT induction in DCIS-Epn3 cells is indeed TGFβ-dependent. Representative

images in **Fig. 10F** show a clear change in cell morphology: Epn3 shEV are sparse cells, whereas Epn3 shTGFβR1 are more compact and rounded cells, especially in the areas carrying the highest levels of the shRNA, evaluated by the fluorescence of the GFP reporter gene.

Thus, the DCIS cell model system recapitulates at least in part the phenotypes observed in non-tumorigenic MCF10A cells: Epn3 overexpression induces the EMT transcriptional program, which is TGFβ-dependent. Therefore, these cells could represent a good model system for further *in vivo* studies.

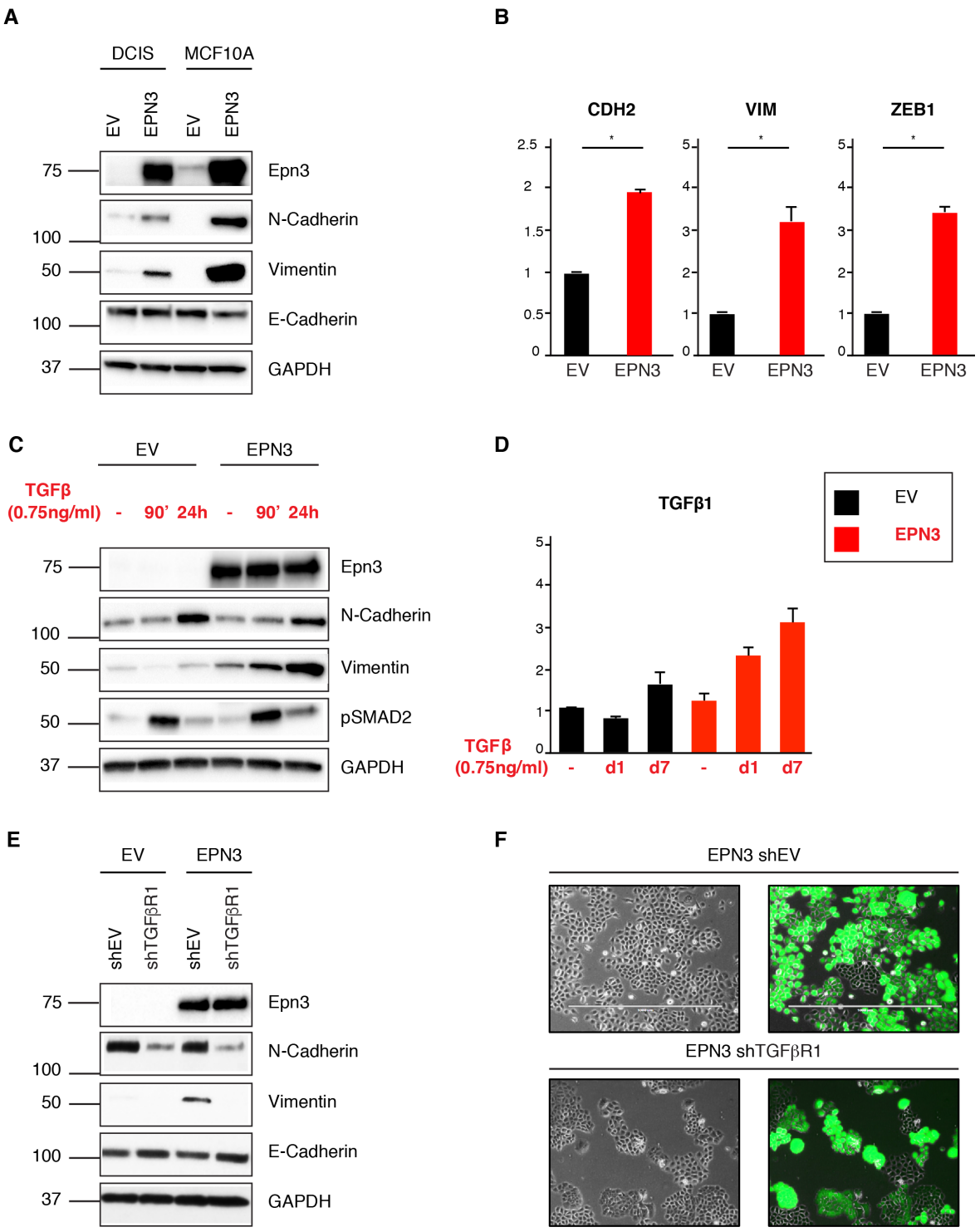


Fig. 10: Epn3-overexpressing MCF10DCIS.com partially recapitulate the phenotypes observed in non-tumorigenic MCF10A.

Continued on the next page

Fig. 10: Epn3-overexpressing MCF10DCIS.com partially recapitulate the phenotypes observed in non-tumorigenic MCF10A.

(A) MCF10DCIS.com (indicated as DCIS) cells were infected with empty vector (EV) or vector expressing Epn3 (EPN3). WB analysis of Epn3 and EMT markers (N-Cadherin, Vimentin, E-Cadherin) in DCIS *vs.* MCF10A EV or Epn3-overexpressing cells. GAPDH was used as a loading control. MW markers are shown on the left.

(B) RT-qPCR of EMT markers N-Cadherin (*CDH2*), Vimentin (*VIM*) and ZEB1 (*ZEB1*) in DCIS EV and Epn3-overexpressing cells described in (A). Data were normalized on the average expression of *18S-ACTB-GAPDH* and results are reported as relative mRNA expression compared to EV ± standard deviation of two biological replicates, each performed in technical triplicates; p-value <0.05 *.

(C) WB analysis of Epn3, EMT markers (N-Cadherin, Vimentin, E-cadherin) and of TGFβ signaling (pSMAD2) in DCIS EV and Epn3-overexpressing cells not treated (-) or treated for the indicated time points with 0.75 ng/ml recombinant TGFβ1. GAPDH was used as a loading control. MW markers are shown on the left.

(D) RT-qPCR of TGFβ ligand 1 (*TGFβ1*) in DCIS EV and Epn3-overexpressing cells treated with TGFβ, as described in (C). Data were normalized on the average expression of *18S-ACTB-GAPDH* and results are reported as relative mRNA expression compared to EV ± standard deviation of technical triplicates.

(E) WB analysis of Epn3 and EMT markers (N-Cadherin, Vimentin, E-cadherin) of DCIS EV and Epn3-overexpressing cells infected with an shRNA targeting TGFβR1 (shTGFβR1) or shEV and carrying a GFP reporter gene. GAPDH was used as a loading control. MW markers are shown on the left.

(F) Phase contrast microscopy of shEV or shTGFβR1 DCIS Epn3-overexpressing cells described in (E); bar, 250 μm.

3.2 Epn3 overexpression in tumorigenic HCC1569 breast cancer cells partially recapitulates Epn3-dependent phenotypes observed in MCF10A cells and increases their *in vivo* tumorigenic potential

MCF10A cells are non-tumorigenic cells and Epn3 or Twist overexpression alone is not sufficient for a full transformation, as indicated by their inability of forming colonies in the soft agar colony formation assay, commonly used to evaluate cellular transformation *in vitro* (data not shown, Iavarone *et al.*, in preparation). Therefore, to

investigate the contribution of Epn3 to breast tumorigenesis together with other genetic lesions, among which *HER2* amplification, we overexpressed Epn3 (or EV as a control) in the HCC1569 BC cell line using a lentiviral expression system. These cells are HER2-positive, ER- and PR-negative, and p53-negative[203]. Moreover, they already express much higher levels of mesenchymal marker Vimentin and lower levels of epithelial marker E-Cadherin, compared to MCF10A-EV cells, but only slightly higher levels of mesenchymal marker N-Cadherin. Nevertheless, upon Epn3 overexpression, N-Cadherin levels were further increased in HCC1569 cells similar to the levels observed in MCF10A-Epn3 cells, whereas Vimentin was only slightly up-regulated (**Fig. 11A**), suggesting that Epn3 is potentiating the EMT that is already present in HCC1569 cells.

We aimed to explore if also in this cell line TGFβ stimulation was able to enhance Epn3-dependent EMT. Therefore, we stimulated HCC1569-EV and -Epn3 cells with high dose TGFβ (5 ng/ml) in a time-course experiment, for 2 and 7 days, and then analyzed cell lysates by WB. As shown in **Fig. 11B**, TGFβ stimulation progressively increased the levels of N-Cadherin and, to a lower extent, of Vimentin, particularly in Epn3-overexpressing cells, suggesting that also in this BC cell line there might be some form of cooperation between Epn3 overexpression and TGFβ signaling, although the extent of interplay, as in the case of DCIS cells, is milder as compared to MCF10A non-tumorigenic cells.

Taking advantage of the fact that HCC1569 are tumorigenic, we used them for *in vivo* transplantation assays. In particular, we injected 10^6 or 10^5 EV or Epn3-overexpressing cells into the mammary fat pads of NSG mice.

First of all, we verified the levels of Epn3-overexpression in tumors generated *in vivo*. Therefore, we collected HCC1569-derived tumors and lysed them using of

magnetic beads (see **Materials and Methods**, paragraph **4 In vivo xenografts** for details) and performed WB analysis. As shown in **Fig. 11C,** Epn3-overexpression levels *in vivo* were similar to those observed in the parental cells *in vitro* (**Fig. 11C**, **second lane**); however, greater variability in expression levels was observed in tumors compared to cells grown in 2D culture, probably due to the contamination of other cell types not-overexpressing Epn3 in cell lysates (blood cells, fat cells, immune cells and other cells deriving from tumor stroma), since we did not purify epithelial cells upon tumor digestion.

Importantly, HCC1569-Epn3-derived tumors generated from both injections at 10^6 or 10^5 were considerably larger than control tumors (**Fig. 11D**), meaning that high levels of Epn3 are able to increase *in vivo* tumor potential of HCC1569 cells upon orthotopic transplantation.

Taken together, these data suggest that Epn3-overexpressing HCC1569 cells partially recapitulate Epn3-dependent phenotypes observed in MCF10A cells, implying that the mechanisms identified for non-tumorigenic cells MCF10A upon Epn3 overexpression, may be responsible for the acquisition of some features frequently observed in tumorigenic cells (EMT and CSC properties); however, the switch from non-tumorigenic to tumorigenic, or, the further increase in tumorigenic potential, probably requires the cooperation of Epn3 with other genetic lesions, for example p53 loss and/or HER2 overexpression, as in the case of HCC1569. Indeed, although HCC1569 cells have other genetic lesions, Epn3 overexpression provides a further advantage to the cells for tumor growth.

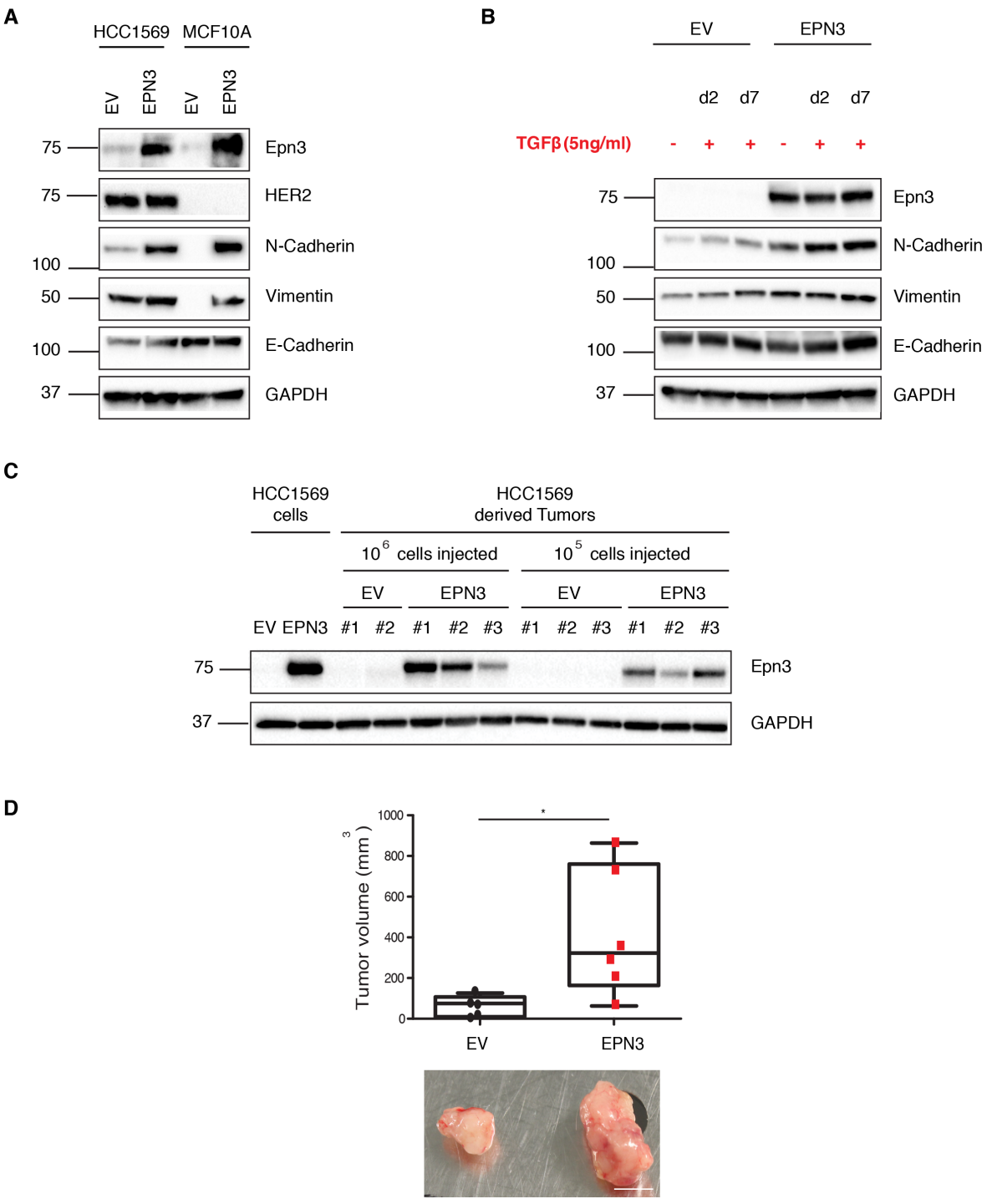


Fig. 11: Epn3-overexpression in HCC1569 breast cancer cells partially recapitulates Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells and increases their *in vivo* tumorigenic potential.

Continued on the next page

Fig. 11: Epn3-overexpression in HCC1569 breast cancer cells partially recapitulates Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells and increases their *in vivo* tumorigenic potential.

(A) HCC1569 cells were infected with empty vector (EV) or vector expressing Epn3 (EPN3). WB analysis of the expression of Epn3 and EMT markers (N-Cadherin, Vimentin, E-Cadherin) in HCC1569- *vs.* MCF10A-EV or -Epn3 cells was performed. GAPDH was used as a loading control. MW markers are shown on the left.

(B) WB analysis of the expression of EMT markers (N-Cadherin, Vimentin, E-Cadherin) in HCC1569-EV and -Epn3 cells, not treated (-), or treated for the indicated time points with 5 ng/ml recombinant TGFβ1. GAPDH was used as a loading control. MW markers are shown on the left.

(C) 10^6 or 10^5 HCC1569 EV or Epn3-overexpressing cells were injected into the mammary fat pads of NSG mice. Mice were sacrificed after 44 days (in the case of the injection of 10^6 cells) or 64 days (in the case of the injection of 10^5 cells) and tumors were lysed in order to obtain protein extract.
WB analysis of Epn3 expression in *in vitro* HCC1569 cell cultures (left two lanes) and in HCC1569-derived tumors (#1, #2, or #3). GAPDH was used as a loading control. MW markers are shown on the left.

(D) *Upper panel*: Tumor size, calculated according to the standard formula (length x width^2)/2, is reported as box-plot for all tumors generated (10^5 or 10^6 samples), described in (C).
Bottom panel: Representative picture of a HCC1569 EV-derived tumor (left) or HCC1569 Epn3-derived tumor (right); bar, 5mm; p-value <0.05 *.

DISCUSSION

1. Epn3 overexpression induces EMT and expansion of the SC-like compartment in non-tumorigenic mammary epithelial cells

1.1 Dissection of the molecular mechanisms: proposed model.

Following-up on our evidence of gene amplification and overexpression of the endocytic protein Epn3 in human BC (see **Preliminary Data**), we initially investigated the mechanisms through which overexpression of Epn3 can contribute to tumorigenesis. To this end, we overexpressed Epn3 in non-tumorigenic mammary epithelial cells, MCF10A, and studied the eventual phenotypes leading to cellular transformation. The sum of the results obtained with this first set of experiments allowed us to propose the model shown in **Fig. 1**. We speculate that the first event induced by Epn3 overexpression in MCF10A cells is enhanced E-Cadherin endocytosis and destabilization of the cell-to-cell junctions (**A**). This event allows the detachment of β-catenin from the junctions, its translocation and accumulation in the nucleus, and the activation of β-catenin/TCF-LEF-mediated transcription (**B**). This, in turn, leads to the induction of EMT master regulators (ZEB1, Snail, Twist), consequent up-regulation of mesenchymal markers, N-Cadherin, Vimentin, ECM components, and the proteases Fibronectin and Serpine, and, importantly, the

transcription of TGFβ receptors and ligands (**C**). TGFβ ligands are secreted into the cell medium (**D**) where they bind to their respective receptors on the cell surface, creating a positive feedback loop acting in an autocrine fashion. This feedback loop, sustains - through the TGFβ signaling pathway - E-Cadherin endocytosis (**E**), TGFβ ligand production (**F**) and Epn3-dependent EMT (**G**). All these events contribute to the maintenance of the EMT phenotype and to the consequent generation of cells with CSC-like features (**H**), two events that we showed to be causally linked, causing the expansion of the SC-like compartment in Epn3-overexpressing MCF10A cells.

In addition, we also have data to support the hypothesis that the phenotypes induced by Epn3-overexpression are an exaggeration of the physiological endocytic function of Epn3, acting on E-Cadherin endocytosis. Indeed, Epn3 co-localizes with E-Cadherin in endocytic vesicles and Epn3 KD is able to block E-Cadherin at the PM (data not shown, Iavarone *et al.*, in preparation).

Finally, we speculate that similar mechanisms might be acting in human breast tumors where Epn3 gene amplification and/or overexpression contribute to malignancy together with other genetic lesions. Indeed, based on these observations and the known correlation between EMT, CSCs and tumorigenesis, we reasoned that Epn3 is a novel oncogene in BC, whose overexpression could lead to the acquisition of SC-like properties through EMT activation, and, ultimately, to the emergence of CSCs.

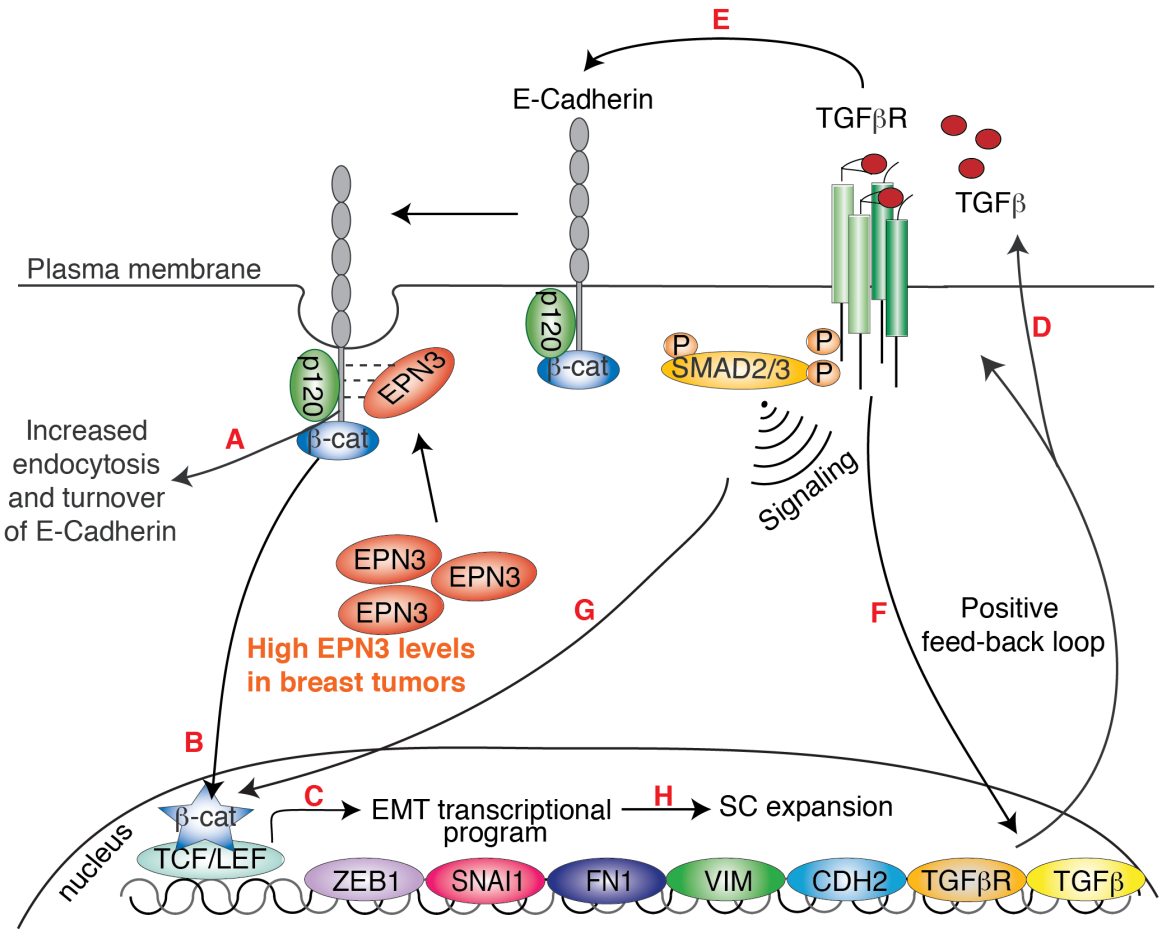


Fig. 1: Proposed model for Epn3 overexpression-dependent phenotypes in non-tumorigenic MCF10A cells and breast cancer cells.

See text for a detailed description of the proposed model.

1.2 Partial EMT and its implications.

Epn3 overexpression in MCF10A cells causes morphological changes and up-regulation of mesenchymal markers, N-Cadherin, Vimentin, EMT master regulators (ZEB1, Snail, Twist), and ECM components and proteases, e.g., Fibronectin and Serpine. However, a full EMT is characterized also by loss of E-Cadherin[92], which

does not occur upon Epn3 overexpression (for a more detailed discussion of E-Cadherin upon Epn3 overexpression see **Discussion** paragraph **1.6 E-Cadherin in EMT and breast cancer**). In addition, Epn3-dependent EMT does not recapitulate the more advanced phenotype obtained upon Twist overexpression, both at the transcriptional and at the morphological level. Accordingly, Epn3-overexpressing cells appear to resemble the so-called *partial EMT state*, recently reported in several papers, which is characterized by the concomitant presence of both epithelial and mesenchymal features[118,119]. In particular, the partial EMT state is characterized by medium levels of E-Cadherin and Vimentin, whereas the full mesenchymal state is characterized by low levels of E-Cadherin and high levels of Vimentin[204], which is what we observed in Epn3- and Twist-overexpressing cells, respectively (see **Results**, **Fig. 1**).

The result of partial EMT is therefore the mixture of epithelial (e.g., adhesion) and mesenchymal (e.g., migration) properties. This peculiarity allows cells to move collectively as clusters. In fact, experimental evidences show that if these clusters reach the bloodstream intact, they can give rise to and be detected as clusters of circulating tumor cells (CTCs)[119,181]. Moreover, cells in a partial EMT state have the highest plasticity, i.e., the ability to switch between invasive and proliferative modes, and between epithelial and mesenchymal behavior. Therefore, it has been proposed that these cells are more prone to metastasize to distant organs[180]. In agreement, it is now well-established that EMT favors invasion and migration to reach distant organs, however, effective-tumor initiation requires the opposite transition, MET, to regain epithelial characteristics and form secondary tumors or macrometastases. In fact, EMT and MET enable solid tumor formation, 90% of which are of epithelial nature (carcinomas)[179].

As a consequence, Epn3 overexpression, by inducing a plastic partial EMT state might favor metastatic colonization of tumor cells. Consistently, Epn3 overexpression positively correlates with, and is a prognostic factor of, the occurrence of distant metastasis in human BC (see **Aim of the project**, **Fig. 2**). Furthermore, also a recent study of publicly available databases aimed at identifying genes linked to metastatic recurrence after surgery, found that Epn3 is prognostic for both BC and non-small cell lung cancer[34].

1.3 EMT and the generation of cells with CSC properties.

Another consequence of EMT is the generation of CSCs[165]. MCF10A cells are non-tumorigenic, nevertheless, through EMT upon Epn3 overexpression, differentiated cells are able to acquire some CSCs properties, such as the ability to form mammospheres at a higher efficiency and the expression of CSC markers ($CD44^{high}/CD24^{low}$) (see **Results**, **Fig. 2**). These markers correlate both with the SC-like population and with the mesenchymal population (see **Results**, **Fig. 3**), therefore, suggesting that Epn3 overexpression contributes to breast tumorigenesis through EMT and CSC appearance.

Indeed, we were able to show that the establishment of the EMT phenotype and the generation of cells with CSC properties are two events causally linked. Firstly, by WB and RT-qPCR analysis of FACS-sorted cells, we demonstrated that Epn3 (or Twist) cells expressing CSC markers are the same cells expressing mesenchymal markers (see **Results**, **Fig. 3**). Moreover, we showed that the reversion of the EMT phenotype in Epn3 cells through the abrogation of TGFβ signaling is accompanied

by reversion of the expansion of SC-like compartment and, accordingly, a partial reversion of the EMT phenotype for Twist cells is accompanied by a partial reversion of the CSC markers expression (see **Results**, **Fig. 6**). We speculate that the same effect would be achieved through abrogation of the TCF-LEF signaling pathway, although we have not tested yet it for CSC markers.

Furthermore, by a time course analysis of the expression of CD44/CD24 cell surface markers on FACS-sorted $CD44^{high}/CD24^{low}$ and $CD44^{low}/CD24^{high}$ populations in Epn3- and Twist-overexpressing cells, we determined that while the mesenchymal population, once sorted from bulk population, is metastable after 7 days in culture, the epithelial population is highly plastic and after 7 days in culture it reaches (in the case of Epn3-, or almost reaches in the case of Twist-overexpressing cells) the equilibrium of $CD44^{high}/CD24^{low}$ *vs.* $CD44^{low}/CD24^{high}$ populations of unsorted cultures (see **Results**, **Fig. 4**). This finding suggests that Epn3 (or Twist)-overexpression might maintain the metastable mesenchymal population and drive differentiated epithelial cells to a more naïve, SC-like state with CSC properties, suggesting once again that Epn3 overexpression could contribute to breast tumorigenesis through promoting the appearance of CSCs.

Importantly, CSCs are widely believed to be responsible for cancer recurrence, therapy failure and metastasis of different cancers, including BC[175]. This hypothesis is again in agreement with the fact that Epn3 expression in human BC correlated with metastasis, rather than loco-regional relapse, suggesting that Epn3 might act by different mechanisms, i.e., by promoting a plastic EMT status and by the generation of CSCs, to favor metastatic colonization.

However, we have not yet investigated whether the expansion of the SC-like compartment upon Epn3 overexpression is only due to the generation of SC-like cells from differentiated progenitors via EMT, or whether it can also be caused by

the skewing of the modality of SC division from ACD to SCD. This issue merits further investigation.

1.4 The TGFβ signaling pathway in EMT and breast cancer.

Epn3-overexpressing MCF10A cells show an increased level of transcription of TGFβ ligands and receptors compared to EV cells (~15-fold increase in TGFβ1, ~4-fold increase in TGFβ2, and ~2-fold increase in TGFβR1 and TGFBR2). We found that this up-regulation of TGFβ ligands and receptors in Epn3-overexpressing cells is able to create an autocrine positive feedback loop and to cooperate with Epn3 overexpression itself to establish and sustain EMT and the expansion of the SC-like compartment. Indeed, we found that the levels of endogenous TGFβ1 secreted in the cell culture media harvested by Epn3-overexpressing cells are higher than those of EV cells (see **Results**, **Fig 5**). Moreover, stimulation of cells with limiting doses of recombinant TGFβ1 is able to enhance Epn3-induced EMT, and importantly, the transcription of endogenous TGFβ1 itself, demonstrating the existence of a transcriptional, self-sustaining, positive feedback loop, which sustains the mesenchymal state (see **Results**, **Fig 5**). Indeed, the Epn3-dependent phenotypes, EMT and expansion of the SC-like compartment, are dependent on TGFβ and both were reverted upon abrogation of TGFβ signaling, both through KD of TGFβR1 or through chemical inhibition of the kinase activity (see **Results**, **Fig. 6**).

Among the multiple signaling pathways, the TGFβ pathway has a predominant role in EMT activation and sustainment, ECM remodeling, cell migration and invasion, and cancer development[127,189]. In cancer, and specifically in BC, TGFβ

signaling has a dual role, which led to the "*good cop bad cop"* metaphor. In the early stages of carcinogenesis, TGFβ has a tumor suppressor role, supporting cell cycle arrest, apoptosis, inhibition of cell proliferation and activation of immune cells to eradicate transformed cells (*good cop*). However, through genetic and epigenetic changes involving other pathways and not yet clearly deciphered, in the more advanced stages of carcinogenesis, TGFβ switches to an oncogenic role, sustaining the tumorigenesis by contributing to immune escape, thypoxic environment, angiogenesis, EMT, invasion, migration and metastatization (*bad cop*)[205]. These latter functions of TGFβ are the ones that cooperate with Epn3 in promoting EMT and acquisition of invasive phenotypes.

1.5 Feedback loops in EMT and breast cancer.

We found that the up-regulation of TGFβ in Epn3-overexpressing cells creates an autocrine positive feedback loop that could be reverted only with the abrogation of the TGFβ pathway itself (see **Results**, **Fig. 6**) or of the TCF-LEF signaling pathway (see **Results**, **Fig. 8**). Importantly, and maybe not so surprisingly, this feedback loop, once established, cannot be reverted by Epn3 knockdown itself (data not shown, Iavarone *et al.*, in preparation). Indeed, once feedback loops are established, they become self-sustaining and independent of the initial event/genetic lesion, for example in the case of TGFβ.

It is well known that TGFβ is an EMT-inducing factor that activates transcription factors, namely Snail and ZEB1[189], both of which are indeed up-regulated in Epn3-

overexpressing cells. To date, several feedback loops involved in TGFβ-dependent EMT progression have been identified. These feedback loops also include the post-transcriptional regulation of EMT master regulators mediated by microRNAs. In particular, a double negative feedback loop regulates the expression of Snail and members of miR-34 family[206]. Another double negative feedback loop regulates ZEB1 and members of miR-200 family[117]. Lastly, an autocrine TGFβ-miR-200 feedback loop can also be established[207]. In other words, the miR-200 family represses ZEB1 and ZEB1 represses the two genetic loci encoding for all 5 members of the miR-200 family. Similarly, the miR-34 family represses Snail and Snail represses the two genetic loci encoding all 3 members of the miR-34 family. The finely balanced levels of the members of the double negative feedback loops determines the state of the cell: the miR-200 and miR-34 families predominate in epithelial cells and, by repressing the expression of ZEB1 and Snail, they allow high expression of E-Cadherin; *vice versa*, ZEB1 and Snail predominate in mesenchymal cells repressing the expression of E-Cadherin and other epithelial genes.

Despite the fact that the exact mechanism by which these interlinked feedback loops coordinately modulate EMT is still controversial, different models have been proposed. In the so-called CBS model (cascading bistable switches), EMT proceeds in a two-step program: a first reversible switch to the partial EMT state is due to the Snail/miR-34 feedback loop with increased levels of Snail only; the ZEB1/miR-200 feedback loop with increased levels of ZEB1 is responsible for the second switch to an irreversible mesenchymal state[208]. In the so called TCS model (ternary chimera switch), TGFβ-dependent EMT progression is a ternary switch in which the ZEB1/miR-200 feedback loop is responsible on its own for the 3 steady states: the epithelial, the partial EMT and the mesenchymal state, with low, medium and high

levels of ZEB1, respectively, and the opposite levels of the miR-200 family members. In contrast, the Snail/miR-34 feedback loop is linear and has a role in the integration of intracellular and extracellular signals; the authors define it as a *noise-buffering integrator*[209].

Moreover, also a positive feedback loop between CD44 and ZEB1 (both of which are up-regulated in Epn3-overexpressing cells) has been identified. In this loop, *ZEB1* regulates the splicing of the *CD44* gene by repression of the gene *ESRP1* in breast and pancreatic cancer cells, resulting in a shift from variant CD44 forms to standard CD44. This shift in turn activates the expression of ZEB1, creating a positive feedback loop[167].

We speculate that besides the TGFβ autocrine loop, also miRNA- or splicing-dependent loops could be acting in MCF10A Epn3-overexpressing cells. We have not yet tested the levels nor the impact of the miR-34 and miR-200 family members or CD44 splicing in our setting, an issue that would be interesting to investigate in the future.

1.6 E-Cadherin in EMT and breast cancer.

A hallmark in EMT and cancer progression is loss of E-Cadherin through different mechanisms (transcriptional repression, protein degradation, inactivating somatic mutations), which allow the detachment of cell-to-cell junctions, favoring migration, invasion and extravasation steps[104]. As previously discussed, our results show a partial EMT in Epn3-overexpressing cells, with no significant repression of E-Cadherin transcription in not sorted cells (see **Results**, **Fig. 1**) and only a partial

decrease in E-Cadherin protein expression, not evident in all the experiments (compare WB in **Results**, **Fig. 1**, **Fig. 5**, **Fig. 6**, **Fig. 8**, **Fig. 9**). Of note, only FACS-sorted $CD44^{high}/CD24^{low}$ Epn3-overexpressing cells displayed a substantial decrease of E-Cadherin not only at the protein but also at the mRNA level, as opposed to the unsorted population (compare **Results**, RTqPCR for *CDH1* in **Fig. 3D** and **Fig. 1D**), suggesting therefore that only a minor fraction of Epn3-overexpressing cells have transcriptionally lost E-Cadherin expression and that post-transcriptional mechanisms are responsible for its down-regulation at the protein level in the unsorted population. Moreover, a strong decrease at the protein and mRNA levels could be obtained by stimulating Epn3-overexpressing cells with TGFβ at high dose (5 ng/ml) for long time points (14 days) (see **Results**, WB in **Fig. 5**, mRNA not shown), suggesting that TGFβ treatment is able to convert the partial EMT state of Epn3-overexpressing cells to a more advanced, full, EMT state.

Thus, our data are in agreement with Epn3 overexpression inducing a partial EMT state, in which E-Cadherin is transcriptionally lost only in a minor fraction of cells, whereas the majority of cells display normal transcriptional levels, but decreased protein levels and, importantly, down-regulation of E-Cadherin specifically at the PM, i.e., the site where E-Cadherin exerts its main function at junctions. This loss of E-Cadherin at the PM results in the same outcome as total loss of the protein. Thus, endocytosis, which has a major role in protein recycling and re-localization to finely regulate the composition of the PM in response to intra- and extra-cellular stimuli[133,138], appears to represent an alternative mechanism for the induction of EMT or partial EMT.

As we discussed, the partial EMT state, with normal levels of E-Cadherin protein expression is the one with the highest plasticity that favors the switch from mesenchymal back to epithelial in the reverse MET process during metastatic

colonization. Indeed, there is evidence that deletion of E-Cadherin in primary mammary murine epithelial cells leads to the disruption of junctions but not to dissemination of cells, despite the reduction of multiple members of the adherens junction complex. Conversely, Twist1 overexpression is able to induce robust dissemination of primary mammary epithelial cells and this process requires E-Cadherin, since E-Cadherin KD blocks Twist-dependent dissemination of cells[210]. Once again, as previously mentioned, the high levels of plasticity favored by the presence of E-Cadherin and the partial EMT state might be important for Epn3 in human BC, since Epn3 is a prognostic factor of the occurrence of distant metastasis in patients.

Lastly, a recent study has also correlated partial and complete EMT states to the different molecular subtypes in BC cell lines Luminal A, luminal B and normal-like subtypes correlated with a partial EMT state with persistent expression of E-Cadherin at the mRNA level and protein re-localization into the cytoplasm. The basal-like subtype instead had loss of E-Cadherin expression both at transcript and protein level. Moreover, the type of EMT had an impact on cell migratory phenotypes, with collective migration in the case of partial EMT and luminal cancers, and single cell migration in the case of complete EMT and basal cancers[120]. Therefore, it would be interesting to investigate the migratory phenotypes of Epn3-overexpressing cells or Epn3-amplified/overexpressed BC, also according to their specific molecular subtypes.

1.7 β-catenin/TCF-LEF axis in EMT and breast cancer.

As a consequence of the destabilization of cell-cell junctions, we have evidence for a subsequent activation of β-catenin-dependent transcription leading to EMT. Indeed, we were able to score translocation of active (i.e., dephosphorylated) β-catenin to the nucleus (see **Results**, **Fig. 7**), where it is known to bind to the TCF/LEF transcription factors and act as a co-activator of transcription of several genes, including those activating the EMT transcriptional program. Although we were not able to obtain reversion of the EMT phenotype through β-catenin KD, the fact that we were able to revert the EMT phenotype upon TCF4 KD points to the involvement of the β-catenin/TCF-LEF pathway in the establishment of EMT mediated by Epn3 (see **Results**, **Fig. 8**). We speculate that this difference is due to the fact that the majority of β-catenin is not in its active form in the nucleus, but is instead in its phosphorylated inactive form at adherent junctions, which are consequently destabilized upon KD. Indeed, β-catenin KD resulted in the destabilization of cell-to-cell junctions in Epn3-overexpressing cells, which therefore grew very sparsely and could not revert to an epithelial-like morphology (see **Results**, **Fig. 8**). Instead, TCF4 KD acts specifically on the β-catenin/TCF complex in the nucleus, responsible for EMT activation, leaving intact the β-catenin pool at the junctions and allowing for the reacquisition of the epithelial phenotype.

TCF4 (gene name *TCF7L2*) is the most characterized member of the TCF/LEF family. It has a fundamental role in embryogenesis, especially in the gut and the central nervous system, and *TCF7L2* KO results in neonatal death. In the adult, it is the most expressed family member in the intestine and colon. To date, according to RefSeq, there are 16 protein-coding isoforms in human cells, which differ especially

in the composition of the C-terminus. Besides its role as a developmental transcription factor necessary for progenitor cell development, TCF4 is the primary transforming factor in human cancers, such as colorectal cancer, where the WNT/β-catenin/TCF-LEF pathway is frequently altered[211]. Mutations in the WNT pathway genes are rare in human BC, yet there is growing evidence of the importance of the WNT signaling pathway in BC. The role of TCF4 in BC is poorly characterized, however, recently, two *TCF7L2* polymorphisms have been found to be associated with familial BC risk[212]. Of note, CD44 is a direct target of TCF4[166], underlying its importance in our cell model system.

The role of β-catenin in BC, instead, is clearer and it is well known that its activation is associated with triple-negative and basal-like BCs, even in the absence of gene mutations[213]. Indeed, nuclear expression of β-catenin is also associated with shorter metastasis-free and overall survival[214]. It would be interesting therefore to investigate also the levels and the subcellular localization of β-catenin in our cohort of BC patients and to check if there is a correlation with the levels of expression of Epn3.

1.8 Crosstalk between the TGFβ and the TCF-LEF axis in EMT and breast cancer.

We found that single TGFβR1 KD and TCF4 KD are both able to revert Epn3-induced EMT (see **Results**, **Fig. 8**). In addition, upon TCF4 KD, MCF10A Epn3-overexpressing cells were strongly impaired in TGFβ-dependent EMT activation,

both at the protein and transcriptional level (see **Results**, **Fig. 9**). These results suggest that the TGFβ signaling pathway regulates EMT through the TCF4 transcription factor, presumably together with the SMAD transcription factors.

Interestingly, many levels of crosstalk exist between the WNT and the TGFβ signaling pathway. Firstly, ligand production for these two pathways is under the regulation of the other pathway. For example, the expression of bone morphogenetic protein 4 (BMP4) in human colon cancer cells HCT116, which contain an heterozygous-activating mutation of β-catenin (in-frame deletion of serine 45, therefore, constitutively dephosphorylated/active β-catenin) is indeed dependent on the expression of the active β-catenin allele[215]. A second level of crosstalk occurs in the cytoplasm where the inhibitory SMAD7 can directly bind to β-catenin and induce its degradation via SMURF2[216] or it can bind to the member of the β-catenin destruction complex Axin, inducing disassembly of the destruction complex and subsequent stabilization of β-catenin[217]. Moreover, the transcription factor SNAI1 and SNAI2, once up-regulated at the mRNA and protein levels by TGFβ1 and TGFβ2, promote the formation of the β-catenin/TCF4 complex that binds to TGFβ3 promoter[218]. Finally, and most importantly, in the nucleus, SMAD members and β-catenin/TCF-LEF regulate shared target gene expression[219,220] and, notably, also E-Cadherin repression in palate medial-edge epithelial cells has been shown to require a functional complex of SMAD2/SMAD4/LEF1 induced by TGFβ3[221].

We speculate that a similar crosstalk can occur also in MCF10A Epn3-overexpressing cells in the transcription of EMT related genes, since both pathways are active and both pathways are needed to establish EMT. Moreover, our data are also in line with the idea that a network of mutually reinforcing autocrine signals

involving the TGFβ and WNT pathways is required to ensure continued expression of EMT transcription factors[222].

1.9 Mechanisms inducing EMT that are active in Epn3-overexpressing MCF10A cells.

Different mechanisms are acting and cooperating in MCF10A Epn3-induced EMT. Firstly, there is a transcriptional regulation by EMT master regulators (ZEB1, Snail, Twist), since their own transcription and the transcription of their targets are increased, and this event is critical to finally establish the EMT phenotype in Epn3-overexpressing cells. In addition, post-transcriptional mechanisms, presumably mediated by microRNAs, are known to regulate the TGFβ autocrine loop and could also act in our system. Also post-transcriptional mechanisms of alternative splicing are known to be involved in the regulation of CD44, generating several variant forms (CD44v), which have been shown to be directly involved in EMT and invasion, and to correlate with many malignant tumors and with metastatic progression, and also used as a CSC marker[114,167]. These mechanisms are presumably involved in Epn3-induced EMT, although we have not investigated specific isoforms, since the antibody used in our tests recognizes epitope 1, which is present in all isoforms of CD44.

However, we think that these transcriptional and post-transcriptional mechanisms and loops that are finally established are a consequence of post-translational endocytic mechanisms induced by Epn3, regulating E-Cadherin protein re-localization and turnover. Indeed, our data strongly suggest that E-Cadherin

endocytosis and consequent destabilization of the cell-cell junctions is the first event induced by Epn3 overexpression in MCF10A cells, therefore underlying the importance of an endocytic-based mechanism in the subsequent activation of the other mechanisms involved in EMT induction.

To date, there is no direct link of a causal role of endocytosis in EMT induction and acquisition of migratory and invasive phenotypes. Indeed, many lines of evidence point to the importance of endocytosis in the regulation of protein recycling and re-localization, finely regulating the composition of the PM in response to intra- and extra-cellular stimuli, including E-Cadherin, which plays a pivotal role in EMT[133,138]. However, to date, no mechanism has been clearly described. Therefore, we believe our results provide one of the first examples of an endocytic-based mechanism responsible for the induction of EMT or partial EMT in cancer.

2. Epn3 overexpression in breast cancer cells partially recapitulates phenotypes observed in non-tumorigenic MCF10A-Epn3 cells and increases their *in vivo* tumorigenic potential

2.1 Epn3-overexpressing MCF10DCIS.com and HCC1569 breast cancer cells partially recapitulate Epn3-dependent phenotypes observed in non-tumorigenic MCF10A cells.

Epn3 overexpression in BC cells, MCF10DCIS.com and HCC1569, partially recapitulates Epn3-dependent phenotypes observed in non-tumorigenic MCF10A-Epn3 cells. Indeed, although both cell types already express a higher level of mesenchymal markers, Epn3 overexpression is able, at least in part, to further up-regulate these mesenchymal markers. In particular, MCF10DCIS.com cells, display higher levels of N-Cadherin and comparable levels of Vimentin; the opposite is true for HCC1569 cells, which show very high levels of Vimentin, whereas the levels of N-Cadherin are comparable to those of MCF10A cells. As a consequence, the most regulated marker by Epn3 is Vimentin in MCF10DCIS.com and N-Cadherin in HCC1569 cells. Moreover, both cell lines show at least a partial cooperation of Epn3 expression with TGFβ stimulation (see **Results**, **Fig. 10** and **Fig. 11**), and for MCF10DCIS.com, we could show EMT reversion upon abrogation of the TGFβ signaling pathway (not tested for HCC1569).

It is important to underline that the biology of transformation of these two cell lines is however different. HCC1569 are BC cells that are HER2-positive, ER- and PR-negative and p53-negative. In contrast, MCF10DCIS.com belong to a unique progression model of MCF10 isogenic cell lines, commonly used in the study of BC progression, to investigate molecular changes and the generation of tumor heterogeneity on a common genetic background[88-90]. MCF10A are non-tumorigenic immortal mammary epithelial cells (*normal*), spontaneously derived by mortal MCF10M cells grown in low calcium medium and then switched to normal calcium levels[86]. MCF10AT, a pre-malignant breast disease, is an ER-positive cell line derived by transformation with H-RAS[223]. The cell lines MCF10DCIS.com, a comedo ductal carcinoma *in situ*, and MCF10CA, metastatic breast carcinomas, were both derived from a xenograft originating from premalignant MCF10AT cells that were injected into NGS mice[193,224]. Thus, BC progression, *in vivo* circuitries and oncogene addiction are most likely different for MCF10DCIS.com and HCC1569 cells.

Nevertheless, our data seem to indicate that similar mechanisms to those identified in Epn3-overexpressing MCF10A cells might be acting in human BCs where Epn3 gene amplification and/or overexpression could contribute to malignancy together with other genetic lesions.

2.2 High Epn3 levels increase the *in vivo* tumorigenic potential of HCC1569 breast cancer cells.

Epn3 overexpression in MCF10A cells expands the mammary epithelial SC-like compartment, with the generation of cells with CSC properties, for example, the ability to form mammospheres at a higher efficiency and the expression of CSC markers ($CD44^{high}/CD24^{low}$). The main property of CSCs is their tumorigenicity; in fact, according to the CSC hypothesis, tumors are generated from one or very few CSCs, also known as TICs, able to generate the entire heterogeneity of the tumor[146].

Given the correlation between CSCs and tumor development, we hypothesized that Epn3-induced SC expansion might have a significant impact on tumor-initiation and/or metastasis. However, MCF10A are non-tumorigenic cells and Epn3 (or Twist) overexpression alone is not sufficient for a full transformation, as indicated by their inability of forming colonies in the soft agar colony formation assay, commonly used to evaluate cellular transformation *in vitro* (data not shown, Iavarone *et al.*, in preparation). Therefore, to investigate the contribution of Epn3 to breast tumorigenesis together with other genetic lesions, among which HER2 amplification, we overexpressed Epn3 in the BC cell line HCC1569. Interestingly, although these cells have already undergone some form of EMT, high levels of Epn3 were able to increase the *in vivo* tumorigenic potential of these cells in orthotopic transplantation experiments in NSG mice (see **Results**, **Fig. 11**). These data suggest that there are circuitries acting *in vivo* that cooperate with Epn3 and increase the tumorigenic potential of HCC1569 cells. Determining whether these circuitries are dependent on TGFβ signaling, as in the case of MCF10A cells *in vitro* (and at least partially of HCC1569 and MCF10DCIS.com cells), will require further investigation. We plan to

perform orthotopic transplantation experiments of tumorigenic cells (HCC1569 or MCF10DCIS.com), stably overexpressing an shRNA targeting TGFβR1 to abrogate TGFβ signaling and investigate this aspect.

Moreover, the importance of high levels of Epn3 in *in vivo* tumorigenic potential is also particularly evident from experiments of orthotopic transplantation in NSG mice of BT474, $HER2^+$ BC cells, that harbor Epn3 gene amplification and overexpression. Indeed, injection of Epn3 KD BT474 cells into murine mammary fat pad, resulted in significantly reduced tumor growth *in vivo*, despite the fact that this cell line is dependent also on the oncogene HER2, confirming that the overexpression of Epn3 contributes to enhance the tumor burden *in vivo* together with other lesions (see **Aim of the project**, **Fig. 2**).

Thus, since HCC1569 and BT474 are both $HER2^+$ BC cell lines, but at the same time responsive to high Epn3 levels in *in vivo* tumor formation, it would be interesting to use HER2 targeted treatments (e.g. Trastuzumab) to investigate whether these cells are still able to form tumors in the absence of HER2 and whether Epn3 is involved in resistance to Trastuzumab, which occurs in $HER2^+$ patients.

2.3 Future directions: MCF10DCIS.com and human breast cancers: analysis of invasive phenotypes.

MCF10DCIS.com cells injected into immunodeficient mice form comedo-type DCIS that spontaneously progress to IDC[193]. Therefore, these cells represent a great model to study *in vivo* invasion, the first step leading to the metastasis cascade. Indeed, high levels of Epn3 were found to be correlated with higher risk of distant metastasis,

rather than with loco-regional relapse, consistent with its role in EMT, migration and generation of CSCs. Therefore, we plan to perform orthotopic transplantation experiments with MCF10DCIS.com EV and Epn3-overexpressing cells and to analyze the derived tumors at different weeks from injection and analyze whether Epn3-overexpression accelerates the switch from DCIS to IDC compared to EV cells.

Moreover, we plan to investigate the eventual differences in DCIS and IDC in human BCs. A preliminary analysis on a sub-cohort of 52 breast tumors that displayed both DCIS and IDC areas, available at the IEO hospital, revealed that Epn3 is up-regulated at the invasive front of DCIS progressing towards IDC, infiltrating the basement membrane in 9 (~17%) of these cases (data not shown, Iavarone *et al.*, in preparation). These data underline the importance of Epn3 in conferring to the cells a more aggressive phenotype, i.e., cells able to infiltrate, invade and most probably start metastasizing and may give us an explanation as to why high levels of Epn3 correlate with a poor prognosis even in the absence of gene amplification or general overexpression in the entire primary tumor.

3. Epn3 as a novel prognostic marker in breast cancer and possible therapies

3.1 Epn3 as a putative novel oncogene in BC.

Taken together, the data presented establish an oncogenic role for Epn3 in BC. Indeed, ~10% of the cohort of breast tumors analyzed showed *EPN3* gene amplification and ~50% of the same cohort showed overexpression both at the transcript and protein level. Moreover, Epn3 represents an independent predictor of metastasis and unfavorable outcome of BC patients (see **Preliminary Data**). In addition, overexpression of Epn3 in non-tumorigenic mammary epithelial cells MCF10A led to the acquisition of some tumorigenic features *in vitro* (EMT and generation of cells with CSC markers), whereas Epn3 overexpression in the BC cell line, HCC1569, increased the *in vivo* tumorigenic potential of these cells. Finally, ~17% of a sub-cohort of human breast tumors showed pathological up-regulation of Epn3 specifically at the invasive front of DCIS progressing to IDC.

The characterization of this novel oncogene in BC could have important implications for the stratification and treatment of BC patients. Moreover, a better understanding of Epn3 function, particular in regard to EMT and the emergence of CSCs, might also reveal common mechanisms used by cells in tumorigenesis and in CSC expansion that could be targeted and lead to the development of novel anti-cancer drugs for BC.

3.2 Epn3 and HER2: two oncogenes in BC on the same amplicon.

The *EPN3* gene is located on chromosome 17, in region 17q21, about 10 Mb apart from the most frequently amplified BC oncogene *ERBB2* (also known as HER2). Despite this close proximity of the *EPN3* and *ERBB2* genes, *EPN3* amplification can occur independently of *ERBB2* amplification in ~50% of the cases, whereas *ERBB2* was found to be amplified in 18% of the BCs analyzed, and co-amplification with Epn3 occurred in ~25% of them (see **Aim of the project**, **Fig. 1**). Moreover, regardless of gene amplification counting only for ~10% of breast tumors analyzed, ~50% of the same cohort showed overexpression both at the transcript and protein level; high levels of Epn3 correlate with ERBB2-positivity in more than ~60% of the tumors of the cohort analyzed (see **Aim of the project**, **Table 1**).

Consequently, the cooperation of Epn3 and HER2 in BC appears to be interesting and further experiments are needed to clarify their molecular relationship. These studies may also be useful in the treatment of HER2-positive tumors with Trastuzumab, a monoclonal antibody directed against HER2. Trastuzumab, alone or in combination with chemotherapy, has substantially improved the treatment of HER2-positive primary breast tumors. However, although Trastuzumab has a high affinity for HER2 and high doses can be administered (because of its low toxicity), treatment of metastatic HER2-positive patients remains still a challenge, in which resistance to the treatment develops rapidly[225]. In this case scenario, Epn3 may help to better stratify the patients in which the treatment might be efficient and in those that could develop resistance to the drug. In addition, ideally the development of anti-Epn3 specific drugs may improve the survival in combination with Trastuzumab and/or chemotherapy in HER2-positive patients.

Nevertheless, Epn3 is prognostic of distant metastasis also in HER2-negative patients, indicating that the role of Epn3 in BC might be independent of HER2 and therefore, Epn3 could represent a potential biomarker and molecular target in HER2-negative patients. This is especially relevant in the case of triple-negative BC, where the three key and targetable receptors (ER, PR and HER2) are lacking and specific molecular targets for therapies have not been identified yet. Indeed, Epn3 is significantly associated with luminal B and triple-negative BC, suggesting its role as a possible novel target.

3.3 Anti-TGFβ signaling therapies.

The fact that up-regulation of TGFβ in Epn3-overexpressing cells creates an autocrine positive feedback loop that could be reverted with the abrogation of the TGFβ pathway itself suggests that treating Epn3-overexpressing BCs with anti-TGFβ signaling therapies could be effective.

Several TGFβ signaling inhibitors are already in clinical trials and are giving promising results. For example, the dual type II and type I receptor kinase inhibitor, LY2109761, which we have also used *in vitro* in our MCF10A cell model system, gave impressive inhibition of metastasis in an orthotopic pancreatic cancer model[200].

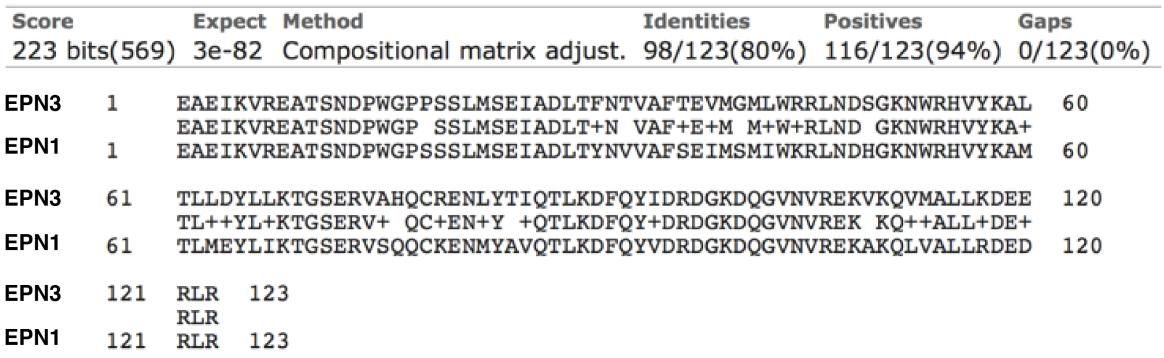
However, the dual role of TGFβ in cancer, acting as a tumor suppressor in early stages and as an oncogene in advanced stages of tumorigenesis, and its fundamental role also in a physiological context, represent a limitation, suggesting that anti-TGFβ signaling therapies may be used only in specific tumors and at late metastatic stages.

In this scenario, Epn3 may be a biomarker helping to predict which patients are most likely to benefit from targeted anti-TGFβ signaling therapies.

3.4 Future directions: Epn3 and Epn1: chimeric approach.

Our *in vitro* results of overexpression of epsin family members in non-tumorigenic MCF10A cells revealed that Epn1 does not phenocopy Epn3-dependent phenotypes of EMT and SC-like expansion, indicating that these are phenotypes acquired specifically because of the presence of Epn3. Indeed, while some papers have pointed out a potential role for Epn1 and Epn2 in other types of cancer (e.g., prostate cancer)[31,53], they are amplified only in 2% of BCs, as predicted on TCGA.

Epn3 displays a high level of sequence similarity and similar structural domain organization with other Epsin family members. Its ENTH domain shares 80% sequence identity and >90% sequence similarity with the Epn1 ENTH domain, meaning that the majority of amino acids are either identical or with similar physicochemical properties and only very few amino acids are different (6%).



Score	Expect	Method	Identities	Positives	Gaps
223 bits(569)	3e-82	Compositional matrix adjust.	98/123(80%)	116/123(94%)	0/123(0%)

```
EPN3   1     EAEIKVREATSNDPWGPPSSLMSEIADLTFNTVAFTEVMGMLWRRLNDSGKNWRHVYKAL  60
            EAEIKVREATSNDPWGP SSLMSEIADLT+N VAF+E+M M+W+RLND GKNWRHVYKA+
EPN1   1     EAEIKVREATSNDPWGPSSSLMSEIADLTYNVVAFSEIMSMIWKRLNDHGKNWRHVYKAM  60

EPN3   61    TLLDYLLKTGSERVAHQCRENLYTIQTLKDFQYIDRDGKDQGVNVREKVKQVMALLKDEE  120
            TL++YL+KTGSERV+ QC+EN+Y +QTLKDFQY+DRDGKDQGVNVREK KQ++ALL+DE+
EPN1   61    TLMEYLIKTGSERVSQQCKENMYAVQTLKDFQYVDRDGKDQGVNVREKAKQLVALLRDED  120

EPN3   121   RLR  123
            RLR
EPN1   121   RLR  123
```

Fig. 2: Alignment of the conserved ENTH domain in Epn3 and Epn1 proteins.

Alignment of the ENTH domain in Epn3 and Epn1 protein sequence performed using the blast2seq tool available on NCBI (https://blast.ncbi.nlm.noh.gov).

Figure shows the score attributed to the alignment and the probability of having the same alignment by chance (expect) and the method used to align the two sequences.

The number and the relative percentage of identical (identities), similar (positives) residues, as well as gaps inserted in the alignment, also are shown.

Following the ENTH domain, epsins have 2 or 3 UIMs. The region between the UIMs and the C-terminal of the protein is mostly unstructured and is the region where the proteins differ the most. However, there are short signature motifs common between the members of the epsin family: the DPW motifs (five DPW motifs for Epn3, compared with eight in Epn1), two clathrin-binding boxes and three NPF motifs (as with Epn1).

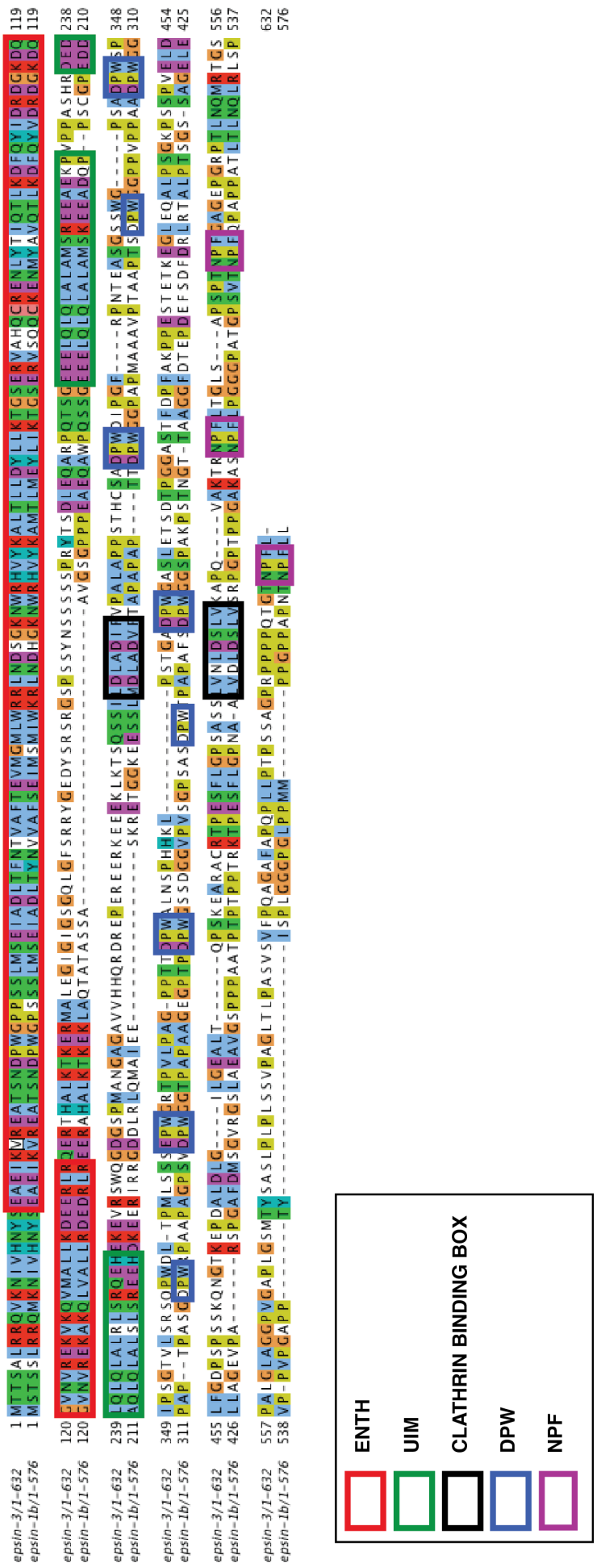


Fig. 3: Alignment of Epn3 and Epn1 protein sequences.

Top panel: Alignment of the Epn3 and Epn1 protein sequences performed using the Clustal Omega tool available on EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/) and visualized with Jalview software.

Residues with similar chemical and physical properties are represented by the same color.

Bottom panel: Legend of colored boxes used in the alignment (see top panel) to highlight specific domains and motifs of epsin proteins.

The high percentage of sequence similarity and similar structural domain organization of Epn1 and Epn3 represents for us a huge advantage in the identification of the minimal region of Epn3 responsible for the above-described phenotypes. Indeed, we plan to engineer chimeric molecules, swapping the domains of Epn3 and Epn1, and overexpress them in MCF10A cells to investigate which combination is sufficient to give similar phenotypes to the full-length Epn3. In particular, we will start by swapping large regions and then we will progressively move to smaller amino acid sequences.

This chimeric approach may help us identify the specific domain(s) responsible for phenotypes and consequently the function of that domain, thereby indicating a specific mechanism of action of Epn3 on its interactors. Moreover, it could represent the starting point for developing specific Epn3 drugs targeting that region.

REFERENCES

1 Sigismund, S. *et al.* Endocytosis and signaling: cell logistics shape the eukaryotic cell plan. *Physiol Rev* **92**, 273-366, doi:10.1152/physrev.00005.2011 (2012).

2 Mayor, S. & Pagano, R. E. Pathways of clathrin-independent endocytosis. *Nat Rev Mol Cell Biol* **8**, 603-612, doi:10.1038/nrm2216 (2007).

3 Swanson, J. A. Shaping cups into phagosomes and macropinosomes. *Nat Rev Mol Cell Biol* **9**, 639-649, doi:10.1038/nrm2447 (2008).

4 Kerr, M. C. & Teasdale, R. D. Defining macropinocytosis. *Traffic* **10**, 364-371, doi:10.1111/j.1600-0854.2009.00878.x (2009).

5 McMahon, H. T. & Boucrot, E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat Rev Mol Cell Biol* **12**, 517-533, doi:10.1038/nrm3151 (2011).

6 Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R. & Owen, D. J. Molecular architecture and functional model of the endocytic AP2 complex. *Cell* **109**, 523-535 (2002).

7 Kirchhausen, T., Owen, D. & Harrison, S. C. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. *Cold Spring Harb Perspect Biol* **6**, a016725, doi:10.1101/cshperspect.a016725 (2014).

8 Slepnev, V. I. & De Camilli, P. Accessory factors in clathrin-dependent synaptic vesicle endocytosis. *Nat Rev Neurosci* **1**, 161-172, doi:10.1038/35044540 (2000).

9 Piper, R. C. & Katzmann, D. J. Biogenesis and function of multivesicular bodies. *Annu Rev Cell Dev Biol* **23**, 519-547, doi:10.1146/annurev.cellbio.23.090506.123319 (2007).

10 de Laurentiis, A., Donovan, L. & Arcaro, A. Lipid rafts and caveolae in signaling by growth factor receptors. *Open Biochem J* **1**, 12-32, doi:10.2174/1874091X00701010012 (2007).

11 Mayor, S., Parton, R. G. & Donaldson, J. G. Clathrin-independent pathways of endocytosis. *Cold Spring Harb Perspect Biol* **6**, doi:10.1101/cshperspect.a016758 (2014).

12 Caldieri, G. *et al.* Reticulon 3-dependent ER-PM contact sites control EGFR nonclathrin endocytosis. *Science* **356**, 617-624, doi:10.1126/science.aah6152 (2017).

13 Grimes, M. L. *et al.* Endocytosis of activated TrkA: evidence that nerve growth factor induces formation of signaling endosomes. *J Neurosci* **16**, 7950-7964 (1996).

14 Palfy, M., Remenyi, A. & Korcsmaros, T. Endosomal crosstalk: meeting points for signaling pathways. *Trends Cell Biol* **22**, 447-456, doi:10.1016/j.tcb.2012.06.004 (2012).

15 Di Guglielmo, G. M., Le Roy, C., Goodfellow, A. F. & Wrana, J. L. Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. *Nat Cell Biol* **5**, 410-421, doi:10.1038/ncb975 (2003).

16 Tsukazaki, T., Chiang, T. A., Davison, A. F., Attisano, L. & Wrana, J. L. SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. *Cell* **95**, 779-791 (1998).

17 Taelman, V. F. *et al.* Wnt signaling requires sequestration of glycogen synthase kinase 3 inside multivesicular endosomes. *Cell* **143**, 1136-1148, doi:10.1016/j.cell.2010.11.034 (2010).

18 Fortini, M. E. & Bilder, D. Endocytic regulation of Notch signaling. *Curr Opin Genet Dev* **19**, 323-328, doi:10.1016/j.gde.2009.04.005 (2009).

19 Wang, W. & Struhl, G. Drosophila Epsin mediates a select endocytic pathway that DSL ligands must enter to activate Notch. *Development* **131**, 5367-5380, doi:10.1242/dev.01413 (2004).

20 Sigismund, S. *et al.* Clathrin-mediated internalization is essential for sustained EGFR signaling but dispensable for degradation. *Dev Cell* **15**, 209-219, doi:10.1016/j.devcel.2008.06.012 (2008).

21 Polo, S. & Di Fiore, P. P. Endocytosis conducts the cell signaling orchestra. *Cell* **124**, 897-900, doi:10.1016/j.cell.2006.02.025 (2006).

22 Lanzetti, L. & Di Fiore, P. P. Endocytosis and cancer: an 'insider' network with dangerous liaisons. *Traffic* **9**, 2011-2021, doi:10.1111/j.1600-0854.2008.00816.x (2008).

23 Du, Z. & Lovly, C. M. Mechanisms of receptor tyrosine kinase activation in cancer. *Mol Cancer* **17**, 58, doi:10.1186/s12943-018-0782-4 (2018).

24 Thien, C. B. & Langdon, W. Y. Cbl: many adaptations to regulate protein tyrosine kinases. *Nat Rev Mol Cell Biol* **2**, 294-307, doi:10.1038/35067100 (2001).

25 Peschard, P. & Park, M. Escape from Cbl-mediated downregulation: a recurrent theme for oncogenic deregulation of receptor tyrosine kinases. *Cancer Cell* **3**, 519-523 (2003).

26 Le Borgne, R., Bardin, A. & Schweisguth, F. The roles of receptor and ligand endocytosis in regulating Notch signaling. *Development* **132**, 1751-1762, doi:10.1242/dev.01789 (2005).

27 Guo, M., Jan, L. Y. & Jan, Y. N. Control of daughter cell fates during asymmetric division: interaction of Numb and Notch. *Neuron* **17**, 27-41 (1996).

28 Cheng, K. W. *et al.* The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. *Nat Med* **10**, 1251-1256, doi:10.1038/nm1125 (2004).

29 Caswell, P. T. *et al.* Rab25 associates with alpha5beta1 integrin to promote invasive migration in 3D microenvironments. *Dev Cell* **13**, 496-510, doi:10.1016/j.devcel.2007.08.012 (2007).

30 Dozynkiewicz, M. A. *et al.* Rab25 and CLIC3 collaborate to promote integrin recycling from late endosomes/lysosomes and drive cancer progression. *Dev Cell* **22**, 131-145, doi:10.1016/j.devcel.2011.11.008 (2012).

31 Tessneer, K. L. *et al.* Endocytic adaptor protein epsin is elevated in prostate cancer and required for cancer progression. *ISRN Oncol* **2013**, 420597, doi:10.1155/2013/420597 (2013).

32 Wang, Y., Dai, Z., Sadee, W. & Hancock, W. S. A pharmacoproteomics study of the cancer cell line EKVX using capillary-LC/MS/MS. *Mol Pharm* **3**, 566-578, doi:10.1021/mp060002b (2006).

33 Pawlowski, K. M. *et al.* Comparison of cellular and tissue transcriptional profiles in canine mammary tumor. *J Physiol Pharmacol* **60 Suppl 1**, 85-94 (2009).

34 Hellwig, B. *et al.* Epsin Family Member 3 and Ribosome-Related Genes Are Associated with Late Metastasis in Estrogen Receptor-Positive Breast Cancer and Long-Term Survival in Non-Small Cell Lung Cancer Using a Genome-Wide Identification and Validation Strategy. *PLoS One* **11**, e0167585, doi:10.1371/journal.pone.0167585 (2016).

35 Sen, A., Madhivanan, K., Mukherjee, D. & Aguilar, R. C. The epsin protein family: coordinators of endocytosis and signaling. *Biomol Concepts* **3**, 117-126, doi:10.1515/bmc-2011-0060 (2012).

36 De Camilli, P. *et al.* The ENTH domain. *FEBS Lett* **513**, 11-18 (2002).

37 Tian, X., Hansen, D., Schedl, T. & Skeath, J. B. Epsin potentiates Notch pathway activity in Drosophila and C. elegans. *Development* **131**, 5807-5815, doi:10.1242/dev.01459 (2004).

38 Chen, H. *et al.* Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis. *Nature* **394**, 793-797, doi:10.1038/29555 (1998).

39 Rosenthal, J. A. *et al.* The epsins define a family of proteins that interact with components of the clathrin coat and contain a new protein module. *J Biol Chem* **274**, 33959-33965 (1999).

40 Ko, G. *et al.* Selective high-level expression of epsin 3 in gastric parietal cells, where it is localized at endocytic sites of apical canaliculi. *Proc Natl Acad Sci U S A* **107**, 21511-21516, doi:10.1073/pnas.1016390107 (2010).

41 Spradling, K. D., McDaniel, A. E., Lohi, J. & Pilcher, B. K. Epsin 3 is a novel extracellular matrix-induced transcript specific to wounded epithelia. *J Biol Chem* **276**, 29257-29267, doi:10.1074/jbc.M101663200 (2001).

42 Chen, H. *et al.* Embryonic arrest at midgestation and disruption of Notch signaling produced by the absence of both epsin 1 and epsin 2 in mice. *Proc Natl Acad Sci U S A* **106**, 13838-13843, doi:10.1073/pnas.0907008106 (2009).

43 Mills, I. G. *et al.* EpsinR: an AP1/clathrin interacting protein involved in vesicle trafficking. *J Cell Biol* **160**, 213-222, doi:10.1083/jcb.200208023 (2003).

44 Saint-Pol, A. *et al.* Clathrin adaptor epsinR is required for retrograde sorting on early endosomal membranes. *Dev Cell* **6**, 525-538 (2004).

45 Horvath, C. A., Vanden Broeck, D., Boulet, G. A., Bogers, J. & De Wolf, M. J. Epsin: inducing membrane curvature. *Int J Biochem Cell Biol* **39**, 1765-1770, doi:10.1016/j.biocel.2006.12.004 (2007).

46 Itoh, T. *et al.* Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science* **291**, 1047-1051, doi:10.1126/science.291.5506.1047 (2001).

47 Polo, S. *et al.* A single motif responsible for ubiquitin recognition and monoubiquitination in endocytic proteins. *Nature* **416**, 451-455, doi:10.1038/416451a (2002).

48 Sigismund, S. *et al.* Clathrin-independent endocytosis of ubiquitinated cargos. *Proc Natl Acad Sci U S A* **102**, 2760-2765, doi:10.1073/pnas.0409817102 (2005).

49 Drake, M. T., Downs, M. A. & Traub, L. M. Epsin binds to clathrin by associating directly with the clathrin-terminal domain. Evidence for cooperative binding through two discrete sites. *J Biol Chem* **275**, 6479-6489 (2000).

50 Yamabhai, M. *et al.* Intersectin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains. *J Biol Chem* **273**, 31401-31407 (1998).

51 Morinaka, K. *et al.* Epsin binds to the EH domain of POB1 and regulates receptor-mediated endocytosis. *Oncogene* **18**, 5915-5922, doi:10.1038/sj.onc.1202974 (1999).

52 Tessneer, K. L. *et al.* Epsin Family of Endocytic Adaptor Proteins as Oncogenic Regulators of Cancer Progression. *J Can Res Updates* **2**, 144-150, doi:10.6000/1929-2279.2013.02.03.2 (2013).

53 Pasula, S. *et al.* Endothelial epsin deficiency decreases tumor growth by enhancing VEGF signaling. *J Clin Invest* **122**, 4424-4438, doi:10.1172/JCI64537 (2012).

54 Meloty-Kapella, L., Shergill, B., Kuon, J., Botvinick, E. & Weinmaster, G. Notch ligand endocytosis generates mechanical pulling force dependent on dynamin, epsins, and actin. *Dev Cell* **22**, 1299-1312, doi:10.1016/j.devcel.2012.04.005 (2012).

55 Fischer, J. A., Leavell, S. K. & Li, Q. Mutagenesis screens for interacting genes reveal three roles for fat facets during Drosophila eye development. *Dev Genet* **21**,

167-174, doi:10.1002/(SICI)1520-6408(1997)21:2<167::AID-DVG6>3.0.CO;2-5 (1997).
56 Aguilar, R. C. *et al.* Epsin N-terminal homology domains perform an essential function regulating Cdc42 through binding Cdc42 GTPase-activating proteins. *Proc Natl Acad Sci U S A* **103**, 4116-4121, doi:10.1073/pnas.0510513103 (2006).
57 Coon, B. G., Burgner, J., Camonis, J. H. & Aguilar, R. C. The epsin family of endocytic adaptors promotes fibrosarcoma migration and invasion. *J Biol Chem* **285**, 33073-33081, doi:10.1074/jbc.M110.124123 (2010).
58 Csikos, G. *et al.* A novel role for the Drosophila epsin (lqf): involvement in autophagy. *Autophagy* **5**, 636-648 (2009).
59 Liu, Z. & Zheng, Y. A requirement for epsin in mitotic membrane and spindle organization. *J Cell Biol* **186**, 473-480, doi:10.1083/jcb.200902071 (2009).
60 Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
61 Kerbel, R. S. Tumor angiogenesis. *N Engl J Med* **358**, 2039-2049, doi:10.1056/NEJMra0706596 (2008).
62 Tessneer, K. L. *et al.* Genetic reduction of vascular endothelial growth factor receptor 2 rescues aberrant angiogenesis caused by epsin deficiency. *Arterioscler Thromb Vasc Biol* **34**, 331-337, doi:10.1161/ATVBAHA.113.302586 (2014).
63 Sigismund, S., Avanzato, D. & Lanzetti, L. Emerging functions of the EGFR in cancer. *Mol Oncol* **12**, 3-20, doi:10.1002/1878-0261.12155 (2018).
64 Gjorevski, N. & Nelson, C. M. Integrated morphodynamic signalling of the mammary gland. *Nat Rev Mol Cell Biol* **12**, 581-593, doi:10.1038/nrm3168 (2011).
65 Manavathi, B., Samanthapudi, V. S. & Gajulapalli, V. N. Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development. *Front Cell Dev Biol* **2**, 34, doi:10.3389/fcell.2014.00034 (2014).
66 Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993-997, doi:10.1038/nature04496 (2006).
67 Torre, L. A., Islami, F., Siegel, R. L., Ward, E. M. & Jemal, A. Global Cancer in Women: Burden and Trends. *Cancer Epidemiol Biomarkers Prev* **26**, 444-457, doi:10.1158/1055-9965.EPI-16-0858 (2017).
68 Malhotra, G. K., Zhao, X., Band, H. & Band, V. Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther* **10**, 955-960 (2010).
69 Li, C. I., Uribe, D. J. & Daling, J. R. Clinical characteristics of different histologic types of breast cancer. *Br J Cancer* **93**, 1046-1052, doi:10.1038/sj.bjc.6602787 (2005).
70 Elston, C. W. & Ellis, I. O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**, 403-410 (1991).
71 Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752, doi:10.1038/35021093 (2000).
72 Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27**, 1160-1167, doi:10.1200/JCO.2008.18.1370 (2009).
73 Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* **12**, R68, doi:10.1186/bcr2635 (2010).
74 Ades, F. *et al.* Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. *J Clin Oncol* **32**, 2794-2803, doi:10.1200/JCO.2013.54.1870 (2014).
75 Hudis, C. A. Trastuzumab--mechanism of action and use in clinical practice. *N Engl J Med* **357**, 39-51, doi:10.1056/NEJMra043186 (2007).

76 Jhaveri, K. & Esteva, F. J. Pertuzumab in the treatment of HER2+ breast cancer. *J Natl Compr Canc Netw* **12**, 591-598 (2014).

77 Khosravi-Shahi, P., Cabezon-Gutierrez, L. & Custodio-Cabello, S. Metastatic triple negative breast cancer: Optimizing treatment options, new and emerging targeted therapies. *Asia Pac J Clin Oncol* **14**, 32-39, doi:10.1111/ajco.12748 (2018).

78 Ratsch, S. B., Gao, Q., Srinivasan, S., Wazer, D. E. & Band, V. Multiple genetic changes are required for efficient immortalization of different subtypes of normal human mammary epithelial cells. *Radiat Res* **155**, 143-150 (2001).

79 Shay, J. W. & Wright, W. E. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* **26**, 867-874, doi:10.1093/carcin/bgh296 (2005).

80 Itahana, K., Dimri, G. & Campisi, J. Regulation of cellular senescence by p53. *Eur J Biochem* **268**, 2784-2791 (2001).

81 Brenner, A. J., Stampfer, M. R. & Aldaz, C. M. Increased p16 expression with first senescence arrest in human mammary epithelial cells and extended growth capacity with p16 inactivation. *Oncogene* **17**, 199-205, doi:10.1038/sj.onc.1201919 (1998).

82 Van der Haegen, B. A. & Shay, J. W. Immortalization of human mammary epithelial cells by SV40 large T-antigen involves a two step mechanism. *In Vitro Cell Dev Biol* **29A**, 180-182 (1993).

83 Wazer, D. E., Liu, X. L., Chu, Q., Gao, Q. & Band, V. Immortalization of distinct human mammary epithelial cell types by human papilloma virus 16 E6 or E7. *Proc Natl Acad Sci U S A* **92**, 3687-3691 (1995).

84 Gao, Q. *et al.* Mutant p53-induced immortalization of primary human mammary epithelial cells. *Cancer Res* **56**, 3129-3133 (1996).

85 Elenbaas, B. *et al.* Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* **15**, 50-65 (2001).

86 Cowell, J. K. *et al.* Molecular characterization of the t(3;9) associated with immortalization in the MCF10A cell line. *Cancer Genet Cytogenet* **163**, 23-29, doi:10.1016/j.cancergencyto.2005.04.019 (2005).

87 Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**, 256-268 (2003).

88 Worsham, M. J. *et al.* High-resolution mapping of molecular events associated with immortalization, transformation, and progression to breast cancer in the MCF10 model. *Breast Cancer Res Treat* **96**, 177-186, doi:10.1007/s10549-005-9077-8 (2006).

89 Marella, N. V. *et al.* Cytogenetic and cDNA microarray expression analysis of MCF10 human breast cancer progression cell lines. *Cancer Res* **69**, 5946-5953, doi:10.1158/0008-5472.CAN-09-0420 (2009).

90 Kadota, M. *et al.* Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines. *PLoS One* **5**, e9201, doi:10.1371/journal.pone.0009201 (2010).

91 Blackshear, P. E. Genetically engineered rodent models of mammary gland carcinogenesis: an overview. *Toxicol Pathol* **29**, 105-116, doi:10.1080/019262301301418919 (2001).

92 Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J Clin Invest* **119**, 1420-1428, doi:10.1172/JCI39104 (2009).

93 Micalizzi, D. S., Farabaugh, S. M. & Ford, H. L. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *J Mammary Gland Biol Neoplasia* **15**, 117-134, doi:10.1007/s10911-010-9178-9 (2010).

94 Vicovac, L. & Aplin, J. D. Epithelial-mesenchymal transition during trophoblast differentiation. *Acta Anat (Basel)* **156**, 202-216 (1996).

95 Thiery, J. P. & Sleeman, J. P. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol* **7**, 131-142, doi:10.1038/nrm1835 (2006).

96 Duband, J. L. & Thiery, J. P. Appearance and distribution of fibronectin during chick embryo gastrulation and neurulation. *Dev Biol* **94**, 337-350 (1982).

97 Kalluri, R. & Neilson, E. G. Epithelial-mesenchymal transition and its implications for fibrosis. *J Clin Invest* **112**, 1776-1784, doi:10.1172/JCI20530 (2003).

98 Thiery, J. P. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**, 442-454, doi:10.1038/nrc822 (2002).

99 Shapiro, L. & Weis, W. I. Structure and biochemistry of cadherins and catenins. *Cold Spring Harb Perspect Biol* **1**, a003053, doi:10.1101/cshperspect.a003053 (2009).

100 Itoh, M. & Bissell, M. J. The organization of tight junctions in epithelia: implications for mammary gland biology and breast tumorigenesis. *J Mammary Gland Biol Neoplasia* **8**, 449-462, doi:10.1023/B:JOMG.0000017431.45314.07 (2003).

101 Mese, G., Richard, G. & White, T. W. Gap junctions: basic structure and function. *J Invest Dermatol* **127**, 2516-2524, doi:10.1038/sj.jid.5700770 (2007).

102 Green, K. J. & Jones, J. C. Desmosomes and hemidesmosomes: structure and function of molecular components. *FASEB J* **10**, 871-881 (1996).

103 Peinado, H., Portillo, F. & Cano, A. Transcriptional regulation of cadherins during development and carcinogenesis. *Int J Dev Biol* **48**, 365-375, doi:10.1387/ijdb.041794hp (2004).

104 Hirohashi, S. Inactivation of the E-cadherin-mediated cell adhesion system in human cancers. *Am J Pathol* **153**, 333-339, doi:10.1016/S0002-9440(10)65575-7 (1998).

105 Hazan, R. B., Qiao, R., Keren, R., Badano, I. & Suyama, K. Cadherin switch in tumor progression. *Ann N Y Acad Sci* **1014**, 155-163 (2004).

106 Theveneau, E. & Mayor, R. Cadherins in collective cell migration of mesenchymal cells. *Curr Opin Cell Biol* **24**, 677-684, doi:10.1016/j.ceb.2012.08.002 (2012).

107 Mendez, M. G., Kojima, S. & Goldman, R. D. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB J* **24**, 1838-1851, doi:10.1096/fj.09-151639 (2010).

108 Wu, Y., Sarkissyan, M. & Vadgama, J. V. Epithelial-Mesenchymal Transition and Breast Cancer. *J Clin Med* **5**, doi:10.3390/jcm5020013 (2016).

109 Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-196, doi:10.1038/nrm3758 (2014).

110 Dave, N. *et al.* Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition. *J Biol Chem* **286**, 12024-12032, doi:10.1074/jbc.M110.168625 (2011).

111 De Craene, B. & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* **13**, 97-110, doi:10.1038/nrc3447 (2013).

112 Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:10.1016/j.cell.2006.02.041 (2006).

113 Yanagisawa, M. *et al.* A p120 catenin isoform switch affects Rho activity, induces tumor cell invasion, and predicts metastatic disease. *J Biol Chem* **283**, 18344-18354, doi:10.1074/jbc.M801192200 (2008).
114 Brown, R. L. *et al.* CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest* **121**, 1064-1074, doi:10.1172/JCI44540 (2011).
115 Ru, P. *et al.* miRNA-29b suppresses prostate cancer metastasis by regulating epithelial-mesenchymal transition signaling. *Mol Cancer Ther* **11**, 1166-1173, doi:10.1158/1535-7163.MCT-12-0100 (2012).
116 Zhang, J. *et al.* miR-30 inhibits TGF-beta1-induced epithelial-to-mesenchymal transition in hepatocyte by targeting Snail1. *Biochem Biophys Res Commun* **417**, 1100-1105, doi:10.1016/j.bbrc.2011.12.121 (2012).
117 Burk, U. *et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* **9**, 582-589, doi:10.1038/embor.2008.74 (2008).
118 Chao, Y., Wu, Q., Acquafondata, M., Dhir, R. & Wells, A. Partial mesenchymal to epithelial reverting transition in breast and prostate cancer metastases. *Cancer Microenviron* **5**, 19-28, doi:10.1007/s12307-011-0085-4 (2012).
119 Lecharpentier, A. *et al.* Detection of circulating tumour cells with a hybrid (epithelial/mesenchymal) phenotype in patients with metastatic non-small cell lung cancer. *Br J Cancer* **105**, 1338-1341, doi:10.1038/bjc.2011.405 (2011).
120 Aiello, N. M. *et al.* EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Dev Cell* **45**, 681-695 e684, doi:10.1016/j.devcel.2018.05.027 (2018).
121 Moustakas, A. & Heldin, C. H. Non-Smad TGF-beta signals. *J Cell Sci* **118**, 3573-3584, doi:10.1242/jcs.02554 (2005).
122 Li, X. *et al.* Snail induction is an early response to Gli1 that determines the efficiency of epithelial transformation. *Oncogene* **25**, 609-621, doi:10.1038/sj.onc.1209077 (2006).
123 Niessen, K. *et al.* Slug is a direct Notch target required for initiation of cardiac cushion cellularization. *J Cell Biol* **182**, 315-325, doi:10.1083/jcb.200710067 (2008).
124 Yadav, A., Kumar, B., Datta, J., Teknos, T. N. & Kumar, P. IL-6 promotes head and neck tumor metastasis by inducing epithelial-mesenchymal transition via the JAK-STAT3-SNAIL signaling pathway. *Mol Cancer Res* **9**, 1658-1667, doi:10.1158/1541-7786.MCR-11-0271 (2011).
125 Yang, M. H. *et al.* Direct regulation of TWIST by HIF-1alpha promotes metastasis. *Nat Cell Biol* **10**, 295-305, doi:10.1038/ncb1691 (2008).
126 Shi, Y. & Massague, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700 (2003).
127 Miyazono, K. Transforming growth factor-beta signaling in epithelial-mesenchymal transition and progression of cancer. *Proc Jpn Acad Ser B Phys Biol Sci* **85**, 314-323 (2009).
128 Kavsak, P. *et al.* Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. *Mol Cell* **6**, 1365-1375 (2000).
129 Ebisawa, T. *et al.* Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. *J Biol Chem* **276**, 12477-12480, doi:10.1074/jbc.C100008200 (2001).
130 Nelson, W. J. Regulation of cell-cell adhesion by the cadherin-catenin complex. *Biochem Soc Trans* **36**, 149-155, doi:10.1042/BST0360149 (2008).
131 Stamos, J. L. & Weis, W. I. The beta-catenin destruction complex. *Cold Spring Harb Perspect Biol* **5**, a007898, doi:10.1101/cshperspect.a007898 (2013).

132 Kim, K., Lu, Z. & Hay, E. D. Direct evidence for a role of beta-catenin/LEF-1 signaling pathway in induction of EMT. *Cell Biol Int* **26**, 463-476 (2002).

133 Corallino, S., Malabarba, M. G., Zobel, M., Di Fiore, P. P. & Scita, G. Epithelial-to-Mesenchymal Plasticity Harnesses Endocytic Circuitries. *Front Oncol* **5**, 45, doi:10.3389/fonc.2015.00045 (2015).

134 Seto, E. S. & Bellen, H. J. Internalization is required for proper Wingless signaling in Drosophila melanogaster. *J Cell Biol* **173**, 95-106, doi:10.1083/jcb.200510123 (2006).

135 van der Bliek, A. M. & Meyerowitz, E. M. Dynamin-like protein encoded by the Drosophila shibire gene associated with vesicular traffic. *Nature* **351**, 411-414, doi:10.1038/351411a0 (1991).

136 Blitzer, J. T. & Nusse, R. A critical role for endocytosis in Wnt signaling. *BMC Cell Biol* **7**, 28, doi:10.1186/1471-2121-7-28 (2006).

137 Yamamoto, H., Komekado, H. & Kikuchi, A. Caveolin is necessary for Wnt-3a-dependent internalization of LRP6 and accumulation of beta-catenin. *Dev Cell* **11**, 213-223, doi:10.1016/j.devcel.2006.07.003 (2006).

138 Bruser, L. & Bogdan, S. Adherens Junctions on the Move-Membrane Trafficking of E-Cadherin. *Cold Spring Harb Perspect Biol* **9**, doi:10.1101/cshperspect.a029140 (2017).

139 de Beco, S., Gueudry, C., Amblard, F. & Coscoy, S. Endocytosis is required for E-cadherin redistribution at mature adherens junctions. *Proc Natl Acad Sci U S A* **106**, 7010-7015, doi:10.1073/pnas.0811253106 (2009).

140 Miyashita, Y. & Ozawa, M. Increased internalization of p120-uncoupled E-cadherin and a requirement for a dileucine motif in the cytoplasmic domain for endocytosis of the protein. *J Biol Chem* **282**, 11540-11548, doi:10.1074/jbc.M608351200 (2007).

141 Peglion, F. & Etienne-Manneville, S. p120catenin alteration in cancer and its role in tumour invasion. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130015, doi:10.1098/rstb.2013.0015 (2013).

142 Fujita, Y. *et al.* Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nat Cell Biol* **4**, 222-231, doi:10.1038/ncb758 (2002).

143 Palacios, F., Tushir, J. S., Fujita, Y. & D'Souza-Schorey, C. Lysosomal targeting of E-cadherin: a unique mechanism for the down-regulation of cell-cell adhesion during epithelial to mesenchymal transitions. *Mol Cell Biol* **25**, 389-402, doi:10.1128/MCB.25.1.389-402.2005 (2005).

144 Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84-88, doi:10.1038/nature04372 (2006).

145 Morrison, S. J. & Kimble, J. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068-1074, doi:10.1038/nature04956 (2006).

146 Clarke, M. F. & Fuller, M. Stem cells and cancer: two faces of eve. *Cell* **124**, 1111-1115, doi:10.1016/j.cell.2006.03.011 (2006).

147 Owens, T. W. & Naylor, M. J. Breast cancer stem cells. *Front Physiol* **4**, 225, doi:10.3389/fphys.2013.00225 (2013).

148 Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-3988, doi:10.1073/pnas.0530291100 (2003).

149 Cicalese, A. *et al.* The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. *Cell* **138**, 1083-1095, doi:10.1016/j.cell.2009.06.048 (2009).

150 Honeth, G. *et al.* The CD44+/CD24- phenotype is enriched in basal-like breast tumors. *Breast Cancer Res* **10**, R53, doi:10.1186/bcr2108 (2008).

151 Hwang-Verslues, W. W. *et al.* Multiple lineages of human breast cancer stem/progenitor cells identified by profiling with stem cell markers. *PLoS One* **4**, e8377, doi:10.1371/journal.pone.0008377 (2009).
152 Ghebeh, H. *et al.* Profiling of normal and malignant breast tissue show CD44high/CD24low phenotype as a predominant stem/progenitor marker when used in combination with Ep-CAM/CD49f markers. *BMC Cancer* **13**, 289, doi:10.1186/1471-2407-13-289 (2013).
153 Aruffo, A., Stamenkovic, I., Melnick, M., Underhill, C. B. & Seed, B. CD44 is the principal cell surface receptor for hyaluronate. *Cell* **61**, 1303-1313 (1990).
154 Kadmon, G., Eckert, M., Sammar, M., Schachner, M. & Altevogt, P. Nectadrin, the heat-stable antigen, is a cell adhesion molecule. *J Cell Biol* **118**, 1245-1258 (1992).
155 Litvinov, S. V. *et al.* Epithelial cell adhesion molecule (Ep-CAM) modulates cell-cell interactions mediated by classic cadherins. *J Cell Biol* **139**, 1337-1348 (1997).
156 Seita, J. & Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med* **2**, 640-653, doi:10.1002/wsbm.86 (2010).
157 Friedrichs, K. *et al.* High expression level of alpha 6 integrin in human breast carcinoma is correlated with reduced survival. *Cancer Res* **55**, 901-906 (1995).
158 Hynes, R. O. Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**, 11-25 (1992).
159 Mizrak, D., Brittan, M. & Alison, M. CD133: molecule of the moment. *J Pathol* **214**, 3-9, doi:10.1002/path.2283 (2008).
160 Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555-567, doi:10.1016/j.stem.2007.08.014 (2007).
161 Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* **2**, 78-91, doi:10.1016/j.stemcr.2013.11.009 (2014).
162 Dontu, G. *et al.* In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* **17**, 1253-1270, doi:10.1101/gad.1061803 (2003).
163 Pece, S. *et al.* Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140**, 62-73, doi:10.1016/j.cell.2009.12.007 (2010).
164 Lanzkron, S. M., Collector, M. I. & Sharkis, S. J. Hematopoietic stem cell tracking in vivo: a comparison of short-term and long-term repopulating cells. *Blood* **93**, 1916-1921 (1999).
165 Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715, doi:10.1016/j.cell.2008.03.027 (2008).
166 Wielenga, V. J. *et al.* Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. *Am J Pathol* **154**, 515-523, doi:10.1016/S0002-9440(10)65297-2 (1999).
167 Preca, B. T. *et al.* A self-enforcing CD44s/ZEB1 feedback loop maintains EMT and stemness properties in cancer cells. *Int J Cancer* **137**, 2566-2577, doi:10.1002/ijc.29642 (2015).
168 Horwitz, S. B. Taxol (paclitaxel): mechanisms of action. *Ann Oncol* **5 Suppl 6**, S3-6 (1994).
169 Grem, J. L. Mechanisms of Action and Modulation of Fluorouracil. *Semin Radiat Oncol* **7**, 249-259, doi:10.1053/SRAO00700249 (1997).

170 Shervington, A. & Lu, C. Expression of multidrug resistance genes in normal and cancer stem cells. *Cancer Invest* **26**, 535-542, doi:10.1080/07357900801904140 (2008).
171 Yin, H. & Glass, J. The phenotypic radiation resistance of CD44+/CD24(-or low) breast cancer cells is mediated through the enhanced activation of ATM signaling. *PLoS One* **6**, e24080, doi:10.1371/journal.pone.0024080 (2011).
172 Diehn, M. *et al.* Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* **458**, 780-783, doi:10.1038/nature07733 (2009).
173 Sladek, N. E. Human aldehyde dehydrogenases: potential pathological, pharmacological, and toxicological impact. *J Biochem Mol Toxicol* **17**, 7-23, doi:10.1002/jbt.10057 (2003).
174 Tanei, T. *et al.* Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin Cancer Res* **15**, 4234-4241, doi:10.1158/1078-0432.CCR-08-1479 (2009).
175 Visvader, J. E. & Lindeman, G. J. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer* **8**, 755-768, doi:10.1038/nrc2499 (2008).
176 Zhang, M. & Rosen, J. M. Stem cells in the etiology and treatment of cancer. *Curr Opin Genet Dev* **16**, 60-64, doi:10.1016/j.gde.2005.12.008 (2006).
177 Gomez-Cuadrado, L., Tracey, N., Ma, R., Qian, B. & Brunton, V. G. Mouse models of metastasis: progress and prospects. *Dis Model Mech* **10**, 1061-1074, doi:10.1242/dmm.030403 (2017).
178 Nguyen, D. X. & Massague, J. Genetic determinants of cancer metastasis. *Nat Rev Genet* **8**, 341-352, doi:10.1038/nrg2101 (2007).
179 Christiansen, J. J. & Rajasekaran, A. K. Reassessing epithelial to mesenchymal transition as a prerequisite for carcinoma invasion and metastasis. *Cancer Res* **66**, 8319-8326, doi:10.1158/0008-5472.CAN-06-0410 (2006).
180 Jolly, M. K. *et al.* Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. *Front Oncol* **5**, 155, doi:10.3389/fonc.2015.00155 (2015).
181 Paterlini-Brechot, P. & Benali, N. L. Circulating tumor cells (CTC) detection: clinical impact and future directions. *Cancer Lett* **253**, 180-204, doi:10.1016/j.canlet.2006.12.014 (2007).
182 Furthauer, M. & Gonzalez-Gaitan, M. Endocytosis, asymmetric cell division, stem cells and cancer: unus pro omnibus, omnes pro uno. *Mol Oncol* **3**, 339-353, doi:10.1016/j.molonc.2009.05.006 (2009).
183 Santolini, E. *et al.* Numb is an endocytic protein. *J Cell Biol* **151**, 1345-1352 (2000).
184 Berdnik, D., Torok, T., Gonzalez-Gaitan, M. & Knoblich, J. A. The endocytic protein alpha-Adaptin is required for numb-mediated asymmetric cell division in Drosophila. *Dev Cell* **3**, 221-231 (2002).
185 Colaluca, I. N. *et al.* NUMB controls p53 tumour suppressor activity. *Nature* **451**, 76-80, doi:10.1038/nature06412 (2008).
186 Tosoni, D. *et al.* The Numb/p53 circuitry couples replicative self-renewal and tumor suppression in mammary epithelial cells. *J Cell Biol* **211**, 845-862, doi:10.1083/jcb.201505037 (2015).
187 Pece, S. *et al.* Loss of negative regulation by Numb over Notch is relevant to human breast carcinogenesis. *J Cell Biol* **167**, 215-221, doi:10.1083/jcb.200406140 (2004).

188 Tosoni, D. *et al.* Pre-clinical validation of a selective anti-cancer stem cell therapy for Numb-deficient human breast cancers. *EMBO Mol Med* **9**, 655-671, doi:10.15252/emmm.201606940 (2017).
189 Moustakas, A. & Heldin, C. H. Mechanisms of TGFbeta-Induced Epithelial-Mesenchymal Transition. *J Clin Med* **5**, doi:10.3390/jcm5070063 (2016).
190 Veronesi, U. *et al.* A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. *N Engl J Med* **349**, 546-553, doi:10.1056/NEJMoa012782 (2003).
191 Confalonieri, S. *et al.* Alterations of ubiquitin ligases in human cancer and their association with the natural history of the tumor. *Oncogene* **28**, 2959-2968, doi:10.1038/onc.2009.156 (2009).
192 Wolff, A. C. *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* **31**, 3997-4013, doi:10.1200/JCO.2013.50.9984 (2013).
193 Miller, F. R., Santner, S. J., Tait, L. & Dawson, P. J. MCF10DCIS.com xenograft model of human comedo ductal carcinoma in situ. *J Natl Cancer Inst* **92**, 1185-1186 (2000).
194 Hu, M. *et al.* Regulation of in situ to invasive breast carcinoma transition. *Cancer Cell* **13**, 394-406, doi:10.1016/j.ccr.2008.03.007 (2008).
195 Malinverno, C. *et al.* Endocytic reawakening of motility in jammed epithelia. *Nat Mater* **16**, 587-596, doi:10.1038/nmat4848 (2017).
196 Tosoni, D., Di Fiore, P. P. & Pece, S. Functional purification of human and mouse mammary stem cells. *Methods Mol Biol* **916**, 59-79, doi:10.1007/978-1-61779-980-8_6 (2012).
197 Yang, J. *et al.* Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927-939, doi:10.1016/j.cell.2004.06.006 (2004).
198 Chaffer, C. L. *et al.* Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc Natl Acad Sci U S A* **108**, 7950-7955, doi:10.1073/pnas.1102454108 (2011).
199 Li, W., Wei, W. & Ding, S. TGF-beta Signaling in Stem Cell Regulation. *Methods Mol Biol* **1344**, 137-145, doi:10.1007/978-1-4939-2966-5_8 (2016).
200 Melisi, D. *et al.* LY2109761, a novel transforming growth factor beta receptor type I and type II dual inhibitor, as a therapeutic approach to suppressing pancreatic cancer metastasis. *Mol Cancer Ther* **7**, 829-840, doi:10.1158/1535-7163.MCT-07-0337 (2008).
201 Janda, E. *et al.* Raf plus TGFbeta-dependent EMT is initiated by endocytosis and lysosomal degradation of E-cadherin. *Oncogene* **25**, 7117-7130, doi:10.1038/sj.onc.1209701 (2006).
202 Hrckulak, D., Kolar, M., Strnad, H. & Korinek, V. TCF/LEF Transcription Factors: An Update from the Internet Resources. *Cancers (Basel)* **8**, doi:10.3390/cancers8070070 (2016).
203 Gazdar, A. F. *et al.* Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. *Int J Cancer* **78**, 766-774 (1998).
204 Zhang, J. *et al.* TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci Signal* **7**, ra91, doi:10.1126/scisignal.2005304 (2014).
205 Gulei, D. *et al.* The "good-cop bad-cop" TGF-beta role in breast cancer modulated by non-coding RNAs. *Biochim Biophys Acta* **1861**, 1661-1675, doi:10.1016/j.bbagen.2017.04.007 (2017).

206 Siemens, H. *et al.* miR-34 and SNAIL form a double-negative feedback loop to regulate epithelial-mesenchymal transitions. *Cell Cycle* **10**, 4256-4271, doi:10.4161/cc.10.24.18552 (2011).
207 Gregory, P. A. *et al.* An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Mol Biol Cell* **22**, 1686-1698, doi:10.1091/mbc.E11-02-0103 (2011).
208 Tian, X. J., Zhang, H. & Xing, J. Coupled reversible and irreversible bistable switches underlying TGFbeta-induced epithelial to mesenchymal transition. *Biophys J* **105**, 1079-1089, doi:10.1016/j.bpj.2013.07.011 (2013).
209 Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N. & Ben-Jacob, E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc Natl Acad Sci U S A* **110**, 18144-18149, doi:10.1073/pnas.1318192110 (2013).
210 Shamir, E. R. *et al.* Twist1-induced dissemination preserves epithelial identity and requires E-cadherin. *J Cell Biol* **204**, 839-856, doi:10.1083/jcb.201306088 (2014).
211 Lan, F. *et al.* Genome-wide identification of TCF7L2/TCF4 target miRNAs reveals a role for miR-21 in Wnt-driven epithelial cancer. *Int J Oncol* **40**, 519-526, doi:10.3892/ijo.2011.1215 (2012).
212 Wang, F. *et al.* Association between TCF7L2 polymorphisms and breast cancer susceptibility: a meta-analysis. *Int J Clin Exp Med* **8**, 9355-9361 (2015).
213 Geyer, F. C. *et al.* beta-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. *Mod Pathol* **24**, 209-231, doi:10.1038/modpathol.2010.205 (2011).
214 Zhang, Y., Liu, B., Zhao, Q., Hou, T. & Huang, X. Nuclear localizaiton of beta-catenin is associated with poor survival and chemo-/radioresistance in human cervical squamous cell cancer. *Int J Clin Exp Pathol* **7**, 3908-3917 (2014).
215 Kim, J. S. *et al.* Oncogenic beta-catenin is required for bone morphogenetic protein 4 expression in human cancer cells. *Cancer Res* **62**, 2744-2748 (2002).
216 Han, G. *et al.* Smad7-induced beta-catenin degradation alters epidermal appendage development. *Dev Cell* **11**, 301-312, doi:10.1016/j.devcel.2006.06.014 (2006).
217 Tang, Y., Liu, Z., Zhao, L., Clemens, T. L. & Cao, X. Smad7 stabilizes beta-catenin binding to E-cadherin complex and promotes cell-cell adhesion. *J Biol Chem* **283**, 23956-23963, doi:10.1074/jbc.M800351200 (2008).
218 Medici, D., Hay, E. D. & Olsen, B. R. Snail and Slug promote epithelial-mesenchymal transition through beta-catenin-T-cell factor-4-dependent expression of transforming growth factor-beta3. *Mol Biol Cell* **19**, 4875-4887, doi:10.1091/mbc.E08-05-0506 (2008).
219 Labbe, E., Letamendia, A. & Attisano, L. Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-beta and wnt pathways. *Proc Natl Acad Sci U S A* **97**, 8358-8363, doi:10.1073/pnas.150152697 (2000).
220 Labbe, E. *et al.* Transcriptional cooperation between the transforming growth factor-beta and Wnt pathways in mammary and intestinal tumorigenesis. *Cancer Res* **67**, 75-84, doi:10.1158/0008-5472.CAN-06-2559 (2007).
221 Nawshad, A., Medici, D., Liu, C. C. & Hay, E. D. TGFbeta3 inhibits E-cadherin gene expression in palate medial-edge epithelial cells through a Smad2-Smad4-LEF1 transcription complex. *J Cell Sci* **120**, 1646-1653, doi:10.1242/jcs.003129 (2007).
222 Scheel, C. *et al.* Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. *Cell* **145**, 926-940, doi:10.1016/j.cell.2011.04.029 (2011).

223 Dawson, P. J., Wolman, S. R., Tait, L., Heppner, G. H. & Miller, F. R. MCF10AT: a model for the evolution of cancer from proliferative breast disease. *Am J Pathol* **148**, 313-319 (1996).

224 Santner, S. J. *et al.* Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast Cancer Res Treat* **65**, 101-110 (2001).

225 Orphanos, G. & Kountourakis, P. Targeting the HER2 receptor in metastatic breast cancer. *Hematol Oncol Stem Cell Ther* **5**, 127-137, doi:10.5144/1658-3876.2012.127 (2012).

ACKNOWLEDGMENTS

First of all, I would like to thank my Ph.D. supervisor, the P.I., Prof. Pier Paolo Di Fiore, who gave me the opportunity to join his lab and carry out this exciting project for a Ph.D. program in Molecular Oncology at the IFOM-IEO campus, with the European School of Molecular Medicine (SEMM). In particular, I would like to thank him for his mentoring, for his scientific discussions and for always having the door opened for me, if needed.

I would also like to thank my supervisor, Dr. Sara Sigismund, for her mentoring as well, for scientific discussions and for her daily supervision in the lab and help with many experimental and technical issues. On a personal note, I would also like to thank her for her support and encouragement throughout these years, both in joyful and stressful moments, making me realize that both are part of the life of a scientist and I must persevere to succeed.

I am very grateful for the suggestions and feedback of my external advisor Prof. Marcos Gonzáles-Gaitán, from the University of Genève, and my internal advisor Dr. Francesco Nicassio, Research Director of the Center for Genomic Science of IIT@SEMM. I also acknowledge the help of Rosalind Gunby for critically reading my thesis.

A special thanks goes to my colleague Giovanni Giangreco, because Epn3 is a team project and we worked together side by side, always helping each other, and it would have not been the same without him. I also thank Manuela Vecchi and Claudia Iavarone that started this project and all the people that shared their expertise: Giovanni Bertalot and all the members of the Molecular Pathology Unit for IHC experiments; Stefano Confalonieri for his help with statistical analysis and bioinformatics; Chiara Tordonato for her help with the *in vivo* experiments.

A personal thanks goes to Marilena Bicchieri, who mentored me at the beginning of this long journey, taught me a lot thanks to her own personal experience and constantly supported me throughout these four years. A huge thanks is for Alexia Conte, for all the support inside and outside the lab, for sharing the same passions with me and for always being there for me, both as a colleague and especially as a friend. I also thank my colleagues Elisa Barbieri and Giusi Caldieri, because all together we work as a fantastic *endo* group.

Finally, I also would like to thank all the people that make our daily work easier, in particular all the technical facilities available at the campus and the coordinators of the SEMM Ph.D. program, Francesca Fiore and Veronica Viscardi.

Last but not least, I would like to thank my family for supporting me throughout my entire academic carrier, for believing in me and for thinking that I am able to do anything I set my mind to, for always helping and encouraging me.