THE ROLE OF EpCAM IN BREAST CANCER STEM CELLS

STRATEGIES TO INVESTIGATE THE ROLE OF EpCAM IN CANCER STEM CELLS OF BREAST CANCER CELL LINES

By:

Ali Lessan, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements for the Degree

Master of Science

McMaster University

Master of Science (2012)

Biochemistry and Biomedical Sciences

McMaster University

Hamilton, Ontario

AUTHOR: Ali Lessan, B.Sc.

SUPERVSOR: Doctor John A. Hassell

NUMBER OF PAGES: xii, 97

Abstract

Based on the cancer stem cell (CSC) model, tumors develop as a hierarchy, much like normal tissue. At the apex, CSCs are capable of self-renewal and differentiating into non-tumorigenic cells that form the bulk of the tumor. CSCs have been isolated for many types of cancer based on cell surface marker expression. For instance in human breast carcinoma, the $CD44^{+}$ $CD24^{-/low}$ $EpCAM^{+}$ population are enriched in CSCs. These cells are more resistant to traditional chemo and radiation therapy relative to the bulk of the tumor. As such, many believe CSCs to be responsible for relapse and metastasis events. Hence, a more targeted therapy towards breast CSCs can prove to be very effective. EpCAM has been shown to be a reliable, albeit not exclusive, marker of breast CSCs. A more thorough understanding of EpCAM's function can provide new angles for designing therapeutic agents. Originally thought to be a mere adhesion molecule, EpCAM is now known to derive a signalling pathway that promotes transcription in its target genes.

We aimed to provide further insight into this novel pathway, by manipulating EpCAM's expression and function in MCF7 breast cancer cell line. We designed a dominant negative allele of the EpCAM protein to hinder the activity of the endogenous EpCAM. However, this transgene offered no significant effect on our cell line. Furthermore, we used RNA interference technology to reduce EpCAM expression. The introduction of the EpCAM short hairpin RNA constructs into our cell line had inhibitory effects on transcription reporter expression, cellular growth, adherent colony and unattached sphere formation. The inhibition of EpCAM in MCF7 spheres was followed by signs of cell

death, differentiation and an epithelial mesenchymal transition process. Sphere cultures are used because they are enriched in cancer stem cells. The adverse effects of EpCAM inhibition in MCF7 spheres provides further evidence as to the role of this protein in cancer stem cells.

Acknowledgements

I would like to begin by thanking my supervisor Dr. John Hassell for giving me this incredible opportunity and for his invaluable guidance in developing this thesis. He was most understanding during the course of this prolonged thesis, for which I shall be eternally grateful. I am also most thankful to the other members of my committee; Dr. Bernardo Trigatti, Dr. Jonathan Bramson, and Dr. Karen Mossman whose criticism was most constructive in shaping my project. I would also like to extend my gratitude to Bonnie Bojovic and Adele Girgis-Gabardo, our wonderful lab technicians, whose instructions were crucial in my education. Bonnie and Alison Nixon also designed and prepared plasmid vectors that were used in this project, for which I am most appreciative. Furthermore, I would like to thank Andrew Giacomelli, Craig Aarts and Robin Hallett, whose advice pointed me in the right direction in carrying out my experiments. I would also like to acknowledge the contributions of the other members of the Hassell lab; Anna Dvorkin, Maria Kondratyev, Salma Trabelsi and Dora Ileiva. I would also like to mention members of the Biochemistry Department, whose continuous helps shall not be forgotten; Lisa Kush, Jennifer Seager and Dr. Brian Coombes. Last but not least, I will be forever indebted to my dear family, especially my incredible parents, for their unwavering support and sacrifices.

Table of Contents

List of Figures

List of Abbreviations

ALDH= aldehyde dehydrogenase

AML= acute myloid leukemia

bFGF= basic fibroblast growth factor

CFE= colony forming efficiency

CSC= cancer stem cell

CTC= circulating tumor cell

C-terminus= carboxyl terminus

ΔNEpCAM= dominant negative EpCAM

EGF= epidermal growth factor

EMT= epithelial mesenchymal transition

EpCAM= epithelial cell adhesion molecule

EpEx= EpCAM extracellular domain

EpICD= EpCAM intracellular domain

EpSH= EpCAM shRNA

EtOH= Ethanol

FHL2= four and a half LIM domains protein 2

ICD= intracellular domain

iPS= induced pluripotent stem cell

MET= mesenchymal epithelial transition

miRNA= micro RNA

MUC1= Mucin 1

NK= natural killer (cells)

PBS= phosphate buffered saline

PBST= PBS+ 0.03% Tween 20

RIP= regulated intramembrane proteolysis

shRNA= short-hairpin RNA

TACE= TNFα converting enzyme

TRE= Tetracycline response element

YFP= yellow fluorescent protein

Declaration of Academic Achievement

Ali Lessan designed, performed and analyzed the results of all the experiments. Ali Lessan wrote all the chapters of this manuscript. Dr. John Hassell provided the initial concept of this project and approved the final design of all the experiments. Dr. John Hassel reviewed and edited all the chapters of this manuscript. Bonnie Bojovic carried out the pilot studies that formed the basis for this project. Bonnie Bojovic and Allison Nixon designed and prepared plasmid vectors that were used during the course of this project.

1 Introduction

1.1 Cancer Stem Cell Model

There is, as of yet, no definitive explanation for the initiation and the development of cancers and how they often develop into a heterogeneous population of cells. There are two widely accepted models that attempt to explain this phenomenon (Figure 1). The more traditional model, known as the clonal evolution model relies on accumulation of stochastic genetic mutations in somatic cells resulting in unregulated proliferation and differentiation that produces cancer cells[1]. On the other hand, the cancer stem cell (CSC) model proposes a phenotypically distinct subset of cells at the apex of a hierarchy with the ability to self-renew and differentiate into non-tumorigenic cells that form the bulk of the tumor[2]. The idea of cancer stem cells, also called tumor initiating cells, was first suggested in 1977[3]. However, no real supporting evidence was put forth until 1994 with the work of John Dick's group on acute myloid leukemia (AML), who using flow cytometry techniques separated subgroups of cells with distinct cell surface marker patterns and studied their tumorigenicity by xenotransplantation into immunocompromised NOD/SCID mice. The leukemia initiating cells were shown to be a rare subgroup of $CD34^+$ $CD38^-$ cells[4,5]. The first solid tumor that was shown to follow the CSC model was breast cancer, with the tumor initiating cells being of a $CD44^+$ $CD24^{-/low}$ phenotype[6]. Subsequently, tumor initiating cells were isolated and characterized by their specific cell surface markers for many other solid tumors including brain[7], colon[8,9], pancreatic[10], melanoma[11] and ovarian[12].

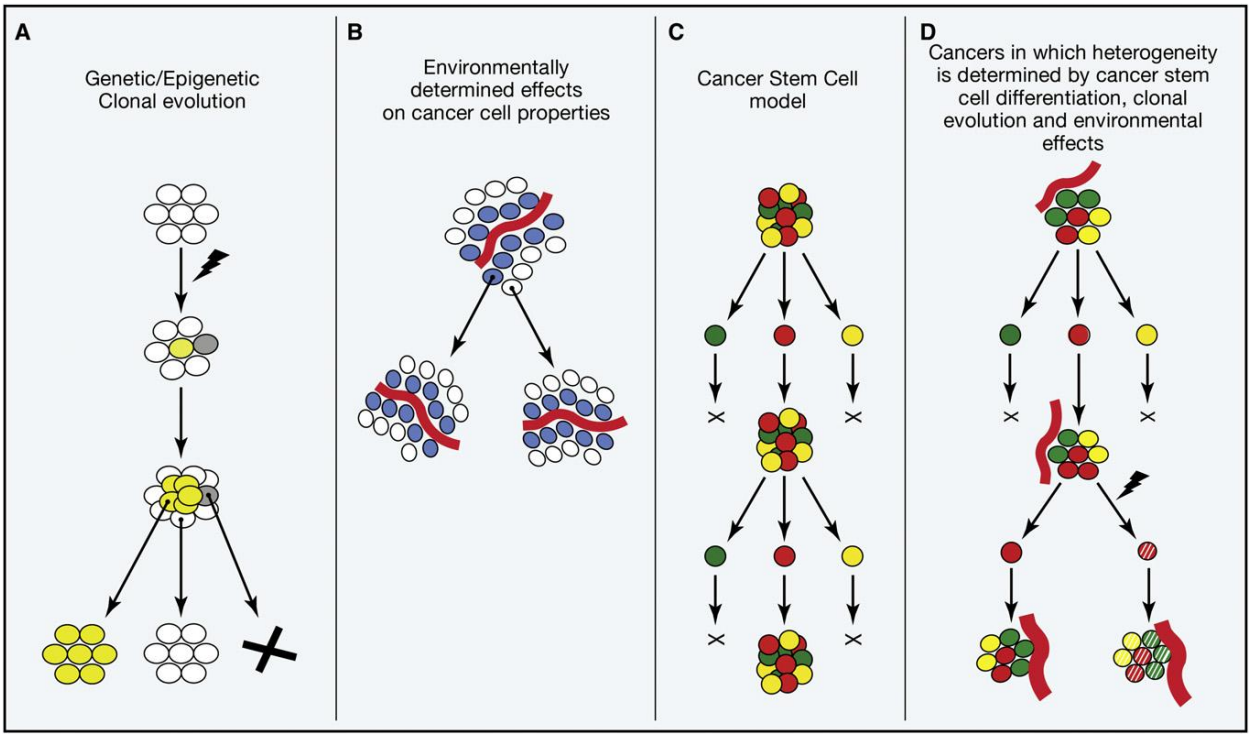


Figure 1: Models of cancer development and heterogeneity

a) The clonal evolution model explains tumor formation to be the result of accumulation of genetic mutations in population of cells. b) Factors associated with the microenvironment also affect the diversity of tumor cells. c) Cancer stem cell model defines tumor as hierarchy of cells, much like normal tissue, with the cancer stem cells, with self-renewal and differentiation capabilities, at the apex. d) Many believe the combination of all three to be true mechanism of carcinogenesis. (Adapted from Magee *et al.*[13])

The cancer stem cell model does not address the cell-of-origin issue. In some cancers the CSC may arise from normal stem cells with deregulated self-renewal and differentiation mechanisms. Alternatively, restricted progenitor or terminally differentiated cells that have gained such abilities can be the proverbial cell-of-origin in other cases. Knowing the cell-of-origin is likely to have implications on the properties of the cancer stem cell. The cell-of-origin may provide insight into the type of genetic mutations and signalling pathway changes that transformed it into a cancer stem cell[13].

There is debate on whether the cancer stem cell model can be analogously applied to all types of cancer. One criticism to this model is the universality of the cancer stem cell markers. It has been suggested that not all the individuals with same type of cancer present the same cell surface marker profile on the cancer stem cells. For instance, whereas leukemia initiating cells were previously shown to be $CD34^+$ $CD38^-$, more recent studies determined presence of cancer stem cells in the $CD34^-$ and the $CD38^+$ fractions in some AMLs[14,15]. In breast cancer as well, the $CD44^+$ $CD24^{-/low}$ phenotype does not always distinguish cancer stem cells from the bulk of the tumor[6]. CD133 was originally reported as marker for brain stem cells[7], but later studies found tumorigenic potential in the $CD133^-$ population as well[16,17]. A similar observation was made for the use of CD61 in isolating cancer stem cells in murine breast cancer. Whilst CD61 is effective in identifying cancer stem cells in Wnt1 driven breast cancer, it is not so for Her2/Neu tumors[18]. This discrepancy is not limited to the cancer stem cell phenotype but also its frequency within the tumor. Mouse AML resulting from a Pten deletion was shown to be very rare in CSC frequency[19]. Conversely, cancer stem cells are very common in MLL-

AF9 induced mouse AML[20]. All such data may attest to diversity between different individual tumors of the same type when it comes to the cancer stem cell model, may it be the surface marker phenotype, the frequency or even the extent to which they follow this model. The initial driver mutation can be the defining factor in the way one tumor behaves versus another[13]. Another explanation for such inconsistencies between experiments is the concept of plasticity. The differentiation of cancer stem cells into non-tumorigenic cells may be reversible. This may occur by an epithelial mesenchymal transition (EMT) which has been shown to induce stem cell phenotypes in normal or cancerous epithelial cells[21,22]. If cancer stem cells and the differentiated bulk tumor cells could indeed interconvert, it would explain the difficulties in reproducing the data related to the frequency and cell surface markers of the cancer stem cells.

Another point of controversy is the use of immunocompromised mice in the studies that have provided evidence for this model. Many claim that xenotransplantation assays hardly provide a microenvironment for the tumors resembling the human tissues[23]. There are differences in the stromal cells and growth factors, but mostly the immune conditions. The use of such mice that have virtually no immune system is problematic, since human body is in no way devoid of immunity. Despite xenotransplantation assays are in no way a good mimic for the human tissue, they are currently our only tool. Using immunocompromised NOD/SCID mice is the only way to escape the severe xenogeneic response by the mouse immune system. However, when NOD/SCID IL2Rγ^{null} (NOG) mice, which lack natural killer (NK) cells in addition to B and T cells, are used, the frequency of the cancer stem cells can significantly increase. For instance, this frequency

in human melanoma was originally suggested to be as low as one in a million in NOD/SCID mice[11]. However, by using NOG mice and optimizing several other factors this number rose to 1 in 4[24]. Furthermore, the frequency of cancer stem cell in mouse AML was shown to be as high as 1 in ten[23], whereas they had been previously thought to be much rarer[4,5]. Such high frequencies question the rarity of cancer stem cells, which despite not being part of the definition seems to be implied as part of the model, since it entails a hierarchal structure.

Such contradictions in universality of the cancer stem cell markers and frequency may provide a caveat to the cancer stem cell model. One can argue that at such high frequencies and in absence of a definitive marker, cancer stem cell model is flawed and the genetic mutations as per the clonal evolution model and factors in the microenvironment of the tumor are sufficient in explaining tumor heterogeneity. The most likely explanation is that the cancer stem cell model and the clonal evolution model are not mutually exclusive and both can operate independently or co-dependently. This means some types of cancer follow one model and some the other and sometimes the two models work in collaboration (Figure 1).

1.2 Breast Cancer Stem Cell

As mentioned previously, breast cancer stem cells were identified by surface markers Lin^{-} $CD44^{+}$ $CD24^{-/low}$[6]. In the same article it was suggested that the breast cancer stem cell population can be further enriched by selecting for the $EpCAM^{+}$ fraction of the

$CD44^+$ $CD24^{-/low}$ cells. 50,000 unsorted cells, 1000 $CD44^+$ $CD24^{-/low}$ cells or 200 $EpCAM^+$ $CD44^+$ $CD24^{-/low}$ cells were required to produce xenotransplants in mice. The tumors arising from the $EpCAM^+$ $CD44^+$ $CD24^{-/low}$ cells fully recapitulated the heterogeneity of the original tumor[6]. Aldehyde dehydrogenase (ALDH) activity is another approach for enriching breast CSCs. As few as 20 $EpCAM^+$ $CD44^+$ $CD24^{-/low}$ $ALDH^+$ cells are sufficient for generating a tumor[25]. These cells are also capable of forming tumorsphere *in vitro*. In turn tumorsphere cells are also capable of forming tumors in xenotransplantation assays[26]. The $EpCAM^+$ $CD44^+$ $CD24^{-/low}$ and $ALDH^+$ surface marker phenotypes also apply to numerous breast cancer cell lines[27,28]. Cancer stem cells have also been identified in many murine breast cancers such as Her2[29], Wnt1[30] and P53-null[31] driven breast cancers.

The cell-of-origin of breast cancer stem cells is unknown much like any other type of cancer. Normal stem cells are the most suspect cell of origin for cancer stem cells, due to their ability to accumulate mutations during their long lifespan enabling them to undergo transformation. It is also possible that the breast cancer stem cells develop from more differentiated cells undergoing an EMT, because they express many genes associated with EMT[21]. An EMT event can activate the Ras-MAPK pathway, which converts human mammary epithelial cells into breast CSCs[32].

Several pathways typically associated with normal mammary stem cells are also active in breast cancer stem cells, i.e. Wnt[33], Notch[34], Hedgehog[35] and PI3-K[36] pathways. These pathways are intimately involved in self-renewal mechanisms. For example, a

PTEN knock-down in MCF7 cells induces Akt, which in turn phosphorylates GSK3β and activates the Wnt pathway by β-Catenin accumulation in cytoplasm[36]. This leads to an increase in $ALDH^+$ cancer stem cell population, increased mammosphere formation and increased tumorigenicity. Other signalling moieties such as Her2 have been shown to increase the cancer stem cell frequency[37]. Conversely, the tumor suppressor P53 inhibits the symmetric division of cancer stem cells, which leads to a decrease in cancer stem cell population[38]. The breast CSCs also express Oct-4, a pluripotency marker in normal stem cells, a feat that enables these cells to differentiate and form heterogeneous tumors[39]. We introduced the concept of plasticity in cancer stem cell differentiation before. Such a phenomenon has also been observed in breast cancer. For instance, certain human basal epithelial cells spontaneously dedifferentiate into stem-like cell. Such occurrences are even more enhanced in malignant tissue[40]. IL-6 presence has also been shown to generate CSCs from non-tumorigenic cancer cells[41]. This implies a dynamic equilibrium between CSCs and differentiated cells.

Chemo and radiation therapy have been shown to enrich for breast cancer stem cells, confirming the resistant nature of these cells relative to the bulk of the tumor[42]. This is again due to the activation of self-renewal pathways Wnt, Notch and Hedgehog, which are also involved in cell fate decisions. Notch for instance induces expression of Survivin[43], which deregulates multiple cell cycle check points and prevents radiation or drug induced apoptosis[44]. Cyclin D1, also an important regulator of the cell cycle and governs cell survival, is also a target of the Notch[43] and Wnt pathways[45]. Cyclin D1 is also necessary for self-renewal and differentiation of breast cancer stem cells and

promotes tumorigenicity[46]. There is a direct correlation between breast cancer stem cells' survival and cancer stage, metastasis, relapse and decreased survival[47]. The resistance of breast cancer stem cells to classic treatments necessitates development of therapeutics directly targeting these cells. Such therapies can target cell surface markers of breast cancer stem cells, interfere with the vital signalling pathways or go after the drug resistance mechanisms onboard these cells. We will discuss EpCAM targeting antibodies in the next section.

1.3 Epithelial Cell Adhesion Molecule

The function of many cell surface markers for cancer stem cells and how or whether they are involved in tumorigenesis is not fully understood. This is unfortunate, because a better functional understanding of these proteins would provide a better platform for targeting CSCs. CD44, as one of the most common cancer stem cell markers, has been shown to promote adhesion and initiate a signalling pathway[48]. Upon activation, CD44 undergoes a regulated intramembrane proteolysis (RIP). The CD44 intracellular domain (ICD) relocates to nucleus where it regulates the transcription of gene with TPA-responsive elements[49,50]. Similarly, EpCAM was originally defined as an adhesion molecule until recently it was shown to drive its own signalling pathway[51]. As the main focus of our study, we shall go into further details about EpCAM.

Epithelial cell adhesion molecule (EpCAM) or CD326 is a 39-42 kDa, glycosylated type I membrane protein. This molecule consists of an extracellular domain

(EpEx) with epidermal growth factor (EGF) and thyroglobulin repeat-like domains, a single transmembrane domain, and a short intracellular domain (EpICD) comprising 26-amino acid residues[52]. EpCAM was first described as a homophilic cell-cell adhesion molecule by Litvinov *et al* in 1994[53]. The intracellular domain of EpCAM connects to the cytoskeleton via α-actinin[54]. The same group discovered that EpCAM expression reduces cadherin mediated intercellular interactions[55] via PI3 kinase[56]. EpCAM is expressed in variety of cells such as in human epithelial tissue, progenitors and stem cells as well as cancer cells. EpCAM contributes to the maintenance of murine embryonic stem cell phenotype[57]. EpCAM null mice die in utero, and the embryos are small, developmentally delayed, and display prominent placental abnormalities[58].

EpCAM is expressed on essentially all human adenocarcinomas, on certain squamous cell carcinomas, on retinoblastomas, and on hepatocellular carcinomas[59]. Specifically, EpCAM was shown to be over-expressed in breast cancer[60], with much higher expression in metastatic tumors relative to the corresponding primary tumor[61]. Expression of EpCAM is indeed recognised as a marker of poor prognosis in node-positive invasive breast carcinoma[62] and node negative, including triple negative, breast carcinoma[63]. By contrast, high EpCAM expression in pancreatic cancer cells was shown to suppress proliferation and indicated low malignancy and a good prognosis[64]. EpCAM is also expressed at high levels on other solid tumor types, including prostate, colon, gastric, ovarian, and lung cancer[65,66]. The distinction of EpCAM between normal epithelial cells and cancer cells is not limited to its expression level, but also its distribution across the membrane. Whereas EpCAM is mainly distributed in baso-lateral

intracellular spaces with tight junctions in epithelial cells, its distribution is homogeneous on cancer cells' surfaces[52]. The localization of EpCAM to the baso-lateral membrane in normal tissue is explained by its association with claudin-7 and CD44[67]. The claudin-7-EpCAM complex encourages proliferation in addition to up-regulating antiapoptotic and drug resistance proteins, which leads to cancer progression and metastasis[68].

Involvement of EpCAM in tumorigenesis was further corroborated when EpCAM was proven to have a direct impact on cell cycle and proliferation, by rapidly up-regulating the proto-oncogene *c-myc* and cyclin A/E[69]. It is also highly glycosylated in tumor cells relative to healthy tissue[70], which leads to an increase in its half-life[71]. EpCAM inhibition experiments, using siRNA, decreased the proliferation and migration of breast cancer cell lines[60]. Thus, EpCAM was recognized to be involved in regulation of proliferation, but nothing was known about its possible mechanism of action. The first clue was provided by Maaser and Borlak in 2008, who demonstrated regulation of genes involved in cell cycle and proliferation by EpCAM, which suggested a role in signal transduction pathways[72]. However, it was in 2009 that in a seminal paper by Maetzel *et al.* clearly demonstrated the EpCAM signalling pathway. EpCAM undergoes a regulated intramembrane proteolysis (RIP), triggered upon binding of the extracellular domain to another EpCAM molecule (Figure 2). This cleavage is sequentially catalyzed by TNFα converting enzyme (TACE), which leads to release of EpEx into the extracellular matrix, and presinilin-2 γ-secretase, which causes shedding of the intracellular domain into the cytoplasm. This mechanism explains the necessity of cell-cell contact to provide the

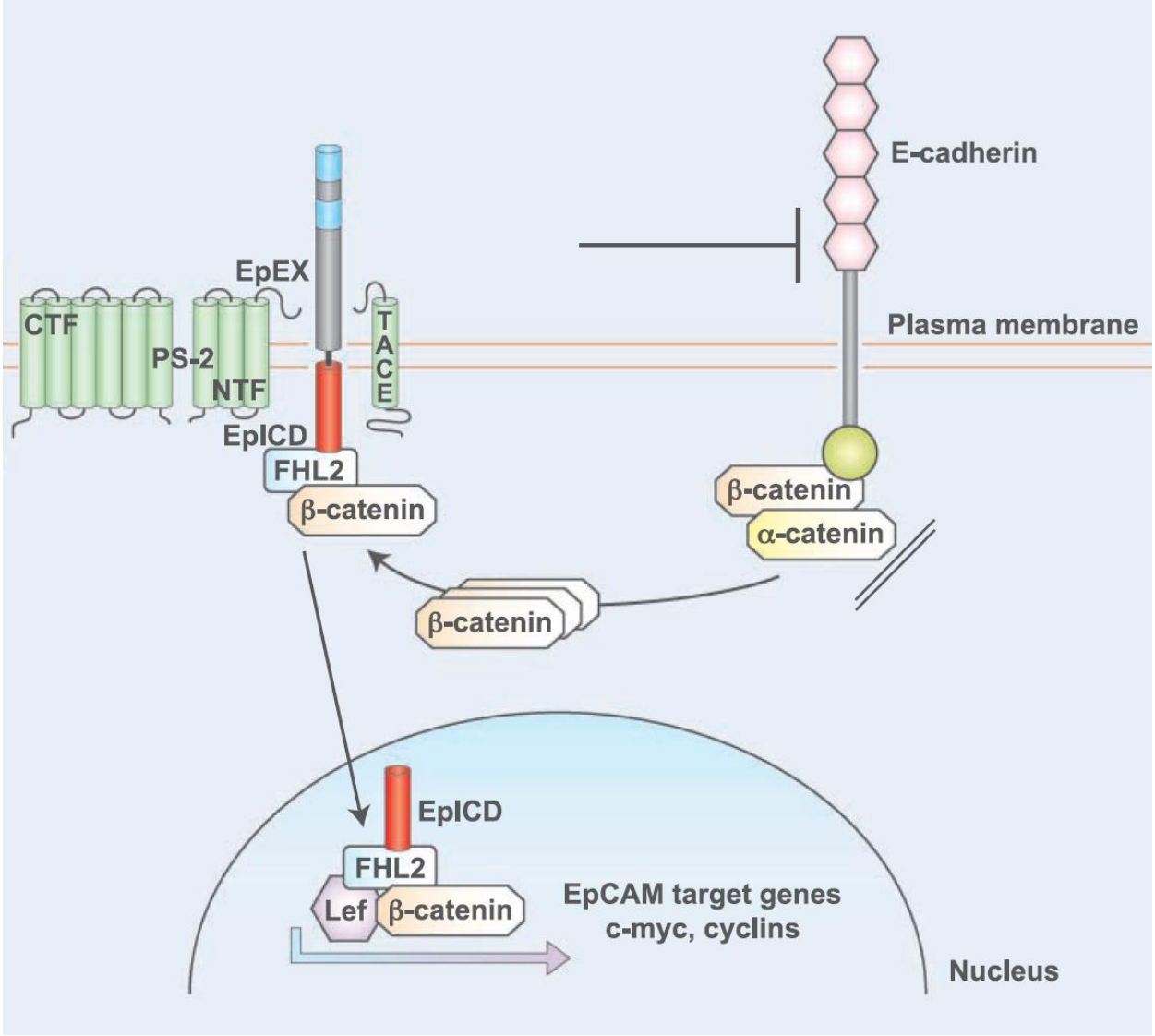


Figure 2: EpCAM signalling Pathway

Cell adhesion molecule EpCAM undergoes a RIP by enzymes TACE and presinilin-2 γ-secretase complex. The Ep-Ex promotes cleavage of other EpCAM molecules. The EpICD forms a complex with FHL2 and β-catenin translocates to the nucleus. LEf1 then joins the signalosome complex which enhances expression of target genes with a Lef1 recognition sequence. (Adapted from Munz *et al.*[52])

initial signals of cleavage to activate EpCAM signalling, which can be further propagated via soluble EpEx in an auto- and paracrine fashion providing a positive feedback loop[73]. The EpICD associates with a scaffolding protein four and a half LIM domain protein 2 (FHL2) and the co-activator β-catenin and also the transcription factor Lef-1 to form a nuclear complex. The EpCAM signalosome then activates transcription of genes with the Lef-1 consensus sequence[51]. An active EpCAM pathway, represented by high nuclear and cytoplasmic localization of EpCAM was observed in multiple types of cancer including breast, prostate, head and neck, esophagus, lung, colon, liver, bladder, pancreatic, and ovarian[74]. A high EpCAM nuclear localization has been associated with the aggressiveness of thyroid tumors and indicates a poor prognosis[75].

As part of a transcriptional regulation pathway, EpCAM expression and activation levels can be very consequential. To begin with, EpCAM over-expression can lead to up-regulation of its target genes such as the previously discussed proto-oncogene c-myc[69]. EpCAM up-regulation was also associated with an increase in cyclins A and E. It has recently been shown that this is through the induction of cyclin D1, which is also an EpCAM target gene. It seems EpCAM's control of cell cycle is via the classical cyclin-regulated pathway. Logically, EpCAM expression positively correlates with the expression of proliferation marker Ki-67[76]. Furthermore, knocking down EpCAM in lung cancer cell lines reduced proliferation, clonogenicity and invasive properties and induced apoptosis[77]. EpCAM expression is mandatory for the self-renewal and the pluripotency of human embryonic stem cells[78], as several key transcription factors paramount to pluripotency in embryonic stem cells are targets of EpCAM. These include *oct4*, *sox2*,

klf4 and *nanog* in addition to the previously mentioned *c-myc*[79]. Thus, EpCAM seems to play a central role in reprogramming of differentiated cells into induced pluripotent stem cells (iPS)[80]. An argument can be made for EpCAM signalling being partially responsible for the undifferentiated and proliferative phenotype of $EpCAM^+$ cells such as embryonic, adult, cancer and induced pluripotent stem cells and tissue progenitor cells[81]. Expression of EpCAM on hepatic stem and progenitor cells, despite complete loss of EpCAM in mature hepatocytes, is supportive of this notion[82]. The expression of EpCAM in most normal epithelia in no way negates this hypothesis, since activation and nuclear localization of EpCAM is lacking in these cells[74].

The considerable downstream effects of EpCAM elevate the importance of the factors that regulate the expression of EpCAM. The extent of knowledge about the upstream regulators of EpCAM is limited. Wnt signalling was shown to induce EpCAM expression by TCF4 in hepatocellular carcinoma[83]. A cross-talk between Wnt and EpCAM can also be seen at the protein level, since EpICD binds β-catenin and Lef1. Additionally, TNFα is known to down-regulate EpCAM through the activation of NF-kB[84]. Tumor suppressor p53 protein also binds to a response element within the EpCAM gene and negatively regulates EpCAM expression, which contributes to p53 control of breast cancer invasion[85].

Over-expression of EpCAM on many tumor types makes it a good candidate for targeted therapeutic practices. Expression of EpCAM on normal epithelia does not limit this option, since EpCAM is localized to the baso-lateral membrane of these cells. This

makes EpCAM far less accessible to therapeutic antibodies relative to cancer cells, which have a homogenous EpCAM distribution across their membranes[52]. This was clearly proven in a transgenic mouse model, where the monoclonal antibody preferentially bonded with the xenograft cells versus the healthy mouse cells expressing human EpCAM[86]. EpCAM directed monoclonal antibody therapy of cancer has been of interest for many years. In fact, EpCAM was first detected in 1979 as a colorectal carcinoma-specific antigen using murine monoclonal antibody 17-1A before the EpCAM molecule was even characterized[87]. The same antibody otherwise known as Edrecolomab was used in treatment of colon[88] and breast cancer patients[89]. In spite of initial optimism, the trials with Edrecolomab were stopped when its clinical benefits were questioned and eventually refuted[90]. The clinical trials of the fully human EpCAM antibody Adecatumumab (MT201)[91] provided some response to the $EpCAM^{high}$ breast tumors in terms of metastasis, but no clear tumor regression[92]. Most recent development is Catumaxomab, a trifunctional antibody with dual specificity for EpCAM and CD3, a marker for cytotoxic T cells[93]. Catumaxomab was recently approved in Europe for the treatment of patients with malignant ascites that arise as a result of epithelial cancers. The results of the clinical trials demonstrated a clear advantage for the patients in terms of overall survival[94]. Despite the effectiveness of many therapeutic antibodies in treating cancers, there are many problems attributed to them as well including safety profiles and serum half-life. Future projects should investigate treatment options that interfere with the EpCAM signalling cascade, for instance by blocking the RIP of EpCAM protein via inhibition of the TACE or presinilin-2 γ-secretase enzymes.

Another major application of the EpCAM protein in the field of cancer research and therapy is to detect circulating tumor cells (CTC) in the peripheral blood. The commercially available CTC detection system CellSearch employs the use of anti-EpCAM antibody covered beads as the initial step in enriching for these cells[95]. This may be used as prognostic tool for individuals with metastatic breast cancer[96]. However, although EpCAM is expressed in almost all epithelial cancer cells, there are rare tumor cells with low or no EpCAM expression[97]. It is known that many CTCs with metastatic properties have undergone an EMT and express markers such as vimentin or fibronectin[98,99]. Such migratory mesenchymal cells have lost or down regulated many properties of epithelial cells such as EpCAM expression. Therefore it is possible that EpCAM dependent CTC detection systems might very well miss these important circulating tumor cells[100,101]. Yet co-expression of epithelial marker EpCAM and mesenchymal markers has also been observed on CTCs of advance breast and prostate cancer[102]. Mature differentiated cells were previously mentioned as the cell of origin for cancer stem cells as they may be able to revert back to a cancer stem cell phenotype by EMT. CTCs and CSCs can be similar in this regard. CSCs are EpCAM expressing cells whose formation may be initiated by EMT. Similarly, tumor cells undergo EMT to produce CTCs. At the onset of EMT EpCAM expression may decrease to ease detachment. But eventually CTCs up-regulate EpCAM to promote proliferation and invasion to facilitate formation of metastasis.

The relationship between EMT and EpCAM expression seems to be dynamic. Proliferative epithelial $EpCAM^+$ cells readily interconvert with migratory mesenchymal

$EpCAM^-$ cells in squamous cell carcinoma[103]. The same plasticity between the EMT and the MET states was shown for breast cancer cells and was revealed to be regulated by microRNAs[104]. Furthermore, EMT markers' expression is the known hallmark of migratory cells. EpCAM also seems to contribute to such invasiveness. Knocking out EpCAM in Langerhans dendritic cell of mice stripped them of their migratory abilities[105]. As an adhesion molecule, it seems odd for EpCAM to be pro-migration. Nonetheless, EpCAM's activation of JNK/AP-1 pathway may explain this EpCAM dependent invasive behaviour[106]. Moreover, EpCAM is not only an adhesive molecule. Once activated, EpCAM initiates a signalling cascade that enhances the expression of many target genes. Indeed, EpCAM signalling pathway was shown to be highly activated in the budding cells at the invasive front of colorectal cancer[107]. EpCAM activation implies EpCAM cleavage. Hence, in such migratory cells EpCAM is no longer functioning as an adhesion molecule, but it can stimulate many processes that contribute to such migratory characteristics through the EpCAM signalling cascade. EpCAM's ability to inhibit cadherin mediated junction must also be taken into account[55], as it provides further insight into how EpCAM contributes to cell migration.

2 Experimental Design

We observed an increase in EpCAM expression in the breast tumors of the MMTV-Neu mouse model relative to the normal mammary tissue. To investigate the possible function of EpCAM in tumorigenicity of this mouse model, we decided to manipulate the expression and activity of EpCAM in breast cancer cell lines. The choice of starting our experiments with cell lines in place of primary cultures was mainly due to the limited longevity of primary cultures. Our hope was to perform our proof of principle experiments in cell lines before proceeding to primary cultures. We began by increasing EpCAM expression using previously designed over-expression vectors[51]. We also prepared vectors expressing shRNA constructs targeting EpCAM to inhibit its expression. Simultaneously, we designed a dominant negative allele of EpCAM. This construct, when introduced into breast cancer cell lines, can partially replace the endogenous EpCAM protein as part of its signalling complex and thus hinder its activity. Our goal was to subject these manufactured cell lines with EpCAM over expression, under-expression or blocked EpCAM activity to functional assays. By studying the effects of EpCAM manipulation on the activity of the cell lines, we hoped to illustrate EpCAM's function. We hypothesize that EpCAM is highly involved in processes that bestow self-renewal and pluripotency capabilities to breast cancer stem cells.

3 Material and Methods

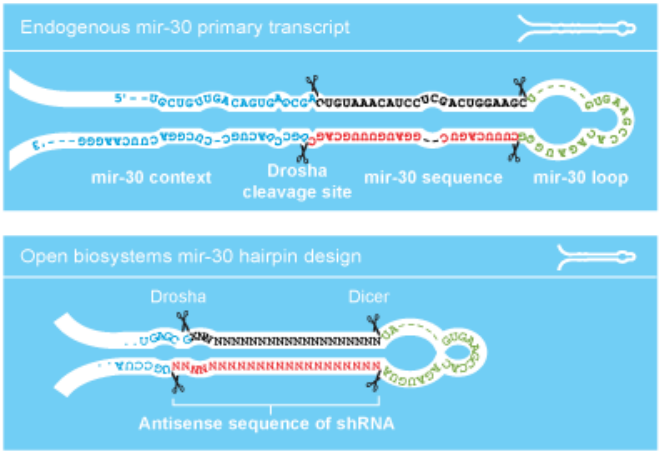
3.1 Dominant Negative EpCAM

To inhibit EpCAM activity, a dominant negative allele of this protein was designed. This allele was designed by fusing the repression domain (amino acids 2-298) of the Drosophila Engrailed protein to the Carboxyl(C)-terminus of the EpICD domain (the last 26 amino acids at the EpCAM C-terminus). This dominant negative EpCAM (ΔNEpCAM) fusion protein was cloned into the inducible pSLIK vector[108] and packaged into lenviral particles.

3.2 EpCAM Short Hairpin RNA

Two validated short-hairpin RNA sequences[109] against human EpCAM were chosen from the online Open Biosystems library[i]. The backbone of the mir-30 micro-RNA was, and the acquired sequence fragments were inserted in place of the mature miRNA sequence within the template (Figure 3a). The use of a micro RNA template allows for processing of the transcribed shRNA by Drosha and Dicer enzymes[110]. This facilitates the active loading of the processed siRNA into the RISC complex, which binds to the target mRNA (EpCAM in this case) and blocks its translation (Figure 3b). The use of the miRNA template has another advantage, since miRNAs are transcribed by RNA polymerase II, which is necessary for the use of the inducible pSLIK system. The shRNA sequences are as follows:

a)



b)

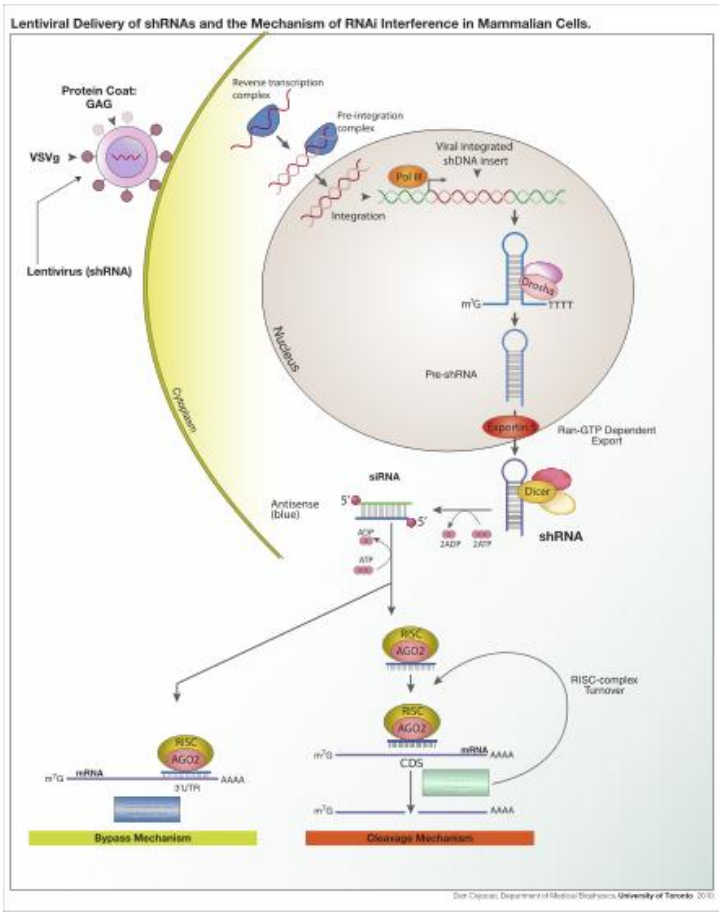


Figure 3: The design and mechanism of function of shRNA constructs

a) shRNA sequences are inserted into a mir-30 backbone. b) shRNA is processed in the cells and the resulting siRNA blocks translation of the target as part of the RISC complex. (Adapted from Open Biosystems website[i])

EPsh12:

5’
CGGTCATGAGTTTGTTAGTTTATAGTGAAGCCACAGATGTATAAACTAACAAA
CT CATGACCT 3’

EpSH25:

5’
ACCGTAAACTGCTTTGTGAATATAGTGAAGCCACAGATGTATATTCACAAAG
CAG TTTACGGC 3’

These sequences were cloned into our inducible vector pSLIK and packaged them into lentiviruses.

3.3 Plasmid Vectors

Four plasmids were used as the fundamental blocks of our work. peYFP-N1was the backbone of our EpCAM over-expression plasmids[51]. pSLIK-Neo was the doxycycline inducible vector that was used to express our ΔNEpCAM and EpCAM shRNA constructs[108]. 7XTFP was the TOPFlash TCF reporter plasmid expressing firefly luciferase[111]. pLenti-CMV Renilla was the plasmid was used as a general transcription reporter[112].

3.4 Lentivirus Preparation

Maxipreps of the pSLIK plasmid containing each of our constructs (ΔNEpCAM, EpSH12, EpSH25) were prepared using the PureLink HiPure plasmid Maxiprep kit (Invitrogen catalogue number (cat#) K2100-07). The vectors were to be packaged using the Invitrogen ViraPower Lentiviral Expression Systems from Invitrogen[ii]. Maxipreps of the three packaging plasmids pLP, pLP1 and pLP2 was also prepared. $5.0x10^6$ 293FT cells were seeded 24 hours in advance in 10 cm tissue culture plates in DMEM medium with 10% FBS supplemented with 0.1 mM non-essential amino acids (Invitrogen 11140-050) and 2 mM L-glutamine. On the day of transfection, the existing media was replaced with fresh complete antibiotic-free medium. The cells in each 10 cm plate were transfected with 1.4 pmol of each Lenti vector plus 0.7 pmol of each of the three packaging plasmids using Lipofectamine 2000 reagent (Invitrogen 11668). The following day the media was changed to fresh complete medium with antibiotics. After 48 hours the cells were removed from the plate by pipetting the media. The remaining cellular debris was pelleted down by centrifugation at 3000 rpm for 15 minutes at 4^0C. The viral supernatant was filtered through a 0.45μm filter and then ultracentrifugated at 28,000rpm for 90min at 4^0C. The viral pellet was resuspended in PBS+1% BSA and stored at -80°C. To titer the viral lysates MCF7 cells were pre-seeded in 6 well dishes at 100,000 cells/well. 10 fold serial dilutions of the viral lysate were prepared from 1:1000 to $1:10^6$. Each dilution was added to one well in addition to the mock infection. After 48 hours fresh media with the selection agent (G418) was added to the wells. The selection continued until no cells are left in the mock infected well. The average titer of the lysate

was calculated using the number of resistant colonies remaining in the well. For the subsequent infections, the multiplicity of infection used was 1:10.

3.5 DNA Transfection

For the transfection experiments cells were seeded in 24 well dishes at 200,000 cells per well. The following day, DNA samples (Refer to the Results section for the specific amount of DNA used from each transfection experiment) were mixed and then combined with Lipofectamine 2000 reagent (Invitrogen 11668-019) and added to the cell in antibiotic free media. 3 μl of lipofectamine reagent was used for each 1 μg of DNA. Each transfection condition was done in triplicate. 24 hours later the cells were trypsinized and reseeded in opaque 96 well plates at 20,00 cells per well.

3.6 Cell Viability and Luminescence Assays

The cells were seeded in opaque 96 well dishes at 20,000 cells per well. They were grown (and induced with doxycycline at 1 μg/ml) for 4 days. On day 4 AlamarBlue was added to the wells and incubated for 4 hours. Readings for relative fluorescence units were read using a fluorometer for one second. The fluorescence readings are due to resazurin reduction which is a measure of cell viability. Subsequently, luminescence assay was performed using the Steady Glo reagent (Promega E2520) to measure firefly

luciferase activity. In other cases, firefly and renilla luciferase activities are measured in the same assay using the Dual Glo reagent (Promega E2920).

3.7 Quantitative RT-PCR

Cells were seeded in 10 cm plates and induced with doxycycline at 1 μg/ml concentration for 4 days before being lysed for RNA extraction, which was done using PureLink RNA Mini kit (Invitrogen 12183018A). cDNA was prepared using 2.5μg RNA as a template, oligo-dT primers, and SuperScript II RT enzyme (Invitrogen 11904-018). Expression levels of genes were determined using the FastStart DNA Master SYBR Green I kit (Roche 03003230001) on a LightCycler (Roche). The results were normalized using the house keeping gene GAPDH. The sequences of the primers used are:

Engrailed: Forward 5' CGCCTCGTCCACGCCGCTGA 3'
Reverse 5' GCGTCCGCCCGCCGACGTATC 3'

EpCAM: Forward 5' GCGGCTCAGAGAGACTGTG 3'
Reverse 5' CCAAGCATTTAGACGCCAGTTT 3'

3.8 Western Blot

Cells were seeded in 10 cm plates and induced with doxycycline at 1 μg/ml concentration for 4 days before being lysed, using a modified NP40 lysis buffer, for protein extraction. Lysate concentrations were measured using Bradford assay and 50 mg protein was loaded on Mini-Protean TGX gels (Biorad 456-1021). The gels were ran at 200V for one hour

using a Mini-Protean Tetra Cell (Biorad cat#165-8000). After the run, the gels were transferred to PVDF membranes overnight at 4°C. A Mini Trans-Blot transfer cell (Biorad 170-3930) was used. The membranes were then placed in blocking buffer (3% BSA in PBST (PBS with 0.1% Tween 20)) for 4 hours at room temperature or overnight at 4°C. The blocked membranes were then incubated overnight at 4°C in the first primary antibody diluted in blocking buffer. The membranes were then thrice washed in PBST for 10 minutes. Alexa Fluor 680 secondary antibody diluted in blocking buffer (supplemented with 0.01% SDS) was added to the membranes at room temperature for one hour. A LI-COR Odyssey infrared imaging system was used for fluorescence detection. The second primary antibody was used subsequently.

The list of antibodies used is as follows: EpCAM (Epitomics cat#1144-1), Engrailed (DSHB 4D9).

Modified NP40 lysis buffer recipe:
50 mM Tris-HCl, pH 7.4
250 mM NaCl
0.5 % NP40 (IGEPAL)
10% glycerol
4 mM EDTA (pH 8.0)

3.9 Tissue Culture

All the media used in our experiments (with the exception of transfection

experiments) were supplemented with antibiotics penicillin (100 units/ml), streptomycin (100 μg/ml) (Invitrogen cat#15140-122) and antimycotic amphotericin B (1 μg/ml) (Invitrogen Fungizone 15290-018). The adherent MCF7 cells were grown in tissue culture treated 20 ml flasks in DMEM medium with 10% fetal bovine serum (FBS). The cells were passaged approximately every 4 days, or upon confluence. To passage, the media was siphoned off, 1 ml of trypsin was added to the flask and incubated at 37°C for 5 minutes. After ensuring detachment, 9 ml of fresh complete medium was used to completely remove the cells. For routine sub-culturing the cells were split 1:4 to 1:8. To seed at a precise density the cells were counted with Trypan blue dye using a haemocytometer.

The adherent MCF7 cultures were converted to spheres by growing them in human stem cell media (hSCM). To make human stem cell media, mix low glucose DMEM with HAM's F12 media in a (3:1) ratio. Add the following supplements to the corresponding final concentration:

Human EGF (20 ng/ml)
Human bFGF (10 ng/ml)
Insulin (5 μg/ml)
Heparin (4 ng/ml)
B27 supplement 50X (Invitrogen 17504-044) (1:500 dilution)

The trypsinized adherent cells were washed with HAM's F12 to remove any remaining serum. The cells were then seeded in human stem cell media at 30,000 cells/ml. To

establish a sphere culture, the cells were initially grown in hSCM for 7 days. Further sub-culturing was done every 5 days. To passage the sphere cultures, the cells were pelleted by centrifugation at 1000 rpm for 5 minutes. 250 µl of trypsin was added to the pellet from a 20 ml flask. The cells were mixed by pipetting and incubated for 30 seconds at 37°C. Next, 1 ml of HAM's F12 was added and the dissociated cells were pelleted by centrifuge. The cells were resuspended in fresh human stem cell media and seeded at 30,000 cells/ml. The spheres need to be established for at least 3 passages before performing the experiments.

3.10 Colony Forming Assay

MCF7 cells were seeded in 6 well dishes at 200 cells per well in DMEM medium with 10% FBS and grown for 7 days. On day 7 the media was siphoned off and 1 ml of methanol was gently added to each well, to avoid the colonies lifting off. The cells were fixed with methanol for 20 minutes at room temperature. After removing the methanol, the cells were washed with PBS for 1 minute. Colonies were then stained with Wright's Giemsa stain for 20 minutes at room temperature. Finally, the plates were gently rinsed with ddH2O and air-dried overnight.

3.11 Sphere Forming Assay

For the sphere forming assays the dissociated sphere cells were seeded in 96 well

plates at 30,000 cells/ml (200 µl per well). To induce the expression of the integrated inducible lentivectors doxycycline was added at 1 µg/ml. Spheres were counted on day 5.

3.12 Preparing Sphere Blocks

A 2% agar solution was prepared by adding cell culture grade agar in isosmotic PBS. The mixture was microwaved to dissolve the agar and subsequently kept hot on a hot plate to prevent solidification. To prepare sphere blocks, $5.0x10^5$ to $1.0x10^6$ cells are ideal. The sphere cultures were pelleted at 500 rpm. The pellet was gently resuspended in 100 to 500 µl of paraformaldehyde (PFA) (depending on the number of cells) using a large bore pipette tip. After 10 minutes, equal amount of the 2% agar solution was added to the cell in PFA inside a 1.5 ml microtube and mixed gently to avoid bubbles. The tubes were kept on ice for 15 minutes to solidify the agar. The agar block was then removed from the tube and kept in a tissue cassette and fixed in formalin for 48 hours. The cassettes were then stored in 70% ethanol. The cell blocks were embedded in paraffin wax and sectioned and mounted onto glass slides at the McMaster core histology facility.

3.13 Immunofluorescence

Since the sections are embedded in paraffin, they first need to be dewaxed. De-waxing entails submerging the slides in solvents according to the following steps: 5 min in xylene (2 times), 3 min in 100% EtOH (2 times), 2 min in 95% EtOH and 2 min in 70

EtOH. An antigen retrieval process was employed to expose the target antigens. The (1:100) diluted antigen unmasking solution (Vector Labs H-3301) was microwaved for 3 minutes to boil. The slides were submerged in the solution and leave for 10 minutes. The slides were then washed with PBS for 5 minutes (2 times). 150ul of blocking buffer (3% normal goat serum in PBST) was applied to the slides for 45 minutes. The blocking buffer was replaced with the primary antibody diluted in blocking buffer (based on the manufacturer's recommendation) and left overnight at 4°. The next day the slides were washed with PBS for 5 minutes (3 times), and then the diluted Alexa Flour 2° antibody in blocking buffer was added and incubated for 45 minutes. The slides have to be protected from light after adding the 2° antibody. The slides were then washed with PBS for 5 minutes 2 times), followed by a wash with PBST (PBS with 0.03% Tween 20) for 5 minutes and finally another wash with PBS for 5 minutes. After a final rinse with PBS a drop of mounting media with DAPI was added and a coverslip was placed on each slide. The slides were imaged after 15 minutes.

The list of antibodies used is as follows: EpCAM (Epitomics cat#1144-1), CK8 (DSHB TROMA-I), α-SMA (Sigma A2547), Ki-67 (Abcam AB15580), Cyclin D1 (Abcam ab16663), β-Catenin (Abcam Ab32572), E-Cadherin (BD 610182), Vimentin (Sigma V5255), MUC1 (Neomarker HM-1630-P1).

4 Results

4.1 Endogenous EpCAM Expression

To determine the relevance of the EpCAM molecule in our breast cancer model system, we investigated the expression level of EpCAM in mammary gland isolated from wild-type FVB mice and mammosphere driven as a primary culture from these glands, as well as primary tumor samples isolated from our breast cancer mouse model that carry the MMTVneu allele[113], and the resulting tumorspheres. Using multiple techniques, we observed the expression of EpCAM from different aspects. qRT-PCR analysis was done to measure the mRNA levels of EpCAM in these samples. Western blot was performed to examine the protein levels. Immunofluorescence was employed to examine the expression pattern of EpCAM on a cellular level. On both mRNA and protein levels, EpCAM was shown to be highly over-expressed in primary tumor samples. Mammary gland and tumorsphere samples similarly showed moderate levels of EpCAM expression (Figure 4). The expression level in mammosphere samples was very low. The Immunofluorescence results corroborated these data. EpCAM expression was homogenous in both primary tumor and tumorsphere samples, but at a much higher level in primary tumor. In mammary gland EpCAM was highly expressed in luminal cells, but not on myoepithelial cells (Figure 5).

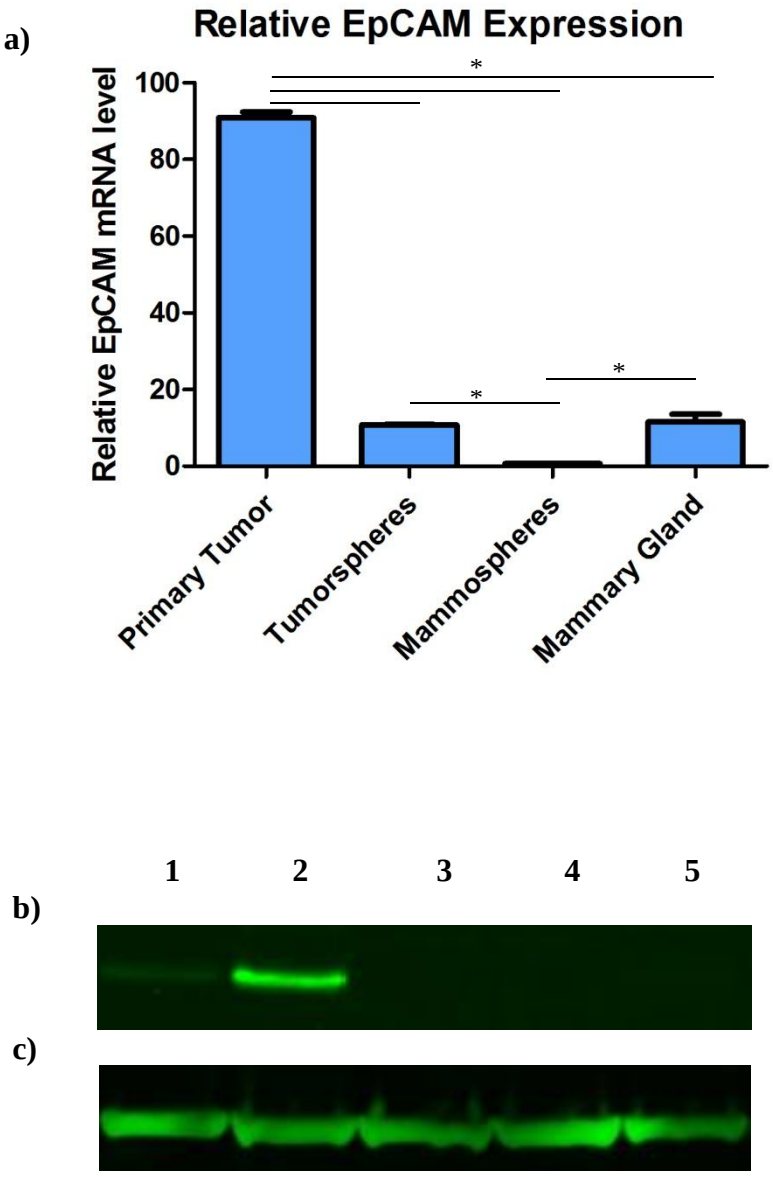


Figure 4: EpCAM expression in breast tissue and cell culture

a) Relative expression of EpCAM at RNA level measured by quantitative RT-PCR (One-way post-Tukey, $P<0.0001$). b) Western blot analysis with b) anti-EpCAM antibody, c) anti- α-tubulin antibody. Samples: 1. mammary gland, 2. primary tumor, 3/4. mammospheres, 5. tumorspheres.

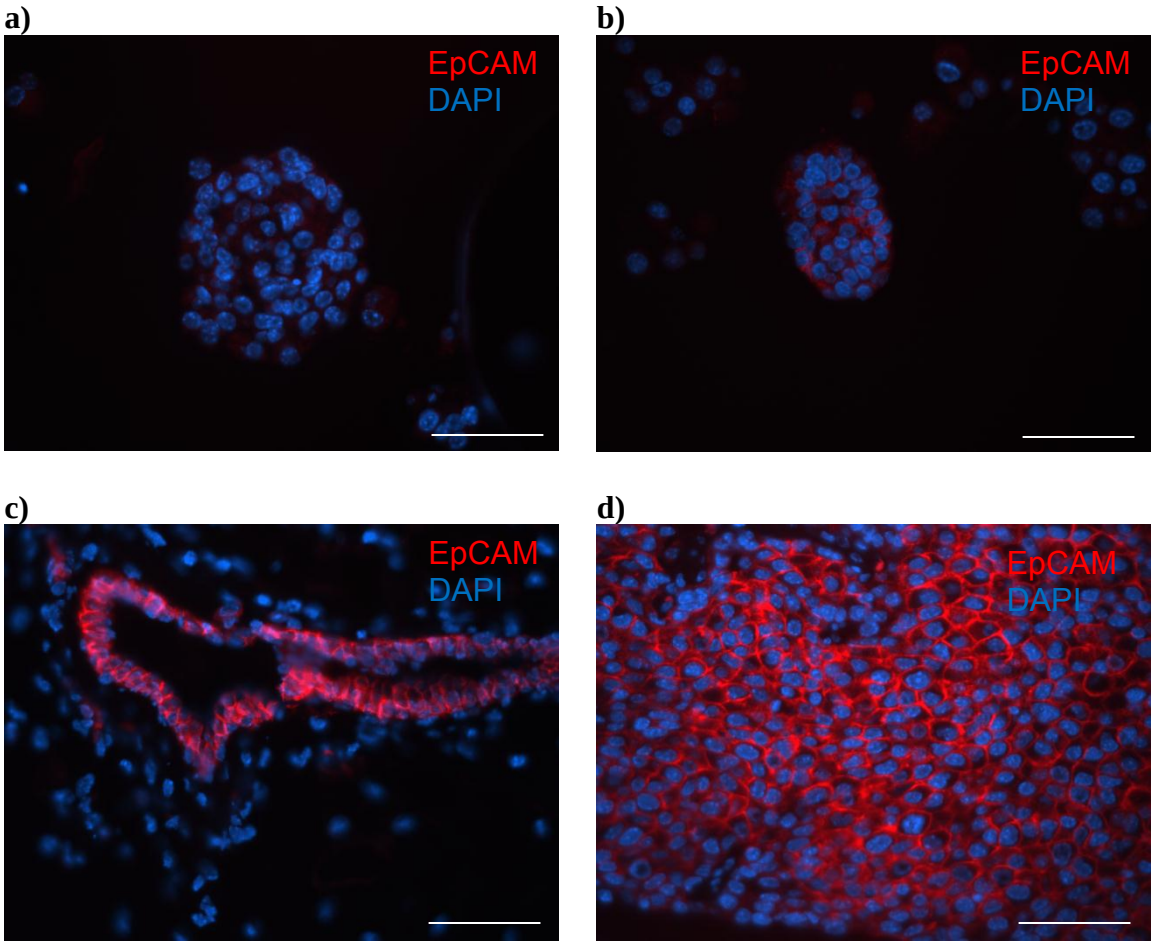


Figure 5: EpCAM expression in breast tissue and cell culture

Immunofluorescence staining with anti-EpCAM antibody(red) and DAPI(blue) of Zn-fixed sections of a) mammospheres, b) tumorspheres, c) mammary gland and d) primary tumor. 400X magnification, scale bar 50 um.

4.2 EpCAM Over-expression

To study the effects of EpCAM over-expression on tumor initiating cells we used plasmid vectors, which express either the full length EpCAM or the EpICD signalling domain fused to a yellow fluorescent protein (YFP) reporter, under the control of a CMV promoter[51]. These plasmids were used for transient transfection assays. We also inserted EpCAM-YFP and EpICD-YFP constructs into a pSLIK plasmid[iii]. This is an inducible vector since the inserted genes are expressed under a CMV2 promoter controlled by a Tet response element (TRE). As such the EpCAM can be inducibly expressed in the presence of doxycycline, which is a tetracycline variant. The pSLIK plasmids were packaged into lentiviral vectors and used to infect cells, which stably integrated into the cell genome under a G418 selection regiment.

We needed a reporter system to elucidate the changes in EpCAM activity. We chose a TOPFlash reporter system[111] that responds to TCF promoter elements, notably TCF4, most often used as a reporter for the Wnt pathway. We made this choice based on the fact that the EpCAM pathway targets genes with Lef1 recognition sequences, and Lef1 is part of the TCF family of transcription factors, also known as TCF10. Additionally, a previous study had shown a drop in TOPFlash reporter levels by knocking down EpCAM[51]. This reporter plasmid expresses firefly luciferase from a 7XTcf promoter, which includes seven repeats of the TCF recognition element. It also contains a puromycin resistance cassette expressed from a SV40. These 7XTFP reporter plasmids

were also packaged into lentiviral vectors to be used for infecting cells[iv]. This reporter system was used in many experiments during the course of the project.

Two cell lines with different level of endogenous Wnt activity were used to investigate the effects of EpCAM over-expression. The MCF7 cell line is a breast cancer line that has a low level of endogenous Wnt activity[v]. The HCT116 cells are on the other hand colon cancer cell lines with a high level of endogenous Wnt activity. In HCT116 cells, the serine 45 in one of the β-catenin alleles is mutated[114]. This mutation prevents phosphorylation by GSK-3β which prevents the proteosomal degradation and leads to accumulation of β-catenin in cell, which in turn provides a high Wnt activity. This selection provided us with two different baselines in our experiments to observe the effects of EpCAM over-expression on the activity of our reporter system.

4.2.1 Stable Infection of EpCAM Over-expressing Inducible Lentivirus

MCF7 and HCT116 cell lines with the integrated reporter were infected with the EpICD-YFP pSLIK lentivirus and infected cells were selected for with G418 to achieve a population of cells with a stable integration of the transgene. In these cell lines EpICD can be overexpressed using doxycycline. To illustrate the expression of EpICD-YFP in these cells, they were induced with doxycycline and observed and photographed under the fluorescence microscope's green filter, which can also show the fluorescence emitted by YFP. The results showed existence of fluorescent (YFP^+) cells in presence of doxycycline (Figure 6).

These doubly infected cell lines were induced with various concentrations of

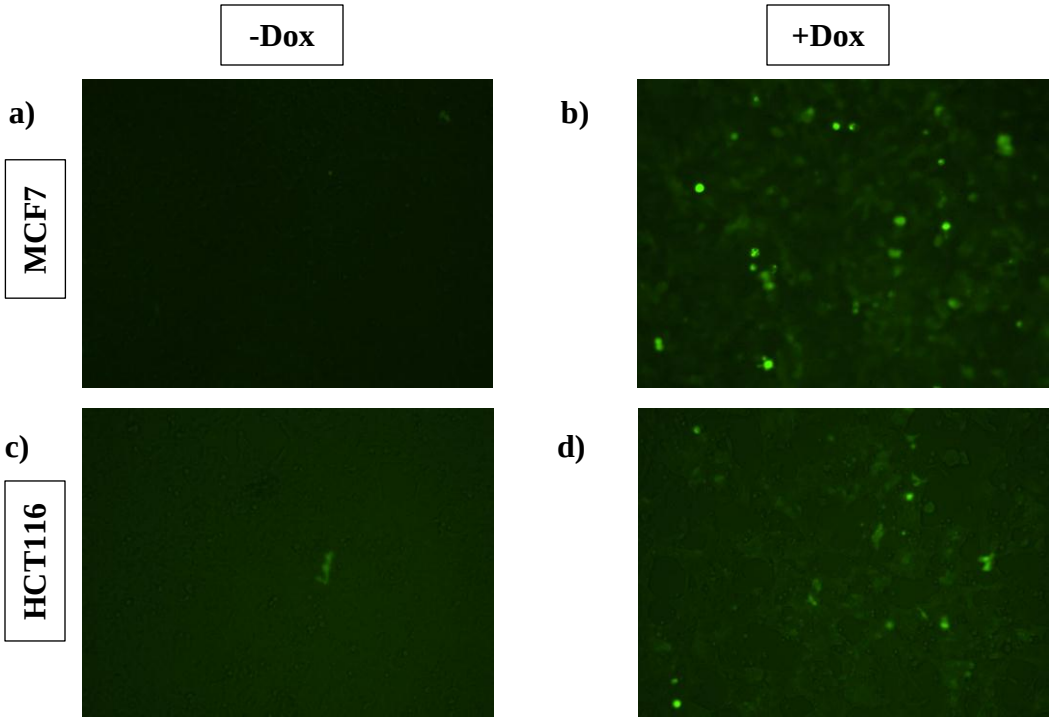


Figure 6. EpICD-YFP expression induction by doxycycline
Inducible EpICD-YFP lentivirus infected a) MCF7 cells without and b) with doxycycline, c) HCT116 cells without and d) with doxycycline.

doxycycline and subjected to a luciferase assay to measure the activity of the integrated reporter. AlamarBlue proliferation assay[115] was employed as an internal control to normalize the luciferase activity. AlamarBlue is commercial name for the blue dye Resazurin. Itself non-fluorescent, Resazurin is reduced inside cells to highly red fluorescent Resorufin. This assay if often used as a measure of cell viability. The results showed an increase in the reporter activity, becoming more pronounced at higher concentrations of doxycycline, especially for HCT116 cells (Figure 7a). The effect was more significant for MCF7 after normalization with AlamarBlue. The 1 μg/ml concentration of doxycycline was chosen to conduct future experiments, as it was the highest concentration that showed no toxic effect. It is of note that the spike in reporter activity was more substantial in the HCT116 cells relative to the MCF7 cells.

4.2.2 Transient Transfection of EpCAM Over-expressing Vectors

The MCF7 and HCT116 cell lines were stably integrated with the 7XTFP reporter system. A population of infected clones resistant to puromycin were isolated. These cells were transfected with 0.02, 0.1, 0.5 and 1 μg DNA of overexpression plasmids for EpCAM-YFP, EpICD-YFP and β-catenin. We hoped to get a more substantial increase in the reporter activity by flooding the cells with numerous copies of the over-expression plasmids relative to a controlled infection which only provides a single copy integrated into the genome. All transfection mixes were topped off to 1 μg of total DNA using salmon sperm DNA. Compound chir99021 was used as a positive control. This compound is a GSK inhibitor which induces accumulation of β-catenin and high Wnt

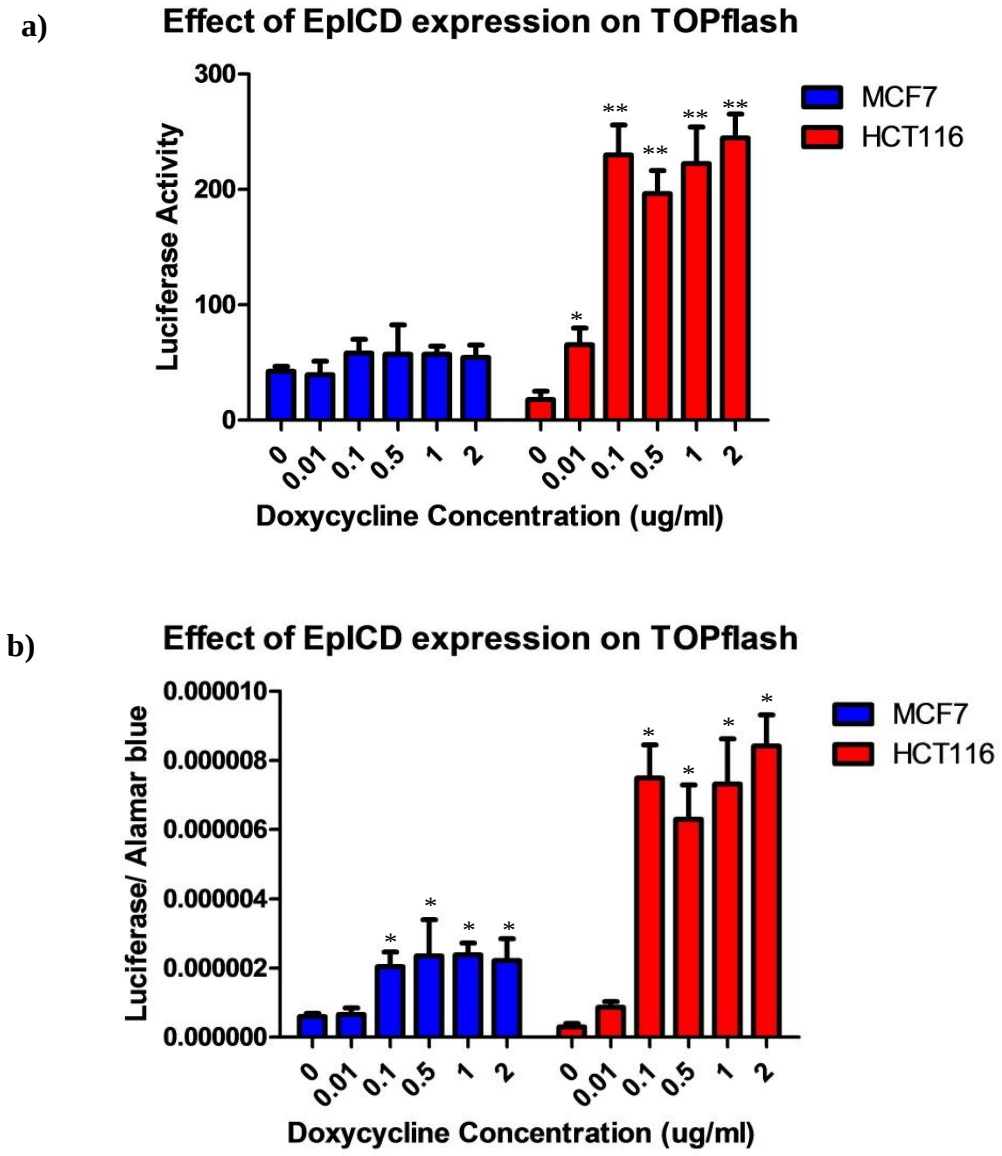


Figure 7: Effect of EpCAM over-expression on TCF reporter activity
MCF7 and HCT116 cells were stably infected with a TOPFlash TCF reporter lentivirus and the EpICD over-expression lentivirus and subjected to variable concentrations of doxycycline. a) Luciferase assay results, b) Luciferase assay results normalized by AlamarBlue reduction assay readings. (One-way, post-Tukey $P<0.0001$)

activity.

The reporter activity values of the transfected samples were normalized against the non-transfected cells. This means a residual activity of 1 for the control non-transfected cells. The AlamarBlue assay numbers were used as an internal control to normalize the luciferase assay readings. Unfortunately, no significant change was observed as a result of the transfections of any of the three over-expression plasmids into either cell line. The positive control using the chir99021 showed the expected increase in TOPFlash activity (Figure 8). Hence, the transfection seems to have been unsuccessful.

4.3 Dominant Negative EpCAM

To inhibit EpCAM activity, we designed a dominant negative allele of this protein. This allele was produced by fusing the repression domain (amino acids 2-298) of the Drosophila Engrailed protein to the C-terminus of the EpICD (the last 26 amino acids at EpCAM C-terminus). This fusion protein can replace the endogenous EpICD in its siganolsome complex with Lef1 and inhibit transcription due to the Engrailed repression domain (Figure 9). This dominant negative EpCAM (ΔNEpCAM) fusion protein was cloned into the pSLIK inducible vector and packaged into lenviral particles. Populations of MCF7 and HCT116 cell lines that were previously stably infected with the 7XTFP reporter were infected with the ΔNEpCAM virus and a population of resistant clones were selected for using G418.

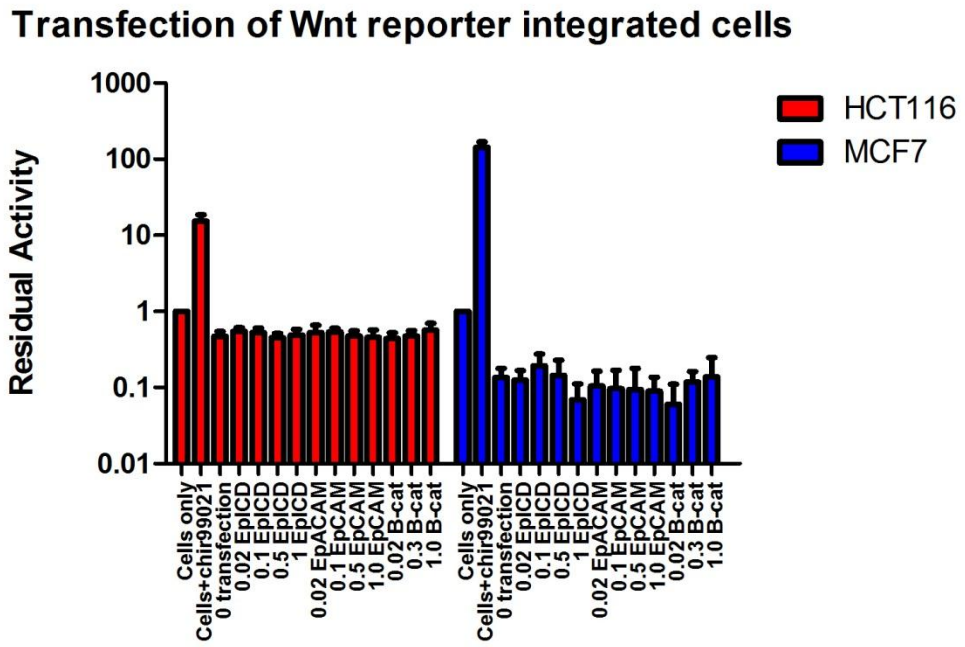


Figure 8: Effect of EpCAM over-expression on TCF reporter activity

MCF7 and HCT116 cells stably infected with a TOPFlash TCF reporter lentivirus are transfected with varying amounts of one of the three over-expression plasmid: EpICD, EpCAM or β-Catenin. A Wnt activator compound (chir99021) is included as positive control.

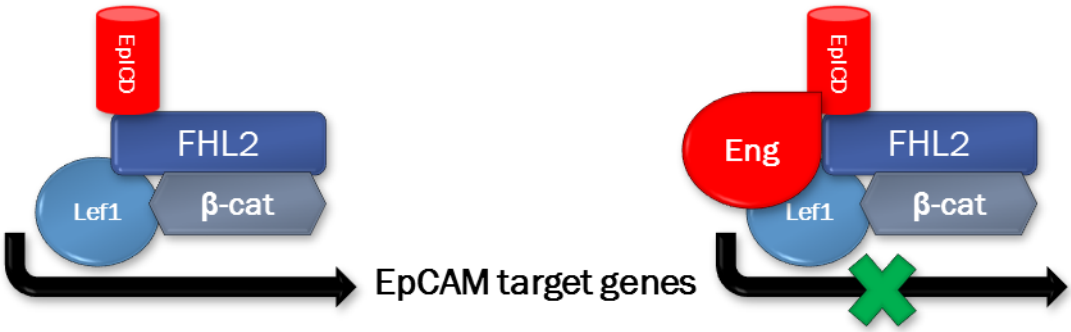


Figure 9: Functional mechanism of the dominant negative EpCAM
The engrailed repression domain fused to EpICD hinders the activity of its signalosome and prevents the transcription of the EpCAM target genes.

4.3.1 Dominant Negative EpCAM Expression

The expression of ΔNEpCAM was tested at RNA level using Quantitative RT-PCR. MCF7 and HCT116 cells that had been infected with the ΔNEpCAM virus were grown in the absence and presence of doxycycline for 48 hours. Uninfected parental cell lines were also included as controls. qRT-PCR of these samples, using primers designed to detect Drosophila Engrailed cDNA, showed about a 100 fold above background expression of the EpICD-engrailed fusion at RNA level, by doxycycline induction. No significant difference was observed between the uninfected parental cell line and the uninduced infected cells. This observation rejects the possibility of our inducible system being leaky (Figure 10).

To test the expression of ΔNEpCAM at the protein level we isolated protein lysates from the same cell samples that were used for the qRT-PCR. The western blot, using the anti-EpCAM antibody, only revealed a band representing the endogenous EpCAM (Figure 11a). No extra bands were seen in the samples to demonstrate the presence of EpICD-engrailed fusion protein, which is very similar in size to the EpCAM protein. It is of note that the epitope of this antibody is within the EpICD domain. A second western blot was performed using and anti-engrailed antibody. The same protein samples were used. In addition, a cell lysate from a HEK293FT cell line was used as negative control (Figure 11b- Lane 7), whilst HEK293FT transfected with the plasmid expressing the engrailed protein was used as positive control (Figure 11b- Lane 8). Once again, none of the lysates from the induced cells showed a distinct band representing the

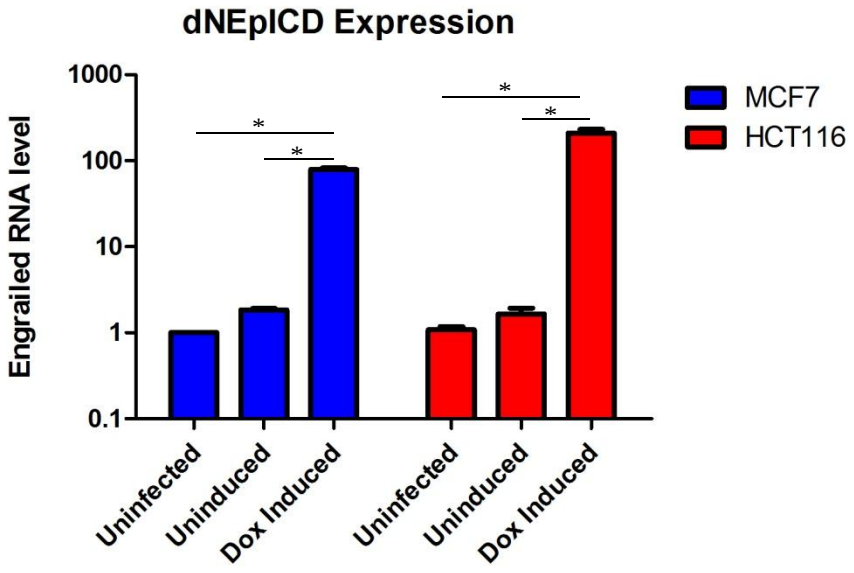


Figure 10: Expression of dominant negative EpCAM at RNA level
Quantitative RT-PCR of RNA extracts from MCF7 and HCT116 cells infected with the inducible lentivirus expressing EpICD-engrailed. Relative RNA levels were normalized against the house-keeping gene GAPDH. (One-way, post-Tukey $P<0.0001$)

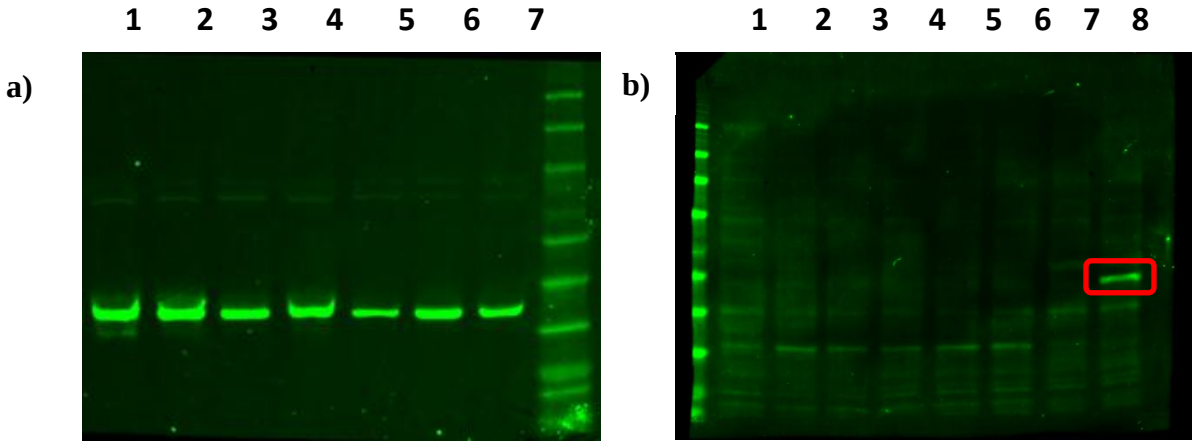


Figure 11: Dominant Negative EpCAM Expression at protein level
Western blot analysis with a) Anti-EpCAM antibody on 1- HCT116-ΔNEpCAM+Dox, 2- HCT116-ΔNEpCAM-Dox, 3- HCT116-Dox, 4- MCF7-ΔNEpCAM+Dox, 5- MCF7-ΔNEpCAM-Dox, 6- MCF7+Dox, 7- MCF7-Dox cell lysate, and b) Anti-Engrailed antibody on 1- MCF7-Dox, 2- MCF7-ΔNEpCAM-Dox, 3- MCF7-ΔNEpCAM+Dox, 4- HCT116-Dox, 5- HCT116-ΔNEpCAM-Dox, 6- HCT116-ΔNEpCAM+Dox, 7-FT293, 8- FT293+ Engrailed (TT) cell lysate.

ΔNEpCAM protein. However, the included positive control provided a band for the engrailed protein (Figure 11b- Lane 8). As such, western blots with neither antibody succeeded in confirming the expression of the ΔNEpCAM protein. These results indicate a problem with the translation of the ΔNEpCAM protein.

4.3.2 Dominant Negative EpCAM Activity

Difficulties in demonstrating the expression of the dominant negative EpCAM persuaded us to pursue another course of action to this time test for the activity of ΔNEpCAM construct. We tested our MCF7 and HCT116 cell lines with the integrated reporter and the ΔNEpCAM construct for any activity of the dominant negative EpCAM. We used doxycycline to induce production of the ΔNEpCAM. We also included BIO, a GSK inhibitor, to increase the endogenous reporter activity. We hoped to pronounce any inhibitory effect the dominant negative might impart by raising the baseline. However, the results showed no difference in the luciferase activity as a result of the dominant negative, whether in presence or absence of BIO (Figure 12).

We repeated the previous experiment, performing our assay at various time points instead of the standard 48 hours used previously. We expected to see a more dramatic effect at a different time point. The MCF7 cells showed no effect in absence of BIO. In the presence of BIO some decrease in activity was seen at later time points. For HCT116 cells earlier time points showed a slight difference in absence of BIO, as well as in later time points in the presence of BIO (Figure 13). However, none of the observed effects were significant.

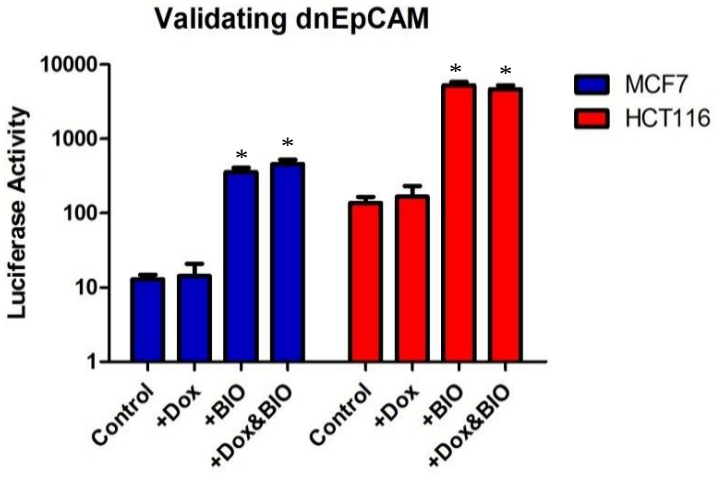


Figure 12: Validating the activity of the dominant negative EpCAM using a TOPFlash reporter system. MCF7 and HCT116 cells are co-infected with a TOPFlash TCF reporter lentivirus and the inducible ΔNEpCAM lentivirus. Cells were grown and tested in the presence and absence of inducer doxycycline and BIO, which inhibits β-Catenin degradation. (One-way, post-Tukey $P<0.0001$)

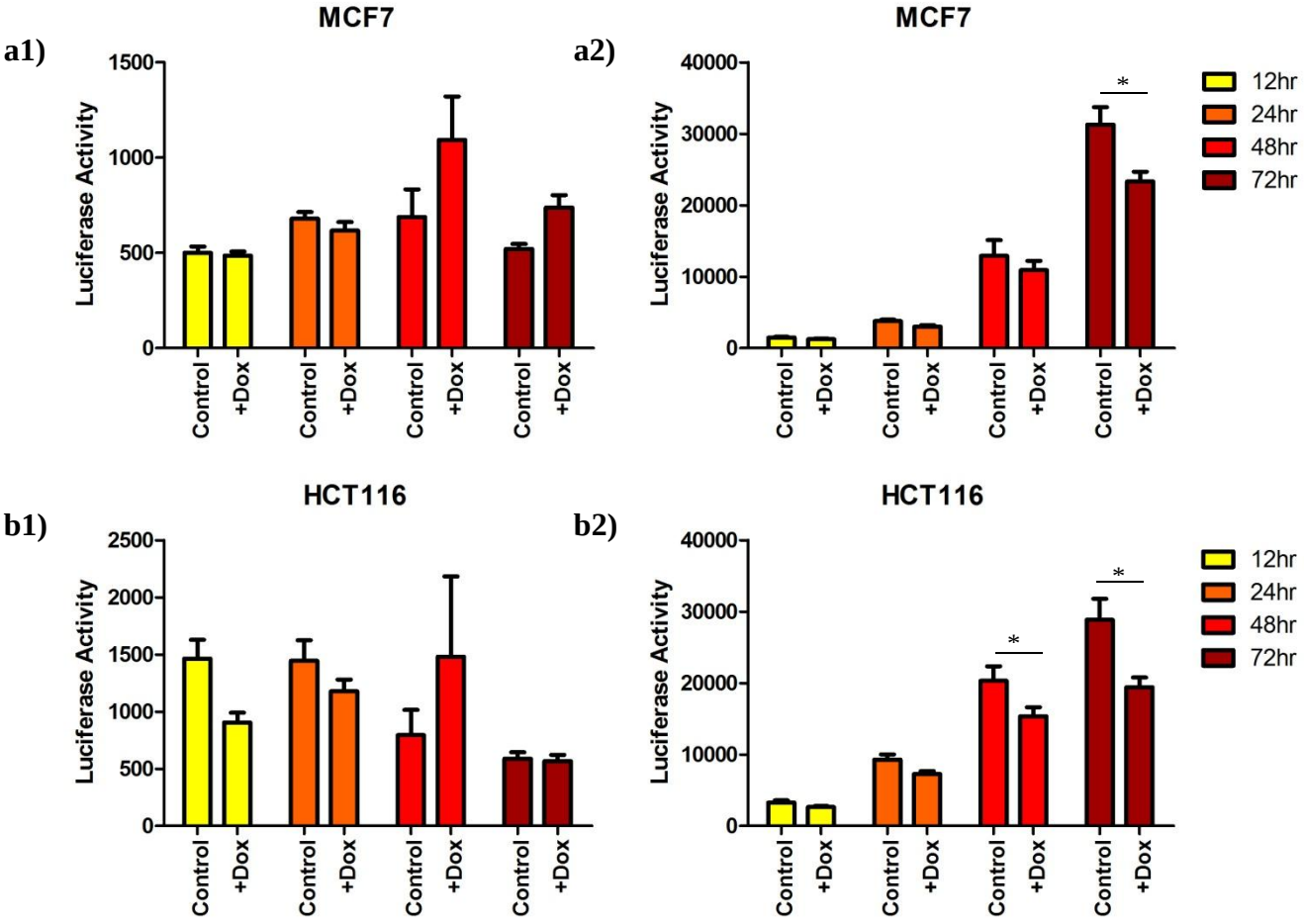


Figure 13: Monitoring the activity of the dominant negative EpCAM at multiple time points. Luciferase activity in a) MCF7 cells (1-control, 2-with 4uM BIO) and b) HCT116 cells (1-control, 2-with 4uM BIO) co-infected with a TOPFlash TCF reporter lentivirus and the inducible ΔNEpCAM lentivirus. Cells were grown and tested in presence and absence of inducer doxycycline for 12, 24, 48 and 72 hours. (Unpaired t-test, $P<0.0001$)

To amplify the effect of our dominant negative we chose to transiently transfect the plasmid containing the construct into highly transfectable HEK293FT cells at various concentrations. This would provide many fold higher copies of this gene relative to the possibly single copy in the stably infected cells. In parallel we also ran a dominant-negative TCF construct[112]. In addition to our 7XTFP reporter, we transfected a plasmid with a cmv-renilla luciferase as a reporter for the general transcription activity to normalize the firefly luciferase activity. The experiment was done in three parts: first for the endogenous reporter activity, second with BIO compound, an inducer of the WNT pathway, and third with a β-Catenin over-expression plasmid to increase the endogenous levels. At the endogenous level we observed a decrease in reporter activity only at the highest amount of the ΔNEpCAM, whereas the dnTCF was effective at lower concentrations (Figure 14a). The inclusion of BIO compound increased the baseline of the reporter activity in the non-transfected cells (Figure 14C). This seemed to accentuate the effects of both dominant negative constructs in reducing the reporter activity level. However, the same results were not achieved by the β-Catenin over-expression vector (Figure 14b).

For a more functional examination of our ΔNEpCAM construct, we performed colony forming assays on the cells with the integrated construct. The results showed to significant effect on colony formation in either MCF7 or HCT116 cells once induced by doxycycline (Figure 15).

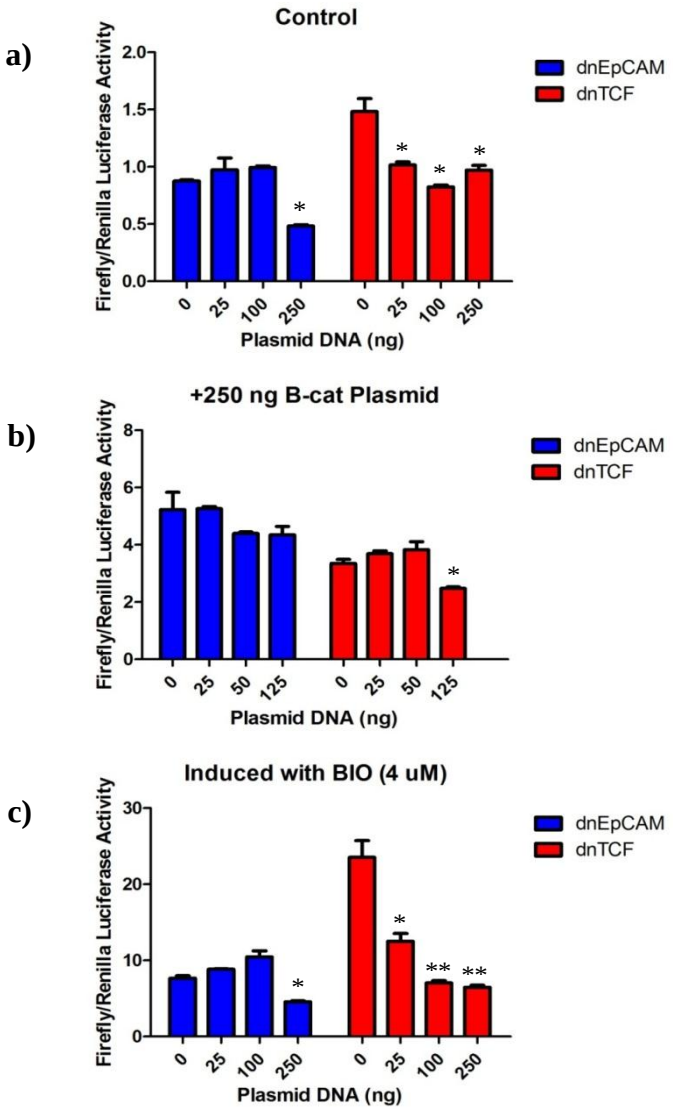


Figure 14: Validating the activity of the dominant negative EpCAM using a TOPFlash reporter system. HEK293FT cells are co-transfected with a control endogenous reporter (Renilla luciferase) plasmid, a TCF reporter (Firefly luciferase) plasmid and varying amounts of either the ΔNEpCAM or dnTCF plasmids. As inducers a) nothing (control), b) a β-Catenin expressing plasmid, c) BIO compound were also added. (One-way, post-Tukey $P<0.0001$)

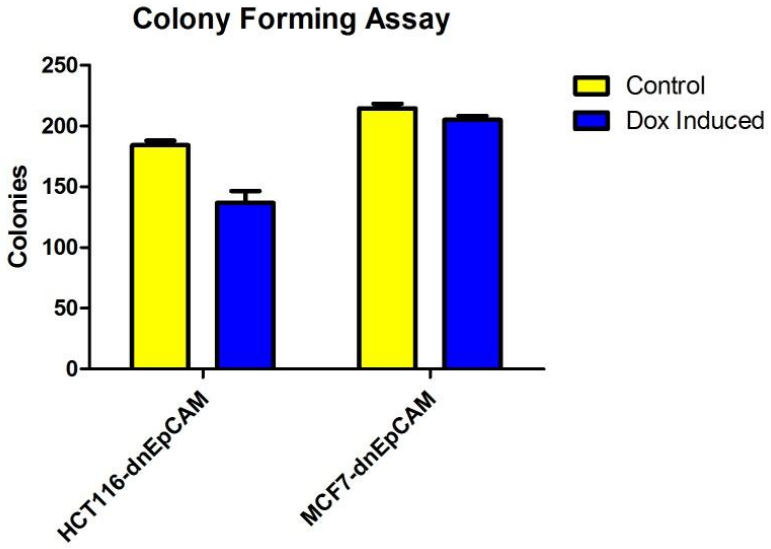


Figure 15: Effect of the dominant negative EpCAM expression on colony formation. MCF7 and HCT116 cells infected with the inducible ΔNEpCAM lentivirus were subjected to a colony forming assay in presence and absence of doxycycline.

4.4 EpCAM Short Hairpin RNA

Two validated short-hairpin RNA sequences against human EpCAM were chosen from the online Open Biosystems library. These sequences, henceforth called EpSH12 and EpSH25, were validated in inducible systems[109]. We cloned these sequences into our inducible vector pSLIK and packaged them into lentiviruses. These viruses were used to stably infect MCF7 cell line.

To validate the functionality of our two shRNAs, we tested the MCF7 cell line infected with the lentiviruses encoding the two shRNA constructs. qRT-PCR analysis, using primers targeting EpCAM, showed that when induced by doxycycline the relative RNA level of EpCAM decreases by about two fold with either shRNA (Figure 16). Western blot by anti-EpCAM antibodies corroborated this effect at protein level, at similar knock-down levels (Figure 17). The uninfected parental MCF7 cell line showed the same EpCAM expression level as the uninduced infected cells (Figure 17- Lanes 1 &2).

4.4.1 Effects of EpCAM Knock-Down on Transcription Reporter Expression

After verifying the activity of our two shRNA constructs in knocking down EpCAM, we next investigated the consequences of this knock-down on various cellular processes. We introduced our shRNA constructs into a population of MCF7 cells lines that were previously stably infected with the 7XTFP reporter. We achieved a population infected with both the reporter and the shRNA constructs through simultaneous selection by puromycin and G418. Then we performed luciferase assays on these cells in presence

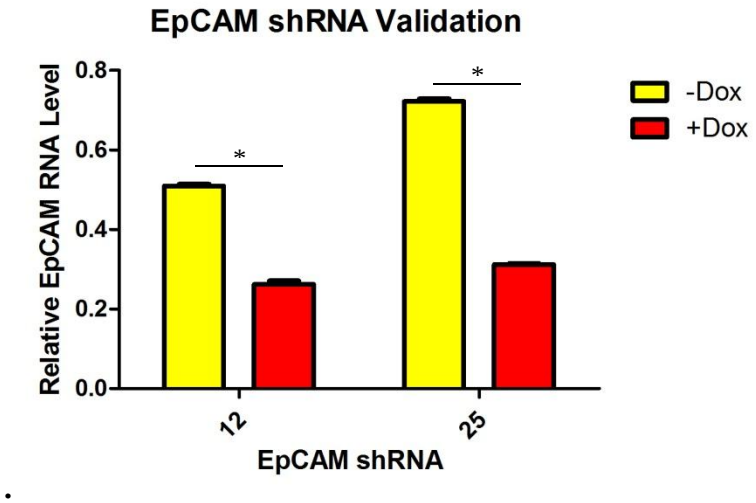


Figure 16: Knock-down of EpCAM expression using short hairpin RNA constructs at RNA level. Quantitative RT-PCR of RNA extracts from MCF7 cells infected with the inducible lentivirus expressing either the EpSH-12 or the EpSH-25 constructs. (Unpaired t-test, $P<0.0001$)

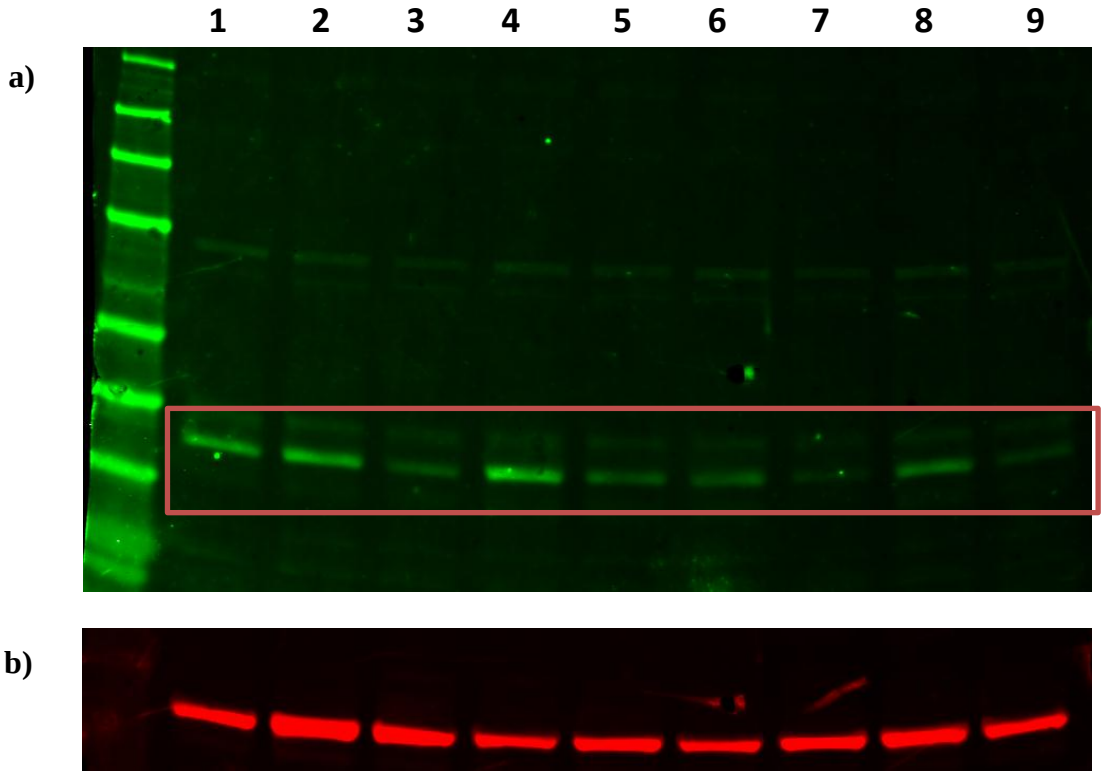


Figure 17 : Knock-down of EpCAM expression using short hairpin RNA constructs at protein level. Western blot analysis with a) an anti EpCAM antibody (green) and b) anti α-Tubulin (red) on 1- MCF7, 2&6- MCF7-EpSH12, 3&7- MCF7-EpSH12+Dox, 4&8- MCF7-EpSH25, 5&9- MCF7-Ep/SH25+Dox cell lysate.

and absence of doxycycline. We also included the parental cell line with only the integrated reporter, lacking the shRNA, as a control. The results paralleled those of our previous experiment. An almost 50% decrease was observed in firefly luciferase activity when shRNA expression was induced by doxycycline, for both our shRNA constructs (Figure 18). The slight increase observed for the control cells that are induced is due to doxycycline toxicity, which decreases the AlamarBlue reduction readings. The experiment was repeated in MCF7 cell that also contained the renilla luciferase endogenous reporter. This reporter can provide a measure of the general transcription activity, not specific to the 7XTFP reporter target genes. This was done to ensure the effect on the firefly luciferase readings was exclusive to EpCAM target genes. The results showed no significant effect on the activity of the renilla luciferase (Figure 19b). As such, normalizing the firefly luciferase readings with the renilla luciferase readings did not affect our previous observations (Figure 19).

4.4.2 Effects of EpCAM Knock-Down on Growth Curve

The MCF7 cell lines containing the integrated shRNA constructs and the parental cell line were subjected to a 7 day growth curve analysis. The results showed a considerable effects on the shRNA infected cell line in presence of doxycycline, whereas the parental cell line suffered no significant disadvantage (Figure 20). The experiment was repeated, this time allowing for the assay to continue for 9 days and also performing a PrestoBlue proliferation assay in parallel. This growth curve again showed no effect for

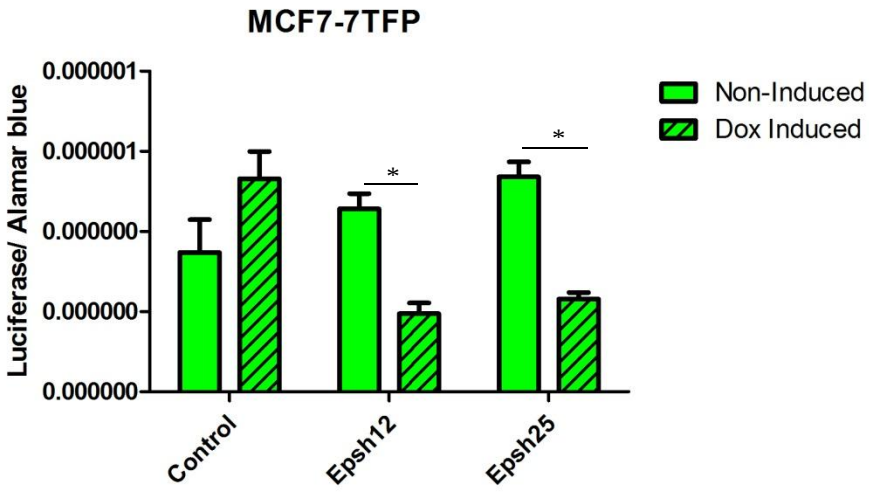


Figure 18: Effect of EpCAM knock-down on TOPFlash TCF reporter activity
MCF7 cells were co-infected with the TOPFlash TCF reporter lentivirus and the inducible lentivirus expressing either the EpSH-12 or the EpSH-25 constructs and selected for to achieve stable cell lines. Infected and uninfected control cells are grown in presence and absence of doxycycline and tested for luciferase activity. (Unpaired t-test, $P<0.0001$)

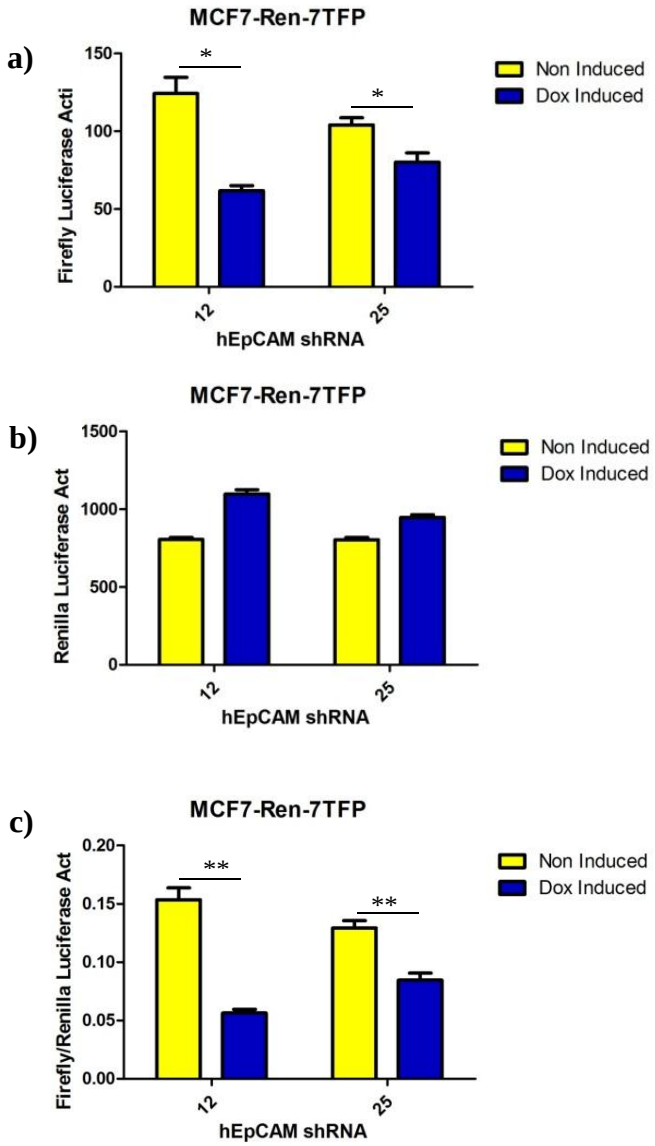


Figure 19: Effect of EpCAM knock-down on TOPFlash TCF reporter activity
MCF7 cells are co-infected with the general transcription reporter (renilla luciferase) lentivirus, the TCF reporter (firefly luciferase) lentivirus and the inducible lentivirus expressing either the EpSH-12 or the EpSH-25 constructs. a) Firefly luciferase activity, b) renilla luciferase activity, c) normalized firefly/renilla luciferase activity. (Unpaired t-test, *P<0.05, **P<0.001)

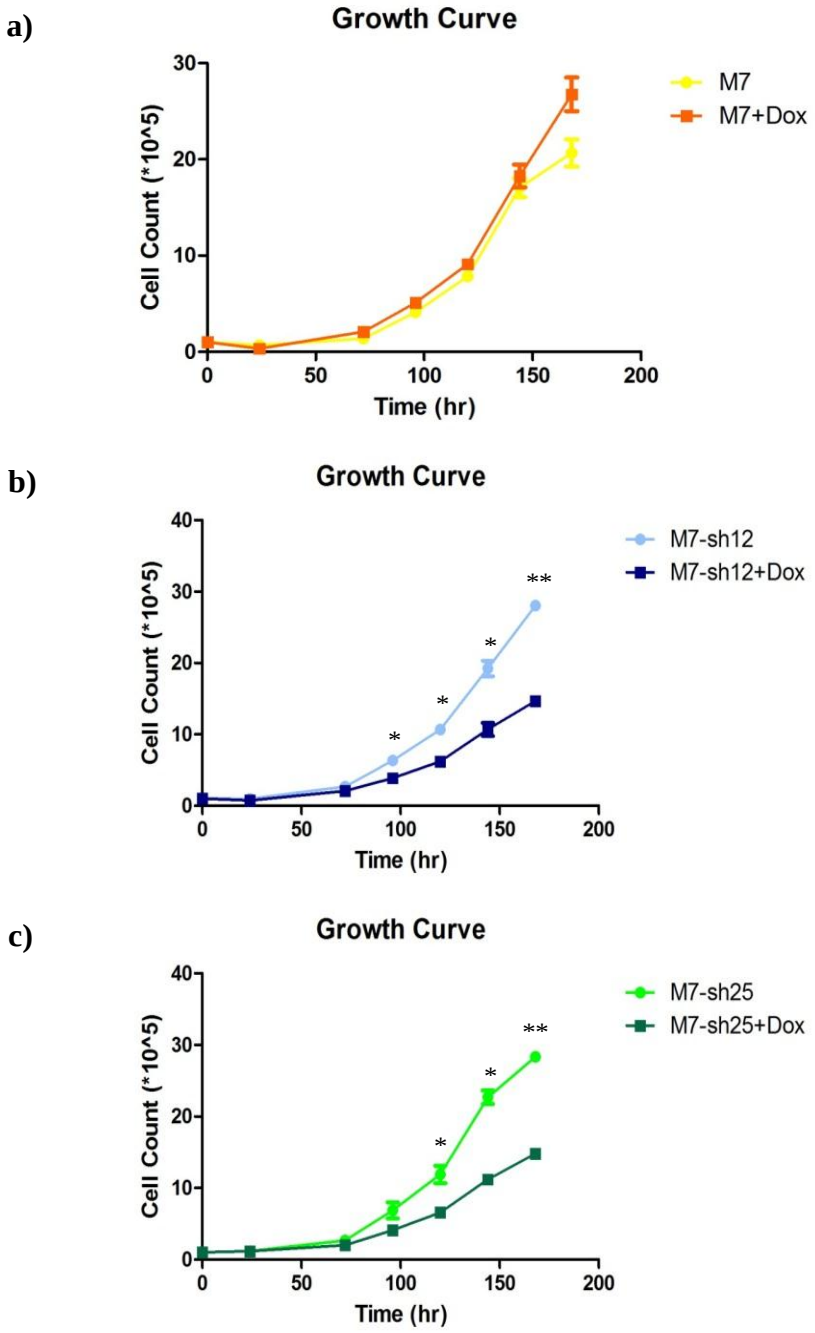


Figure 20: Effect of EpCAM knock-down on the growth curve of MCF7 cells
a) Uninfected MCF7 cells as control, b) MCF7 cells infected with EpCAM short hairpin 12 lentivirus, c) MCF7 cells infected with EpCAM short hairpin 25 lentivirus. (Unpaired t-test, *P<0.05, **P<0.0001)

doxycycline on the parental cell line. The cell lines with the shRNA construct behaved similarly to the previous experiment, but by increasing the duration of the assay the doxycycline induced cells eventually caught up to the control cells. Nonetheless, the induction of the EpSH12 and EpSH25 caused a 1.5 and 1.8 fold increase in the doubling time of the MCF7 cells respectively (Figure 21). The parallel PrestoBlue proliferation assay was corroborative of the growth curve results. The rise in the PrestoBlue readings stopped past day 6 (Figure 22).

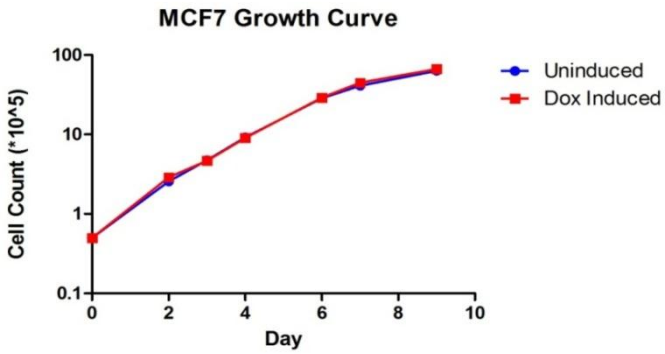
4.4.3 Effects of EpCAM Knock-Down on Colony Formation

The MCF7 cell lines containing the integrated EpCAM shRNA constructs and the parental cell line were subjected to a colony forming assay. Colony forming efficiency (CFE) is expressed as a percentage, which describes the number of colonies per number of plated cells. CFE of the MCF7 cells showed about a 3 fold decrease by inducing the expression of either EpCAM shRNA construct (Figure 23). Such effect was not observed in the parental cell line in presence of doxycycline. The size of the colonies was also affected as most of the colonies formed when the shRNAs were induced were smaller in size. Many of such smaller colonies were less compact in morphology and contained larger cells (Figure 24 a2, b2).

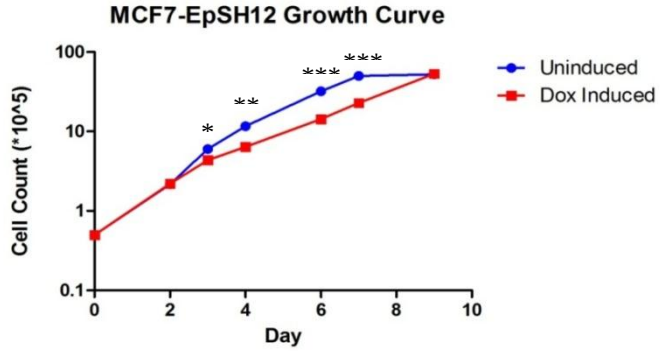
4.4.4 Effects of EpCAM Knock-Down on Sphere Formation

Tumorspheres are characterized as a clonal growth of cultured cancer cells in suspension. As they tend to mirror their corresponding tumor in cellular characteristics such as morphology or gene expression, they are often used as *in vitro* surrogates[116]. We

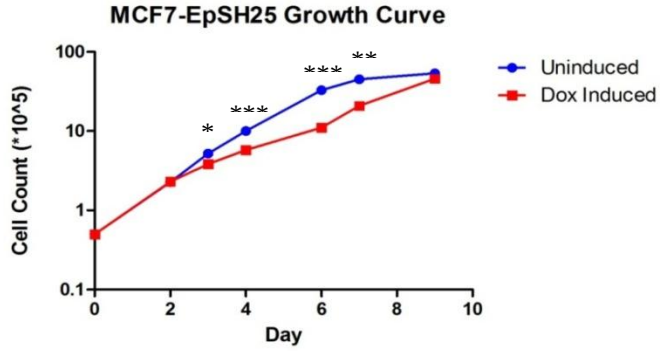
a)



b)



c)



d)

Doubling Time (days)	- Dox	+ Dox
MCF7	1.14	1.18
MCF7-EpSH12	1.06	1.52
MCF7-EpSH25	1.05	1.79

Figure 21: Effect of EpCAM knock-down on the growth curve of MCF7 cells
a) Uninfected MCF7 cells as control, b) MCF7 cells infected with EpCAM short hairpin 12 lentivirus, c) MCF7 cells infected with EpCAM short hairpin 25 lentivirus, d) Table of doubling times (days). (Unpaired t-test, *P<0.05, **P<0.001, ***P<0.0001)

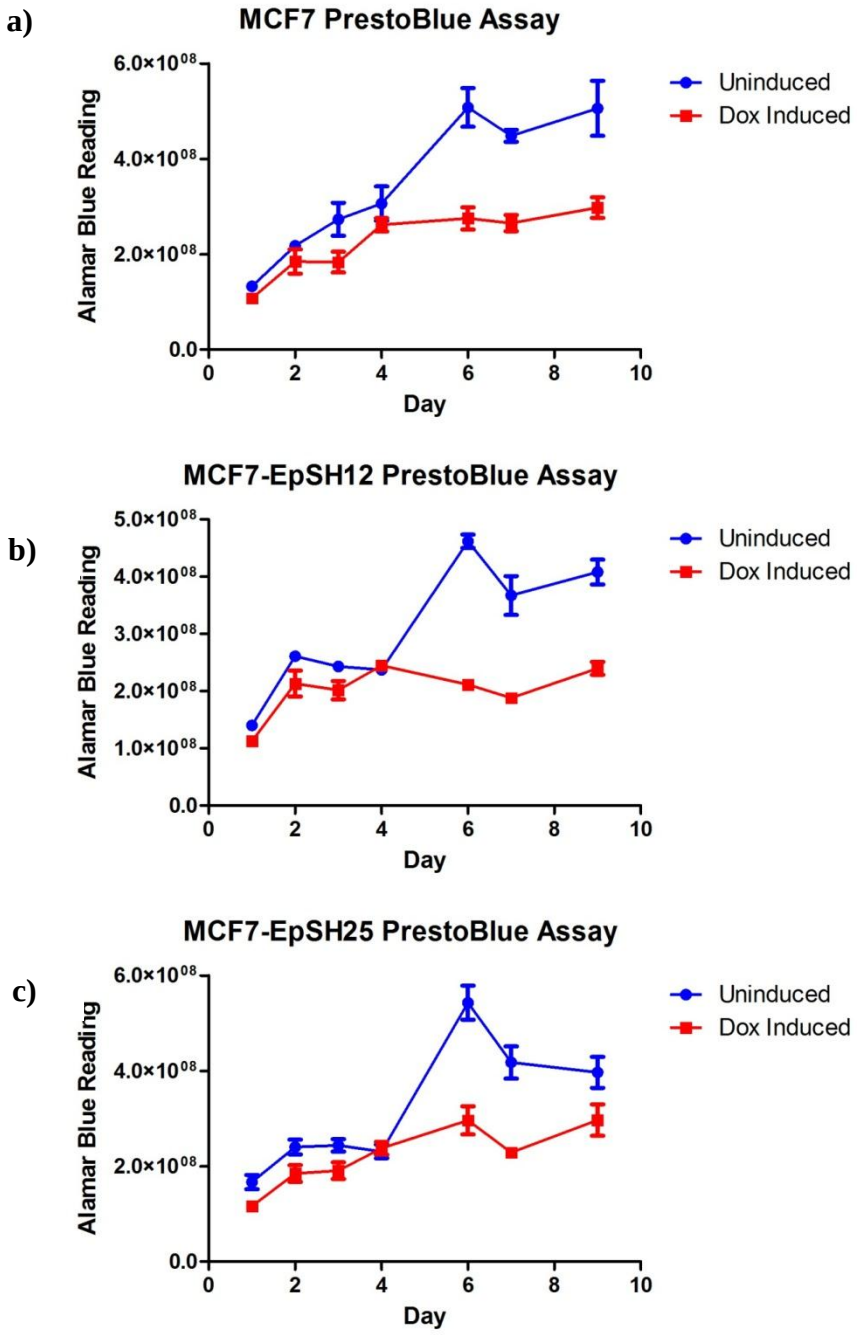


Figure 22: Effect of EpCAM knock-down on PrestoBlue reduction by MCF7 cells
a) Uninfected MCF7 cells as control, b) MCF7 cells infected with EpCAM short hairpin 12 lentivirus, c) MCF7 cells infected with EpCAM short hairpin 25 lentivirus.

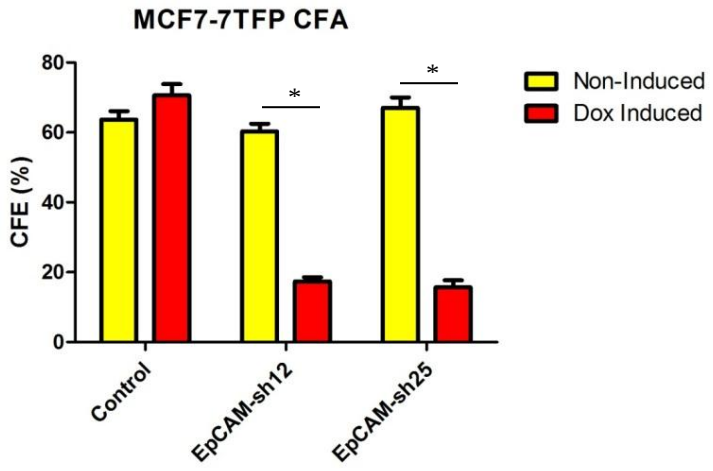


Figure 23: Effect of EpCAM knock-down on colony formation

Colony forming assay of uninfected MCF7 cells as control, MCF7 cells infected with EpCAM short hairpin 12 lentivirus, MCF7 cells infected with EpCAM short hairpin 25 lentivirus in presence and absence of doxycycline. (Unpaired t-test, $P<0.001$)

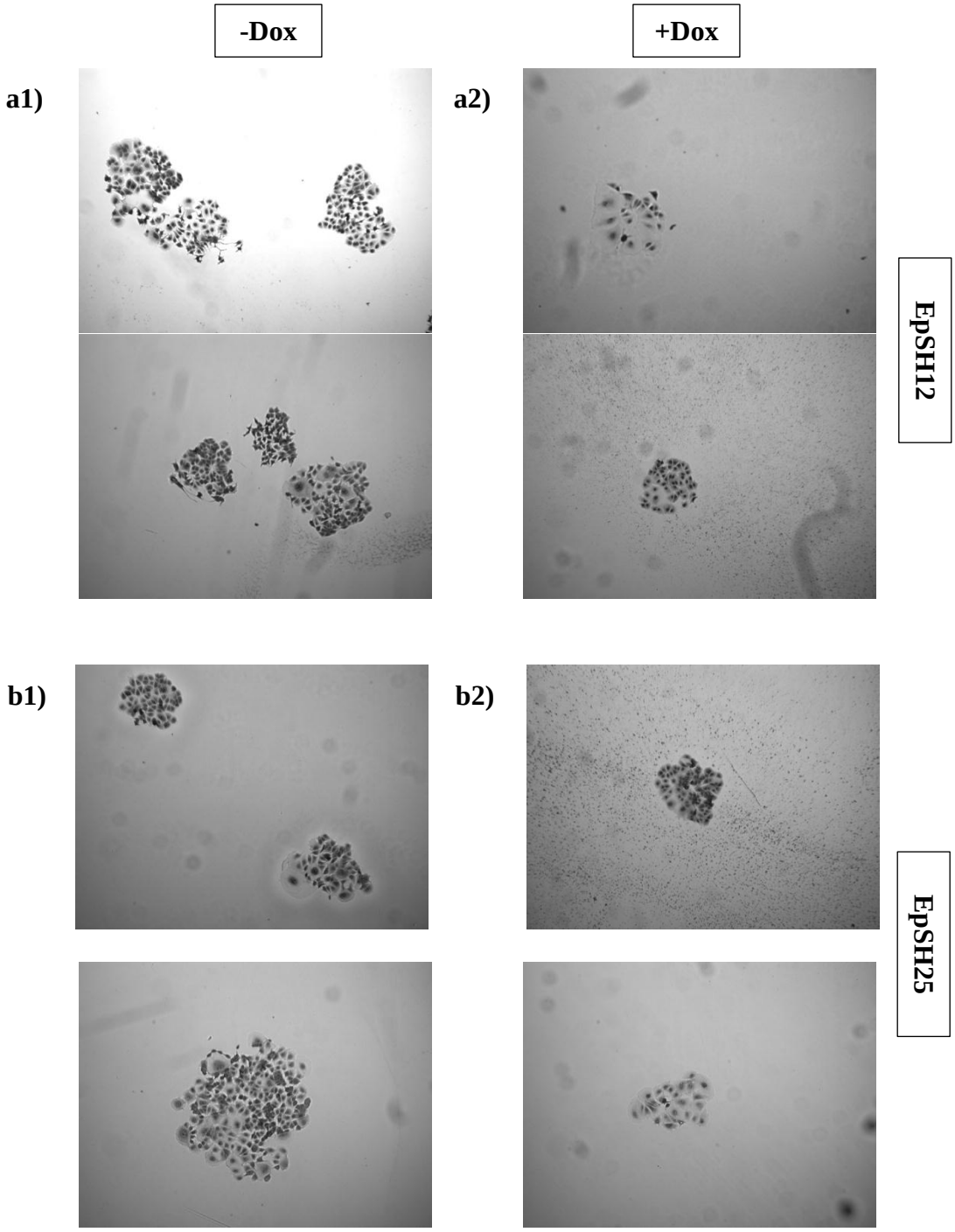


Figure 24: Effect of EpCAM knock-down on colony morphology

Colony forming assay of a) MCF7 cells infected with EpCAM short hairpin 12 lentivirus (1-control, 2-doxycycline induced), b) MCF7 cells infected with EpCAM short hairpin 25 lentivirus (1-control, 2-doxycycline induced).

opted to use this assay since tumorspheres have been shown to be enriched in cancer stem cells. Therefore, sphere forming assay was performed using the MCF7 cells infected with EpCAM shRNA constructs. It is of note that the cells were first established as spheres for 3 passages before being used for this assay. In regular tissue culture flasks, when induced by doxycycline, the cells adhere to the plate to form a monolayer (Figure 25). To investigate whether the cells formed spheres first before attaching to the plate or attach as single cells, we performed a time-lapse experiment. The results showed that the cells do not form sphere earlier on, but begin attaching to the plate as single cells (Figure 26).

Sphere forming assays done in low attachment plates showed similar number of spheres in presence and absence of doxycycline (Figure 27a). Despite the comparable number of spheres, the morphology of the spheres was very different. Whereas the spheres formed in absence of doxycycline resembled normal MCF7 spheres, when induced by doxycycline the spheres resembled abnormal hollow-like structures (Figure 28). Due to the subjective nature of sphere counting, we also performed AlamarBlue reduction assay on the plates from the sphere forming assays to corroborate the sphere counts. The AlamarBlue readings were lower for the induced spheres (Figure 27b).

4.4.5 Effect of EpCAM Knock-down on MCF7 Spheres

We further investigated of the hollow like structures formed when EpCAM is knocked down in MCF7 spheres in low attachment conditions. We prepared blocks of these spheres, along with the uninduced spheres, embedded in paraffin. We then sectioned these blocks and applied to various staining procedures. The

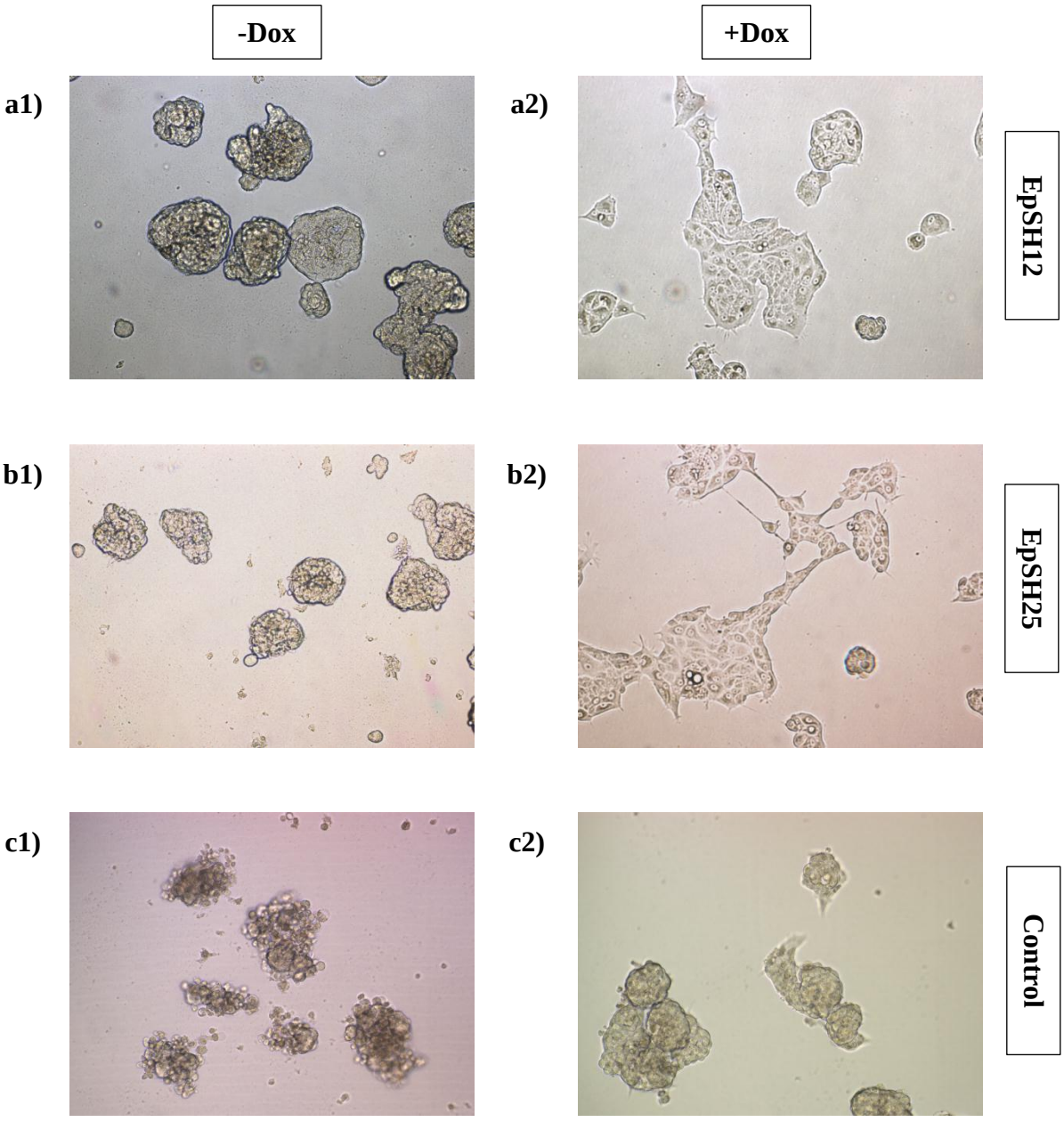


Figure 25: Effect of EpCAM knock-down on sphere formation

Sphere forming assay in regular tissue culture flasks of a) MCF7 cells infected with EpCAM short hairpin 12 lentivirus (1-control, 2-doxycycline induced), b) MCF7 cells infected with EpCAM short hairpin 25 lentivirus (1-control, 2-doxycycline induced), c) uninfected MCF7 cells (1-control, 2-doxycycline induced).

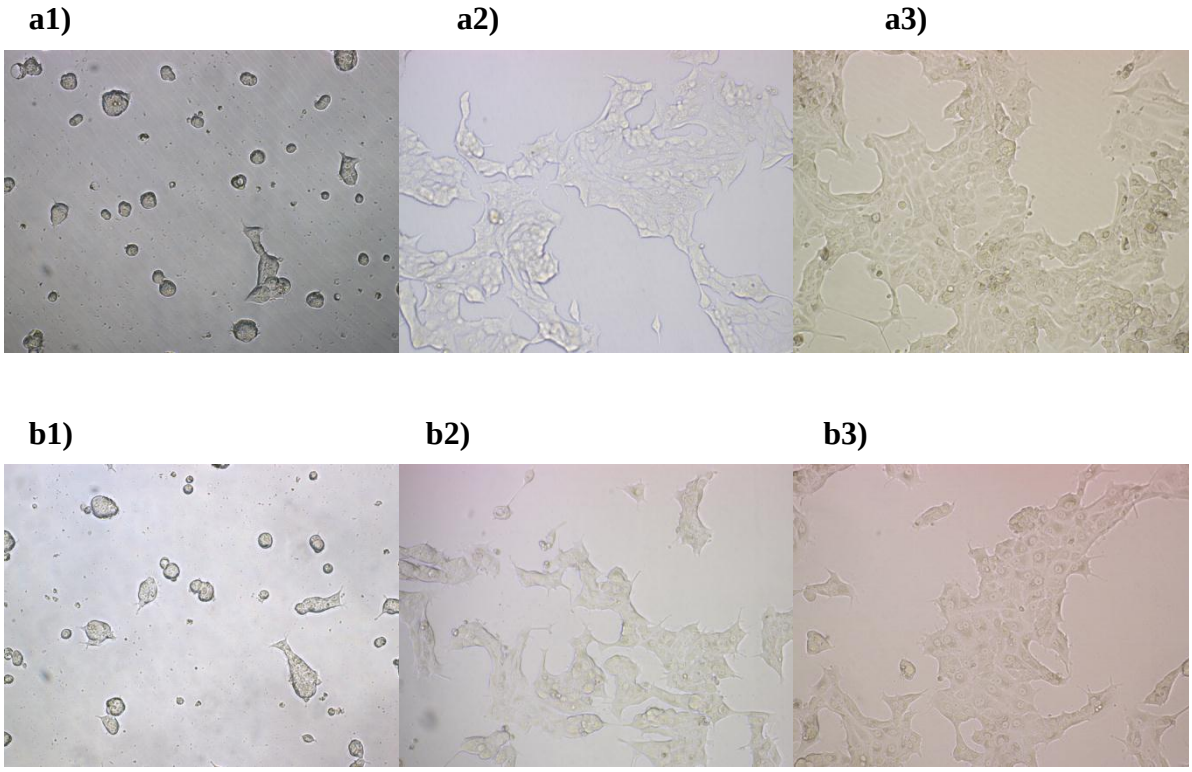


Figure 26: Time-lapse showing the effect of EpCAM knock-down on sphere formation. Sphere forming assay in presence of doxycycline in regular tissue culture flasks of a) MCF7 cells infected with EpCAM short hairpin 12 lentivirus (1-day 2, 2-day 4, 3-day 7), b) MCF7 cells infected with EpCAM short hairpin 25 lentivirus (1-day 2, 2-day 4, 3-day 7).

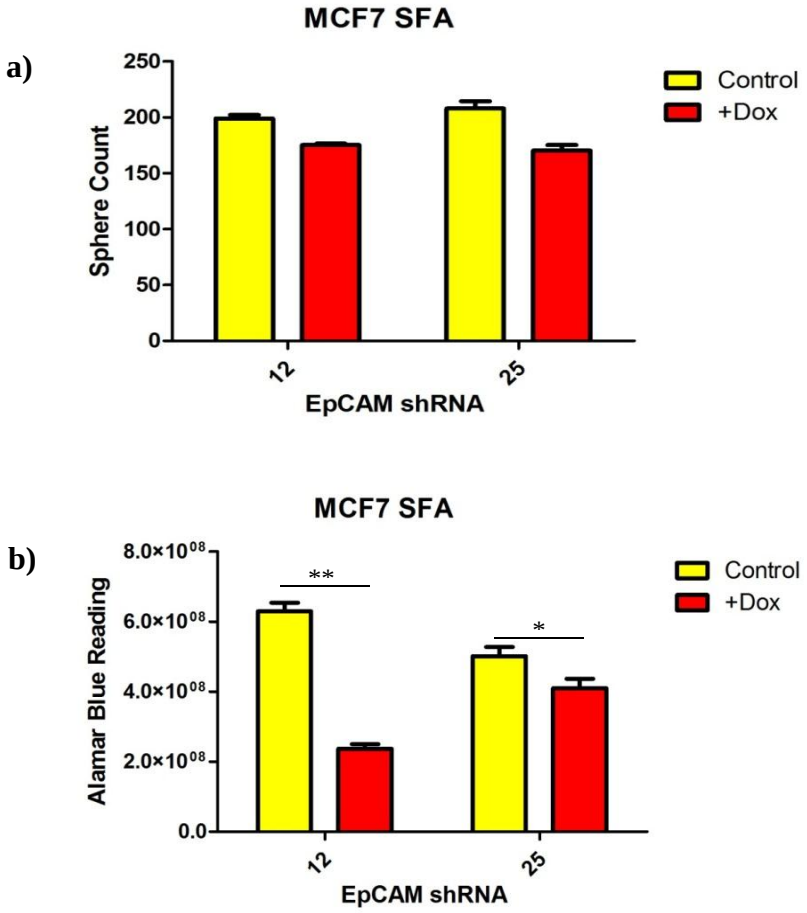


Figure 27: Effect of EpCAM knock-down on sphere formation

Sphere forming assay (in low attachment tissue culture flasks) of MCF7 cells infected with EpCAM short hairpin 12 lentivirus, MCF7 cells infected with EpCAM short hairpin 25 lentivirus in presence and absence of doxycycline. Graph a) number of sphere, b) AlamarBlue assay of the spheres. (Unpaired t-test, *P<0.05, **P<0.0001)

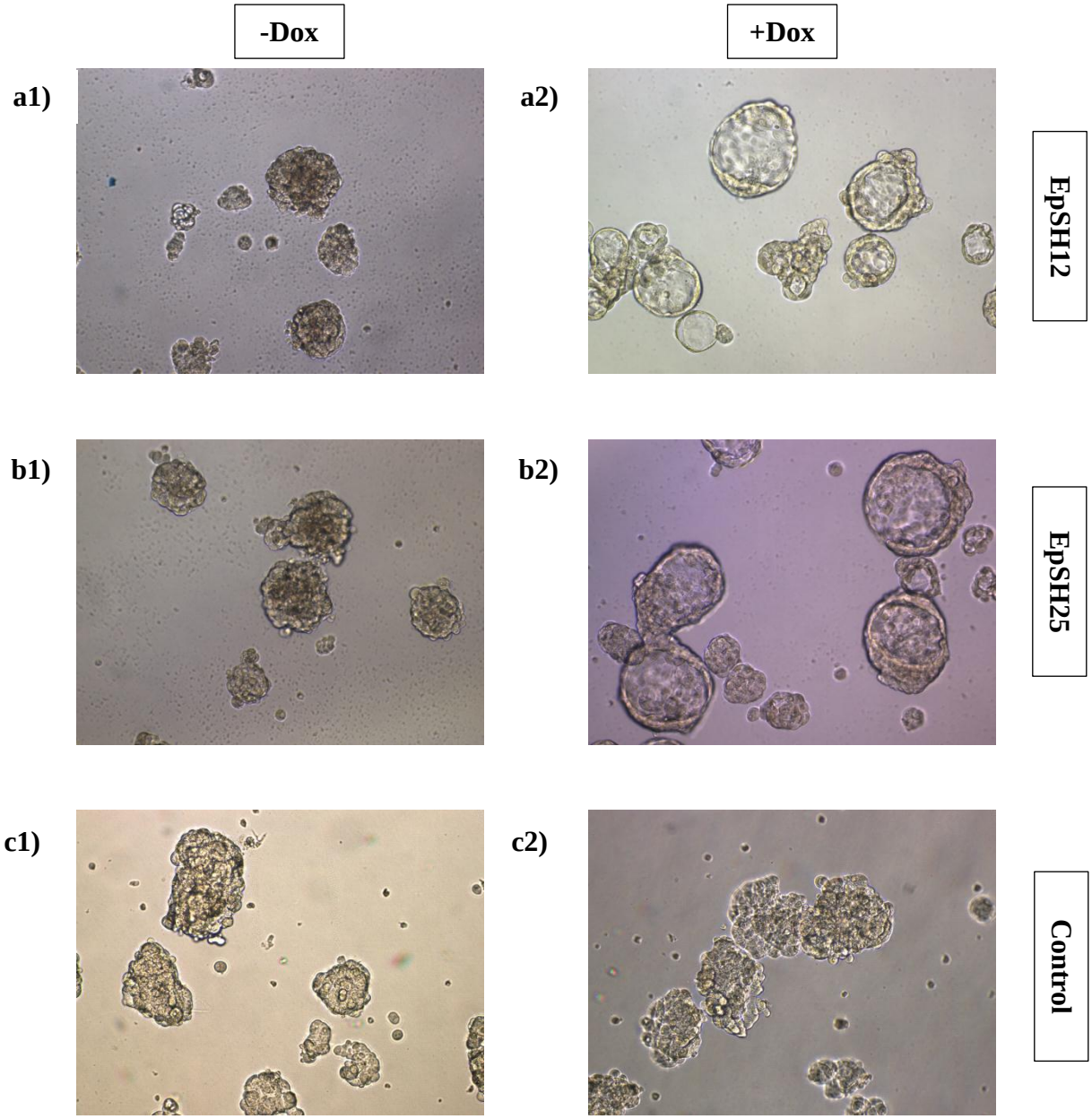


Figure 28: Effect of EpCAM knock-down on sphere morphology

Sphere forming assay in low attachment tissue culture flasks of a) MCF7 cells infected with EpCAM short hairpin 12 lentivirus (1-control, 2-doxycycline induced), b) MCF7 cells infected with EpCAM short hairpin 25 lentivirus (1-control, 2-doxycycline induced), c) uninfected MCF7 cells (1-control, 2-doxycycline induced).

first step was to verify the knock-down in EpCAM expression. We achieved this by immunofluorescence using an anti-EpCAM antibody. The results showed that EpCAM expression is indeed reduced by shRNA expression (Figure 29). Subsequently we tried to attain some information about the morphology of these spheres using H&E immunohistochemistry. In absence of doxycycline the EpCAM shRNA infected spheres resemble normal MCF7 spheres in morphology. Once the expression of the shRNA is induced by doxycycline, the H&E staining shows the spheres to be indeed hollow in the middle in most cases, proving our supposition. It is possible to detect spheres in various stages. Some are just beginning to hollow out (Figure 31b2.bottom), whilst others were only a thin layer of cells surrounding a big cavity (Figure 30b2.top). It seems the more hollow the sphere becomes; the smaller the cells in sphere appear to be. It is also possible to see the remnants of cellular debris in some of the cavities (Figure 30b1.top).

Next, we attempted to clarify the process that EpCAM knock-down initiates, through which the spheres transform into the hollow structures. Firstly, it was also necessary to determine if the cavity within the transformed spheres was indicative of cell death. Hence, we performed a TUNEL assay on the sphere sections. However, no apoptotic TUNEL+ cells were observed in any of the transformed spheres (Figure 32).We then tried to illustrate the level of cellular proliferation in these spheres using the Ki-67 marker. The results clearly showed a significant drop in the number of Ki-67+ highly proliferative cells in presence of the EpCAM shRNA (Figure 33). We then did immunofluorescence staining with some lineage marker antibodies to investigate a possible differentiation program taking place as a result of EpCAM knock-down.

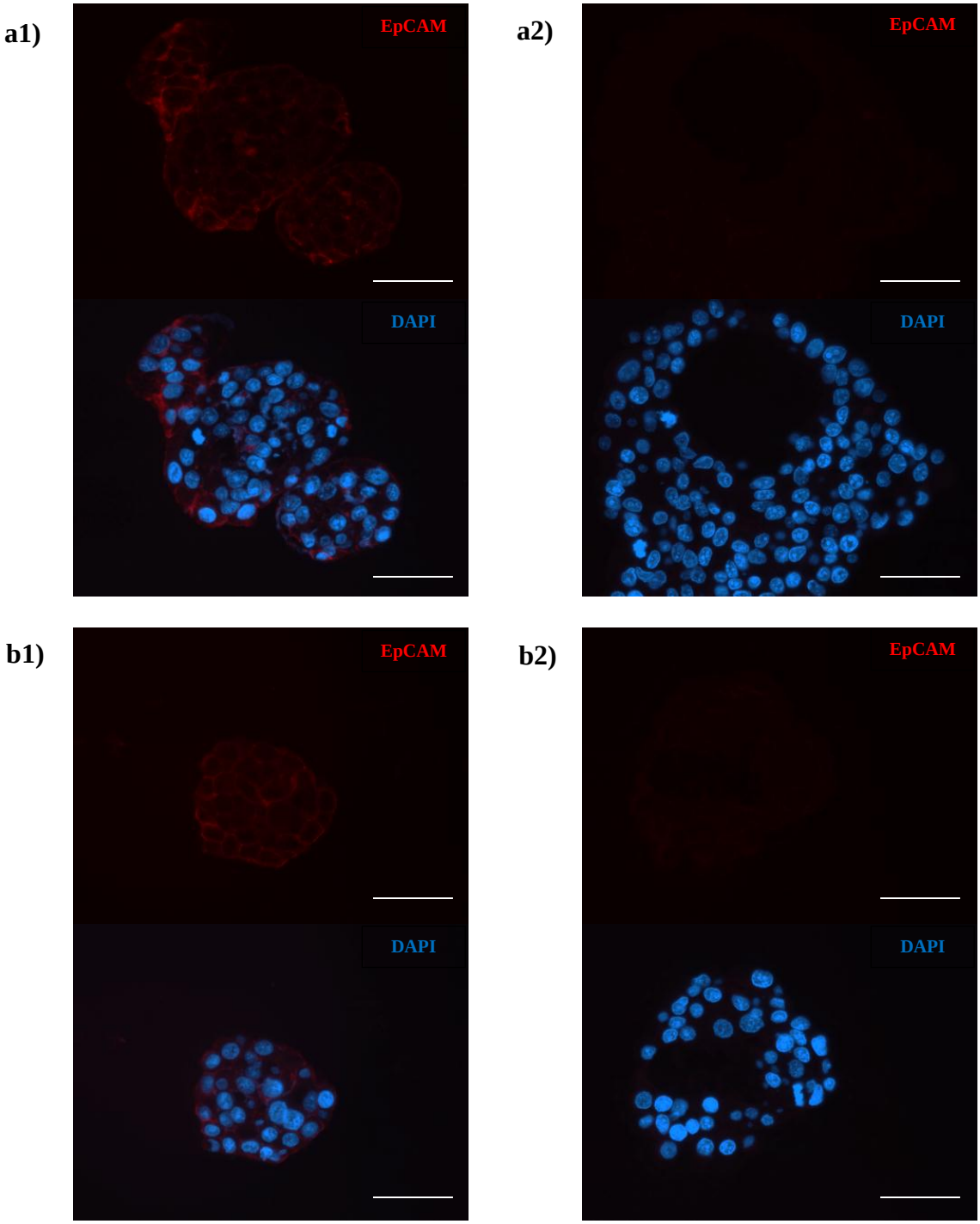


Figure 29: Knocking down EpCAM in MCF7 spheres using short hairpin RNA
MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immunofluorescence staining top: EpCAM(red), bottom: merge EpCAM with DAPI(blue). 400X magnification, scale bar 50 um.

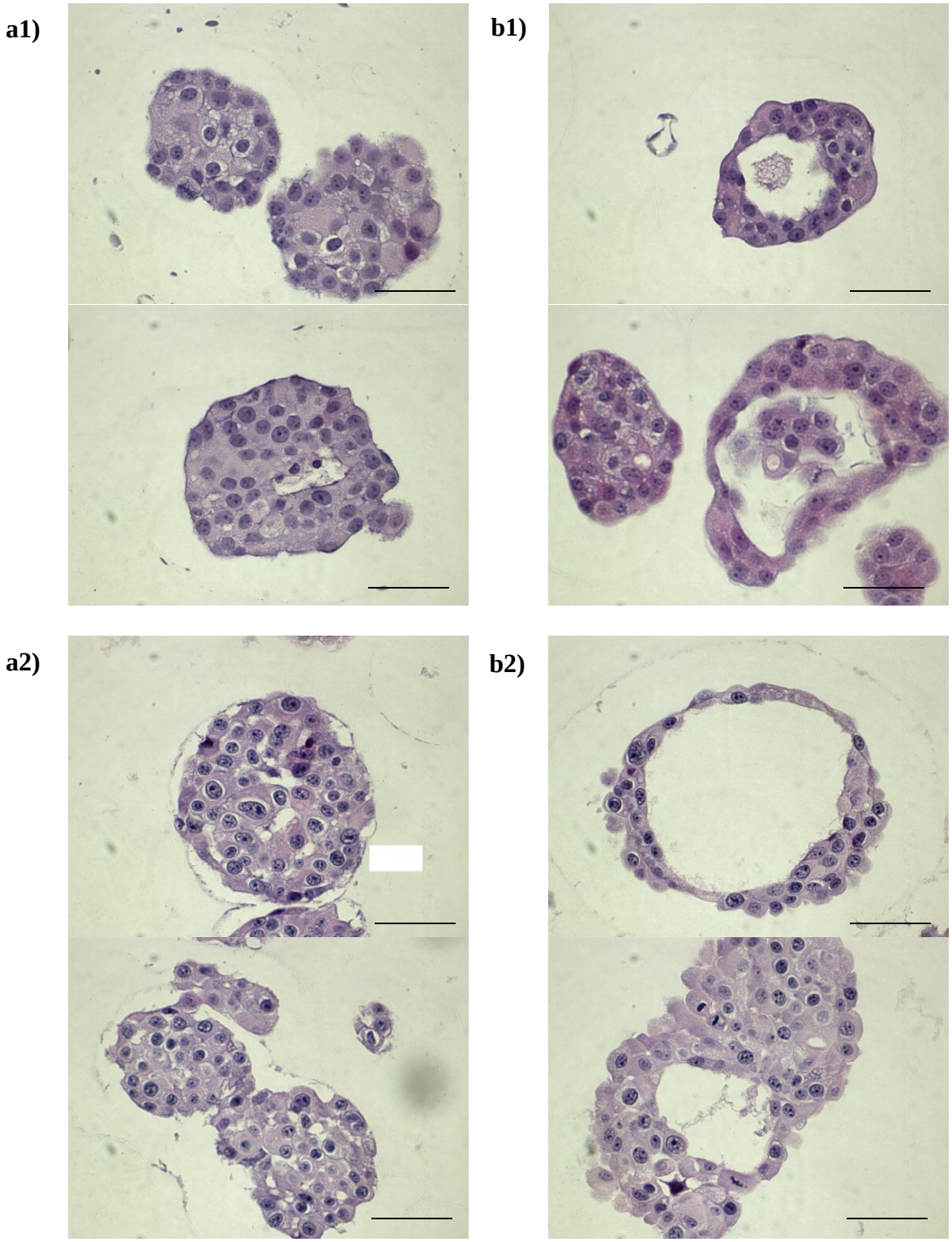


Figure 30: Effect of EpCAM knock-down on MCF7 spheres (short hairpin 12)
H&E stained MCF7 spheres infected with EpSH-12 lentivirus**.** a) Control untreated spheres, b) Doxycycline treated spheres. (a1 and b1 are PFA fixed sections. a2 and b2 are zinc fixed section.) 400X magnification, scale bar 50 um.

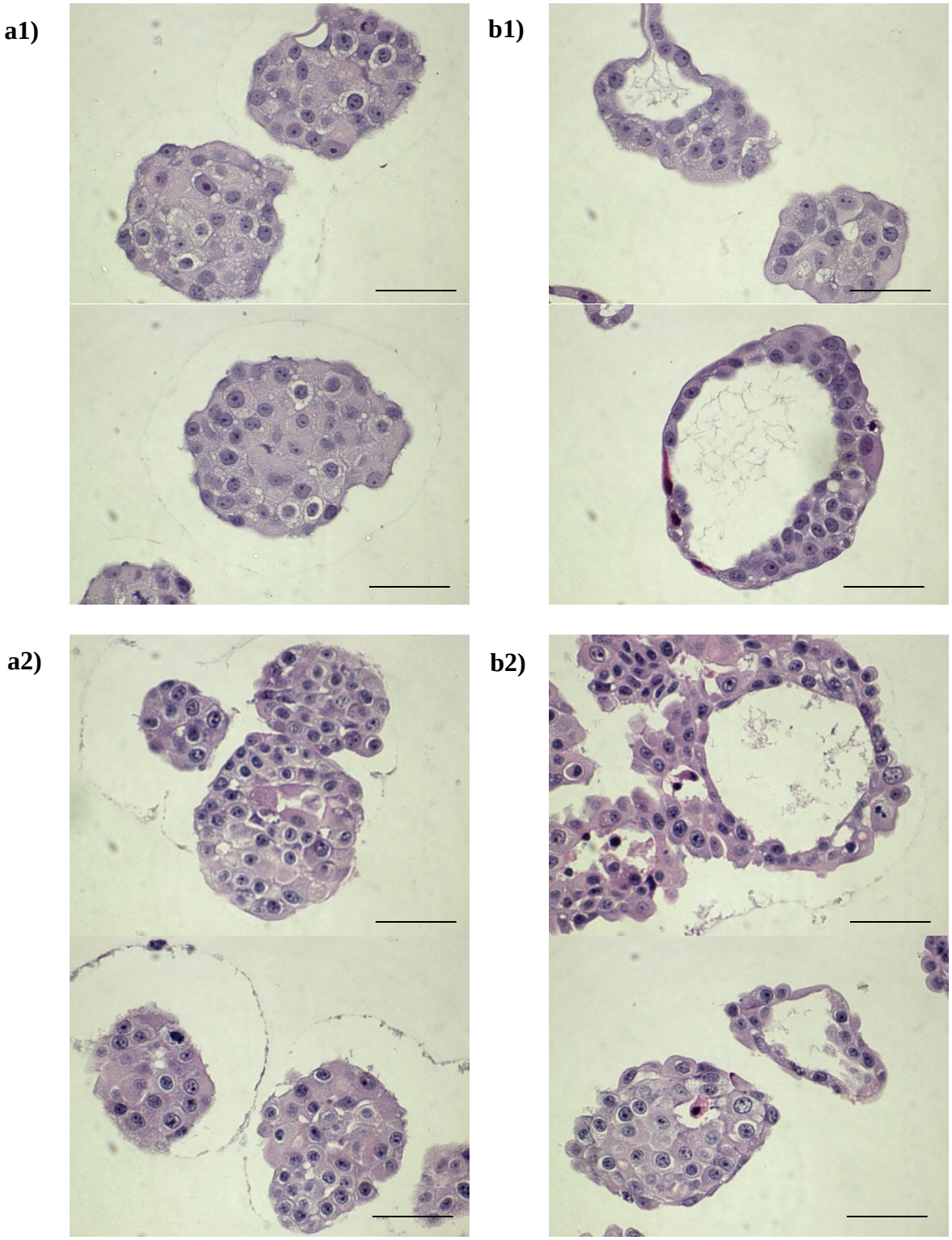


Figure 31: Effect of EpCAM knock-down on MCF7 spheres (short hairpin 25)
H&E stained of MCF7 spheres infected with EpSH-25 lentivirus. a) Control untreated spheres, b) Doxycycline treated spheres. (a1 and b1 are PFA fixed sections. a2 and b2 are zinc fixed section.) 400X magnification, scale bar 50 um.

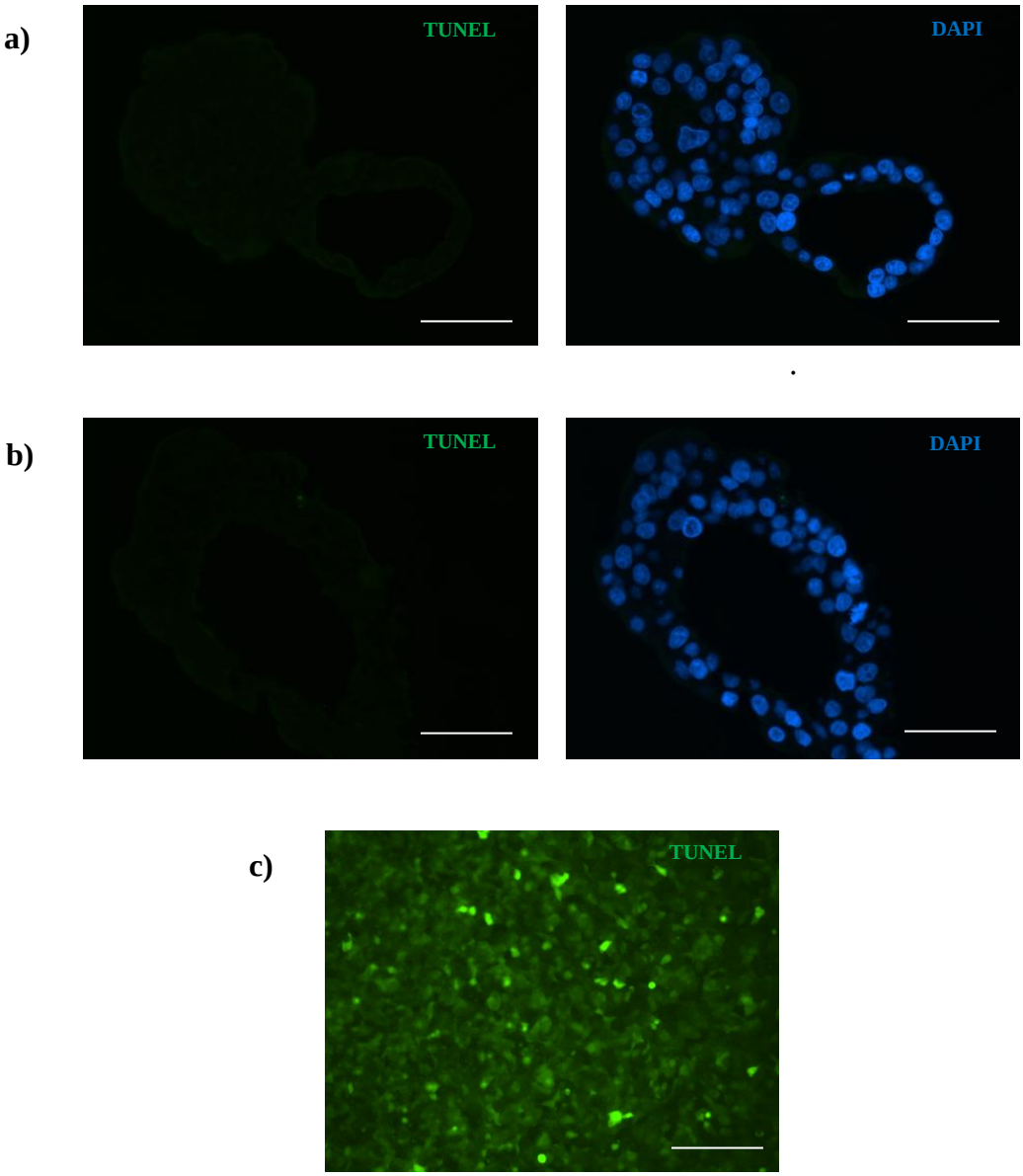


Figure 32: Absence of apoptotic cells in MCF7 spheres with EpCAM knock-down. MCF7 cells infected with a)EpSH-12 lentivirus, b)EpSH-25 lentivirus and induced with doxycycline, c)MMTV-Neu mouse tumor sample treated with Merck-003 (positive control). TUNEL assay (green). 400X magnification, scale bar 50 um.

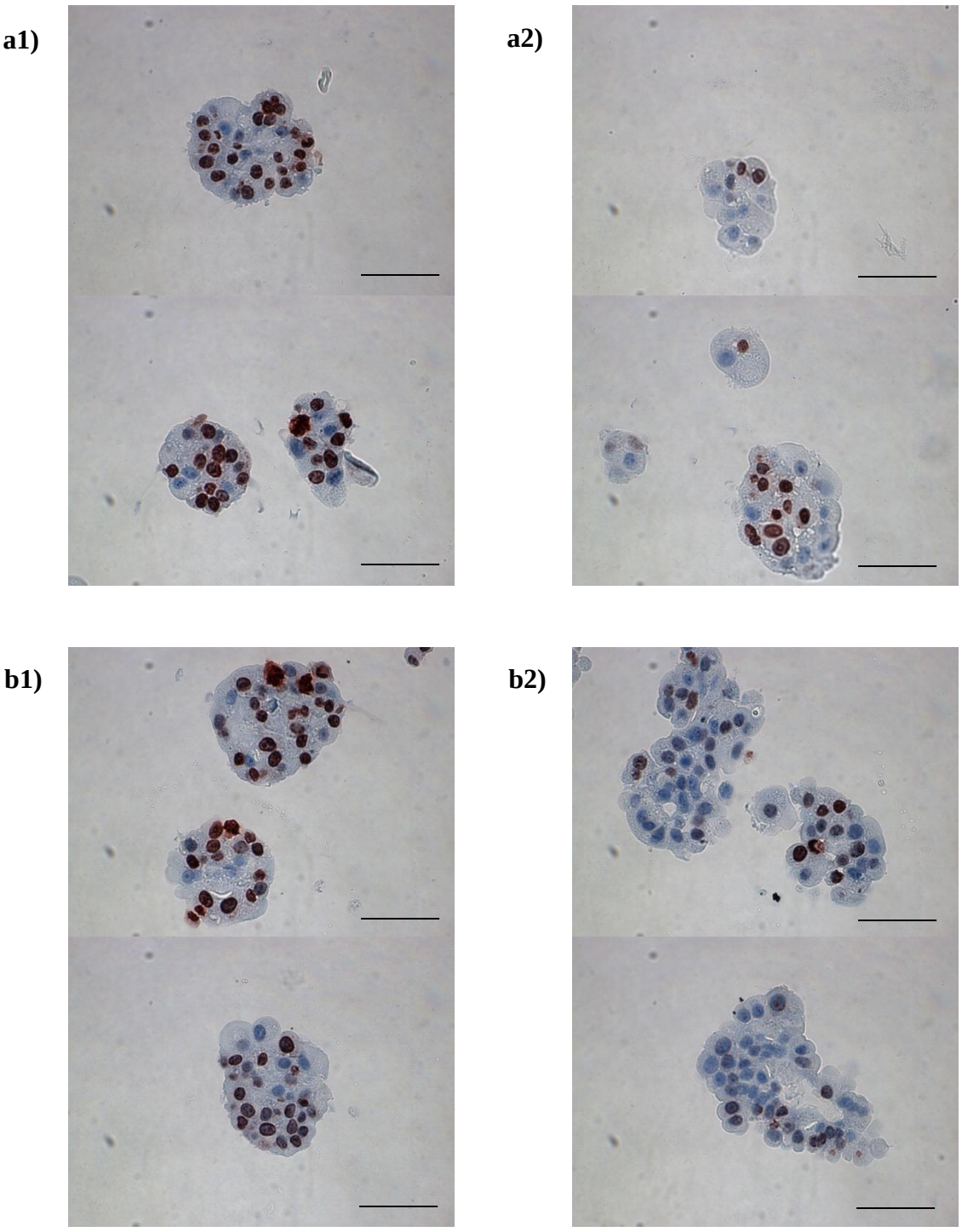


Figure 33: Effect of EpCAM knock-down on Ki-67 expression in MCF7 spheres. MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immuno-staining for Ki-67 (brown). 400X magnification, scale bar 50 um.

Although the staining patterns seemed very close, it was possible to detect a slight increase in the myoepithelial marker α-SMA in many of the transformed spheres in conjunction with a reduction in the luminal marker CK8 strength (Figure 34). Neither form of the spheres stained for CK5 or CK14 markers (data not shown).

The next logical step was to investigate the effect of knocking down EpCAM on the expression patterns of some of the proteins that are connected to the EpCAM pathway. Expectedly, staining for the EpCAM target gene Cyclin D1 visibly demonstrated a decrease in the number of Cyclin D1+ cells (Figure 35). The next protein we tested was β-Catenin, which is part of the EpCAM signalosome. The membranous β-Catenin staining clearly intensified by knocking down EpCAM (Figure 36). We also studied the expression pattern of E-Cadherin, since it has been shown that EpCAM inhibits E-Cadherin mediated connections. E-Cadherin also appeared to have decreased for one shRNA but increased for the other at the in the transformed spheres (Figure 37).

EpCAM expression and epithelial mesenchymal transitions (EMT) are two hall marks of circulating tumor cells, though they do not appear to coincide. Thus we studied the relationship between EpCAM down regulation and some proteins involved in the EMT process. Using an anti-Vimentin antibody, we showed this EMT marker to be up-regulated following EpCAM knockdown (Figure 38). High expression of MUC1 in cancer cells has been shown to induce the EMT process[117]. We saw a clear increase in MUC1 expression in our transformed spheres due to EpCAM knockdown (Figure 39).

The expression pattern is highly polar and towards to the outer edge of the sphere and not the inner lumen.

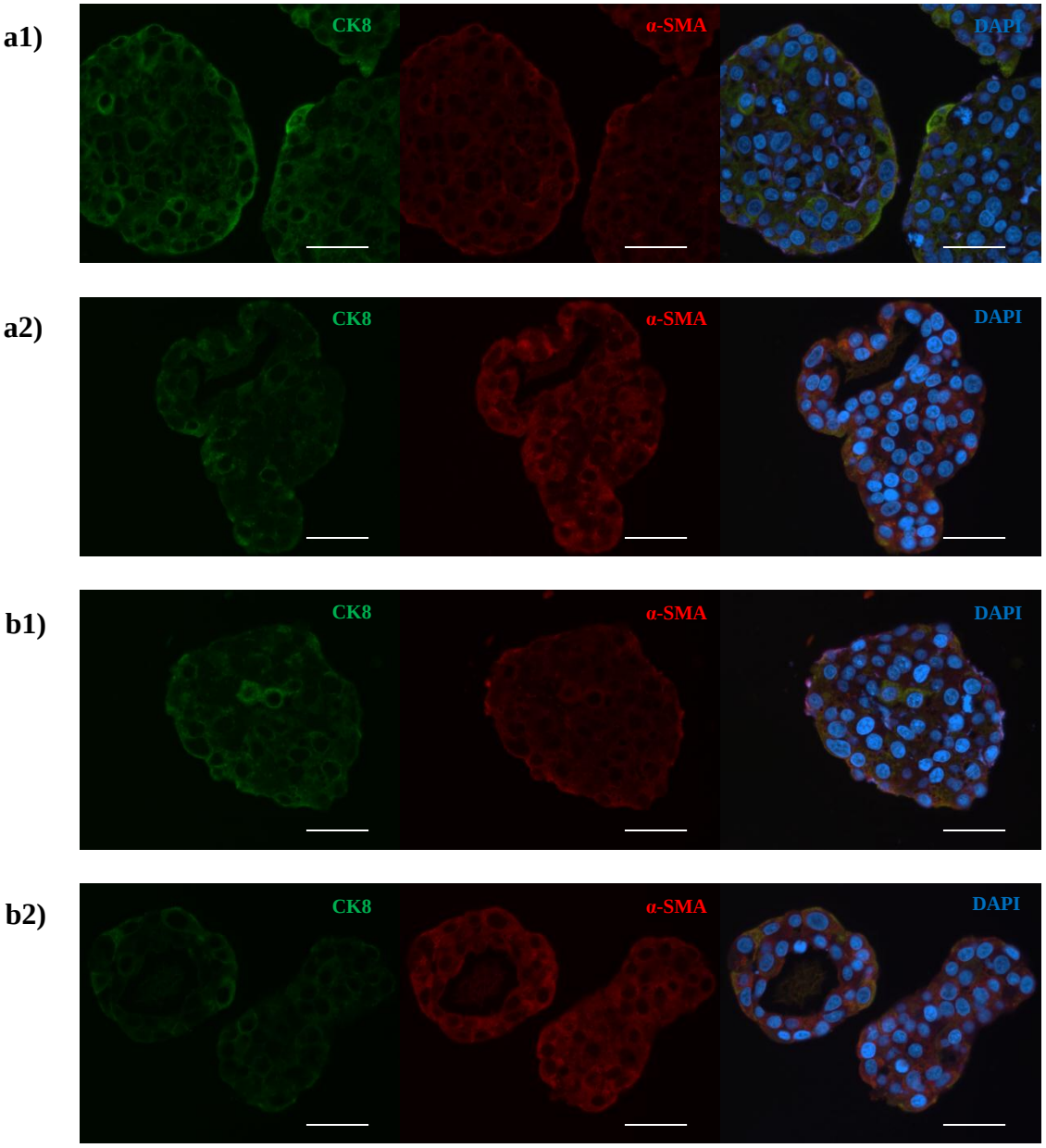


Figure 34: Effect of EpCAM knock-down on the expression of lineage markers in MCF7 spheres. MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immunofluorescence staining left CK8(green), middle α-SMA(red), left CK8 and α-SMA merge with DAPI(blue). 400X magnification, scale bar 50 um.

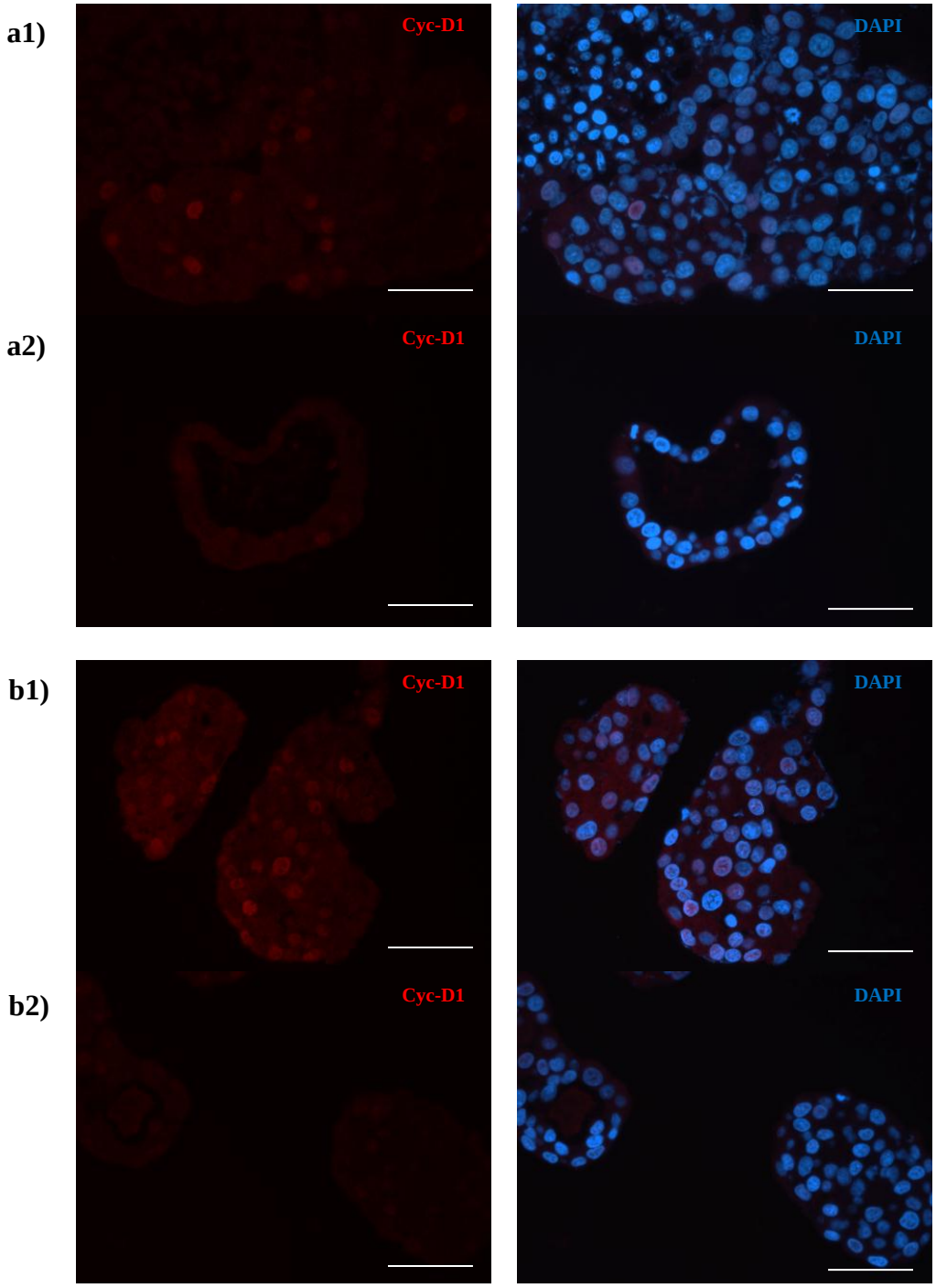


Figure 35: Effect of EpCAM knock-down on the expression of Cyclin-D1 in MCF7 spheres. MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immunofluorescence staining left Cyclin-D1 (red), right Cyclin-D1 merge with DAPI (blue). 400X magnification, scale bar 50 um.

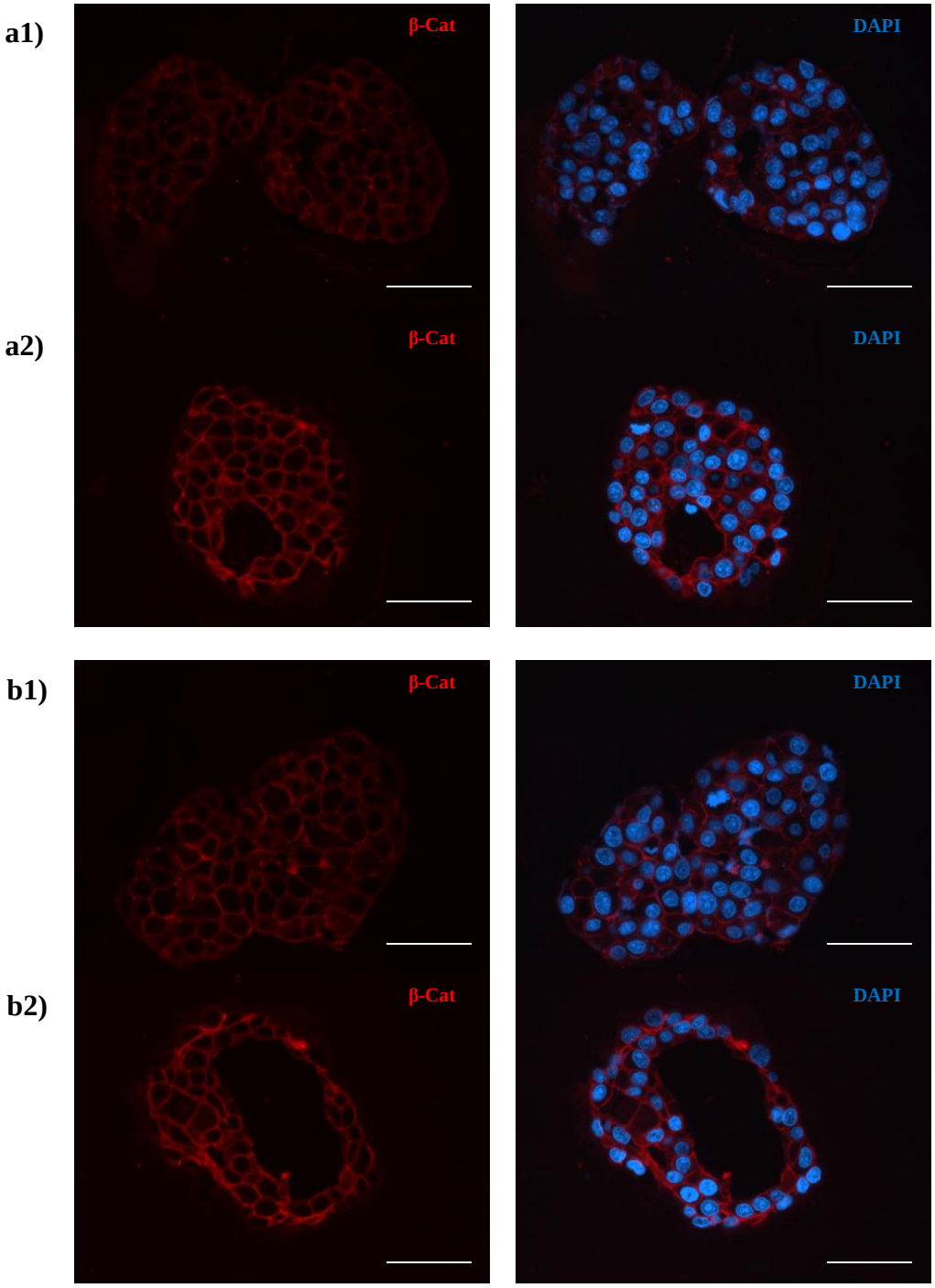


Figure 36: Effect of EpCAM knock-down on the expression of β-Catenin in MCF7 spheres. MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immunofluorescence staining left β-Catenin (red), right β-Catenin merge with DAPI (blue). 400X magnification, scale bar 50 um.

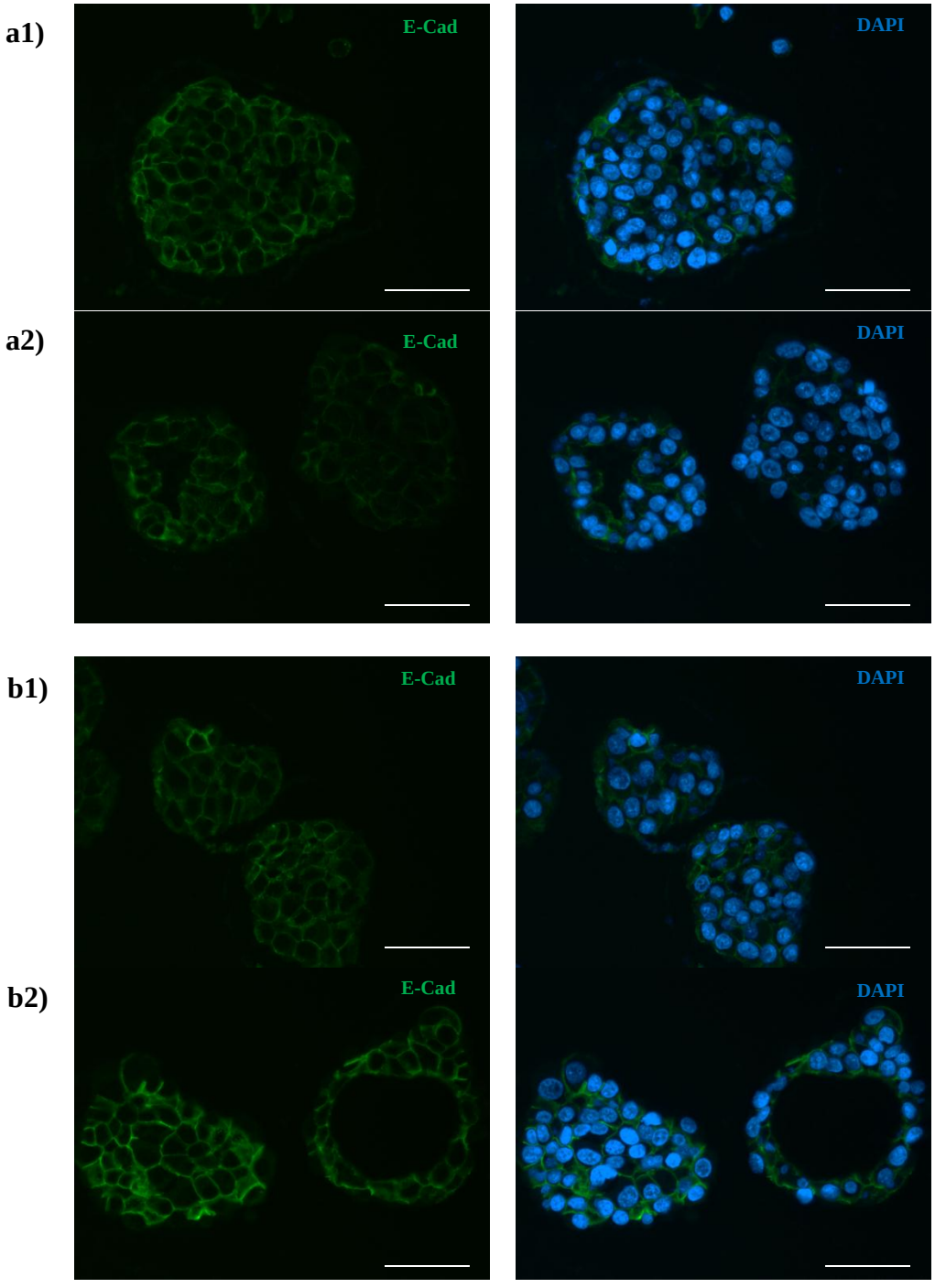


Figure 37: Effect of EpCAM knock-down on the expression of E-Cadherin in MCF7 spheres. MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immunofluorescence staining left E-Cadherin (green), right E-Cadherin merge with DAPI (blue). 400X magnification, scale bar 50 um.

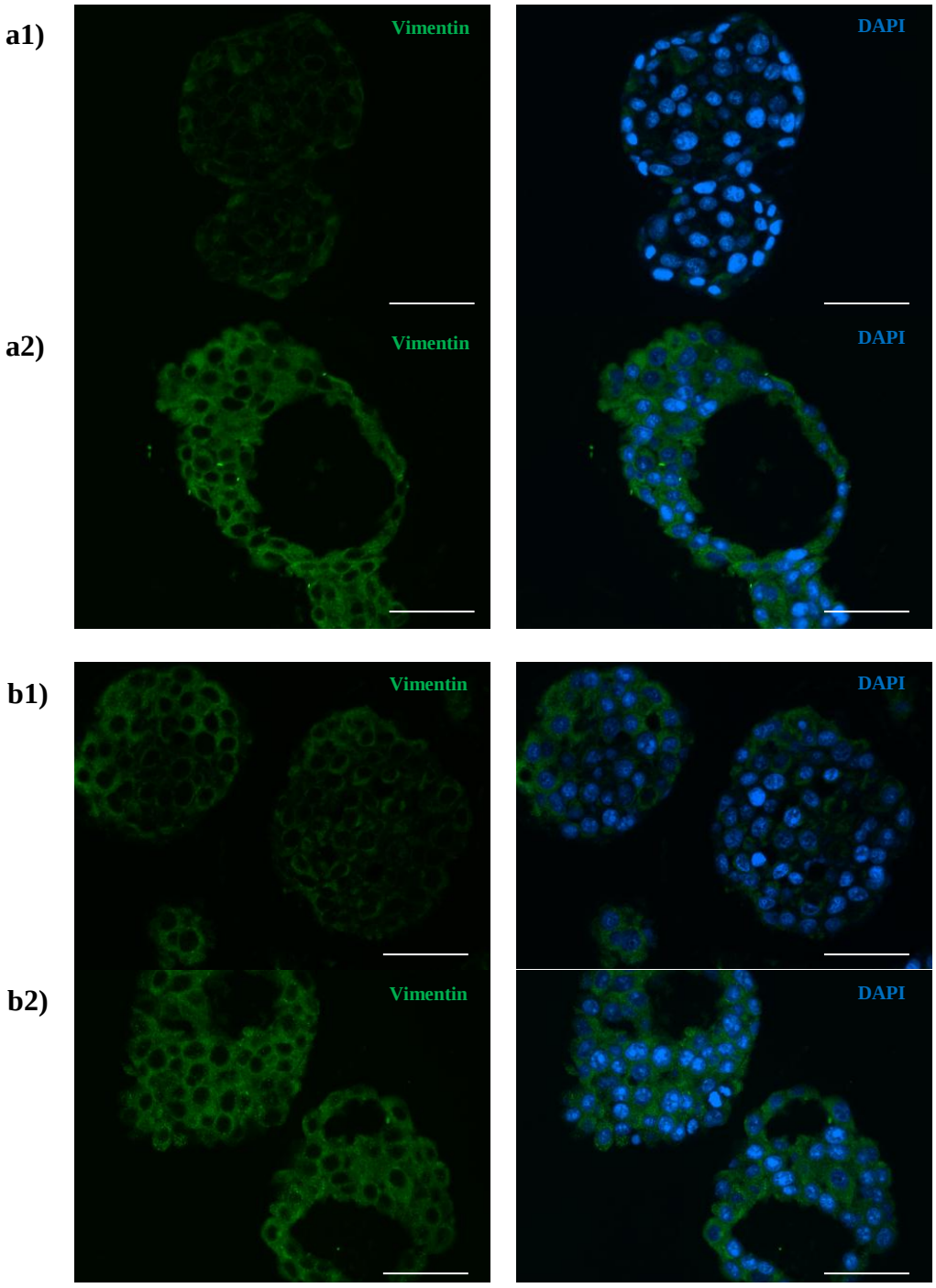


Figure 38: Effect of EpCAM knock-down on the expression of Vimentin in MCF7 spheres. MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immunofluorescence staining left Vimentin (green), right Vimentin merge with DAPI (blue). 400X magnification, scale bar 50 um.

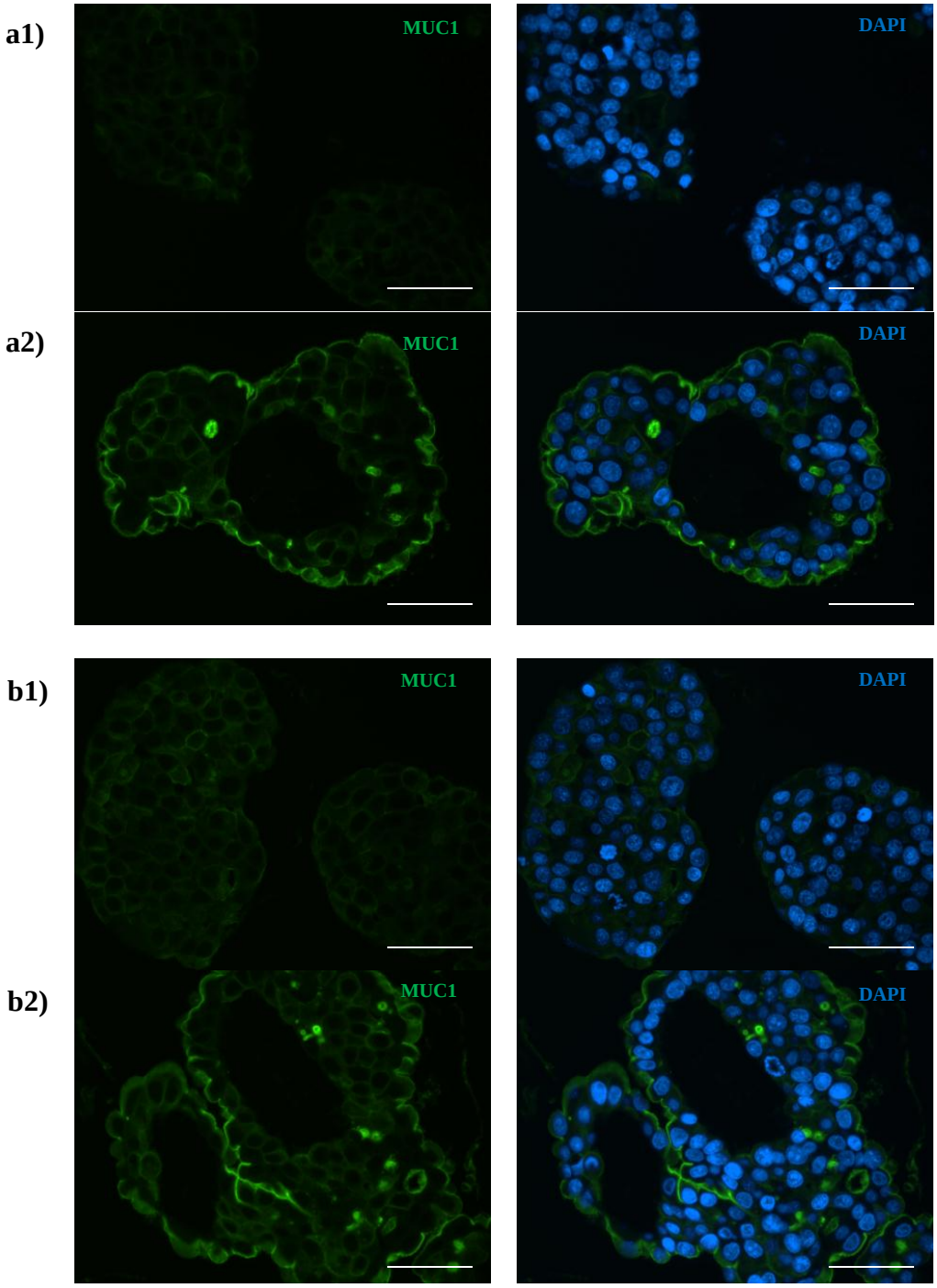


Figure 39: Effect of EpCAM knock-down on the expression of Mucin 1 in MCF7 spheres. MCF7 cells infected with a)EpSH-12 lentivirus (1-control, 2-doxycycline induced), b)EpSH-25 lentivirus (1-control, 2-doxycycline induced). Immunofluorescence staining left Mucin 1 (green), right Mucin 1 merge with DAPI (blue). 400X magnification, scale bar 50 um.

5 Discussion

The high expression level of EpCAM in our murine model of breast cancer was the starting point for this project (Figure 4). In this transgenic mouse model the oncogene Her2/Neu is expressed via the MMTV promoter, which is a mammary epithelial cell specific promoter[113]. These murine tumors closely resemble $Her2^{+}$ human breast tumors[118]. Whereas EpCAM expression was pretty uniform across the cells of the primary tumors (Figure 5d), such expression was limited to the luminal cells of the healthy mammary tissue (Figure 5C). As such total EpCAM expression is much lower in a healthy mammary gland relative to a breast tumor (Figure 4). These data coincide with the published data which show over-expression of EpCAM in human breast cancer[60]. The *in vitro* tumorsphere cultures derived from the primary tumors also displayed a moderate level of EpCAM expression. Correspondingly, *in vitro* mammosphere cultures showed very negligible EpCAM expression (Figure 4). EpCAM seemed to be uniformly expressed across the cells of both types of spheres (Figure 5a/b). The *in vitro* data is significant since the sphere forming cells of mammospheres and tumorspheres are thought to be the *in vitro* counterparts of stem cells and cancer stem cells respectively[116].

To further elucidate the role of EpCAM in our breast cancer mouse model, we began our proof of principle experiments in the MCF7 breast cancer cell line. We engineered multi-clonal populations of MCF7 cells that inducibly over-express EpCAM, express EpCAM shRNAs or a dominant negative EpCAM protein. We aimed to compromise EpCAM function in these cells and subject them to various functional assays

to elucidate EpCAM's contribution to cellular characteristic associated with cancer stem cells such as proliferation, differentiation, cell surface marker expression and *in vitro* clonogenicity.

In studying the effects of EpCAM over-expression we opted to employ a second cell line in addition to MCF7 cells. Because MCF7 cells have relatively low endogenous Wnt activity, a low reporter activity level is implied. Hence, we also used colon cancer cell line HCT116 which have higher Wnt pathway activity. This decision was rather beneficial in explaining the outcome of the experiment. We observed an increase in the TCF reporter activity by over-expressing EpCAM in both cell lines with the integrated over-expression vector, though this effect was more pronounced for HCT116 cells (Figure 7). This is not unexpected since HCT116 cells have much higher concentration of β-catenin in their cytoplasm, due to the mutation in the β-catenin gene. As β-catenin is part of the EpCAM signalling cascade, its accumulation facilitates the increase in the activity of its pathway. This is a confirmation of sorts for EpCAM over-expression being indeed responsible for the hike in the reporter activity. The fluorescence positivity of these cells is another validation, since the EpCAM over-expression constructs are fused to YFP (Figure 6). The use of transfection in introducing the EpCAM over-expression vector into the cells was meant to amplify the effect that we observed by infection. However, no significant increase in the reporter activity was observed in either cell line (Figure 8). As the β-catenin over-expression vector did not prompt any increase either, promoter competition seems to be a reasonable explanation. By pouring so many copies of the genes into these cells we may have exhausted the resources of the transcriptional

machinery. The effectiveness of the Wnt activator compound chir99021, in spite of the failure of the transfections, is corroborative of this hypothesis. Low transfection efficiency could also possibly account for these observations.

The expression of our dominant negative EpCAM at the RNA level (Figure 10) but not at the protein level (Figure 11) could have multiple explanations. It could mean the expressed mRNA is not being translated into protein, which may imply an error in the cloning process inhibiting translation. Nonetheless, sequencing our construct from both directions revealed no mutations and the Kozak consensus sequence, which is essential for the initiation of translation, was included in the design. On the other hand, it is possible that the fusion protein is indeed expressed but we are unable to detect it. The dominant negative fusion protein is very similar in size to the endogenous EpCAM protein (~40kDa). This would bring into question the result of the western blot using the anti-EpCAM antibody (Figure 11a). The epitope of the anti-engrailed antibody was originally thought to be within the repression domain included in the ΔNEpCAM protein, based on the catalogue information. However, upon negative results, we further investigated this issue by consulting with the primary article introducing the 4D9 antibody. The 4D9 epitope was discovered to be residues 487-500[119], whereas the engineered dominant negative EpCAM protein only contains residues 2-298 of the drosophila engrailed protein. The presence of a band for our positive control sample seems to confirm this, since those cells expressed the full length engrailed protein (Figure 11b- lane 8).

Due to the difficulties in obtaining another functional engrailed antibody with the appropriate epitope to demonstrate the expression of our dominant negative protein, we decided to perform some functional assays in the hope of observing some activity. The luciferase reporter assay did not show any decrease by inducing the expression of the dominant negative EpCAM in MCF7 cells. Even HCT116 cells that have a higher endogenous TOPFlash level were not affected by the ΔNEpCAM. Hiking up this baseline even further using BIO did not improve the results either (Figure 12). Repeating the experiment at multiple time points did not produce the desired results, as no truly significant effect was observed in any of the experimental conditions (Figure 13). Transfection techniques were once again employed to possibly enhance any possible effect of our construct, by introducing many copies of the transgene into the cells. Although the dominant negative TCF construct seemed effective, the ΔNEpCAM only functioned at the highest concentration (Figure 14). This single instance of the ΔNEpCAM showing activity is not significant. The lack of activity of ΔNEpCAM was again seen in the colony forming assay, since no significant change was observed by inducing the expression of ΔNEpCAM (Figure 15).

Although the dominant negative EpCAM was shown to be expressed at the transcript level, the absence of tangible proof for the expression or the functionality of our dominant negative EpCAM protein was not reassuring. We can only hypothesise as to the reason. Either the translation process of the protein is erroneous, or even if properly translated it may not be functional. The application of the engrailed repression domain, though effective many times in the past[120], is not foolproof. The three dimensional

structure of the EpCAM signalosome is not fully known. The components are known to be EpICD, FHL2, β-catenin and Lef-1, but the orientation of the binding is unknown[51]. FHL2 is alleged to be a scaffolding protein to which the other three bind. Lef-1 is the DNA binding protein. EpICD's position at the complex may be somehow that the engrailed fusion does not sterically hinder Lef1, in which case the dominant negative EpCAM protein would not be functional. In consequence, the dominant negative EpCAM protein is either not expressed (or rapidly degraded within the cell), or lacks any functional activity.

If the attempt to block EpCAM's function indeed failed, knocking down EpCAM's expression is still a viable option. Therefore, we proceeded with our plans to use shRNAs against EpCAM. We designed two shRNA constructs and confirmed their activity with an almost 2 fold reduction at both mRNA (Figure 16) and protein level (Figure 17). We also succeeded in knocking down the TOPFlash reporter activity using our EpCAM shRNAs (Figure 18). Although this reporter was originally designed for the TCF4 of the Wnt pathway, it is also effective as a reporter for EpCAM pathway. Because Lef1 which is part of the EpICD nuclear complex is also one of the TCFs involved in the Wnt pathway[121]. Maetzel *et al.* observed similar effects in HCT-8 cells[51]. Including the renilla luciferase general transcription reporter to normalize our firefly luciferase TCF reporter activity further confirmed this data, as it showed the observed knock-down to be specific to the TCF controlled genes (Figure 19). The growth curve analysis was performed as a measure of the effects of the EpCAM short hairpins on proliferation. The seven day assay showed a significant difference in the growth of the MCF7 cells infected

with the shRNAs in presence and absence of doxycycline (Figure 20b/c). We proved this to be a result of induction of our shRNAs and not general toxicity of doxycycline by performing the same assay with the parental cell line not infected with the shRNA viruses where little or no effect was observed (Figure 20a). Thus the inhibition of proliferation seems to be a phenotype of knocking down EpCAM. This is consistent with what Osta *et al.* observed regarding the inhibitive effect of knocking down EpCAM on cell proliferation[60]. Cells with the knocked down EpCAM eventually catch up to the control cells on the growth curve. This may be as a result of shortage of space, since the cells reach over-confluence (Figure 21). This is confirmed by the plateauing of the proliferation past day 6 shown in the PrestoBlue reduction assay (Figure 22).

To test the effect of knocking down EpCAM on the clonogenic properties of these MCF7 cells, we performed colony forming assays. We observed a 3 fold reduction in CFE as a result of knocking down EpCAM (Figure 23). Furthermore, colonies formed in presence of doxycycline were generally smaller than the uninduced colonies (Figure 24). Therefore, knocking down EpCAM seems to have had a double effect on MCF7 colony formation. It not only reduced the number of colony forming cells, it also reduced their proliferation capabilities. Colony forming cells are generally accepted to be the *in vitro* counterparts of tissue progenitor cells. As most of the small colonies in the induced group seemed to consist of larger cells, namely myoepithelial colonies, we can deduce that luminal progenitor are more affected by the knock-down of EpCAM (Figure 24a2-right).

In the sphere forming assays in normal tissue culture treated plates rather than forming spheres, the cells seeded in presence of doxycycline formed an adherent monolayer (Figure 25). However, these adherent cells did not resemble normal adherent MCF7 cells grown in the presence of serum. They did not look like spheres that have stuck down to the plate either. In presence of doxycycline, the single cells seem to attach to the plate and enlarge and never form spheres (Figure 26). In conclusion, this may imply the start of a possible differentiation program in these cells, due to the EpCAM knock-down, that inhibits sphere formation and converts them into adherent cell. One explanation for this phenomenon could be through the connection between EpCAM and E-cadherin. EpCAM is known to have an inhibitory effect on cadherin mediated junctions[55]. We may postulate a similar effect EpCAM may emanate on other adhesion molecules such as integrins that are involved in cell adhesion to the extracellular matrix attaching the cells to the plastic surface. This hypothesis may provide a rationalization as to how knocking down EpCAM promotes cell adherence. It is possible that a process like a reverse EMT is taking place. In fact MCF7 sphere cultures have been shown to have mesenchymal properties[122]. Thus it is possible that by knocking down EpCAM, we are putting into place a MET (reverse EMT or mesenchymal epithelial transition) process that turns the mesenchymal like cells in the MCF7 sphere cultures into epithelial adherent cells. Previous studies had shown EpCAM expression to be mandatory for the migratory characteristics of dendritic cells[105] and the EpCAM pathway to be highly active in invasive cancer cells[107]. Our hypothesis is that EpCAM is highly expressed and its pathway is active in epithelial cells with invasive tendencies. This leads to an EMT

process, which expedites migration of these cells. Consequently, by knocking down EpCAM in our cultures we reverse this process, meaning the sphere cells undergo MET to be more epithelial like and attach to the plate.

The sphere forming assay in ultra-low attachment plates provided different results. The similar number of spheres between induced and uninduced cultures, when assayed in the ultra-low attachment plates, seems to have one explanation (Figure 27a). Knocking down EpCAM does not affect the number of sphere forming cells. Seemingly, the reduction of EpCAM expression does not influence the viability of cancer stem cells (*in vivo* equivalents of sphere forming cells[123]). On the other hand, taking into consideration the morphology (Figure 28) and the proliferation levels (Figure 27b) of these abnormal spheres I hypothesized the cells at the centre of the spheres may be dying as a result of EpCAM knock-down. Therefore, when it comes to cellular viability, even if the sphere forming cells are not affected, the more differentiated cells in the sphere may be. The H&E stained sections confirmed the hollowness of the spheres and the presence of some cellular debris seems to support our supposition of the interior cells dying (Figure 30/31). The absence of $TUNEL^+$ cells does not necessarily contradict this notion, as apoptosis is not the only method of cell death (Figure 32). Further experiments can be performed to ascertain whether necrosis or autophagy is taking place. We confirmed the decreased proliferation that we gathered from the AlamarBlue assay by Ki-67 staining. In line with a previous study[76], the number of Ki-67+ cells declined by knocking down EpCAM (Figure 33).

We used staining with lineage markers to try to ascertain if a differentiation program was initiated by EpCAM knock-down. MCF7 cells are known as a luminal epithelial cell line. The staining with the lineage markers seem to hint at a shift towards a more myoepithelial phenotype, though the effect was not extensive (Figure 34). As a target of EpCAM pathway[76], cyclin-D1 expression was reduced by EpCAM knock-down. Cyclin-D1 was shown to be necessary for self-renewal and differentiation of breast cancer stem cells[46]. Henceforth, cyclin-D1 reduction may contribute to the effects that EpCAM knock-down seems to have on MCF7 spheres. By knocking down EpCAM, fewer β-catenin molecules are being translocated to the nucleus as part of the EpICD signalosome. This could explain the increased membranous β-catenin (Figure 36). EpCAM knock down seems to have instigated an EMT based on the increased expression of vimentin (Figure 38). This is substantiated by the marked increase in MUC1 expression in the EpCAM knocked down spheres (Figure 39). MUC1 expression has been shown to induce EMT in pancreatic cells demonstrated by increase in EMT markers such as Vimentin and decrease in epithelial proteins such as E-Cadherin[117]. Accordingly, we saw a slight decrease in E-cadherin expressions by knocking down EpCAM (Figure 37). EpCAM's inhibition of cadherin junctions[55] does not necessitate the alteration of their expression pattern.

Our most revealing results were from the EpCAM knock down experiments in MCF7 cells. We demonstrated the commonality of the EpCAM and Wnt pathway in the use of the TCF family of transcription factors. We established the detrimental effects of EpCAM knock down on cell growth and proliferation. Colony forming (progenitor) cells

were also shown to be dependent on EpCAM. We also learned that whereas the number of sphere forming (cancer stem) cells is not correlated with EpCAM expression, the morphology of spheres most certainly is. Some form of cell death seems to take place in the MCF7 spheres as a result of knocking down EpCAM. A shift of the MCF7 sphere cells from a luminal to a basal phenotype as a result of EpCAM knock-down was also proposed. Furthermore, our interpretation of the data from the sphere forming assays lead us to believe that EpCAM knock down can bring about either EMT or MET depending on the circumstances. We previously introduced the flexible relationship that seems to be present between EpCAM's expression and the EMT process associated with migratory characteristics of cancer cells. The apparent contradiction in our observations can be alleviated by this phenomenon.

6 References

1 Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23-28, (1976).

2 Gupta, P. B., Chaffer, C. L. & Weinberg, R. A. Cancer stem cells: mirage or reality? *Nat Med* **15**, 1010-1012, (2009).

3 Hamburger, A. W. & Salmon, S. E. Primary bioassay of human tumor stem cells. *Science* **197**, 461-463, (1977).

4 Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645-648, (1994).

5 Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* **3**, 730-737, (1997).

6 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, (2003).

7 Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432**, 396-401, (2004).

8 O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106-110, (2007).

9 Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111-115, (2007).

10 Li, C. *et al.* Identification of pancreatic cancer stem cells. *Cancer Res* **67**, 1030-1037, (2007).

11 Schatton, T. *et al.* Identification of cells initiating human melanomas. *Nature* **451**, 345-349, (2008).

12 Curley, M. D. *et al.* CD133 expression defines a tumor initiating cell population in primary human ovarian cancer. *Stem Cells* **27**, 2875-2883, (2009).

13 Magee, J. A., Piskounova, E. & Morrison, S. J. Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer Cell* **21**, 283-296, (2012).

14 Taussig, D. C. *et al.* Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood* **112**, 568-575, (2008).

15 Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nature medicine* **17**, 1086-1093, (2011).

16 Beier, D. *et al.* CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles. *Cancer research* **67**, 4010-4015, (2007).

17 Wang, J. *et al.* CD133 negative glioma cells form tumors in nude rats and give rise to CD133 positive cells. *International journal of cancer. Journal international du cancer* **122**, 761-768, (2008).

18 Vaillant, F. *et al.* The mammary progenitor marker CD61/beta3 integrin identifies cancer stem cells in mouse models of mammary tumorigenesis. *Cancer research* **68**, 7711-7717, (2008).
19 Yilmaz, O. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475-482, (2006).
20 Krivtsov, A. V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818-822, (2006).
21 Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715, (2008).
22 Scheel, C. *et al.* Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. *Cell* **145**, 926-940, (2011).
23 Kelly, P. N., Dakic, A., Adams, J. M., Nutt, S. L. & Strasser, A. Tumor growth need not be driven by rare cancer stem cells. *Science* **317**, 337, (2007).
24 Quintana, E. *et al.* Efficient tumour formation by single human melanoma cells. *Nature* **456**, 593-598, (2008).
25 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, (2007).
26 Ponti, D. *et al.* Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer research* **65**, 5506-5511, (2005).
27 Fillmore, C. M. & Kuperwasser, C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res* **10**, R25, (2008).
28 Charafe-Jauffret, E. *et al.* Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer research* **69**, 1302-1313, (2009).
29 Liu, J. C., Deng, T., Lehal, R. S., Kim, J. & Zacksenhaus, E. Identification of tumorsphere- and tumor-initiating cells in HER2/Neu-induced mammary tumors. *Cancer Res* **67**, 8671-8681, (2007).
30 Cho, R. W. *et al.* Isolation and molecular characterization of cancer stem cells in MMTV-Wnt-1 murine breast tumors. *Stem Cells* **26**, 364-371, (2008).
31 Zhang, M. *et al.* Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. *Cancer research* **68**, 4674-4682, (2008).
32 Morel, A. P. *et al.* Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One* **3**, e2888, (2008).
33 Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. *Nature* **434**, 843-850, (2005).
34 Harrison, H., Farnie, G., Brennan, K. R. & Clarke, R. B. Breast cancer stem cells: something out of notching? *Cancer research* **70**, 8973-8976, (2010).
35 Liu, S. *et al.* Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer research* **66**, 6063-6071, (2006).
36 Korkaya, H. *et al.* Regulation of mammary stem/progenitor cells by PTEN/Akt/beta-catenin signaling. *PLoS Biol* **7**, e1000121, (2009).

37 Korkaya, H., Paulson, A., Iovino, F. & Wicha, M. S. HER2 regulates the mammary stem/progenitor cell population driving tumorigenesis and invasion. *Oncogene* **27**, 6120-6130, (2008).
38 Cicalese, A. *et al.* The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. *Cell* **138**, 1083-1095, (2009).
39 Shipitsin, M. *et al.* Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11**, 259-273, (2007).
40 Chaffer, C. L. *et al.* Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7950-7955, (2011).
41 Iliopoulos, D., Hirsch, H. A., Wang, G. & Struhl, K. Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1397-1402, (2011).
42 Yu, F. *et al.* let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* **131**, 1109-1123, (2007).
43 Stahl, M., Ge, C., Shi, S., Pestell, R. G. & Stanley, P. Notch1-induced transformation of RKE-1 cells requires up-regulation of cyclin D1. *Cancer research* **66**, 7562-7570, (2006).
44 Phillips, T. M., McBride, W. H. & Pajonk, F. The response of CD24(-/low)/CD44+ breast cancer-initiating cells to radiation. *J Natl Cancer Inst* **98**, 1777-1785, (2006).
45 Velasco-Velazquez, M. A., Homsi, N., De La Fuente, M. & Pestell, R. G. Breast cancer stem cells. *Int J Biochem Cell Biol* **44**, 573-577, (2012).
46 Jeselsohn, R. *et al.* Cyclin D1 kinase activity is required for the self-renewal of mammary stem and progenitor cells that are targets of MMTV-ErbB2 tumorigenesis. *Cancer Cell* **17**, 65-76, (2010).
47 Pece, S. *et al.* Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140**, 62-73, (2010).
48 Ponta, H., Sherman, L. & Herrlich, P. A. CD44: from adhesion molecules to signalling regulators. *Nature reviews. Molecular cell biology* **4**, 33-45, (2003).
49 Nagano, O. & Saya, H. Mechanism and biological significance of CD44 cleavage. *Cancer Sci* **95**, 930-935, (2004).
50 Murakami, D. *et al.* Presenilin-dependent gamma-secretase activity mediates the intramembranous cleavage of CD44. *Oncogene* **22**, 1511-1516, (2003).
51 Maetzel, D. *et al.* Nuclear signalling by tumour-associated antigen EpCAM. *Nat Cell Biol* **11**, 162-171, (2009).
52 Munz, M., Baeuerle, P. A. & Gires, O. The emerging role of EpCAM in cancer and stem cell signaling. *Cancer Res* **69**, 5627-5629, (2009).
53 Litvinov, S. V., Velders, M. P., Bakker, H. A., Fleuren, G. J. & Warnaar, S. O. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J Cell Biol* **125**, 437-446, (1994).
54 Balzar, M. *et al.* Cytoplasmic tail regulates the intercellular adhesion function of the epithelial cell adhesion molecule. *Mol Cell Biol* **18**, 4833-4843, (1998).

55 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).
56 Winter, M. J., Cirulli, V., Briaire-de Bruijn, I. H. & Litvinov, S. V. Cadherins are regulated by Ep-CAM via phosphaditylinositol-3 kinase. *Mol Cell Biochem* **302**, 19-26, (2007).
57 Gonzalez, B., Denzel, S., Mack, B., Conrad, M. & Gires, O. EpCAM is involved in maintenance of the murine embryonic stem cell phenotype. *Stem Cells* **27**, 1782-1791, (2009).
58 Nagao, K. *et al.* Abnormal placental development and early embryonic lethality in EpCAM-null mice. *PLoS One* **4**, e8543, (2009).
59 Baeuerle, P. A. & Gires, O. EpCAM (CD326) finding its role in cancer. *Br J Cancer* **96**, 417-423, (2007).
60 Osta, W. A. *et al.* EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. *Cancer Res* **64**, 5818-5824, (2004).
61 Cimino, A. *et al.* Epithelial cell adhesion molecule (EpCAM) is overexpressed in breast cancer metastases. *Breast Cancer Res Treat*, (2009).
62 Spizzo, G. *et al.* High Ep-CAM expression is associated with poor prognosis in node-positive breast cancer. *Breast Cancer Res Treat* **86**, 207-213, (2004).
63 Schmidt, M. *et al.* Prognostic effect of epithelial cell adhesion molecule overexpression in untreated node-negative breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 5849-5855, (2008).
64 Akita, H. *et al.* Ep-CAM is a significant prognostic factor in pancreatic cancer patients by suppressing cell activity. *Oncogene* **30**, 3468-3476, (2011).
65 Went, P. *et al.* Frequent high-level expression of the immunotherapeutic target Ep-CAM in colon, stomach, prostate and lung cancers. *Br J Cancer* **94**, 128-135, (2006).
66 Went, P. T. *et al.* Frequent EpCam protein expression in human carcinomas. *Hum Pathol* **35**, 122-128, (2004).
67 Kuhn, S. *et al.* A complex of EpCAM, claudin-7, CD44 variant isoforms, and tetraspanins promotes colorectal cancer progression. *Mol Cancer Res* **5**, 553-567, (2007).
68 Nubel, T. *et al.* Claudin-7 regulates EpCAM-mediated functions in tumor progression. *Mol Cancer Res* **7**, 285-299, (2009).
69 Munz, M. *et al.* The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation. *Oncogene* **23**, 5748-5758, (2004).
70 Pauli, C. *et al.* Tumor-specific glycosylation of the carcinoma-associated epithelial cell adhesion molecule EpCAM in head and neck carcinomas. *Cancer Lett* **193**, 25-32, (2003).
71 Munz, M., Fellinger, K., Hofmann, T., Schmitt, B. & Gires, O. Glycosylation is crucial for stability of tumour and cancer stem cell antigen EpCAM. *Frontiers in bioscience : a journal and virtual library* **13**, 5195-5201, (2008).

72 Maaser, K. & Borlak, J. A genome-wide expression analysis identifies a network of EpCAM-induced cell cycle regulators. *Br J Cancer* **99**, 1635-1643, (2008).
73 Denzel, S. *et al.* Initial activation of EpCAM cleavage via cell-to-cell contact. *BMC Cancer* **9**, 402, (2009).
74 Ralhan, R. *et al.* Nuclear and cytoplasmic accumulation of Ep-ICD is frequently detected in human epithelial cancers. *PLoS One* **5**, e14130, (2010).
75 Ralhan, R. *et al.* EpCAM nuclear localization identifies aggressive thyroid cancer and is a marker for poor prognosis. *BMC Cancer* **10**, 331, (2010).
76 Chaves-Perez, A. *et al.* EpCAM regulates cell cycle progression via control of cyclin D1 expression. *Oncogene*, (2012).
77 Hase, T. *et al.* Pivotal role of epithelial cell adhesion molecule in the survival of lung cancer cells. *Cancer Sci*, (2011).
78 Ng, V. Y., Ang, S. N., Chan, J. X. & Choo, A. B. Characterization of epithelial cell adhesion molecule as a surface marker on undifferentiated human embryonic stem cells. *Stem Cells* **28**, 29-35, (2010).
79 Lu, T. Y. *et al.* Epithelial cell adhesion molecule regulation is associated with the maintenance of the undifferentiated phenotype of human embryonic stem cells. *J Biol Chem* **285**, 8719-8732, (2010).
80 Huang, H. P. *et al.* Epithelial cell adhesion molecule (EpCAM) complex proteins promote transcription factor-mediated pluripotency reprogramming. *J Biol Chem* **286**, 33520-33532, (2011).
81 Imrich, S., Hachmeister, M. & Gires, O. EpCAM and its potential role in tumor-initiating cells. *Cell Adh Migr* **6**, 30-38, (2012).
82 Yoon, S. M. *et al.* Epithelial cell adhesion molecule (EpCAM) marks hepatocytes newly derived from stem/progenitor cells in humans. *Hepatology* **53**, 964-973, (2011).
83 Yamashita, T., Budhu, A., Forgues, M. & Wang, X. W. Activation of hepatic stem cell marker EpCAM by Wnt-beta-catenin signaling in hepatocellular carcinoma. *Cancer Res* **67**, 10831-10839, (2007).
84 Gires, O. *et al.* Tumor necrosis factor alpha negatively regulates the expression of the carcinoma-associated antigen epithelial cell adhesion molecule. *Cancer* **92**, 620-628, (2001).
85 Sankpal, N. V., Willman, M. W., Fleming, T. P., Mayfield, J. D. & Gillanders, W. E. Transcriptional repression of epithelial cell adhesion molecule contributes to p53 control of breast cancer invasion. *Cancer Res* **69**, 753-757, (2009).
86 McLaughlin, P. M. *et al.* The epithelial glycoprotein 2 (EGP-2) promoter-driven epithelial-specific expression of EGP-2 in transgenic mice: a new model to study carcinoma-directed immunotherapy. *Cancer research* **61**, 4105-4111, (2001).
87 Herlyn, M., Steplewski, Z., Herlyn, D. & Koprowski, H. Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proc Natl Acad Sci U S A* **76**, 1438-1442, (1979).
88 Riethmuller, G. *et al.* Randomised trial of monoclonal antibody for adjuvant therapy of resected Dukes' C colorectal carcinoma. German Cancer Aid 17-1A Study Group. *Lancet* **343**, 1177-1183, (1994).

89 Braun, S. *et al.* Monoclonal antibody therapy with edrecolomab in breast cancer patients: monitoring of elimination of disseminated cytokeratin-positive tumor cells in bone marrow. *Clin Cancer Res* **5**, 3999-4004, (1999).

90 Fields, A. L. *et al.* Adjuvant therapy with the monoclonal antibody Edrecolomab plus fluorouracil-based therapy does not improve overall survival of patients with stage III colon cancer. *J Clin Oncol* **27**, 1941-1947, (2009).

91 Naundorf, S. *et al.* In vitro and in vivo activity of MT201, a fully human monoclonal antibody for pancarcinoma treatment. *Int J Cancer* **100**, 101-110, (2002).

92 Schmidt, M. *et al.* An open-label, randomized phase II study of adecatumumab, a fully human anti-EpCAM antibody, as monotherapy in patients with metastatic breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **21**, 275-282, (2010).

93 Ruf, P. *et al.* Characterisation of the new EpCAM-specific antibody HO-3: implications for trifunctional antibody immunotherapy of cancer. *Br J Cancer* **97**, 315-321, (2007).

94 Heiss, M. M. *et al.* The trifunctional antibody catumaxomab for the treatment of malignant ascites due to epithelial cancer: Results of a prospective randomized phase II/III trial. *International journal of cancer. Journal international du cancer* **127**, 2209-2221, (2010).

95 Riethdorf, S. *et al.* Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13**, 920-928, (2007).

96 Criscitiello, C., Sotiriou, C. & Ignatiadis, M. Circulating tumor cells and emerging blood biomarkers in breast cancer. *Curr Opin Oncol* **22**, 552-558, (2010).

97 Konigsberg, R. *et al.* Detection of EpCAM positive and negative circulating tumor cells in metastatic breast cancer patients. *Acta Oncol* **50**, 700-710, (2011).

98 Raimondi, C. *et al.* Epithelial-mesenchymal transition and stemness features in circulating tumor cells from breast cancer patients. *Breast cancer research and treatment*, (2011).

99 Aktas, B. *et al.* Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients. *Breast cancer research : BCR* **11**, R46, (2009).

100 Gorges, T. M. *et al.* Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition. *BMC Cancer* **12**, 178, (2012).

101 Kasimir-Bauer, S., Hoffmann, O., Wallwiener, D., Kimmig, R. & Fehm, T. Expression of stem cell and epithelial-mesenchymal transition markers in primary breast cancer patients with circulating tumor cells. *Breast cancer research : BCR* **14**, R15, (2012).

102 Armstrong, A. J. *et al.* Circulating tumor cells from patients with advanced prostate and breast cancer display both epithelial and mesenchymal markers. *Molecular cancer research : MCR* **9**, 997-1007, (2011).

103 Biddle, A. *et al.* Cancer stem cells in squamous cell carcinoma switch between two distinct phenotypes that are preferentially migratory or proliferative. *Cancer research* **71**, 5317-5326, (2011).

104 Liu, S., Clouthier, S. G. & Wicha, M. S. Role of microRNAs in the regulation of breast cancer stem cells. *J Mammary Gland Biol Neoplasia* **17**, 15-21, (2012).

105 Gaiser, M. R. *et al.* Cancer-associated epithelial cell adhesion molecule (EpCAM; CD326) enables epidermal Langerhans cell motility and migration in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E889-897, (2012).

106 Sankpal, N. V., Mayfield, J. D., Willman, M. W., Fleming, T. P. & Gillanders, W. E. Activator protein 1 (AP-1) contributes to EpCAM-dependent breast cancer invasion. *Breast cancer research : BCR* **13**, R124, (2011).

107 Gosens, M. J., van Kempen, L. C., van de Velde, C. J., van Krieken, J. H. & Nagtegaal, I. D. Loss of membranous Ep-CAM in budding colorectal carcinoma cells. *Mod Pathol* **20**, 221-232, (2007).

108 Shin, K. J. *et al.* A single lentiviral vector platform for microRNA-based conditional RNA interference and coordinated transgene expression. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 13759-13764, (2006).

109 Stegmeier, F., Hu, G., Rickles, R. J., Hannon, G. J. & Elledge, S. J. A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13212-13217, (2005).

110 Gregory, R. I., Chendrimada, T. P. & Shiekhattar, R. MicroRNA biogenesis: isolation and characterization of the microprocessor complex. *Methods Mol Biol* **342**, 33-47, (2006).

111 Fuerer, C. & Nusse, R. Lentiviral vectors to probe and manipulate the Wnt signaling pathway. *PLoS One* **5**, e9370, (2010).

112 Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *Rna* **4**, 479-486, (1998).

113 Guy, C. T. *et al.* Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 10578-10582, (1992).

114 Sekine, S., Shibata, T., Sakamoto, M. & Hirohashi, S. Target disruption of the mutant beta-catenin gene in colon cancer cell line HCT116: preservation of its malignant phenotype. *Oncogene* **21**, 5906-5911, (2002).

115 Nakayama, G. R., Caton, M. C., Nova, M. P. & Parandoosh, Z. Assessment of the Alamar Blue assay for cellular growth and viability in vitro. *J Immunol Methods* **204**, 205-208, (1997).

116 Reynolds, B. A. & Weiss, S. Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell. *Dev Biol* **175**, 1-13, (1996).

117 Roy, L. D. *et al.* MUC1 enhances invasiveness of pancreatic cancer cells by inducing epithelial to mesenchymal transition. *Oncogene* **30**, 1449-1459, (2011).
118 Marcotte, R. & Muller, W. J. Signal transduction in transgenic mouse models of human breast cancer--implications for human breast cancer. *J Mammary Gland Biol Neoplasia* **13**, 323-335, (2008).
119 Patel, N. H. *et al.* Expression of engrailed proteins in arthropods, annelids, and chordates. *Cell* **58**, 955-968, (1989).
120 Shepherd, T. G., Kockeritz, L., Szrajber, M. R., Muller, W. J. & Hassell, J. A. The pea3 subfamily ets genes are required for HER2/Neu-mediated mammary oncogenesis. *Curr Biol* **11**, 1739-1748, (2001).
121 Boras-Granic, K., Chang, H., Grosschedl, R. & Hamel, P. A. Lef1 is required for the transition of Wnt signaling from mesenchymal to epithelial cells in the mouse embryonic mammary gland. *Dev Biol* **295**, 219-231, (2006).
122 Guttilla, I. K. *et al.* Prolonged mammosphere culture of MCF-7 cells induces an EMT and repression of the estrogen receptor by microRNAs. *Breast cancer research and treatment*, (2011).
123 Cioce, M. *et al.* Mammosphere-forming cells from breast cancer cell lines as a tool for the identification of CSC-like- and early progenitor-targeting drugs. *Cell Cycle* **9**, 2878-2887, (2010).

[i] For more information visit the Open Biosystems website at:
https://www.openbiosystems.com/RNAi/shRNAmirLibraries/GIPZLentiviralshRNAmir/
[ii] For more information visit the Invitrogen website at:
http://tools.invitrogen.com/content/sfs/manuals/virapower_lentiviral_system_man.pdf
[iii] Unpublished work done by Bonnie Bojovic, Centre for Functional Genomics, McMaster Universoty.
[iv] Unpublished work done by Allison Nixon, Centre for Functional Genomics, McMaster University.
[v] Unpublished work done by Robin Hallett, Centre for Functional Genomics, McMaster University.