PhD degree in Molecular Medicine

European School of Molecular Medicine (SEMM)

University of Milan and University of Naples "Federico II"

Faculty of Medicine

A Novel MicroRNA Family as Molecular Determinant in Mammary Stem Cells

Chiara Tordonato

IEO European Institute of Oncology, Milan

27-11-2014

Director of studies: Prof. Pier Paolo Di Fiore, M.D. PhD.
Department of Experimental Oncology
European Institute of Oncology, Milan
IFOM-Firc Institute of Molecular Oncology, Milan

Internal Supervisor: Dr. Giacchino Natoli, M.D Ph.D.
IEO European Institute of Oncology, Milan

External Supervisor: Dr David M. Livingston, M.D Ph.D.
Dana-Farber Cancer Institute, Boston

Academic Year 2014-2015

List of Figures and Tables

List of Abbreviations

Nomenclature

miRNA: microRNA
pre-miRNA: precursor microRNA
pri-miRNA: primary microRNA
RISC: RNA-induced silencing complex
RIP: RNA immunoprecipitation
UTR: untraslated region
MRE: microRNA responsive element
KD: Knock Down
KO: Knock Out
OE: Over Expression
TEBs Terminal End Buds
TLDUs: Terminal Lobular-Ductal structures
CD44: Cluster of Differentiation 44
CD24: Cluster of Differentiation 24
CD49f: Integrin subunit α6
CD29: Integrin β1
CD61: Integrin β3
Sca-1: Stem cell antigen
Lin-: Lineage negative
ER: Estrogen Receptor
PgR: Progesterone Receptor
K5: Cytokeratin 5
K8: Cytokeratin 8
E- Cad: E-cadherin
HER2: Human Epidermal growth factor Receptor 2
TIC: Tumor Initiating Cell
MRU: Mammary Repopulating Unit
MaCFC: Mammary Colony Forming Cells
MaSCs: Mammary Stem Cells
CSC: Cancer Stem Cell
SC: Stem Cell
BCSCs: Breast Cancer Stem Cell
SCD: Symmetric Cell Division
ACD: Asymmetric Cell Division
MMTV: Mouse Mammary Tumour Virus
EMT: Epithelial to Mesechimal Transition
ΔNGFR: Deleted Nerve Growth Factor
SCR: Scrambled
CTRL: Control
FVB: Friend Virus B mice
NSG: NOD Scid Gamma mice
MEC: Mammary Epithelial Cells
PDX: Patient-Derived Xenografts
TCGA: The Cancer Genome Atlas
PCA: Principal Component Analysis
GSEA: Gene-Set Enrichment Analysis
IPA: Ingenuity Pathway Analysis
TRE: Transcriptional Responsive Element

GFP: Green Fluorescent Protein
RFP Red Fluorescent Protein
LNA: Locked Nucleic Acid
RLU: Relative Luminescence Units
DAPI: 4',6-Diamidino-2-Phenylindole
DMEM: Dulbecco's Modified Eagle Medium
EGF: Epidermal Growth Factor
FGF: Fibroblast Growth Factor
FBS: Fetal Bovine Serum
MEBM: Mammary Epithelial Basal Medium
IEO: European Institute of Oncology
PBS: Phosphate buffered salin
WB: Western Blot
WT: Wild Type

ABSTRACT

A novel microRNA family as molecular determinant in mammary stem cells

MicroRNAs (miRNAs) are an evolutionarily conserved class of small (18-22 nucleotides) noncoding RNAs involved in the regulation of a variety of cellular and developmental processes. MiRNAs have recently emerged as key regulators of transcriptional programs that control self-renewal and the cell-fate of stem cells (SCs). In the breast compartment, forced expression of some miRNAs (e.g. miR-200, let-7 and miR-93) was shown to inhibit normal and/or cancer SCs by silencing self-renewal determinants. In fact, these miRNAs are poorly or not expressed in the SC compartment, and their stem-inhibiting function is mainly achieved through the induction of differentiation.

The aim of this project is to identify miRNAs specifically expressed in mammary SCs isolated from normal and cancer samples that could act as markers and/or as regulators of stem cell biology. We will also evaluate their potential as novel therapeutic targets in breast cancer.

Using an innovative technique developed in our lab for the isolation of quasi-pure SCs/CSCs, we identified a miRNA family as being highly enriched in the SC compartment, in both primary normal and tumor samples. Endogenous levels of miRNAs of this family can stratify distinct sub-populations with different SC abilities, and in human cancer high miRNA levels correlate with a basal tumor subtype and with an elevated CSC content. Furthermore, depletion of these miRNAs by a synthetic sponge impair self-renewal of SCs/CSCs, as measured by a serial mammosphere propagation assay in human cancer cell lines and in mouse primary epithelial cells. *In vivo*, miRNA loss reduced the frequency of tumor formation, as well as the mean tumor volume and the frequency of CSCs measured with either an established human basal-like cancer cell line or human patient-derived xenografts (PDXs). To determine the molecular mechanisms underlying

the regulation of SCs/CSCs behavior, we employed an unbiased approach assessing the effects of miRNA loss on global gene expression profile, coupled with the prediction of miRNA targets involved in stem cell biology. We identified 91 stem cell genes that are putatively targeted by this miRNA family, belonging to pathways critically involved in the regulation or in the mainteinance of stem cell traits.

In conclusion, our results suggest that this SC-specific miRNA family could potentially be used as novel therapeutic targets for breast cancer treatment. This study also provides insights into the mechanisms that sustain CSCs growth in the breast, involving complex molecular circuits of coding as well as noncoding RNAs.

1 Introduction

1.1 microRNAs

MicroRNAs (miRNAs) are small, non coding RNAs, approximately 20-22 nt long, that negatively regulate gene expression by base-pairing with the 3' untranslated region (3' UTRs) of target messenger RNAs (mRNAs) (Bartel, 2009). Since the discovery in 1993 of the first two small regulatory RNAs, lin-4 and let-7, able to control the timing of C. elegans larval development (Lee et al., 1993; Reinhart et al., 2000), thousands of different miRNAs have been identified in different organisms including plants, animals and viruses (Lagos-Quintana et al., 2001). According to the miRBase database (release 21), for humans alone, there are a total of 2588 unique sequences of mature miRNAs (Kozomara and Griffiths-Jones, 2014).

MiRNAs are involved in the regulation of a great variety of physiological processes such as the cell cycle, differentiation, apoptosis and development. Therefore, it is not surprising to discover their involvement in pathological conditions, such as cancer (Calin et al., 2004; Ventura and Jacks, 2009).

In the first part of the introduction, I will discuss the biology of miRNAs. In the second part, I will focus on the description of miRNAs as critical regulators of stem cell (SC) biology and on the role of miR-146 in human physiology and pathology, in particular of the breast.

1.1.1. Biogenesis: transcription and nuclear processing

MiRNAs loci are located both in exonic and intronic regions of protein coding genes. They are usually transcribed in parallel with their host genes (intragenic miRNAs), or in some cases they are located, as independent transcriptional units, in intergenic regions or gene deserts (Garzon et al., 2009). Regarding the genomic organization, miRNAs are

occasionally organized in single transcription units called “miRNA clusters”. By definition, these miRNAs are located in close proximity to each other in the genome (<10kb), share the same promoter and usually include two or more miRNA genes: for example, the human miR-17-92 cluster comprises 6 miRNA genes (miR-17, 18a, 19a, 20a, 19b-1 and 92a-1).

MiRNAs were numbered progressively as they were discovered. Generally, miRNAs that display a very high sequence homology share the same number with different letter suffixes (e.g.. miR-146a and miR-146b differ by two bases). Instead, miRNAs that possess the same mature sequence but are derived from different primary transcripts (and different genetic loci) have numbered suffixes (e.g. miR-24-1 and miR-24-2). Lastly, the suffix 5p or 3p is given when mature miRNAa are derived, respectively, from the 5’-3’ or 3’-5’ arm of the precursor miRNA.

MiRNAs are mostly transcribed by RNA polymerase II to generate a stem-loop containing primary miRNA (pri-miRNA), which can range in size from hundreds of nucleotides to tens of kilobases. The pri-miRNA is often polyadenylated and capped, as is standard for pol II transcripts. However, miRNAs with Alu repeats may also be transcribed by RNA pol III (Borchert et al., 2006). Intronic miRNAs are transcribed with their host genes, so their maturation is associated with mRNA splicing.

The current model of maturation of intronic and exonic miRNAs is shown in Figure 1.

After transcription, the pri-miRNA undergoes two subsequent steps of maturation. The first one occurs in the nucleus and generates a ~ 70-100 nt long hairpin RNA molecule, called a precursor miRNAs (“pre-miRNA”). The second one occurs in the cytosol to generate the final 18-24 nt long RNA duplex. The first cleavage is performed by Miprocessor, a nuclear multiprotein complex composed of Drosha and DGCR8. This complex measures the distance from the ss-dsRNA duplex, generating a stem-loop pre-miRNAs of ~ 70 nt. Drosha is a member of the RNase III endonuclease family,

characterized by two tandem RNase III domains (RIIID) that constitute the catalytic site of the protein. It is essential for processing the pri-miRNA and generating a pre-miRNA stem loop with 5' phosphate and a ~2 nt 3' overhang (Han et al., 2006), characteristic of RNase III-mediated cleavage. Furthermore, DGCR8, with its two double-strand RNA binding domains (dsRBD), helps the endonuclease to find the right position for cleavage, recognizing the ssRNA-dsRNA junction at the base of the hairpin (Han et al., 2006).

After nuclear processing, Exportin 5 translocates the pre-miRNA from the nucleus to the cytoplasm. The ~2 nt 3' overhang generated by Drosha cleavage is critical for the export, as it is recognized by Exportin 5 that, in turn, releases the pre-miRNA in the cytoplasm (Lund et al., 2004).

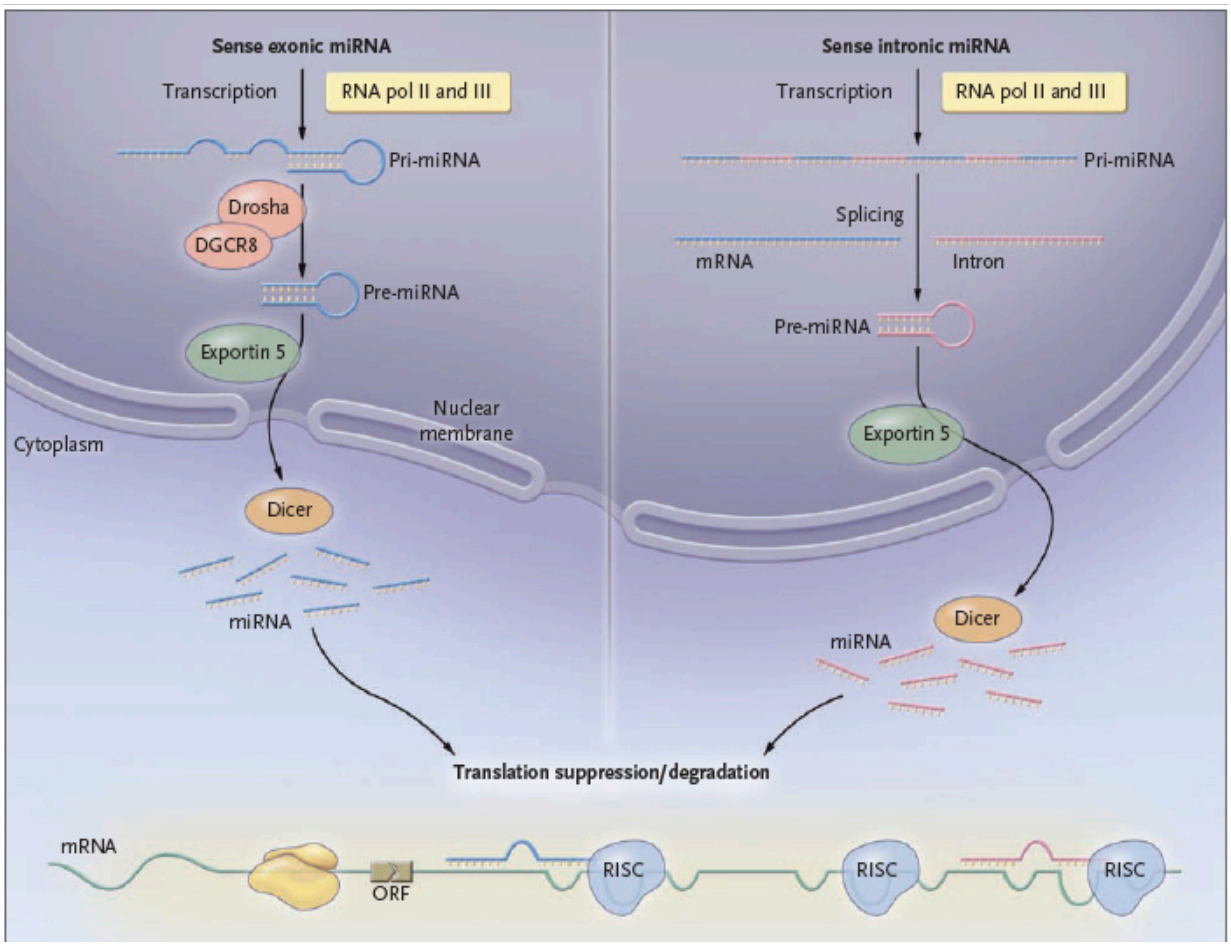


Figure 1- Schematic representation of mechanisms involved in the transcription and maturation of exonic and intronic miRNAs- from (Croce, 2008) ©N Engl J Med.

1.1.2. Cytoplasmic processing

In the cytoplasm, the pre-miRNA is cleaved by Dicer, in complex with TRBP, to generate the mature miRNA. Dicer is an RNase III enzyme (as is Drosha), which cuts away the terminal base pair and the stem loop of the pre-miRNA, leaving another 5' phosphate and a ~2 nt 3' overhang, thus generating a mature ~22 nt miRNA:miRNA* duplex. Overall, the "nuclear cut" of Drosha defines one end of the mature miRNA, while the "cytoplasmic cut" by Dicer defines the other end. It has been demonstrated that the specificity of Drosha during the first cut influences the correct maturation of the precursor and defines both the ends of the mature miRNAs (Lee et al., 2003). TRBP, instead, is involved in the "dicing" event by interacting with N-terminal domain of Dicer and in recruiting the Argonaute protein 2 (Ago2). In this way, the duplex is ready to be loaded into the RNA-induced silencing complex (RISC), as shown in Figure 2, (Winter et al., 2009). In principle, each miRNA:miRNA* duplex could give rise to two different mature miRNAs, but only one of these is correctly loaded on the RISC complex and guides the complex to the target mRNAs. For strand selection, the critical parameter seems to be thermodynamic asymmetry of the duplex. As Ago unwinds the duplex in 3'-5' direction, the miRNA strand with lower stability of base pairing at its 5' end is typically loaded into RISC: this strand is called the "guide strand". Instead, the other strand, defined as the "passenger strand", is more stable and is typically degraded.

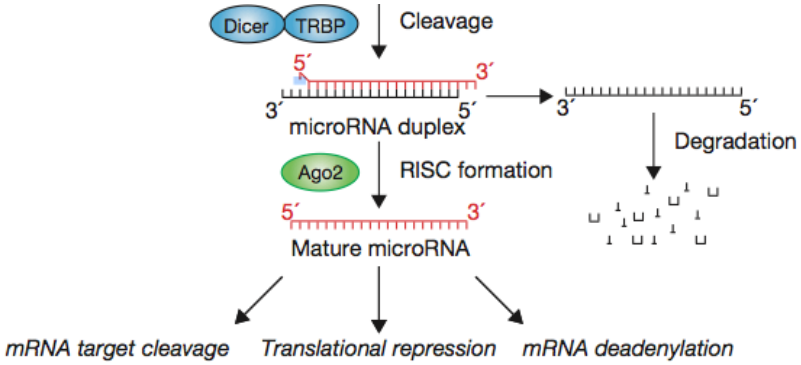


Figure 2- Schematic representation, of the cytoplasmic processing that leads to formation of mature miRNA and loading on the RISC complex- from (Winter et al., 2009), ©Nature press.

1.1.3. Target recognition of miRNAs and target prediction algorithms

The critical determinant of miRNA specificity is the region called "seed" sequence, located in position 2-8 nt at the 5' end of the miRNA. This region is directly involved in Watson-Crick base pairing to the miRNA responsive element (MRE), usually located in the 3' untranslated region (UTR) of the target messenger RNA. According to the number of complementary bases, the seed match can be classified as an 8-mer, a 7-mer or a 6-mer, represented in Figure 3. The degree of sequence complementarity between a mature miRNA and its target mRNA seems also to direct the mechanism of down-modulation of the target. In cases of perfect, or near-perfect, complementarity, mRNAs targets are sliced and subsequently degraded. Ago2 protein, in the RISC complex, is the only subunit in the humans that has the "slicing" activity, cleaving at the nucleotides paired with residues 10-11 of the miRNA, leading to the mRNA degradation. In this case, the miRNA usually remains loaded on the RISC complex and can subsequently guide degradation of additional messengers. When a target matches the miRNA imperfectly, the target down-regulation occurs by multiple mechanisms (Figure 4) involving translational repression of mRNA (inhibition of the cap recognition or the 60S recruitment stage, ribosome drop-off and increasing termination efficiency), or mRNA destabilization (through deadenylation or decapping of the mRNA) (Carthew and Sontheimer, 2009). The ability of miRNAs to interact with targets with different degrees of complementarity, indicated that miRNAs could interact with several targets simultaneously.

Indeed, such is the complexity that target prediction programs are necessary to unravel miRNA:target interactions through computational analysis. Since the discovery of the first algorithm, five tools have been developed for the identification of predicted targets: miRanda, Targetscan, Pictar, DIANA miT and RNAhybrid. All of them share common rules, such as the presence of seed- matched elements, the grade of MRE conservation among different species and the stability of miRNA:target interaction (Maziere and

Enright, 2007). The most widely used algorithms are miRanda and Targetscan. Both were also used in the analyses reported in this thesis and, therefore, will be discussed in details. miRanda identifies regions of complementarity in the 3'UTR of the target gene without requiring an absolute perfect complementarity. It keeps into account the free energy binding contribution and the number of seed-matched sites in the mRNAs. All these parameters contribute to the strength of the miRNA:mRNA predicted interaction measured as a score (miRsvr score). The lower the score, the stronger the predicted interaction. However, miRanda has a high rate of false positve prediction (30-40%).

Targetscan predictions are instead more stringent and usually few. Targetscan searches for biological targets keeping into account 8-mer and 7-mer sites, conserved across species, that bind with perfect complementarity to the MRE in the 3'UTR of mRNAs and subsequently extending this region of complementarity outside the seed. This process aims to filter out false positives as much as possible during the prediction process, thus reducing the estimated false positive rate range to between 20% and 30%. Targetscan calculates a "context score" as measurement of the likelihood of a mRNA to be repressed by a selected miRNA. Unfortunately, many true targets are discarded by the stringent criteria of this algorithm.

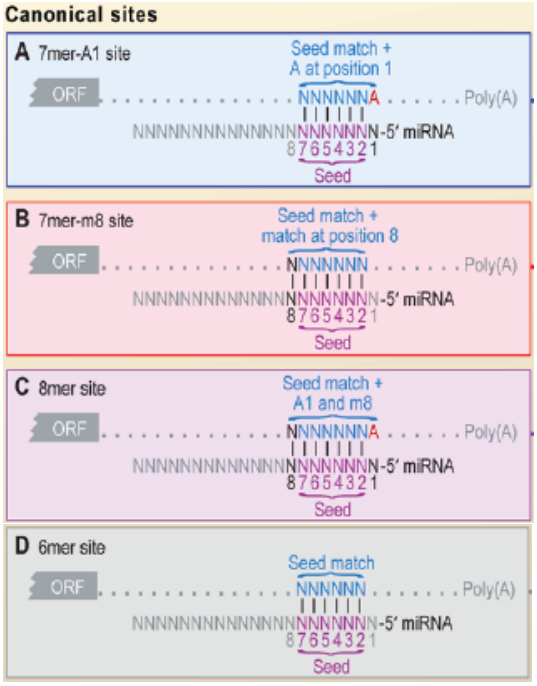


Figure 3- Representation of canonical sites for miRNA binding to target mRNAs – reproduced from (Bartel, 2009) ©Cell press.

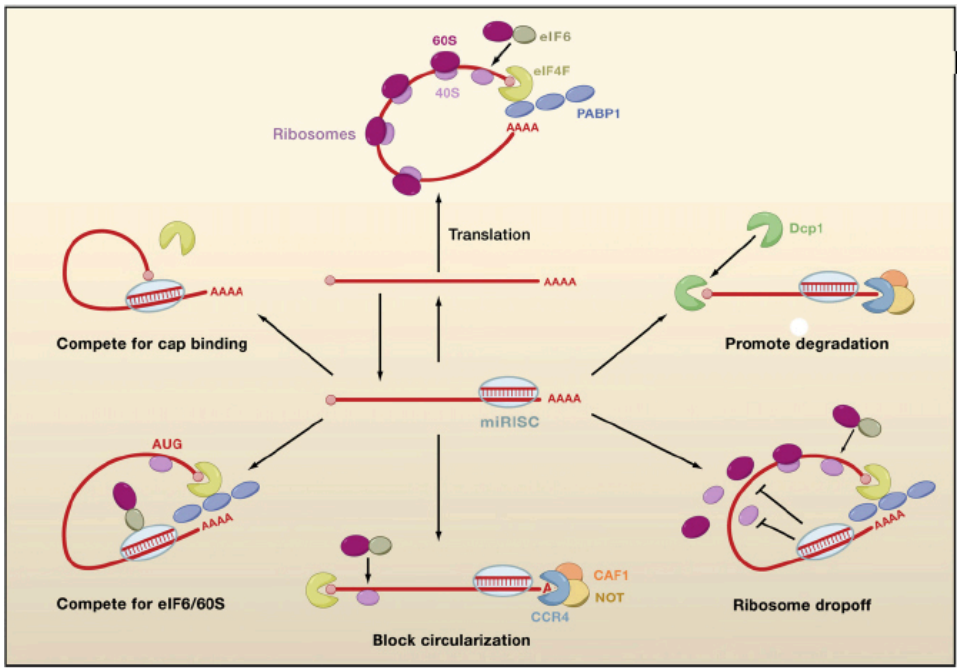


Figure 4- Schematic representation of miRISC-mediated repression on target genes- from (Carthew and Sontheimer, 2009) ©Cell press.

1.1.4. Regulatory functions of miRNAs

Two different modes of target regulation have been observed. miRNAs can act as molecular switches, modulating one specific key target to a level below which it has any biological activity in the cell. Alternatively, they can function as fine tuners, targeting many genes of the same pathway simultaneously, causing target fluctuations and subtle phenotypic effects (Sevignani et al., 2006).

An example of molecular switch is the interaction between miR-200 and ZEB1. ZEB1 is a transcriptional factor that controls the activation of an "epithelial to mesenchymal transition (EMT)" program, physiologically active during embryonic development. ZEB1 can suppress the transcription of all the members of miR-200 family, which are organized in two clusters (miR-200a-200b-429 and miR-200c-141). The miR-200 family members can target ZEB1 itself, generating a mutual negative feedback loop in which the two players are never in equilibrium. This mechanism ensures that once the cell activates a gene expression program towards an epithelial cell state (balanced in favor of miR-200) or a mesenchymal cell state (balanced in favor of ZEB1), the decision is maintained and reinforced over time (Wellner et al., 2009). A similar case occurs within the miR-203/p63 axis in epithelial tissues. Stem cells of the skin express high levels of p63, an essential regulator of SC maintenance, together with low level of miR-203. During skin differentiation, miR-203 levels are strongly induced, leading to the downregulation of p63, and thus promoting cycle exit and differentiation. Hence, by tight control of the p63 SC gene, miR-203 acts as a molecular switch between proliferation and differentiation (Yi et al., 2008).

"Fine-tuning interactions" are the ones in which, rather than causing a binary off/on switch of a single master gene, the miRNA reduces multiple targets to optimal cell levels, thus increasing the robustness of biological circuits. For example, miR-24 inhibits cell proliferation by shutting down multiple cell cycle genes directly regulated by Myc and

E2F2 (Lal et al., 2009). Similarly, in our group, Marzi *et al*. have demonstrated that a panel of miRNAs induced during myotubes differentiation, and defined as being "differentiation-associated' miRNAs, can antagonize the expression of cell cycle genes of the Rb-E2F pathway. In this case, the authors provided evidences of a co-targeting mechanism involving multiple co-regulated miRNAs (induced during differentiation) and different genes belonging to the same pathway (G1/S transition) (Marzi et al., 2012).

1.1.5. Genome-wide approaches to dissect miRNAs function

As previously mentioned, miRNAs can repress hundreds of target genes by decreasing the efficiency of protein translation and/or by reducing mRNA levels. Hence, reasonable genome-wide approaches useful for identifying miRNA targets could include: i) expression profiling analysis of mRNAs and/or proteins upon miRNA modulation and ii) biochemical methods based on the purification of the miRISC complex and the physical interaction between miRNAs and their targets. A schematic representation of these approaches is illustrated in Figure 5.

Lim *et al.* performed the first study of gene expression profile, by microarray analysis, upon muscle-specific miR-1 and brain-specific miR-124 overexpression into Hela cells (Lee P. Lim, 2005). More than 100 mRNAs were downregulated in response to miRNAs overexpression and the vast majority of modulated targets contained MREs for the transfected miRNAs in their 3'UTR. Alternatively, proteomic approaches can be used to fish out miRNA targets by measuring the protein output upon either overexpression or knockdown of miRNAs. One preferred technique is SILAC (stable isotope labeling by amino acids in cell culture) or its variant pSILAC (pulsed SILAC), in which proteins are metabolically labeled and two populations, one that is grown in control conditions and one in which a miRNA has been overexpressed or knocked down are analyzed by mass spectrometry. Two influential publications by Baek and Selbach analyzed the translational

effects of miRNA over expression or miRNA depletion. The authors found that hundreds of proteins were concordantly down- or up- regulated, albeit with mild effects (mean log2fold changes $|0.2\text{-}0.3|$) (Baek et al., 2008; Selbach et al., 2008). By searching for nucleotide motifs in the 3'UTR of mRNAs that correlated with protein changes, they further showed that the presence of a seed accounted for ~70% of the total number of regulated proteins. When the repression of protein translation was more pronounced, a decrease in the corresponding mRNA was concomitantly observed through Affymetrix profiling. Recently, in a study performed in Bartel's lab, the effects of overexpression of two miRNAs (miR-1 and miR-155) were analyzed at both mRNA level and at protein translation through the elegant technique of ribosome profiling (Guo et al., 2010a). In both cases, the predominant effect mediated by miRNAs was the lowering of mRNAs levels that in turn accounted for more than 80% of the decreased protein levels (Guo et al., 2010a). These results reinforced the notion that most of the changes in protein output induced by miRNAs could be inferred by analyzing the mRNA levels.

Biochemical approaches rely on the physical interaction between the miRISC and the target mRNAs. Therefore, immuno-precipitation of the miRISC through one of its subunit (e.g. Ago2) allows the co-purification of any associated mRNAs (RNA immuno-precipitation) and subsequent analysis of their mRNA levels through Affymetrix (Ago-RIP-Chip) or via deep sequencing (Ago-RIP-seq) (Karginov et al., 2007).

The main advantage of this technique is the identification of a large class of target mRNAs whose levels are not changed upon miRNA modulation and that would be missed using other approaches of target identification. Nevertheless, some interactions identified by Ago-RIP are aspecific or artifacts of miR-modulation. Hence, recently a more complex experimental approach, known as HITS-CLIP or PAR-CLIP, was developed to enrich for the direct mRNA:miRNA interactions occurring at the miRISC (Chi et al., 2009; Hafner et al., 2010). HITS-CLIP is based on RNA crosslinking to RNA-binding protein (RBPs)

through UV irradiation. This allows the immunoprecipitation of RBPs in very stringent conditions which, together with the digestion of unbound RNAs, eliminates indirect and aspecific interactions, leaving only RNA fragments that are bound to the RBP and the miRNA, and that can be identified by RNA sequencing (HITS-CLIP) (Chi et al., 2009). In the PAR-CLIP, a photoactivatable ribonucleoside is incorporated into nascent RNAs and UV irradiation causes the crosslinking of photo-reactive nucleoside with RBPs. The crosslinked RNAs/RBPs complexes were immunoprecipitated and subsequently analyzed by sequencing (Hafner et al., 2010). Although very challenging, these approaches are extremely insightful and have provided fundamental experimental evidences on the mechanisms of miRNA:mRNA interaction, highlighting for example that Ago-binding sites are often located in ORFs and not only in the 3' UTRs of mRNAs (Leung et al., 2011).

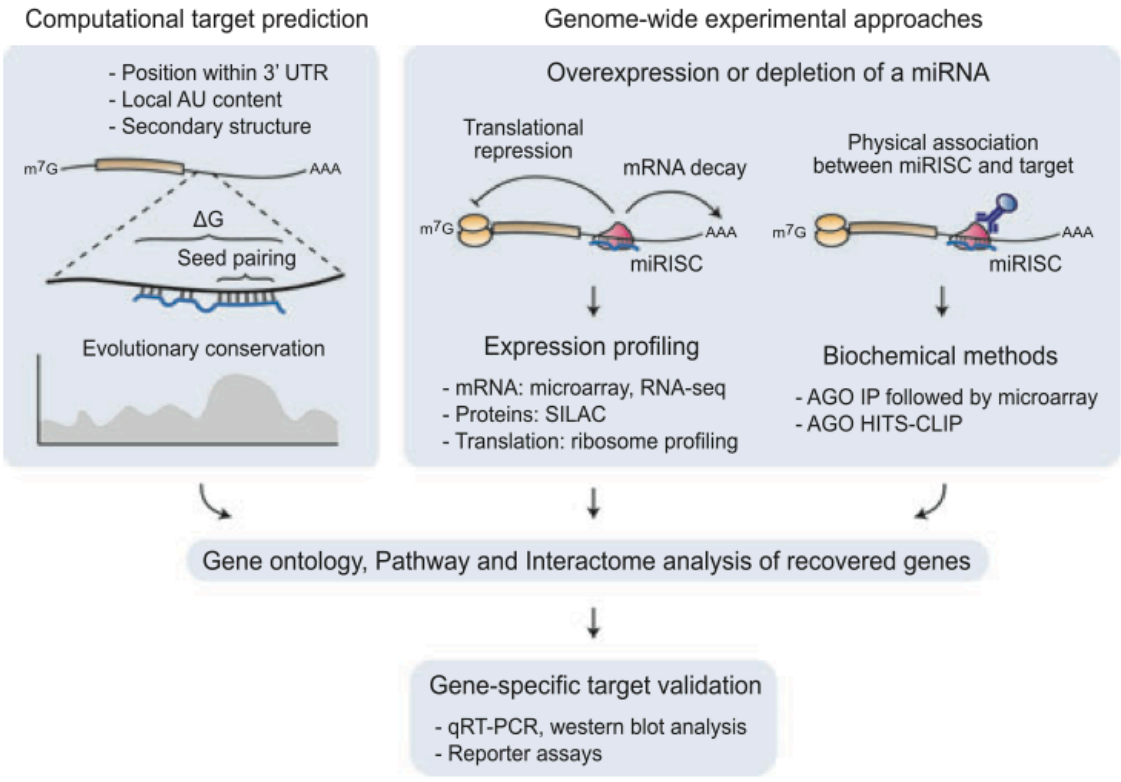


Figure 5- Representation of the current genome-wide approaches to miRNAs target identification- from (Lee and Shin, 2012) ©Annals of the New York Academy of Sciences.

1.2. Mammary stem cells and cancer stem cells

1.2.1. Mammary gland development and mammary stem cells

The mammary gland is a highly dynamic organ, composed of ducts that terminate in buds or lobular structures (TLDUs, terminal lobular-ductal units). TLDUs contain two main cellular types: inner luminal cells that reside around the central lumen, expressing cytokeratins 8 and 18, and outer basal-myoepithelial cells, in direct contact with the basement membrane, which express keratins 5, 14 and smooth muscle actin (SMA). TLDUs are surrounded by a stroma, called mammary fat pad, composed of adipose cells, fibroblasts, blood vessels and nerve terminals.

The mammary epithelium undergoes several important changes during its development: at the embryonic stage, the fetal human breast is derived from an epithelial structure called the primordium that, at 21 weeks, starts to migrate toward the mesenchyme (Jolicoeur, 2005). At the neonatal stage, terminal lobular units and end buds appear; these structures remain immature until puberty, when a series of growth changes occurred, triggered by sex hormones signaling, such as ducts elongation and side-branching. During pregnancy and lactation, the human breast ducts involute, the luminal cells proliferate and differentiate into milk-producing alveoli surrounded by the basal myoepithelial cells that contract to facilitate milk release (Figure 6) (Gjorevski and Nelson, 2011; Howard, 2000). Remarkably, the mammary gland maintains its ability to perform structural remodeling during pregnancy-lactation and involution for several cycles, suggesting the existence of a reservoir able to sustain multiple rounds of this process to generate all the cellular lineages that composes the mammary gland. Pioneering experiments in mice using transplantation of entire section of murine mammary glands (Deome et al, 1959), or even single epithelial FACS-sorted cells (Stingl et al., 2006) into cleared fat pads, proved the existence of a multipotent population of rare cells, called mammary stem cells (MaSCs).

MaSCs possess a number of peculiar characteristics: i) they can regenerate a complete ductal lobular outgrowth, composed of both luminal and basal lineages upon limiting dilution transplantation experiment, slowly becoming the gold standard experiment for MaSCs detection (Stingl et al., 2006; Shackleton et al., 2006); ii) they are able to self-renew, usually through an asymmetric type of cell division that generates one stem cell and one progenitor, and maintains unaltered the stem cell pool; iii) they are quiescent or slowly dividing; and lastly iv) they can withstand anoikis, surviving in anchorage-independent growth.

A great step forward in mammary SC biology came with the work of Gabriella Dontu, who developed an *in vitro* methodology for testing SC self-renewal, called the mammosphere assay (Dontu et al., 2003). Briefly, stem cells maintained in an undifferentiated medium, in anchorage independent conditions, can grow as clonal spheroids, composed of quiescent SCs, progenitors and differentiated cells, representing a tool for enrich *in vitro* a defined cell population for stem cells. Importantly, under these culture conditions, MaSCs maintain their properties: i) they can self-renew, in fact, MaSCs can form a successive generation of spheroids after the dissociation of mammospheres; and ii) they retain the ability to generate differentiated progeny. In fact, under differentiating conditions, MaSCs can form luminal-like and myoepithelial-like mixed colonies, and remarkably, in the presence of prolactin, they can generate acinar structures positive for β-casein, representing one of the first documentation of human breast differentiation *ex vivo* (Dontu et al., 2003).

A number of studies have focused on cell surface markers for the prospective isolation of MaSCs and for their biological characterization. These studies have shown that murine MaSCs reside in the basal/myoepithelial layer of terminal bud end (TEBs, Figure 7) express low levels of Sca-1, CD24 (heat-stable antigen) and high levels of CD49f (α6 integrin) or CD29 (β1 integrin) (Sca-1^{neg}/ CD24$^{med/+}$CD49f^{hi}/ CD29^{hi}) (Stingl et al., 2006)

(Shackleton et al., 2006), and serially transplant *in vivo* (mammary repopulating units or MRUs).

Similarly, several markers in the human breast are associated with MaSCs. These have been shown to be CD49f^{hi}/ Epcam$^{low/neg}$ basal cells (Eirew et al., 2008), although more recent work has identified bipotent progenitors with repopulating properties in both luminal and basal lineages (Keller PJ, 2012). Other groups have exploited the quiescent/slowly dividing nature of MaSCs, to isolate MaSCs from a mammosphere cultures in absence of cell-surface staining. A quasi-pure population of MaSCs was obtained from human and mouse tissues, through the ability of SCs to retain a lipophilic fluorescent dye (PKH26), which is instead diluted upon proliferation of progenitors and differentiated cells (Cicalese et al., 2009; Pece et al., 2010). In both cases, PKH^{pos} MaSCs isolated from mammosphere cultures were able to regenerate the mammary gland outgrowths by fat pad transplantation experiments.

A very recent approach for assessing MaSCs function *in vivo* is based on lineage-tracing experiments, which allow the *in situ* tracking of stem cell/progenitor and the evaluation of their contribution to normal physiology and development. In the first study, based on the use of both basal (K14/K5) or luminal (K18/K8) keratins-inducible mouse models, was drawn a scenario in which unipotent, rather than multipotent, lineage-restricted mammary stem cell drives the development and the expansion of the mammary gland in puberty and adulthood (Van Keymeulen et al., 2011). The recent tracking studies of bipotent mammary stem cells *in vivo* partially confute this preliminary finding. In fact, the use of two independent systems based on Axin2Cre^{ERT2} on one side (van Amerongen et al., 2012) or a multicolor Confetti reporter coupled with a Elf5-inducible system on the other side (Rios et al., 2014), highlights the existence of long-lived multipotent MaSCs that contribute to all the major morphogenetic changes of the postnatal and adult mammary gland.

More recently, an independent group identified the Wnt-responsive gene Procr (protein C receptor) as uniquely expressed in murine multipotent MaSCs, as assessed by both lineage tracing and by transplantation assay (Wang et al., 2014). $Procr^{+}$ MaSCs express low levels of cytokeratins and claudins and high levels of mesenchymal markers, and are located at the top of developmental hierarchy, supporting the model that unipotent and multipotent MaSCs could coexist in the adult mammary gland and reconciling the striking differences found in previous lineage tracing studies.

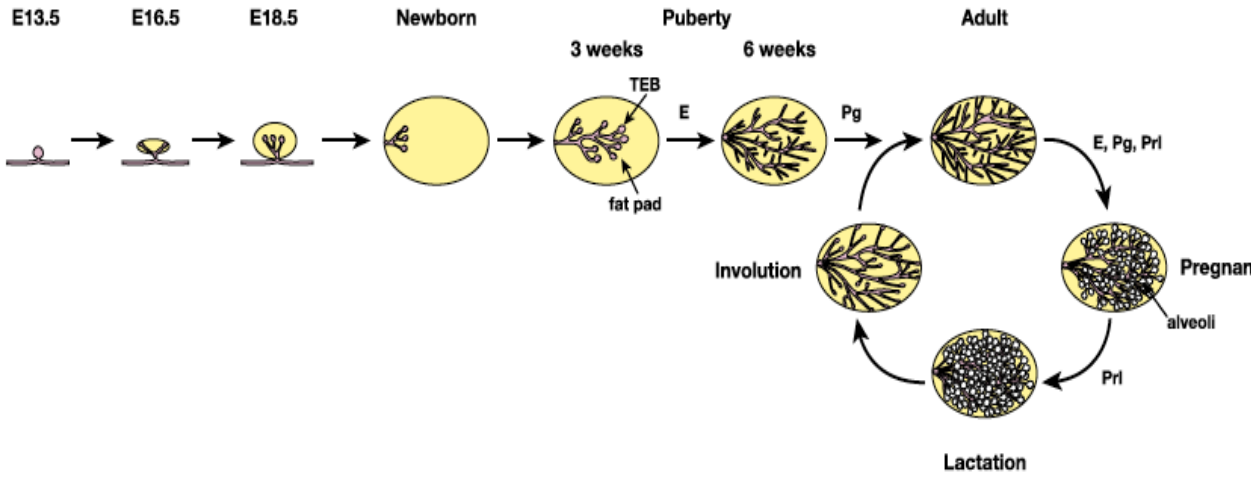


Figure 6- Mammary gland remodeling during murine development- the picture summarizes all the stages of mammary gland development from the embryo to the adult, quite similar to human development, from (Visvader and Stingl, 2014) ©Genes & development.

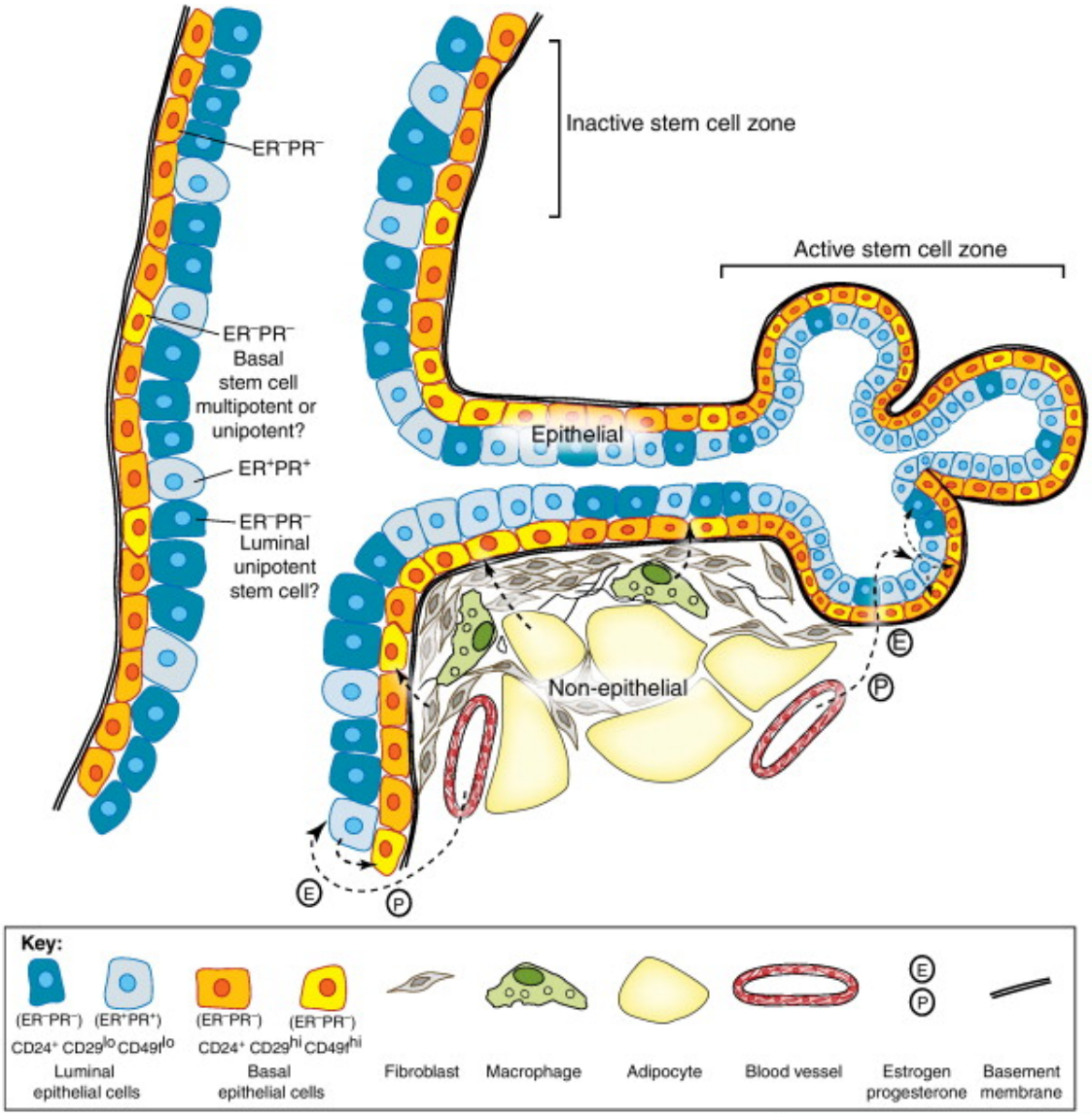


Figure 7- Schematic representation of the MaSC niche- the niche, which includes epithelial and non-epithelial cells, participates in the modulation of SC state. Estrogen and progesterone act on ER+/PR+ luminal cells, that, in turn, transmits the hormonal signal to ER-/PR- MaSCs. Non-epithelial cells include cells such as adipocytes, fibroblast and macrophages- from (Joshi et al., 2012) ©Trends in endocrinology and metabolism.

1.2.2. Breast cancers and breast cancer stem cells (BCSCs)

For decades, human breast tumors were classified histologically according to their hormonal receptor status, based on the immunoreactivity for estrogen receptor (ER), progesterone receptor (PgR), HER2 positivity and proliferative index, as measured by Ki67. On the basis of this characterization, tumors that lack hormonal receptor expression

(so-called triple negative tumors) and posses an high proliferative index (Ki67>10) are termed basal, and are generally associated with a worse prognosis and a more aggressive phenotype. On the contrary, tumors that express the estrogen receptor (ER+) are termed luminal and are associated with a better prognosis and lower recurrence rate. Recently, thanks to genome-wide approaches, both oncologists and biologists have realized that breast cancer is a very heterogeneous disease and have attempted to classify tumors according to their molecular characteristics. This has led to a classification of breast tumors into five different molecular subtypes, according to their peculiar gene expression profile (PAM50): Luminal A, luminal B, HER2 amplified, Basal and Claudin-low (Hu et al., 2006). This classification uncovered a molecular basis for breast tumor heterogeneity, highlighting the possibility that different tumors could arise from different cells-of-origin. For example, by comparing gene expression profiles, the claudin-low subtype emerged as the most similar to the basal/MaSCs signature, suggesting that this subtype could originate directly form SCs. Not surprisingly, Luminal A and Luminal B has the closest molecular profile to that of mature luminal cells. Figure 8 summarizes the possible connection existing between the so-called “cell of origin” and the different subtypes of human breast cancers. The spectrum of genetic alterations associated to breast tumors is also quite different among subtypes. For example, Rb mutations are mainly found in basal-like and Luminal B, but not in the other subtypes, whether p53 or PI3KCA mutations are common hallmarks for all the tumors. Therefore, the initial genetic lesion in concert with the different “cell of origin” could determine the huge diversity existing between the breast tumor molecular subtypes (Visvader and Stingl, 2014).

Indeed, the concept of tumor heterogeneity is further supported by the widely accepted existence of initiating cell (TIC) or cancer stem cell (CSCs). Breast cancer was the first solid tumor from which CSCs were identified and biologically characterized. Using the surface markers CD44 and CD24, Al Hajj *et al.*, demonstrated that as few as 100 cells

from the lin- $CD44^{+}/CD24^{-}$ population coming from 9 human primary tumors could generate tumors in immunocompromised mice, compared to thousands of non tumorigenic $CD44^{-}/CD24^{+}$ counterpart (Al-Hajj, 2003). Importantly, these TICs/CSCs were able to regenerate tumor heterogeneity, and could be serially passaged *in vivo*, both of which are representative features of the self-renewal properties of SCs.

Unfortunately, the question of whether a common tumor SC exists from which all the molecular subtypes are originated, or whether each subtype has its own tumor SC, remained unresolved. What is clear is that CSCs can be considered as "caricatures" of normal SCs, with self-renewing abilities *in vitro* and *in vivo* and aberrant differentiation potential. Of note, tumor recurrence and/or resistance to anti-cancer therapy seem to be linked to CSC number and properties (Visvader and Lindeman, 2008). Indeed, understanding the complex biology of the SCs and the intricate mechanisms that regulate self-renewal, will possibly be one of the most promising strategies to selectively target these rare cells and to definitively eradicate tumor potential for regrowth.

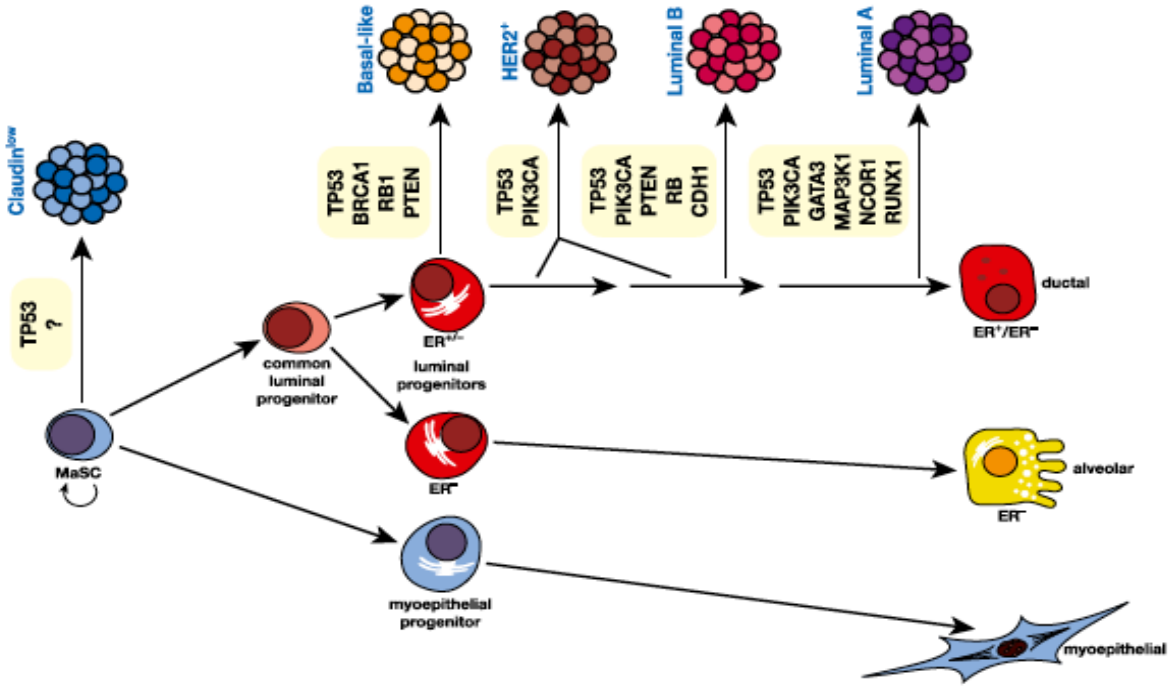


Figure 8- Schematic representation of breast epithelial cell hierarchy and its direct connection with different breast cancer subtypes- from (Visvader and Stingl, 2014) ©Genes & development.

1.3. Signaling pathways involved in SC and CSC maintenance

Several lines of evidence have demonstrated that ligand-dependent signaling pathways, such as Wnt, Notch, TGF-β and Nf-κB are critical regulators of SC biology and their deregulation could sustain oncogenic processes and BCSCs. In this section, I will briefly analyze each pathway in normal SC development and in TICs/CSCs pathogenesis.

1.3.1. Notch signaling

The Notch signaling pathway is initiated at the cell membrane through the interaction between Notch receptors and their ligands. In mammals, there are four different Notch receptors (Notch1-4) and five ligands classed as Delta-like proteins (DLL1-3) or Jagged proteins (Jag 1-2). Notch receptors are transmembrane proteins consisting of an extracellular domain involved in ligand binding and an intracellular domain (ICD) involved in signal transduction. Upon ligand binding, Notch receptor activation occurs through two proteolytic events; firstly, ADAM protease cleavage releases the extracellular portion of the receptor. Subsequently, γ-secretase cleaves the transmembrane portion, and allows the release of active intracellular domain (NCID) (Iwatsubo, 2004). The NCID translocates to the nucleus, where interacts with RBP-jk complex, thus activating the transcription of target genes including Hey and Hes family genes (Iso et al., 2003)

Notch signaling has been associated with normal development and SC self-renewal in a wide range of organisms (Fre et al., 2005). In the breast, Notch activation can produce up to a 10-fold increase in the mammosphere forming ability of human MaSCs, meaning that it affects stem cell self-renewal (Dontu et al., 2004; Dontu et al., 2003). Moreover, constitutive Notch signaling seems to commit cells toward differentiation, but whether Notch effect causes the commitment of MaSCs towards the luminal lineage (inhibiting SC self renewal) (Bouras et al., 2008; Raouf et al., 2008), or the commitment of mammary early progenitors toward myoepithelial differentiation (Dontu et al., 2004) is still a matter

of debate. Several studies have linked Notch signaling to breast tumorigenesis. An *in vivo* mouse model of constitutively active Notch4 showed an impairment of normal mammary gland morphogenesis, followed by mammary tumor development (Uyttendaele H, 1998). The same tumorigenic effects on the mammary gland were also observed for over-activation of Notch1 and Notch3 (Hu et al., 2006). In humans, high levels of Notch1 and Notch4 were detected in basal-like aggressive tumors (Speiser et al., 2012) and loss of Numb-dependent negative regulation of Notch (found in ≈ 50% of the 321 tumors analyzed) contributes to mammary tumorigenesis and correlates with a basal phenotype (Pece, 2004). Several attempts have been made to inhibit Notch signaling in breast CSCs: γ-secretase inhibitor (GSI) treatment and anti-Notch4 shRNA decreased the sphere forming efficiency *in vitro* as well as the number and volume of tumors *in vivo* (Harrison et al., 2010). On the other hands, GSI-induced Notch1 inhibition caused only a decrease in the volume and growth rate of tumors *in vivo,* suggesting a requirement for Notch4 in TIC expansion and a role for Notch1 in tumor growth and proliferation (Harrison et al., 2010). Taken together, it is clear that Notch signaling is involved in the development of the mammary gland and controls the expansion and maintenance of the CSC pool, nonetheless strategies that target Notch signaling through GSI treatment have yet to be improved, in view of severe side effects caused in other organs (e.g. the gastrointestinal system).

1.3.2. Wnt signaling

Wnt signaling is one of the main signal transduction pathways that control embryo development and, not surprisingly, is extensively involved in carcinogenesis. Wnt signaling can be activated through canonical and non canonical pathways: the non canonical pathway plays a role in cell movement, shape and polarity, but its role in tumorigenesis has not be elucidated; conversely, canonical Wnt signaling controls gene expression and is widely reported to play a role in tumorigenesis, thus we focused here

mainly on the canonical pathway. This network involves 19 Wnt ligands, 10 Frizzled (Fdz) receptors and Lrp5 and Lrp6 as co-receptor proteins. The main effector of Wnt signaling is β-catenin, which in the absence of Wnt-ligand is associated with the APC/Axin/GSK3β destruction complex that mediates β-catenin proteasomal degradation. When ligand is present, the protein Disheveled (Dvl) inhibits the β-catenin destruction complex, leading to β-catenin accumulation and translocation to the nucleus, where in association with Tcf/Lef DNA binding proteins, it activates Wnt target genes such as c-Myc, cyclin D1 and others.

The relevance of Wnt signaling in MaSCs have been suggested by a number of studies: Stingl and Shackleton first discovered that murine Sca-1^{neg}/ CD24$^{med/+}$CD49f^{hi}/ CD29^{hi} cells are highly Wnt-responsive and located in the basal layer of mammary ducts, giving the first indication that Wnt could sustained SC self-renewal (Shackleton et al., 2006). Furthermore, mouse mammary glands lacking Lrp5, a co-receptor of Frizzled, show fewer terminal end buds, within which murine MaSCs are located, and impairment in repopulating ability upon limiting dilution assay, confirming a role of Wnt signaling on MaSC self-renewal (Lindvall et al., 2006). Finally, treatment with Wnt3A synthetic peptide of murine MaSCs isolated with surface markers, strongly increases the self-renewal of MaSCs *in vitro*, their ability to differentiate *in vitro* and their ability to sustain mammary gland outgrowth *in vivo*, suggesting that Wnt participates in amplifying the mammary stem cell pool (Zeng and Nusse, 2010). In the context of breast cancer, Wnt was the first mammary oncogene identified in 1982, by identification of the mouse mammary tumor virus (MMTV) within the promoter of Wnt1 gene, in murine mammary adenocarcinomas (Nusse R, 1982). Shackleton *et al.* demonstrated that there was an increased of 6-fold in TICs number in MMTV-Wnt1 transgenic mouse model compared to controls, as a result of an expanded SC pool. The group further showed that mammary outgrowth generated from Lin^{neg}/ CD24$^{med/+}$/ CD29^{hi} of MMTV-Wnt1 model appeared

severely hyperplastic compared to controls, indicating a direct effect on the TIC compartment (Shackleton et al., 2006).

In humans, few mutations of Wnt genes have been found in breast cancer, however up-regulation of the pathway is frequently observed, as for example Wnt ligands and Fzd proteins found to be over-expressed in primary breast tumors as well as Dvl amplification (Ayyanan et al., 2006). Moreover, Wnt pathway inhibition greatly reduces the sphere forming ability of CSCs derived from ER+ and ER- patient-derived samples, prompting the idea that decreasing Wnt pathway activation could significantly affect the CSCs pool amplification in breast cancers (Lamb R and Clarke, 2013).

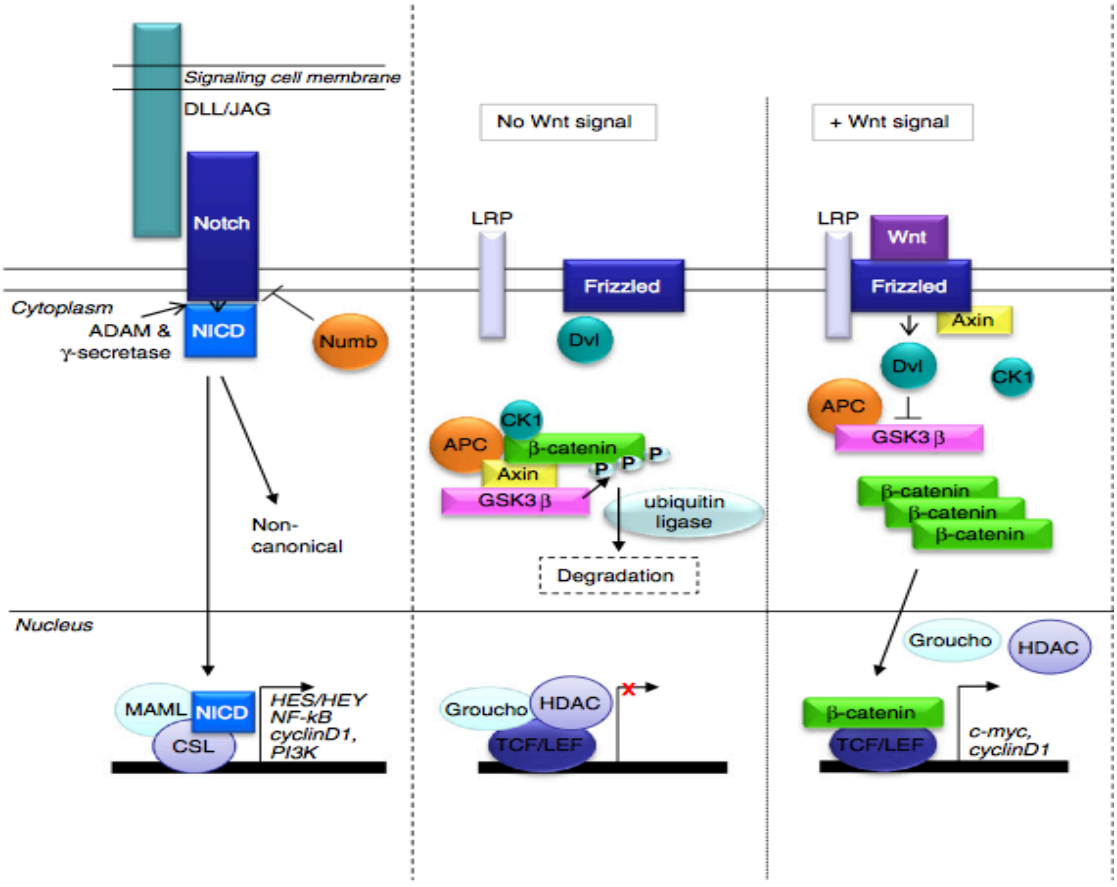


Figure 9- Schematic representation of Notch and Wnt signaling pathways from (Karamboulas and Ailles, 2013) ©Elsevier press.

1.3.3. TGF-β signaling

TGF-β signaling is one of the most complex pathways existing in mammals. It regulates a wide variety of cellular processes, including proliferation, differentiation, apoptosis and extracellular matrix (ECM) deposition. The activation of the pathway starts with binding of the ligand (there are 42 in total) to the extracellular domain of TGF-βI (of which 7 exists in total) and TGF-βII receptors (5 in total). This causes the formation of a tetrameric complex, specific for each ligand, in which type II receptors phosphorylates the intracellular tails of receptor I. Signal transduction can follow two different routes as shown in Figure 10: Smad-mediated signaling and Smad-independent signaling. In the first case, type I receptor transmits the signal through phosphorylation of effector SMADs, that in turn bind to the common-partner Smad4 (Massague et al., 2005). The heteromeric complex of Smads then translocates to the nucleus, interacts with DNA-sequence specific transcription factors and regulates the transcription of several target genes, causing either gene induction or repression. Importantly, two of the genes induced by this signaling are inhibitory Smads (Smad6 and Smad7) that block the TGF-β signaling cascade, thus forming a negative feedback loop (Imamura, 1997). In Smad-independent signaling, TGF-β directly activates other pathways, such as the MAPK pathway (Ras-Raf-Mek-Erk axis), or the JNK pathway through the direct interaction with TRAF6 (Sorrentino et al., 2008) or Akt (Lamouille et al., 2012).

The TGF-β pathway is involved in various aspects of normal mammary gland development: for example, in response to sexual hormone stimulation, TGF-β induction inhibits proliferation in epithelial cells, regulates ECM deposition and causes the remodeling of mammary gland architecture during involution, through the induction of apoptosis (Ewan KB, 2005). Similarly, heterozygous deletion of TGF-β1 (TGF-$\beta1^{+/-}$) in mice causes an accelerated development of mammary gland and an increased proliferation of MECs in response to hormone stimulation (Tang, 1998). Thus, we can speculate that in

normal physiology TGF-β1 acts as cytostatic factor, inducing both cell cycle arrest and apoptosis. Conversely, evidence on the role of TGF-β1 in cancer development and progression is highly conflicting, leading researchers to declare the existence of "TGF-β1 paradox". In the early stages of cancer development, as in normal breast development, TGF-β1 acts as antimitotic agent, inducing the expression of cell cycle inhibitors (p15INK4B which inhibits CyclinD-CDK4/6 complex and p21CIP1 acting on CyclinE-CDK2 complex) and repressing Myc activation (Massague, 2008), thus causing G1 arrest. However, during cancer progression, tumor cells become insensitive to the tumor-suppressive ability of TGF-β1, turning its action in cancer promoting agent, that induces migration, invasion and epithelial-to-mesenchymal transition (EMT). In physiological conditions, EMT exerts tight control over embryogenesis and the formation of different organs and tissues. EMT is regulated by different transcription factors such as Twist, Snail, Slug as well as ZEB1-2, and leads to the loss of E-cadherin and tight junction expression and the acquisition of a more motile and invasive phenotype. The Weinberg laboratory has reported the first association between EMT, TGF-β and a more aggressive CSC phenotype (Mani et al., 2008). The group showed that the SC population coming both from human and mouse breast cells (respectively $CD44^{+}CD24^{-}$ or $CD24^{med/+}/CD49f^{hi}$ cells) posses mesenchymal morphology and express mesenchymal markers (such as vimentin and/or N-cadherin) compared to their non-stem counterparts. Mammary epithelial cells treated with TGF-β1, activate an EMT gene expression program similar to that induced by Snail or Twist transcription factors, showing a mesenchymal morphology, acquisition of SC surface markers positivity and of mammosphere forming ability, suggesting the exiting possibility that EMT was sufficient to endow cells with properties of stem cells (Mani et al., 2008). Whether the TGF-β effect is mediated by the unbalancing the proliferation potential within $CD44^{-}CD24^{+}$ and SC subpopulations or by conversion of $CD44^{-}CD24^{+}$

cells to match their $CD44^{+}CD24^{-}$ counterpart is still matter of debate. Additionally, in mammary epithelial cells, the Wnt pathway acts as a competence factor to generate an environment that is more prone to TGF-β- induced EMT, as reviewed in (Massague, 2012). As reported in Figure 11, the Wnt effector LEF1 occupies, along with Smad3/Smad4, the promoter of E-cadherin, repressing key epithelial cell junction genes and favoring the transition toward a mesenchymal state (Massague, 2012).

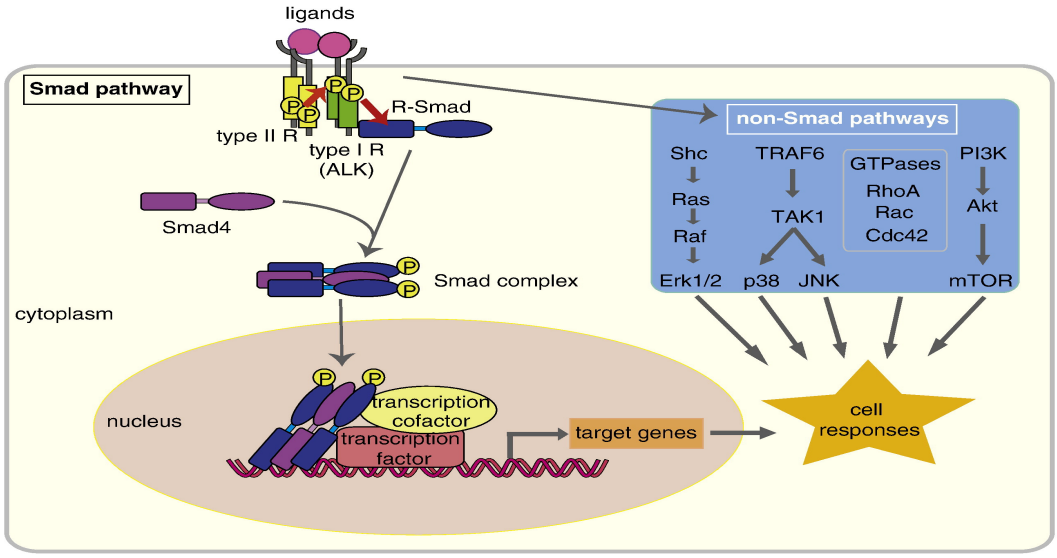


Figure 10- Schematic representation of Smad-dependent and Smad-independent TGF-β signaling pathways, from (Sakaki-Yumoto et al., 2013) ©Elsevier press.

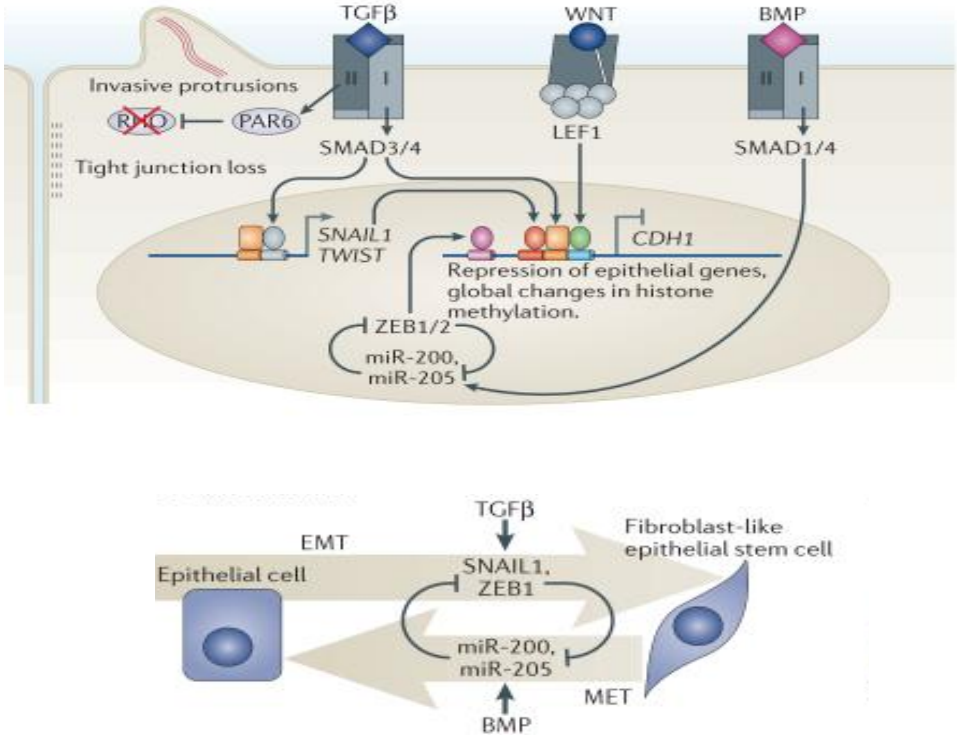


Figure 11- TGF-β signaling in EMT and MET transition- from (Massague, 2012) ©Nature press.

1.3.4. NF-κB signaling

NF-κB is an intracellular signaling pathway that could be activated in response to different extracellular stimuli through two different routes: canonical signaling and non-canonical signaling as reported in Figure 12. In the canonical signaling pathway, pro-inflammatory cytokines, such as TNFα or interleukin-1β, bind to TNFR receptor and, through TRAF2 (TNF-associated factor 2), mediate the recruitment of the IKK complex, which is composed of IKKα, IKKβ and NEMO (NF-κB essential modulator). The activation of IKKβ leads to the phosphorylation and subsequent degradation of IKBα, which, in unstimulated conditions, normally sequesters the p50/p65 heterodimer (NF-kB) within the cytoplasm. Once IKBα is degraded, p50/p65 is free to translocate to the nucleus, where it binds DNA and activates the transcription of genes involved in survival and inflammation. In the non-canonical signaling pathway, the intracellular cascade is mediated by NIK (NF-κB inducing kinase) that activates IKKα dimers (but not the IKK complex), which leads to phosphorylation of the inhibitory molecule p100 and release of p52. P52 is thus free to interact with RelB and translocate to the nucleus where it activates the transcription of genes related to adaptive immunity.

While NF-κB was originally associated with immunity and inflammation, it has emerged more recently as critical regulator of mammary gland morphology, branching and development. The first *in vivo* indication of NF-κB involvement in normal mammary development came from transplantation studies using mammary tissues from IKBα-deficient mice. The lack of this major inhibitor of NF-κB signaling increased lateral ductal branching, stimulating epithelial cell proliferation with local hyperplasia (Brantley et al., 2001; Iso et al., 2003). Additionally, inhibition of NF-κB signaling by loss of IKKα catalytic activity *in vivo* in the mammary glands, caused severe defects in mammary gland architecture, due to impaired epithelial cell proliferation (Cao, 2001). Thus, NF-κB seems to play a key role of signaling in mammary development, and it is not surprisingly that its

deregulation might have profound consequences in breast tumorigenesis. High constitutive NF-κB activation is frequently observed in triple negative breast cancer subtypes cancers and it is required for basal cells proliferation (Yamaguchi et al., 2009). Overexpression of p65 in a normal human mammary cell line (MCF10A) induces loss of epithelial markers and reduces the ability of cells to differentiate *in vitro* as acinar structures, consequently with activation of ZEB1 and ZEB2 transcriptional factors and a switch toward a mesenchymal state, connecting NF-κB deregulation directly to the EMT program (Chua et al., 2007).

NF-κB overexpression can also foster amplification of breast CSCs: human basal-like tumors, but not other breast cancer subtypes, were shown to display non cell-autonomous NF-κB activation that regulates CSCs population by activation of Jag1 and Notch-dependent CSCS expansion (Yamamoto et al., 2013). Additionally, selective ablation of NF-κB in the mammary gland of a HER2-amplified mouse model reduced HER2-dependent tumorigenesis *in vivo;* in these tumors the proportion of TICs was dramatically decreased, cell proliferation, colony formation and the ability to form mammospheres was also impaired (Liu et al., 2010). More recently, using a model of MMTV-HER2-IKK$^{aa/aa}$ mice, in which IKKα is mutated to abolish catalytic activity, IKKα was shown to be a critical mediator of HER2-mediated tumorigenesis, and its inactivation caused a strong reduction in TICs (Zhang et al., 2013). Lastly, NF-κB has been defined as an essential player in the link between inflammation and oncogenic transformation. An epigenetic switch that involves the cooperation between Src, NF-κB and Lin28 has been described. Briefly, the transient activation of Src in human normal mammary cells triggers an inflammatory response, mediated by NF-κB, which directly activates Lin28. Lin28 reduces the levels of let-7 miRNAs, impacting on proliferation and differentiation of breast cells, but also increasing IL6 levels (a direct target of let-7). Since Il6 continues to stimulate NF-

κB activation, this generates a positive feedback loop that maintains the inflammatory response activated and fosters cell transformation (Iliopoulos et al., 2009).

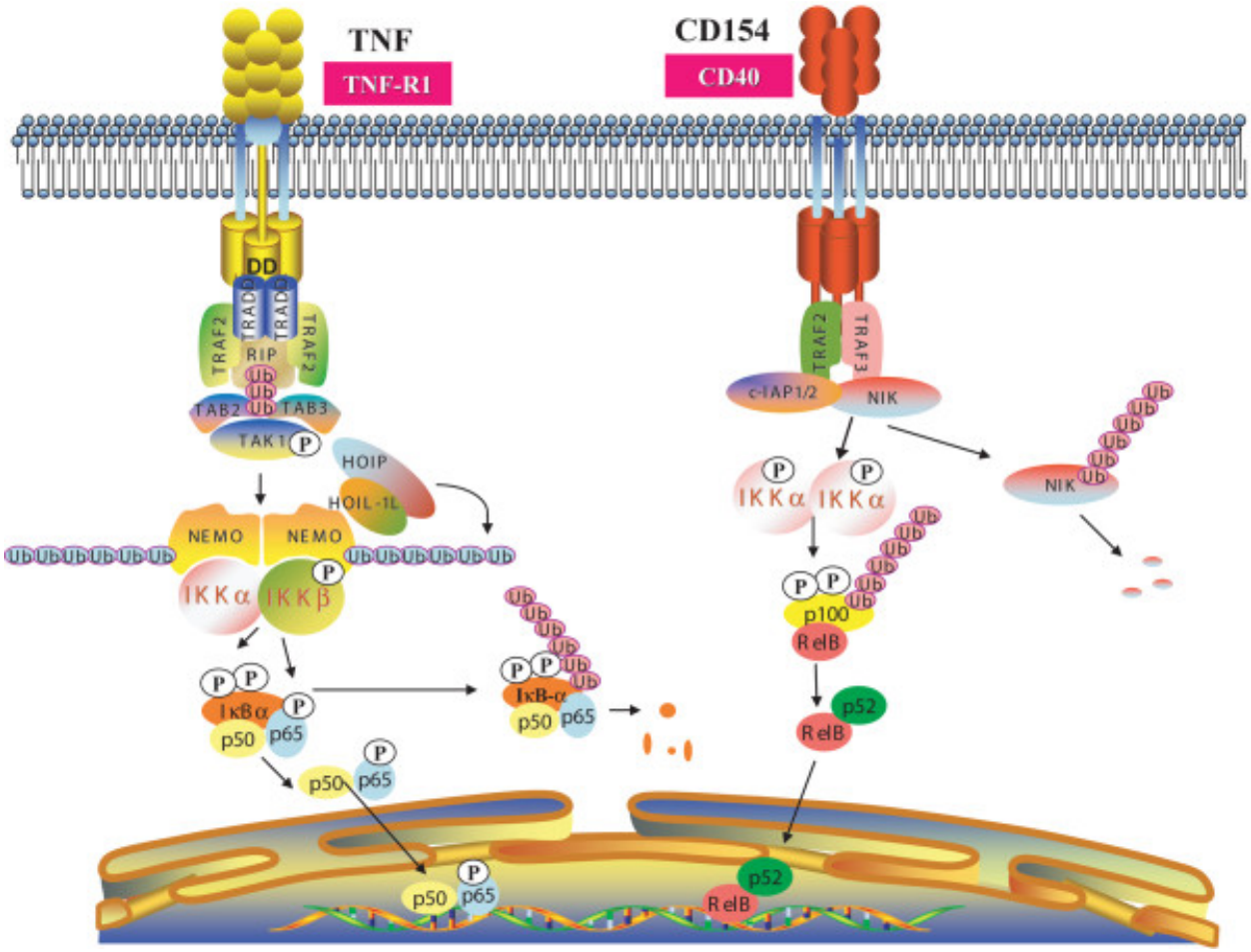


Figure 12- Canonical and non- canonical NF-κB signaling pathways, from (Shostak et al, 2011) ©Breast cancer research press.

1.4. Gene expression profiling of human/mouse normal and cancer stem cells

Understanding the relationship between the normal SCs and the TICs of different breast cancers subtypes is fundamental to gain insight on the molecular mechanisms that sustain breast tumorigenesis and CSCs expansion.

Several groups have attempted to characterize breast normal and tumoral stem cells by analyzing their gene expression profiles. In the first work by Stingl, MaSCs were isolated from murine normal mammary tissues as $Sca1^{neg}/CD49f^{high}/CD24^{med}$ population and proved to be highly efficient as mammary repupulating unit (MRU) (Stingl et al., 2006). The authors directly compared the gene expression of MRUs with the one of the so-called Ma-CFCs (mammary colony forming cells) which represent progenitors cells that under differentiating conditions could form colonies *in vitro*. By this approach, they found that MRUs typically express markers for the basal phenotype (K14 and α–SMA) as opposed to markers for progenitors which express mainly luminal markers (keratin 19 and keratin 18). Moreover, several proteins belonging to the Notch, Wnt, TGF-β and Nf-κB pathways were significantly enriched in MRUs vs Ma-CFCs, for example Jagged1 (the ligand of Notch), the Frizzled7 receptor of the Wnt pathway (recently involved in maintenance of pluripotent state of hESCs, (Fernandez et al., 2014), proteins involved in inflammatory responses (such as interferons and their receptors) and Slug, previously mentioned as a TGF-β-induced mediator of EMT.

A more general view of pathways actived in SCs was provided by the work of Shipitsin *et al.* They compared the gene expression profile of $PROCR^{+}/CD44^{+}/CD24^{-}$ cells (here referred as CD44+) and their $CD44^{-}/CD24^{+}$ counterparts (here referred as $CD24^{+}$) isolated from normal mammary cells vs primary invasive tumor samples (Shipitsin et al., 2007). Several important conclusions were drawn: i) the gene expression profile of $CD44^{+}$ cells coming from normal and tumor samples are more similar to each other than to the $CD24^{+}$ component from the same tissue, suggesting that normal and cancer stem cells share the

same transcriptional program; ii) among the gene ontologies, genes involved in chemotaxis, cell motility and angiogenesis are highly expressed in CD44+ cells from normal and tumor samples, and, among canonical pathways, TGF-β and Wnt pathways and ECM remodeling are the most upregulated in CD44+. These observation are consistent with the fact that stem cells express a more motile, mesenchymal and less proliferative profile than differentiated $CD24^{+}$ cells; iii) CD44+ and $CD24^{+}$ gene signatures correlate with clinical outcome of three independent cohorts of breast primary tumors, suggesting that expression of stem (CD44+) genes is associated with shorter metastasis-free survival times. These results prompted the authors to propose that not all the breast tumors are equal in terms of CSC content and, thus, that tumors could be stratified according to their degree of "stemness", an issue supported by two other independent studies from Visvader and Di Fiores's laboratories.

Indeed, Visvader analyzed murine MaSCs (Lin-/$CD61^{hi}$ /$CD29^{hi}$/ $CD24^{low}$ cells) as compared to luminal progenitors ($CD61^{hi}$ /$CD29^{low}$/ $CD24^{+}$) and mature luminal cells ($CD61^{-}$ /$CD29^{low}$/ $CD24^{high}$) by gene expression microarrays (Lim et al., 2010). These populations were subsequently compared with their corresponding human fractions, to search for conserved genes and pathways. The conserved gene expression pattern upregulated in the MaSC population contains several genes belonging to the Notch, Wnt, TGF-β and Nf-κB pathways including: transcription factors involved in EMT process (such as Slug, Twist2 and Sox11 as well as the EMT inhibitor p63), basal cytokeratins (Krt5 and Krt14), genes responsive to Wnt/β-catenin pathway (such as Lrg6, Frizzled8 and Tcf4), the Notch ligand Jag2 and Interleukin8 inflammatory signaling pathway. Differently, in mature luminal cells were found upregulated genes associated with the process of luminal differentiation including luminal cytokeratins (Krt18), ER and PgR as well as genes belonging to prolactin-induced signaling pathways and high levels of Notch signaling. More importantly, when each signature (MaSCs, luminal progenitors and

luminal differentiated cells) was compared with signatures of mammary tumors isolated from MMTV-Wnt1, MMTV-Neu (Her2 amplified), $p53^{-/-}$and MMTV-Myc mice was found that the MaSCs transcriptional signature was found to highly correlate with MMTV-Wnt1, and $p53^{-/-}$ tumors, indicating that cells within these tumors exibits similarities with basal MaSCs, in contrast the mature luminal signature was found to correlate with tumors derived from MMTV-Myc.

The only study in which surface markers were not used to isolate MaSCs was published by Pece *et al.* Here, stem cells were isolated as quiescent label-retaining (PKH26) cells from mammosphere cultures (Pece et al., 2010). The authors compared the gene expression profile of MaSCs ($PKH26^{pos}$) obtained from a pool of human normal donors to matched progenitors/differentiated cells ($PKH26^{neg}$), and found the following markers enriched in MaSCs: basal markers (such as Krt5, p63 and serpinB5), Notch ligand (Jag1) as well as Notch receptors (Notch2 and Notch3), several cell cycle and cell survival genes (underlying the quiescent status of human SCs) and molecules involved in cell-to-cell and cell-to-ECM contacts. As for the other studies described, the MaSC (PKH^{pos}) signature shows, independently from the tumor-grade, a remarkable correlation with the basal tumor subtypes (and not with LuminalA or B or Her2 amplified) among the different molecular subtypes of breast tumors isolated by Perou et al (Perou CM, 2000) and Sorlie et al (Sorlie et al., 2001). Moreover, among the upregulated genes in SCs was observed a strong enrichment of basal tumor genes, measured by gene set enrichment analysis. These findings support the idea that the degree of similarity of a tumor with the MaSC signature could simply reflect the content of CSCs present in the tumor. Indeed, Pece *et al.* demonstrated by mammosphere and transplantation assay that breast tumors are not equal in term of their CSC content, with high grade ER-negative (G3/ER-) tumors being highly enriched for CSCs, compared to low grade ER-positive (G1/ER+) tumors or normal epithelium.

1.5. miRNAs in stem biology

It is becoming increasingly evident that, apart from the signaling pathways described earlier, another layer of regulation of SC self-renewal, proliferation and differentiation is controlled by miRNAs. Pivotal works analyzed the role of a general impairment of miRNAs levels in embryonic stem (ES) cells through the generation of Dicer-null and DGCR8-deficient ES cells. Dicer is essential to early mouse development, as its absence causes embryonic lethality, likely due to severe developmental defects (Bernstein et al., 2003). In fact, ES cells from Dicer-KO mice display a variety of differentiation defects and a prolonged G0/G1 block in the cell cycle (Bernstein et al., 2003). Similarly to this finding, DGCR8-deficient ES cells exhibited delayed cell cycle progression and failure in the differentiation process, as they failed to silence self-renewal determinants as OCT4, nanog and Sox2 (Wang et al., 2007). These discoveries highlighted that one of the predominant roles of miRNAs in ES cells is to regulate cell cycle progression during stem cell differentiation. In support of this, several miRNA clusters identified as ES cell-specific directly were involved in maintaining pluripotency: for example, the miR-290-295 cluster is specific of ES cells and its levels decrease as the stem cell differentiate. This cluster exerts its role by targeting the repressor of DNA methyl transferase (DMNTs), called RBL2, which in turn is no longer able to inactivate OCT4 by CpG islands methylation, maintaining active a critical effector of self-renewal (Houbaviy et al., 2003). Further, it appears that other miRNAs are indeed necessary to induce ES differentiation. One example is miR-21, which acts by reducing the cellular concentration of key proteins of self-renewal such as Nanog, Sox2 and OCT4, by directly targeting the 3'UTR of these mRNAs (Singh et al., 2008) or miRNAs of the let-7 family (Rybak et al., 2008).

Given the role of miRNAs in balancing self-renewal and proliferation/differentiation of embryonic stem cells, it is not surprising that they also play an important role in the homeostasis of adult SCs. The first evidence came from the hematologic system, in which

specific miRNAs fine-tune each step of hematopoiesis. For example, miR-128a and miR-181a are specifically expressed in early multipotent progenitors (MP) and prevent the immature differentiation process of all lineages; miR-223 and miR-181 are directly involved in the specification of MP cells towards the lymphoid or the myeloid lineages (Chen et al., 2004) and miR-221 and miR-222 block erythropoiesis by directly targeting KIT receptor, an important regulator expressed on hematopoietic SCs (Felli et al., 2005). In the neural tissue, miR-124, one of the most expressed miRNAs in the adult brain was associated to neural differentiation of neural stem cells by direct inhibition of the SC determinants Sox9 and REST (Cheng et al., 2009). In the skin, as previously mentioned, miR-203 acts as a "stemness inhibitor" of epidermal stem cells. This activity is promoted by direct inhibition of p63, a critical stem cell determinant, resulting in limited proliferative potential, forced cell-cycle exit and induction of differentiation (Lena et al., 2008).

Recent data have highlighted the functional role of miRNAs as regulators of "stemness" acquisition and maintenance in CSCs.

For example, the Let-7 family and miR-200c were shown to inhibit breast CSCs. Yu et al. first described that let-7 levels are markedly low in TICs compared to their non-TIC counterparts and that its levels are increased only upon differentiation (Yu et al., 2007). Over-expressing let-7 in TICs caused a striking impairment in proliferation, mammosphere-forming ability as well as tumor formation and metastasis *in vivo*. On the contrary, reducing let-7 levels restored the self-renewing capacity of TICs. Mechanistically, let-7 activity is permitted by direct inhibition of H-Ras and HMGA2, known let-7 targets, involved in self-renewal and multipotency regulation (Yu et al., 2007). Afterwards, Shimono *et al*. investigated the ability of miR-200c in regulating breast stem cell biology. Multiple members of miR-200 family were found to be downregulated in human primary BCSCs, isolated by surface markers (lin-/CD44+/CD24-) as compared to

the non-tumorigenic counterpart (Shimono et al., 2009). These miRNAs were also found to be downregulated in breast stem cells isolated from normal human and murine mammary stem cells. Forced expression of one member of the family, miR-200c, suppressed the colony formation ability of murine BCSCs, normal mammary outgrowths *in vivo* in mammary gland reconstitution assay, and the tumorigenicity of human breast cancer stem cells *in vivo*, by directly targeting BMI, a know regulator of SC self-renewal and differentiation. Hence, miR-200c emerged as critical regulator of CSCs/SCs in the mammary gland (Shimono et al., 2009).

Finally, a recent work described the role for miR-205 in controlling a cancer-associated stem cell phenotype in breast epithelial cells (Chao et al., 2014). The authors showed that loss of miR-205, induced by the interaction of Jag1 with Notch receptors, causes a marked epithelial to mesenchymal transition, loss of epithelial polarization and stimulates symmetric cell division that causes expansion of stem cell population *in vitro* and spontaneous mammary lesions *in vivo*. These data supported that loss of miR-205 activates EMT and promotes a stem-like phenotype, resulting in mammary tumorigenesis (Chao et al., 2014).

1.6. miR-146 in physiology and pathology

The miR-146 family is a conserved vertebrate miRNA family, which consists of two main members, miR-146a and miR-146b. These are located on different chromosomes, (respectively on chromosome 5 and 10), as intergenic transcriptional units. The degree of similarity of these two miRNAs is high, in fact mature sequences differing by only two nucleotides at the 3' end, a region that seems to play only minimal role in target recognition (Bartel, 2009). However, it is not clear if the two miRNAs act in a redundant manner or if they have specific targets. miR-146a was initially discovered by the Baltimore group, who were searching for miRNAs up-regulated in the innate immune response of

human monocytes upon LPS stimulation. By computational analysis of the miR-146a promoter, three Nf-κB consensus binding sites were identified, demonstrating that the LPS-mediated increase of miR-146a occurs in a Nf-κB-dependent manner, validated by ChIP and Luciferese reporter assays (Taganov et al., 2006). Furthermore, TRAF6 (TNF receptor-associated factor 6) and IRAK1 (interleukin-1 receptor-associated kinase 1) were identified by luciferase assay as direct targets of miR-146a/b. Both these genes are key adaptor molecules acting downstream of TLR and cytokine receptors, suggesting a role for miR-146a in controlling the Nf-κB-dependent inflammatory response through a negative feedback loop involving the downmodulation of TRAF6 and IRAK1 (Taganov et al., 2006). A few years later, the same group generated a miR-146a $^{-/-}$ (KO) mouse model to explore the function of miR-146a *in vivo* (Zhao, 2011). Although KO mice showed no apparent abnormalities at birth, they gradually developed uncontrolled myeloproliferation in both spleen and bone marrow (between 18th and 22nd months of ages), and developed sarcomas and lymphomas in various organs. Tumors were transplantable in immunocompromised mice and both spleen, bone marrow and tumors from KO mice showed increased basal levels of Nf-κB activation. The myeloproliferative disease of miR-146 KO mice was completely rescued by *in vivo* deletion of the p50 subunit, which inhibits Nf-κB activation. Thus, miR-146a appears to be a fundamental component of the immune-cell regulation and its critical function is to act as a tumor suppressors, keeping controlled the levels of Nf-κB in the hematopoietic system (Zhao, 2011).

Several studies emphasize also the involvement of miR-146 expression during the process of hematopoietic differentiation, even though its role remains highly controversial. miR-146a was first associated to macrophages differentiation: using an inducible model in which macrophages were stimulated to differentiate by PU.1 transcription factor activation, miR-146a was found to be one of the most upregulated miRNAs (Ghani et al., 2011). This result was further corroborated by direct comparison between freshly-isolated

murine myeloid early progenitors, expressing low levels of miR-146a, and mature murine macrophages, expressing high levels of miR-146a. Forced expression of miR-146a induced the *in vivo* differentiation of hematopoietic stem cells into functional macrophages (Ghani et al., 2011).

Other studies suggest, instead, a role for miR-146 in maintaining the undifferentiated state of hematopoietic progenitor cells (HPCs) (Tang, 1998). The authors observed that miR-146a is highly expressed in human HPCs and downregulated when HPCs were induced to differentaite into megakaryocytes. Forced expression of miR-146a in HPCs impaired megakaryocytic proliferation and differentiation, whreas miR-146a KD restored the expression of megakaryocytic differentiation markers (Tang, 1998). In line with these results, miR-146a was found to be mainly expressed in murine hematopoietic stem/progenitor cells rather than in their differentiated counterparts. Moreover, overexpression of miR-146a in hematopoietic SCs and subsequent bone marrow transplantation resulted in a number of hematological abnormalities such as transient myeloid expansion, decreased erythropoiesis, impaired lymphopoiesis and impaired bone marrow reconstitution capacity in recipient mice, underlying a possible deleterious effect on hematopoietic differentiation (Starczynowski et al., 2011).

Even more debated is the definition of the biological role of miR-146 in human tumorigenesis, since it has been described as being both a tumor suppressor and an oncogenic miRNA. In several cancers, miR-146 expression is downregulated compared to normal tissues. For example, miR-146a expression is lost in pancreatic carcinomas and its re-expression in cancer cells induces the inhibition of EGFR and Nf-κB signaling (Yu et al., 2012); in gastric cancers, stratification of patients according to the levels of miR-146 (miR-146-high and miR-146-low) revealed that low levels of miR-146 correlated with a higher risk of metastasis, venous invasion and poorer prognosis compared to the miR-high (Kogo et al., 2011); lastly, in gliomas, miR-146a expression is induced by EGFR

stimulation, leading in turn to Notch inactivation, and a reduction in proliferation of cancer cells or in the number of glioma stem-like cells (Mei et al., 2011).

Expression data from the Croce laboratory, instead, suggest that both miR-146a and miR-146b are upregulated in three out of six solid tumor types, including breast, pancreas and prostrate cancers (Volinia et al., 2006). In the breast context, miR-146 levels strongly correlate with the basal phenotype (Garcia et al., 2011). Both miR-146a/b were found highly overexpressed in triple-negative vs non-triple negative, in ER-/PgR- versus ER+/PgR+ and in high grade vs low grade breast primary tumors. The authors showed that in the basal subtype, that is often associated with loss/mutation of BRCA1 tumor suppressor gene, miR-146 could directly target BRCA1 at its 3'UTR, causing both an increased cell-proliferation and a strong impairment in BRCA1-mediated homologous recombination. This would represent a novel mechanism through which miR-146 deregulation could confer aggressiveness to breast cancers (Garcia et al., 2011). In papillary thyroid carcinomas, miR-146b was found significantly overexpressed and could contribute to the regulation of the TGF-β pathway, conferring resistance to TGF-β-mediated antiproliferative signals, via SMAD4 inhibition, suggesting an oncogenic potential for miR-146b (Geraldo et al., 2012).

In two more recent papers, the oncogenic potential of miR-146a/b was coupled to its interaction with the Numb protein, a very well-known cell-fate determinant involved in the negative regulation of Notch signaling. In the first of these two studies, miR-146a was shown to be induced in BRAF- and NRAS- mutated melanoma cells (Forloni et al., 2014). Overexpression of miR-146a in human melanoma cells increased their *in vitro* cell proliferation and *in vivo* tumorigenicity, whereas miR-146 inhibition caused the opposite effects. The observed miR-146a oncogenic effects were mediated by Numb, a validated target of miR-146a, responsible for the hyperactivation of Notch signaling and initiation of melanomas. In this context, the expression of a non-targetable form of Numb or treatment

with Notch inhibitors partially rescues the tumorigenic activity of miR-146a (Forloni et al., 2014).

The second study explored the tumorigenic potential of miR-146a in colorectal cancer SCs (Hwang et al., 2014). The authors showed in CSCs that the transcriptional factor Snail, involved in EMT is critical in mediating tumor-initiating ability *in vivo*, symmetric cells division (SCD) and in inducing high levels of miR-146a via the activation of Wnt/β-catenin pathway. The absence of miR-146a in CSCs caused severe impairment in the sphere forming ability, tumorigenicity *in vivo* as well as a switch from the symmetric mode towards the asymmetric mode (ACD) of cell division. Conversely, high levels of miR-146a caused the downregulation of Numb, increased Notch signaling activation and β-catenin stabilization, generating a positive feedback loop between Snail/miR-146a/ β-catenin/ Numb, that is crucial for the expansion of the stem cell population and for controlling symmetric cell division of CSCs (Hwang et al., 2014).

2 Material and Methods

2.1 Cellular biology

2.1.1. Cell lines and tissue culture procedures

SUM159PT cell line was commercially available (Asterand, Detroit, MI) and was cultured in Ham's F12 medium with 5% fetal bovine NA serum, human insulin (5 ug/ml), hydrocortisone (1 ug/ml) and Hepes.

MCF10A cells were purchased from ATCC and cultured in DMEM/ F-12 (1:1) with 5% horse serum, hydrocortisone (500 ng/ml), human insulin (10 ug/ml), cholera toxin (100 ng/ml) and human EGF (20 ng/ml).

Mammary glands from 5-weeks old FVB/Hsd females (Harlan laboratories) were mechanically dissected into small pieces with scissors and enzymatically digested with digestion medium. The composition is: Dulbecco's modified Eagle Medium (DMEM) plus Ham's F12 (1:1) medium supplemented with human insulin 1 ug/ml, hydrocortisone, human EGF (20 ng/ml), 200 U/ml collagenase type 1A (Sigma) and 100 U/ml hyaluronidase (Sigma) in a humidified atmosphere at 5% CO_2 for 3 h at 37° C. When the digestion was complete, the cell suspension was centrifuged at 80xG for 5 minutes to allow precipitation of epithelial cells but not of adipose cells. The cell suspension was washed in PBS and filtered through 100, 70, 40 and 20 um filters to eliminate cell aggregates. Red blood cells were lysed using ACK lysing buffer. Epithelial cells were plated at single cell density to obtain mammospheres.

Mammary epithelial cells from reductive mammoplasties and primary tumors were collected at European Institute of Oncology (Milan, Italy) from patients that had given the informed consent to use of biological materials for scientific purpose. Primary tissues were

digested as described in Dontu et al., 2003.

All the cells were grown in a humidified atmosphere at 5% CO_2 at 37°C, except for SUM159PT that were grown at 10% CO_2.

2.1.2. Mammosphere culture in vitro

For suspension mammosphere culture, SUM159 and MCF10A were plated in low-adherent 150 mm dishes (coated with poly-HEMA, Sigma Aldrich) at density of 1000 cells/ml in serum-free mammary epithelial medium (MEBM, Lonza) supplemented with 5 µg/ml insulin, 0,5 µg/ml hydrocortisone, 2% B27 (Invitrogen), 20 ng/ml EGF and human b-FGF and 4 µg/ml heparin. For mammospheres culture of primary samples, cells were plated at density of 5000 cells/ml for both mouse and human tissues in stem medium. For serial propagation, human mammospheres were collected after 7 days of culture, enzymatically dissociated with trypsin-EDTA (0.025%) and plated at same density for successive generations. Murine mammospheres were collected after 7 days of culture and mechanically dissociated.

For clonal mammospheres culture, 1000 cells were plated in ultra-low adherent 12-well plates in presence of methylcellulose (2% w/v, Sigma Aldrich) diluted 1:1 to stem medium (with 2x of growth factors).

Each generation of mammosphere, that expressed the shRNA or the antimiR along with a fluorescent reporter marker (GFP or RFP), was counted with an automated macro-based system generated in house with Java and run on Fiji software (ImageJ). Briefly, at each generation we collected the images of all the 12-wells and the macro counted only the rounded fluorescent spheres with a diameter > 70 µm for SUM159 and > 50 µm for mammospheres from primary MECs, all the data (as number of spheres, diameter in micron, circularity and mean dimension area) were collected in an excel file for further analysis.

2.1.3. PKH26 staining and FACS-sorting

PKH26 (Sigma) staining was performed on breast cell lines and primary tissues as described on Pece et al., 2010 and Cicalese et al., 2009. Briefly, single cell suspension was resuspended in PBS (at a density of 10^6 cells/ml) and stained with PKH26 diluted 1:20000 for 5 minutes at room temperature, light protected. Then, the reaction was blocked for 1 minute by adding 1% of FBS. Cells are centrifuged at 1500 rpm for 5' and plated at single cell to generate mammospheres PKH-labelled. PKH-labeled mammospheres were collected after 7-8 days, and enzymatically dissociated with trypsin-EDTA to single cell suspension. Cells are filtered with 40 μm cell strainer and subjected to FACS sorting (Influx cell sorter equipped with a 488 nm laser and with a band pass 575/26 nm optical filter for PKH26 fluorescence detection, BD) to collect PKH^{pos} and PKH^{neg} cells in 96-well plate. To exclude the possibility that non-epithelial cells (i.e. endothelial or hematopoietic cells) or dead cells were co-purified by FACS sorting with epithelial cells, each mammosphere preparation was depleted of contaminants with CD31 and CD45 microbeads (MACS technology) and subsequently stained with DAPI (1mg/ml, diluted 1:200 in PBS) for 1 minute at RT to select only living cells. The PKH^{pos} region was defined according to the estimated percentage of SCs in mammospheres. The PKH^{neg} gate was selected according to the basal fluorescence of unstained cells.

Cells infected with miRZIP KD vector were subjected to FACS sorting at 72h post infection to collect GFP-high population. The gate for negative cells was selected according to the basal fluorescence of not infected cells.

Cells infected with miR-146 sensors, at 72h post infection, were blocked with PBS 5% FBS for 10 min at 4° C. Then cells were stained with anti-ΔNGFR conjugated with PE-cy7 (BD Pharmigen) for 15 min at 4°C (at a concentration of 5 μl/10^6 cells), and analyzed with two colors flow cytometry (GFP and PE-cy7) to collect $\Delta NGFR^+$/ GFP^{high} and/or $\Delta NGFR^+$

/GFP^{low} populations.

2.1.4. Lentiviral infections

293T packaging cells were cultured in a humid atmosphere containing 5% CO_2 in DMEM supplemented with 10% FBS, 2 mM glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. For third-generation lentiviral production, 293T cells were transfected with the calcium-phosphate procedure with a mixture of: 10 µg of pRSV-Rev, 10 µg of pMDLg/pRRE (gag&pol), 10 µg of pMD2.G (VSV-G), and 20 µg of the lentiviral vector per plate, 250 µl of 2M $CaCl_2$, in a final volume of 2 ml TE 0.1X. This mix was added drop wise to 2 ml of 2X HBS, by bubbling, and then the calcium-phosphate precipitates were added to the cells at 70% of confluence. The medium of 293T cells was replaced after 12-16 h post-transfection; the viral supernatant was collected at 36h post transfection and filtered with a 0.45 µm syringe-filter. To concentrate the viral stock, we ultra-centrifuged the viral supernatant for 2h at 19800 rpm at 4°C and the viral pellet obtained was resuspended in MEBM medium at 100X concentration. Viral stock was frozen (-80°C) or directly used to infect target cells at ratio: 10 µl virus/100.000 target cells in presence of 1µg/mL polybrene.

2.1.5. Lentiviral vectors

mir-146 KD and non-targeting control (SCR) was obtained by lentiviral infection with commercially available vector (pmiRZIP lentivector) from System Bioscience (SBI, clone MZIP000-PA-1 for SCR and MZIP146b5p-PA-1 for miR-146 KD).

For construction of miR-146a/b sensor, two DNA sequences containing 4 MREs with perfect complementary to miR-146a/b were synthesized (Primm) as reported:

Sensor 146a sense: 5'--→ 3'
CTAGAAAACCCATGGAATTCAGTTCTCACGATAAACCCATGGAATTCAGTTCTCAACCGGTAAA
CCCATGGAATTCAGTTCTCATCACAAACCCATGGAATTCAGTTCTCAC

Sensor 146a antisense: 5'--→ 3'
CCGGGTGAGAACTGAATTCCATGGGTTTGTGATGAGAACTGAATTCCATGGGTTTACCGGTTGA

GAACTGAATTCCATGGGTTTATCGTGAGAACTGAATTCCATGGGTTTT

Sensor 146b sense: 5'--→ 3'
CTAGAAAGCCTATGGAATTCAGTTCTCACGATAAGCCTATGGAATTCAGTTCTCAACCGGTAAG
CCTATGGAATTCAGTTCTCATCACAAGCCTATGGAATTCAGTTCTCAC

Sensor 146b antisense: 5'--→ 3'
CCGGGTGAGAACTGAATTCCATAGGCTTGTGATGAGAACTGAATTCCATAGGCTTACCGGTTGA
GAACTGAATTCCATAGGCTTATCGTGAGAACTGAATTCCATAGGCTT

Sequences were annealed according to this protocol:

1 µl of each oligo 100 µM in a final volume of 50 µl of Annealing Buffer (Promega).

Incubate 4 minutes at 95°C then 10 minutes at 70°C.

Annealed oligos are slowly cooled down to room temperature.

Dilute annealed oligos 1:10 in water and proceed to ligation with 100 ng of lentiviral backbone (Bd.LV.miRT vector), doubled digested with XhoI and XbaI (sites located at the 3'UTR of a destabilized GFP fluorescent protein) and treated with CIP. Ligation protocol was performed with Quick T4 DNA Ligase according to manufacturer's indications (NEB, New England Biolabs). After cloning, each positive clone sequence was verified by DNA sequencing.

Inducible lentiviral vector for RNA silencing of human NUMB (Tet-ON system) was commercially available and purchased from GE Dharmacon (Clone ID: V3THS_397259).

pmiRGLO Dual Luciferase miRNA target expression vector was commercially available and purchased from Promega (Catalog number: E1330).

NUMB 3'UTR (NM_001005743, 1368 nt) was cloned from MCF10A cDNA, using the PCR primers reported, that amplified a region of 1185 nt (MRE for miR-146 is located in position 623-630 nt, as reported from Targetscan 6.2):

NUMBfl	CACCCCATACCAGACAGGGAGCAG
NUMBrl	GAGGTCAAGAGGCTGCAACT

Deletion of MRE for miR-146 was obtained by using the followingPCR primers:

Del146-sense
TCAAAAGCCAAAATCTTTATTTTTATGCATTTAGAATATTTTAAATGATATTAAGAAGTTGTATGAGTTGTAA

Del146-antisense
TTACAACTCATACAACTTCTTAATATCATTTAAAATATTCTAAATGCATAAAAATAAAGATTTTGGCTTTTGA

The cassettes (NUMB 3'UTR or NUMB del) were inserted into the pmiRGLO Dual Luciferase vector. After cloning, each positive clone sequence was verified by DNA sequencing.

2.1.6. Luciferase assay

293T cells were plated in 24-well in triplicate, and were transfected with the indicated 3' UTR reporters (NUMB 3'UTR or NUMB 3'UTR deleted for miR-146 MRE) in presence of miRNA mimics (Qiagen) at a final concentration of 50 nM, using Lipofectamine2000® according to manufacturer's instructions.

The transfection mix was prepared as following:

Mix 1X	24-well
N. cells plated	200.000
DNA	150 ng
miRNA Mimic 20 µM	1.5 µl
Lipofectamine2000	1.5 µl
Optimem	Up to 100 µl
Final volume	600 µl

Incubate the mix for 20 minutes at room temperature to allow the formation of transfection complexes, and then add the mix drop-wise on cells. Cells were collected at 24 hours post transfection for measuring in 96-wells the luciferase activity using the Dual-Glo® Luciferase assay system (Promega) according to manufacturer's instructions, by using the EnVision Multilabel Reader (PerkinElmer). For each transfection, luciferase activity was

averaged from three replicates. The data were normalized on the ratio Firefly Luciferase activity/ Renilla Luciferase activity.

2.1.7. microRNAs overexpression and Knockdown

MCF10A cells were plated in 60 mm dishes and transfected according to fast-forward protocol with miRNA miRNA Mimic (for miR-146 OE we used the mimic MSY0002809 and for control the all star negative control siRNA SI03650318 Qiagen, stock concentration: 20 µM) at a final concentration of 50 nM. The transfection mix was prepared using Hiperfect transfection reagent (Qiagen) as following:

Mix 1X	60 mm
N. cells plated	300.000
Volume	3 ml
miRNA Mimic 20 µM	7.5 µl
HiPerfect	20 µl
Optimem	Up to 100 µl

Incubate the mix for 5-10 minutes at room temperature to allow the formation of transfection complexes, and then add the mix drop-wise on cells. Cells were collected at 16 and 24 hours post transfection for gene expression analysis.

SUM159 cells were plated in 60 mm dishes and transfected according to fast-forward protocol with the miRNA power family inhibitor at a final concentration of 100 nM (hsa-miR-146 miRCURY LNA™ microRNA Power family inhibitor and as control Negative Control A, Exiqon, stock concentration: 20 µM). The transfection mix was prepared using Hiperfect transfection reagent (Qiagen) as following:

Mix 1X	60 mm
N. cells plated	430.000
Volume	3 ml
miRNA Power family inhibitor (20 µM)	7.5 µl
HiPerfect	20 µl
Optimem	Up to 100 µl

Incubate the mix for 5-10 minutes at room temperature to allow the formation of transfection complexes, and then add the mix drop-wise on cells. Cells were collected at 16 and 24 hours post transfection for gene expression analysis.

2.1.8. Immunohistochemistry

Paraffin sections were twice deparaffinized with Bio Clear (Bio-Optica) for 15 minutes and hydrated through graded alcohol series (100%, 95%, 70% ethanol and water) for 5 minutes. Antigen unmasking was performed with 0.1 mM citrate buffer (pH=6) or EDTA (pH=8) for 50 minutes at 95°C. Slides were cooled down for 20 minutes at RT then washed in water and treated with 3% H_2O_2 for five minutes at RT.

Then, slides were pre-incubated with an antibody mixture (2% BSA, 2% normal goat serum, 0.02% Tween20 in TBS) for 20 minutes at room temperature and then stained with primary antibody for 1 h at 37°C. As primary antibodies, we used: human progesterone receptor 1:50, mouse (Dako), human estrogen receptor, 1:40, rabbit (Dako), HER2 1:1600, rabbit (Dako), Ki67 1:200, mouse (Dako), cytokeratin 5 1:200, mouse (Abcam), cytokeratin 8 1:10, mouse (Abcam), E-cadherin 1:250 mouse (Dako), vimentin 1:50, mouse (Dako) and pan-cytokeratin 1:400 mouse (Dako).

After two washes with TBS slides were incubated with a secondary antibody ready to use (DAKO Envision system HRP rabbit or mouse) for 30 minutes at room temperature and

washed twice again in TBS.

The sections were incubated in peroxidase substrate solution (DAB DAKO) from 2 to 10 minutes (1 drop of chromogen/ml buffer) and the reaction was blocked in water. Slides were counterstained with haematoxylin for 5" seconds, dehydrated through graded alcohol series (water and 70%, 95%, 100% ethanol) for 5 minutes each, washed with Bio Clear (Bio-Optica) twice for 1 minute and finally mounted with Eukitt (Kindler GmbH).

2.1.9. Western Blot analysis

Cellular pellets were washed in PBS and lysed in 50 µl of RIPA buffer 1X freshly added with protease inhibitors: PIC (from Calbiochem, 500X) and PMSF (final C: 1 mM). Cells have been centrifuged for 30 minutes at 16.000g at 4°C, and then the total lysate has been extracted and quantified by BioRad assay absorbance. Protein samples (12.5 µg) have been added with Laemmlie 5X and boiled at 95°C for 10 minutes. Then an Acrylammide gel (10% v/v) has been prepared, according to the following recipe (for 10 ml of total volume): 4.1 ml H_2O, 3.3 ml Acrylammide Solution (29:1, from EuroClone), 2.5 ml Separating Buffer 4X (pH 8.8), 100 µl APS, 10 µl TEMED). Denatured samples have been loaded in the gel, and an electrophoretic run has been performed (2 hours, 100 V). Then the total protein lysate has been transferred on a nitrocellulose membrane through Trans-Blot® Turbo™ Transfer system (Biorad). The membrane was incubated with a blocking solution (Milk, 5 % w/v in TBST) and incubated with the primary antibody overnight at 4°C, according to the following conditions:

- anti-Numb, mouse monoclonal, "home-made" clone Ab21, 1:1000 in milk;
- anti-Vinculin, mouse, from Sigma, 1:5000 in milk.

After washing with TBST, the membrane has been incubated with the secondary antibody 1 hour at RT, according to the following conditions:

- anti-mouse antibody from Cell Signaling, 1:10000 in TBST;

Membrane was washed twice for 10 minutes at RT with TBST, then incubated with chemiluminescent substrate (ECL, peroxidase-conjugated) and developed with Chemidoc™ MP system (Biorad).

2.2. *In vivo* experiments and animal manipulation

2.2.1. Intranipple injections into fat pads of NOD-Scid mice

The sample material to be injected (SUM159 or human PDXs treated with SCR or miR-146 KD) was collected into sterile Eppendorf tubes and gently pelleted in microfuge. Cell pellets were resuspended at the appropriate cell density in a mix composed of: 14 µl of PBS and 6 µl of Matrigel® (injection volume= 20 µl). Immunocompromised female mice (NOD scid gamma) of 6-7 weeks-old were anaesthetized by intraperitoneal injections of 2.5% Avertin in PBS (stock solution Avertin: 10 g of tribromoethanol, Sigma, in 10 ml of tertamyl alcohol, Sigma) were injected intra-nipple.

Animals injected with tumor cells were euthanized after 1-5 months (depending on tumor latency), when the tumors were approximately 0.5-1 cm in the largest diameter, in compliance with regulation for use of vertebrate animal in research. Tumors collected were further analyzed.

2.2.2. Transplantation of human PDXs into fat pads of NOD-Scid mice

Immunocompromised female mice (NOD scid gamma) of 4-5 weeks-old were anaesthetized by intraperitoneal injections of 2.5% Avertin in PBS. With a small incision by scissors into the mammary fat pad of inguinal mammary glands, we created a small pocket in which were inserted small pieces of primary tumor biopsies, to allow tumor engraftment into the fat pads. Tumor growth was monitored by palpation and animals were euthanized after 3-5 months when the tumors were approximately 0.5-1 cm in the largest diameter, in compliance with regulation for use of vertebrate animal in research.

2.3. Molecular biology techniques

2.3.1. Total RNAs extraction

Cells were washed with PBS 1X and then TRIzol® reagent (Invitrogen) was added (for less than $1*10^6$ cells add 500 µl, otherwise 1 ml of TRIzol).

The homogenized sample was incubated for 5' at RT, then was added 0,2 ml of chloroform per 1 ml of TRIzol reagent. Shake tubes for 15'' and incubate sample for 2-3' at RT then centrifuge samples at 11.800 x *g* for 15' at 4° C.

Collect aqueous phase and transfer it in a clean tube, then add 1.5 volumes of pure ethanol and mix. Supernatant was transferred into miRNeasy Mini columns and total RNAs was purified according to manufacturers 'instructions.

Accurate total RNA concentration was measured with a NanoDrop spectrophotometer at at λ= 260 nm. 260/280 ratios were above 2 for all samples used in this study.

2.3.2. Reverse transcription

We used miScript II reverse transcription kit (Qiagen) that allows to reverse transcribe miRNAs and mRNAs in a single step process. The kit comprises a poly (A) polymerase, which adds a poly (A) tail to miRNAs, a reverse transcriptase that converts RNA in cDNA, and a buffer that contains Mg^{2+}, dNTPs, oligo-dT primers and random primers. The reverse transcription mix was prepared as following:

miScript Hispec buffer 5X	4 µl
miScript Nucleics mix 10X	2 µl
Template RNA	250-1000 ng
RNasin® Ribonuclease Inhibitor	0,5 µl
miSript Reverse Transcriptase Mix	2 µl
RNase-free water	Up to 20 µl

Incubate sample for 60' at 37° C, then 5' at 95 ° C to inactivate miScript reverse

transcriptase mix, cDNA was diluted at a final concentration of 1 ng/µl for the following Real-Time PCR.

2.3.3. Quantitative RT-qPCR

After the RT, for microRNA expression we used Qiagen miScript Sybr Green PCR kit. We prepared a qPCR reaction master mix as indicated in the scheme (a 10% excess was always added to consider pipetting errors; all reaction were performed in duplicate):

Reagent	Volume/ reaction
2x miScript SYBR Green PCR master mix	10 µl
10x miscript Universal Primer	2 µl
10x miscript Primer assay	2 µl
Template cDNA	5 ng
RNase-free water	Up to 20 µl

qPCR detection of mature microRNAs was assayed with the following assays (Qiagen): miR-146a (MS00003535) miR-146b-5p (MS00003542), as controls we used SNORD61 (MS00033705) or SNORD72 (MS00033719).

2.3.4. miRNA HT profile

We used these reagents for the generation of the "normal stem cells miRNAs signature". We isolated 100 ng of total RNAs from PKH^{pos} and PKH^{neg} cells isolated from murine primary MECs. Total RNAs was reverse-transcribed with Megaplex™ RT Primers mix, which allows the RT of 381 microRNAs at once, by using 381 different stem-loop primers. After RT-reaction, we performed a step of cDNA pre-amplification step, to increase the sensitivity of detection of low expressed microRNAs. Megaplex™ PreAmp Primers (Rodent Pool A, 381 different primers) and Taqman PreAmp master Mix were used for the reaction (Applied Biosystems). For microRNAs HT profiling we used the TaqMan® Low

Density Array Rodent V2.0 (Applied Biosystems).

We isolated 40 PKH^{pos} and PKH^{neg} cells from MCF10A mammospheres, lysed directly in single-cell Lysis Buffer (Ambion). Total RNAs was reverse-transcribed with Human Megaplex™ RT Primers mix (377 different microRNAs and 4 controls). After RT-reaction, we performed a step of cDNA pre-amplification step with Human Megaplex™ PreAmp Primers (Pool A). For microRNAs HT profiling we used the TaqMan® Low Density Array Human V2.1 (Applied Biosystems).

Megaplex™ RT reaction for both the profiles was prepared according to the scheme

Component	Volume/ reaction
100 mM dNTPs	0,2 μl
MultiScribe Reverse Transcriptase (50U/μl)	1.5 μl
10x Reverse transcription buffer	0.8 μl
Megaplex™ RT Primers	0.8 μl
MgCl2	0.9 μl
RNase-inhibitor (20 U/μl)	0,1 μl
Nuclease–free water	0.2 μl

4.5 μl of RT mix was added to 3 μl of total RNAs (1-350 ng). We followed thermal cycling conditions indicated by manufacturers (Applied). Megaplex™ Preamplification reaction was prepared according to the scheme:

Component	Volume/ reaction
TaqMan® Preamp Master mix	12.5 μl
Megaplex™ PreAmp Primers	2.5 μl
Nuclease–free water	7.5 μl

2.5 µl of RT was added to 22.5 µl of Preamp reaction, and we followed thermal cycling conditions indicated by manufacturers (Applied). Then Preamp product was diluted according to manufacturer's protocol.

PCR reaction was prepared as followed:

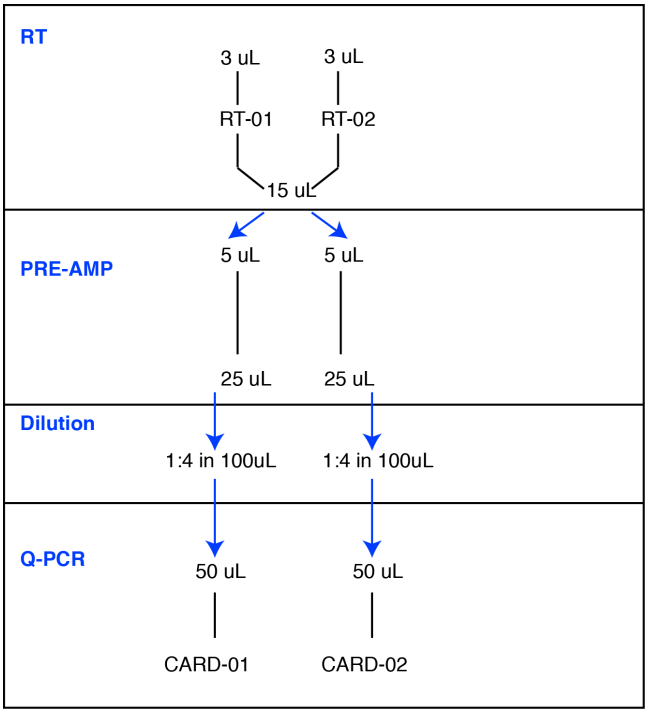
Component	Volume/ reaction
TaqMan® Universal PCR Master mix (no UNG)	450 µl
Diluted PreAmp product	9 µl
Nuclease–free water	441 µl

We loaded 100 µl mix/each port of the card array and we followed thermal cycling conditions indicated by manufacturers (Applied).

2.3.5. Low Sample Input protocol

We set-up this protocol for the reliable detection of 384 miRNAs from SCs/CSCs isolated from human breast primary biopsies. We collected by FACS sorting no more than 50-100 PKH^{pos} and PKH^{neg} cells on 96-well in 10 µl of single cells lysis kit plus DNAse (Ambion®).

Total RNAs was reverse transcribed using Human Megaplex™ RT Primers mix, followed by pre-amplification of the cDNA with Human Megaplex™ PreAmp Primers (Pool A). Then high-throughput (HT) qPCR profiling was performed on TaqMan human platform A V2.1 (Applied Biosystem). We improved the original protocol from Applied Biosystem, as reported in the figure:



Briefly, we increased the starting material for both RT and PreAmp reaction, the number of cycles of Preamplification and the total amount of cDNA loaded in each card. All these steps were performed in technical duplicates.

2.3.6. miRNAs HT data analysis

Raw data were exported on an excel datasheet (Microsoft®) and we discarded all the miRNAs with raw Ct>28 and the not expressed miRNAs (e.g Not Amplified). Expressed microRNAs (Ct<28) for each sample were then normalized with a series of housekeeping controls (RNU44, RNU48 and RNU6B for Human array, MammU6, snoRNA135 and snoRNA202 for Rodent array) present in each microfluidic card.

To keep into account the expression level of each microRNA, data were normalized through a scaling factor procedure. First, the geometric mean of the Ct values of housekeeping genes (HK) was calculated for each sample. Second, as reference we used the average of each HK geomean and the scaling factor for all other samples was

calculated in the following way:

Scaling factor for sample X = Averaged Geomean HK Ct (reference) – Geomean HK Ct (sample X)

The raw Ct value of every microRNA of a given sample was normalized according to the scaling factor.

ΔΔCT analysis: normalized Ct values were compared, using as a reference the median of expression of all the samples (i.e. sample X DDCT = median Ct – Sample X normalized Ct). Then we calculated the Log_2fold difference between PKH^{neg} and PKH^{pos} sample (i.e. PKH^{pos} Log_2fold = PKH^{neg} normalized ΔΔCt – PKH^{pos} normalized ΔΔCt).

We considered as upregulated miRNAs that had > +0.5 log2-fold difference between PKH^{pos} and PKH^{neg} (P-value<0.05) and downregulated miRNAs that had < -0.5 log2-fold difference between PKH^{pos} and PKH^{neg} (P-value<0.05).

2.3.7. mRNAs HT profiling and data analysis

We analyzed the mRNA gene expression of SUM159 WT 2D, WT 3D SCR 3D and miR-146 KD 3D, through the Affymetrix® Human Gene St array 2.1 platform. Analyses of gene expression data were performed using RMAExpress software (available at http://rmaexpress.bmbolstad.com). Microarray spot intensities below the minimum value of 3 Log_2 were excluded, and arrays were then normalized (centered) using the Robust Multi-Array Average normalization, not expressed genes were filtered out.

Principal Component Analysis, used to evaluate the variability among the different samples, was performed using JMP® software (SAS), on the median standardized data.

We identified 1975 genes that were significantly regulated in miR-146 KD vs controls (three independent biological replicas), > +0.25 log_2fold for upregulated genes and <- 0.25 log_2fold for downregulated genes, with p-value < 0.05 by Student's T-test with Benjamini and Hochberg multiple testing correction.

Gene Set Enrichment Analysis (GSEA, http://www.broadinstitute.org/gsea/index.jsp) was performed using the 1975 gene list of miR-146 KD regulated genes. As gene set to calculate the normalized enrichment scores (NES), we used four stem cells signatures (Polyak, Stingl, Pece and Visvader) subdivided in STEM_UP genes and STEM_DOWN genes. P-values were calculated by performing 1.000 random permutations of gene labels to create ES null distribution.

Seed enrichment analysis was performed by Sylamer software (http://www.ebi.ac.uk/research/enright/software/sylamer). All the Ref_seq expressed in SUM159 cells from the Affymetrix® Human Gene St array 2.1 platform (22442) were ordered from the most down-regulated to the most up-regulated upon overexpression of miR-146 in MCF10A cells or upon miR-146 KD in SUM159. This list was uploaded in Sylamer (van Dongen et al., 2008) to compute word enrichment. The analysis was performed with sequences of the 3′ UTR, which usually contains the MREs, as well as with the CDSs or the 5′ UTRs for the human genome (assembly: GRch37/hg19). Word sizes of 6, 7, and 8 were analyzed with standard parameters. The maximal enrichment score ($-\log_{10}$ p-value) observed for heptamers of each over- expressed miRNA was used for comparisons. For MCF10A dataset, the highest enrichment was observed in 3′ UTR sequences within the threshold used to define "down-regulated" genes (-0.2 $\log_2$ ratio), and the most represented words (7mers and 8-mers) exactly matched with the canonical seed of the overexpressed miR-146. We did not score any significant enrichment in 5′ UTR sequences. Conversely, in the 146 KD dataset we did not score any significant enrichment for either the 8-mer or the 7-mer of miR-146a/b in 3'-UTR, 5'UTR or CDS of regulated genes.

Ingenuity pathway analysis was performed using IPA software (Qiagen). The 1975 gene list of miR-146 KD regulated genes was analyzed running a core analysis, and we selected

all the upstream regulators with a z-score> |2| and with a p-value> 10^{-3}. The same analysis was performed for the four SC signatures (Polyak, Stingl, Pece and Visvader), sub selecting those upstream regulators with a p-value of overlap < 10^{-3}.

The lists of miR-146 putative targets were downloaded from Targetscan Human database (release 6.2, http://www.targetscan.org) and from MiRanda database (http://www.microrna.org/microrna/home.do, release Agust 2010).

2.3.8. Clustering and statistical analyses

Heatmaps were generated by Java TreeView software for Mac OS X. Statistical analyses were performed within the statistical software JMP9. Significance of the differences was calculated using ANOVA (in the case of more than two groups) or Welch's t test. Categorical data were analyzed with a contingency analysis within JMP (SAS). P-values were calculated with a Fisher's exact test.

2.3.9. AGO2-RIP

Ago2 RIP experiment was performed using Imprint® RNA Immunoprecipitation kit (Sigma Aldrich). Briefly, 10^7 cells (WT or miR-146 KD) were lysed in mild-lysis buffer (plus Protease Inhibitor Cocktails and Rnase inhibitor) for 15' on ice then the debris were pelleted at 16.000 g, 4°C for 10 minutes. Remove 5% of input from each cell lysate for RNA and protein controls. Protein A magnetic beads were pre-loaded with 2.5 µg Ago2 antibody/RIP (Rat monoclonal, clone 11A9, Sigma Aldrich) or 2.5 µg of IgGs from rat serum/RIP, at RT for 30 minutes with rotation.

Then, RNAs were immunoprecipitated with Ago2 antibodies or rat IgGs, overnight at 4°C with rotation. RIPs were then washed twice with mild wash buffer, RNA was purified with TRIzol® and miRneasy Micro kit and analyzed with RT-qPCR.

The protein lysate input (5%) and the 5% of RIP samples have been loaded in

Laemmli and analyzed on a 10% Acrylammide gel. Then, proteins were transferred on a nitrocellulose membrane through Trans-Blot® Turbo™ Transfer system (Biorad). The membrane was incubated with a blocking solution (Milk, 5 % w/v in TBST) and incubated with the primary antibody overnight at 4°C, according to the following conditions:

- anti-Ago2, rat monoclonal, Sigma Aldrich clone 11A9, 1µg/ml in milk;
- anti-Vinculin, mouse, from Sigma, 1:5000 in milk.

After washing with TBST, the membrane has been incubated with the secondary antibody 1 hour at RT, according to the following conditions:

- anti-mouse antibody from Cell Signaling, 1:10000 in TBST
- anti-Rat antibody from Sigma Aldrich, 1:10000 in TBST

Membrane was washed twice for 10 minutes at RT with TBST, then incubated with chemiluminescent substrate (ECL, peroxidase-conjugated) and developed with Chemidoc™ MP system (Biorad).

3 Results

3.1 Identification of a stem cell microRNAs signature

3.1.1. Isolation of normal human and mouse mammary stem cells

To identify a clear pattern of microRNA (miRNA) expression in normal stem cells (SCs), we took advantage of a technique, previously developed in our lab, to isolate a quasi-pure population of SCs through PKH26 dye dilution (Pece et al., 2010) (Cicalese et al., 2009). In this method, a lipophilic fluorescent dye, PKH26, is used to label mammary epithelial cells that are grown in conditions selective for SCs. When grown in suspension, mammary cells generate clonal spheroids called mammospheres, derived from a single SC (Dontu et al., 2003). PKH26-labeled epithelial cells were grown as mammospheres, so the dye is selectively retained by the slowly dividing/quiescent SCs (PKH^{pos}), while it is lost from actively dividing non-stem cells that constitute the largest proportion of mammosphere (PKH^{neg}). After mammosphere dissociation, stem and progenitor/differentiated cells can be isolated by FACS sorting, according to their differential levels of PKH26 epifluorescence (Pece et al., 2010; Cicalese et al., 2009). We purified PKH^{pos} and PKH^{neg} populations from two independent models: (i) mouse mammospheres, obtained from primary culture of mammary epithelial cells (MEC) depleted for endothelial and/or hematopoietic contaminants (Cicalese et al., 2009) (ii) human mammospheres, from a human normal cell line (MCF10A), which contains a SC-like cell population able to differentiate *in vitro* (Debnath et al., 2003). In order to exclude mammospheres not of clonal origin, human mammospheres were manually selected (> 70 μm of diameter) and checked by visual inspection for PKH26 dilution.

3.1.2. Identification of SC-specific miRNAs

Purification of SCs, even from established cultures, yields very low amount of starting material, ranging in normal mammary epithelial culture grown as mammospheres from the 0.1 to 0.3% of the total cells. Thus, we optimized the protocol for miRNA high-throughput profiling to achieve reliable and efficient detection of up to 384 small RNAs using TaqMan low density array. From human and mouse models, we FACS sorted up to 50 PKH^{pos} cells (and the same amount of PKH negative counterpart) in lysis buffer in biological duplicates, then we proceed with multiplexed reverse-transcription and miRNAs preamplification (Figure 13). As reported in Figure 13, we found 33 miRNAs differentially regulated in mouse stem versus progenitors/differentiated cells (log2fold > |0.5| between PKH^{pos} and PKH^{neg}, p-value< 0.05 two-tailed). In human model, we found five miRNAs that were differentially regulated in PKH^{pos} vs PKH^{neg} (log2fold > |0.5| between PKH^{pos} and PKH^{neg}, p-value< 0.05 two-tailed).

We identified three miRNAs commonly up- or down- regulated between human and mouse (p<0.01), as summarized in Table 1, and we defined these miRNAs "normal stem-cell signature".

miRNA	human MCF10A F.C. ± SD (log)	p-value (10A)	REG (MCF10A)	mouse MEC F.C. ± SD (log)	p-value (MEC)	REG (MEC)
hsa-mir-146a	1.86 ± 0.31	0.0006	UP in PKHpos	0.66 ± 0.15	0.0177	UP in PKHpos
hsa-mir-331-3p	1.47± 0.11	0.0005	UP in PKHpos	0.74 ± 0.04	0.0011	UP in PKHpos
hsa-let-7a	-1.19± 0.31	0.0041	DOWN in PKHpos	-3.22 ± 1.02	0.0285	DOWN in PKHpos

Table 1- The "normal stem cell signature"- the table reports the three miRNAs regulated in common between human and mouse normal stem cells, constituting the "normal stem-cell signature". Data are reported as log2fold difference (> |0.5| log2fold difference, p-value< 0.05) with standard deviation measured on two biological duplicates.

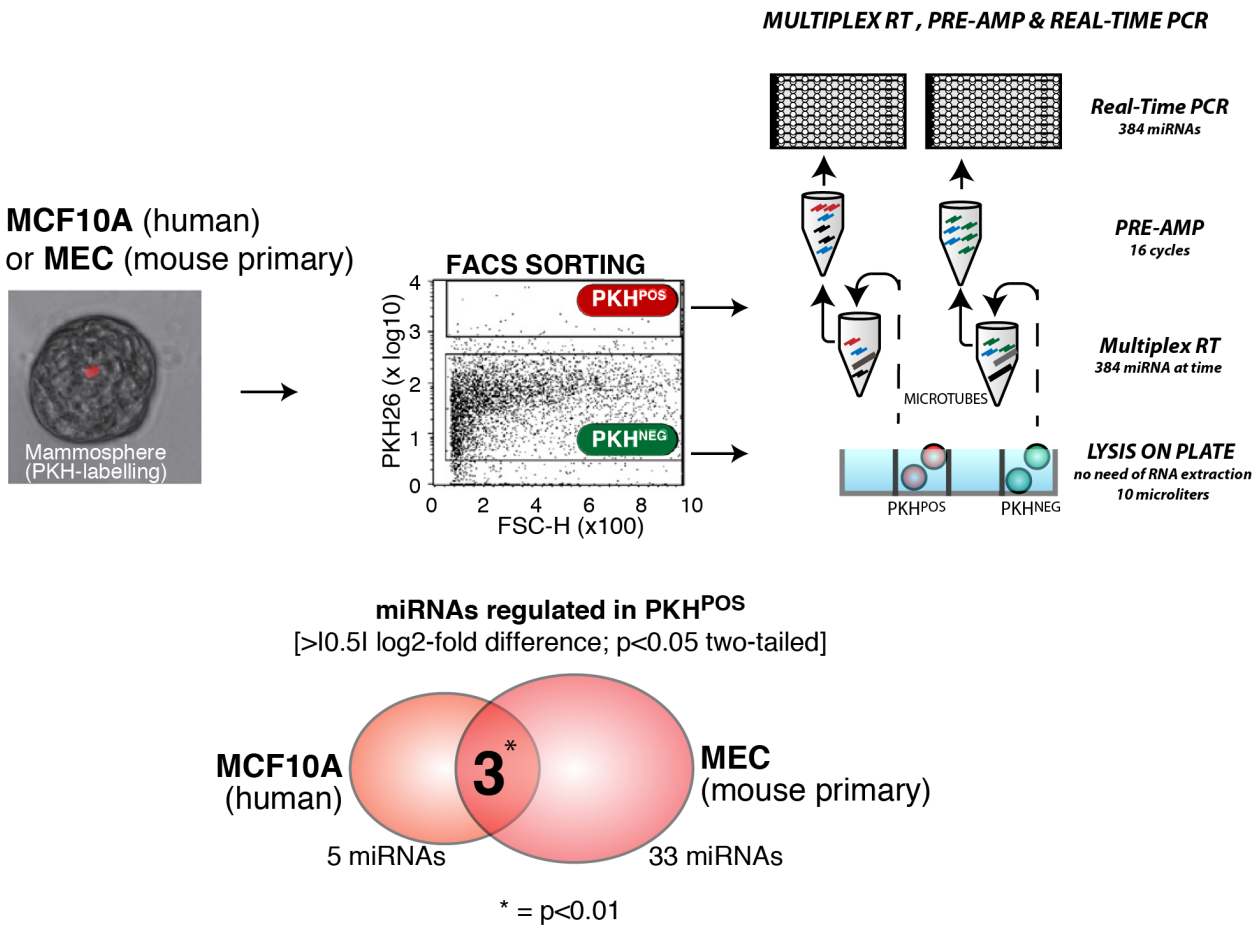


Figure 13- Identification of stem-specific microRNAs - Schematic representation of the method established for miRNAs profile from small number of cells. The same amount of PKH^{pos} or PKH^{neg} cells isolated by FACS were immediately lysed, in a buffer which is compatible with all steps required for quantitative RT-PCR including, multiplexed RT, pre-amplification to increase sensitivity (PreAmp) and Real Time PCR on a microfluidic device. Two independent preparations were used from each sample (human or mouse) to select miRNAs regulated in SCs (PKH^{pos}). The figure also reported the number of miRNAs regulated in PKH^{pos} cells isolated from MFC10A mammospheres and mouse primary mammospheres and the overlap between the two. Asterisk indicates significance of the overlap according to Fisher's Exact Test.

3.1.3. The normal stem cell signature stratifies human breast cancers according to their phenotypic aggressiveness

The cancer stem cell (CSC) hypothesis proposed that human tumors, including breast cancers, are heterogeneous and contain a sub-population of CSCs with properties reminiscent of normal stem cells, such as quiescence, self-renewal and unlimited replicative potential (Al-Hajj, 2003; Visvader and Lindeman, 2012). Furthermore, it has been demonstrated that high-grade, aggressive, estrogen receptor negative (ER-) human breast cancers display a higher frequency of CSCs/TICs compared to less aggressive, low-grade ER+ mammary tumors and normal breast epithelia (Pece et al., 2010). Indeed, the study of Pece and colleagues demonstrated that the transcriptional profile obtained from normal SCs could be used to predict the CSCs content in breast tumors.

Thus, we tested whether the " normal stem cell signature" that we identified, was able to stratify breast cancer patients in published datasets according to their biological characteristics. To address our hypothesis, we retrieved microRNAs-sequencing data generated by The Cancer Genome Atlas (The Cancer Genome Atlas Network TCGA) and we sub-selected 438 breast cancer samples for which histological and molecular characterization was available (ER, PR HER2 status, p53 and/or Myc amplification) and PAM50 classification (LumA, LumB, Basal and Her2 subgroups).

Figure 14 shows that hierarchical cluster analysis according to "normal stem-cell signature" identified three different breast cancers subgroups, named Group 1 (in red), Group 2 (in green) and Group 3 (in blue). Group 1 contains tumors that express low level of let-7a and high levels of miR-146/miR-331-3p, as observed for SCs, conversely Group 3 expresses high level of let-7a and low levels of miR-146/miR-331, similar to progenitors/differentiated cells. Finally, Group 2 contains tumors that express intermediate levels of both miRNAs. We next analyzed the correlation of these groups with histological and molecular subtypes by contingency analysis.

As shown in Figure 15, Group 1 was found: i) significantly enriched for basal-like tumors (61%, n=81, Pearson coefficient <0.001) ii) correlated with ER and PgR negative status ($p<0.001$) iii) enriched for breast tumors bearing p53 mutation (66%, n= 50, $p<0.001$) and Myc amplification (26%, n= 50, $p<0.001$). Conversely, Group 3 is enriched for luminal tumors, with no p53 mutations or Myc amplifications.

We verified these findings using an independent dataset from the METABRIC consortium, which consists of 997 breast cancer samples, analyzed by Illumina HT-12 V3 platform, for which information grade and long-term clinical outcome are available, in comparison with Cancer Genome Atlas dataset. We sub-selected 714 tumors from the total (we excluded the sample without grade information) and we hierarchically clustered them according to "normal stem-cell signature", as reported in Figure 16. This analysis identified one cluster, named cluster A, with high levels of miR-146 and low levels of let-7a, as in normal SCs (unfortunately the levels of miR-331-3p are not validated in this dataset). Contingency analysis on cluster A for group composition revealed clearly that, compared to the other clusters, it was significantly enriched for basal, G3 ER- tumors (95%, n= 135, $p<0.001$) and depleted of G1 ER+ tumors (5%, n=135), further supported by analysis on PAM50 classification (Figure 17).

According to both these analyses, we concluded that "normal stem cell signature" was able to stratify breast tumors based on their molecular subtypes and correlated with known markers of aggressiveness and putatively CSC content.

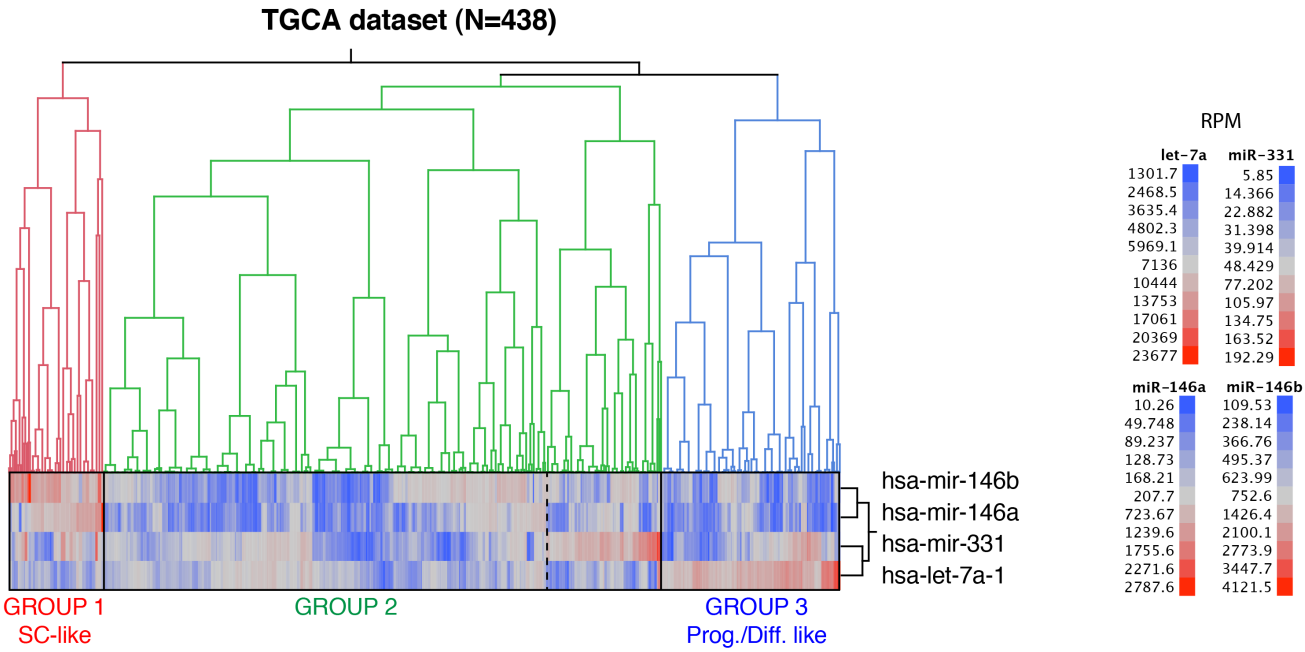


Figure 14- Breast cancer stratification according to "normal stem-cell signature"- Hierarchical clustering (Ward method, with standardized data) of 438 breast primary tumors from TCGA dataset, according to expression of miR-146a/b, miR-331-3p and Let-7a. The data are reported as reads per million mapped (RPM) in every tumor.

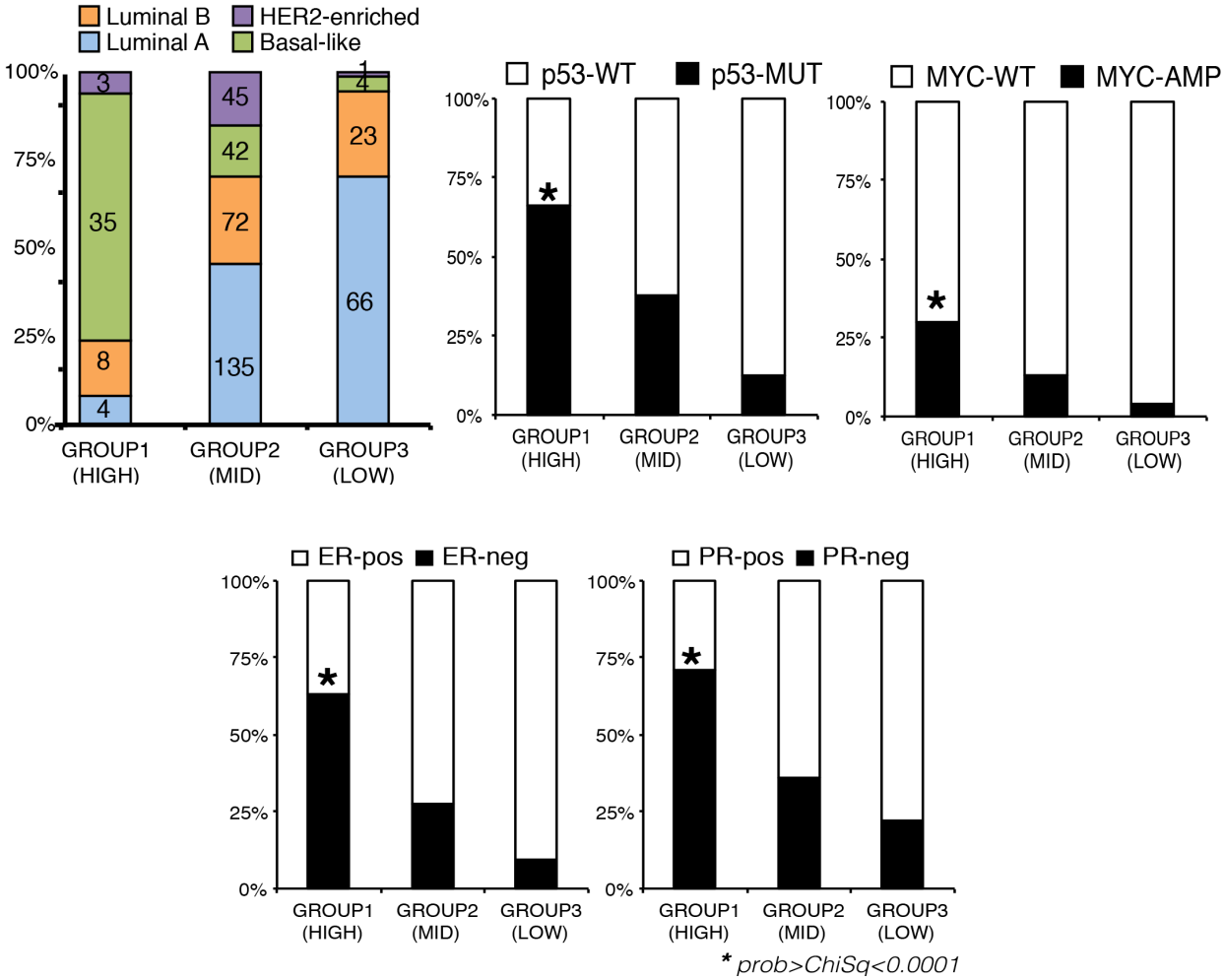


Figure 15- Contingency analysis on group composition - Subgroups identified by hierarchical clustering were analyzed according to PAM50 classification (number of samples are indicated) or p53 mutation/ Myc amplification, and hormone receptor status (ER and PR). Data are reported as percentage of the total number of samples and enrichments observed within group 1 are statistically significant (Chi-square Pearson coefficient <0.001, marked with asterisk).

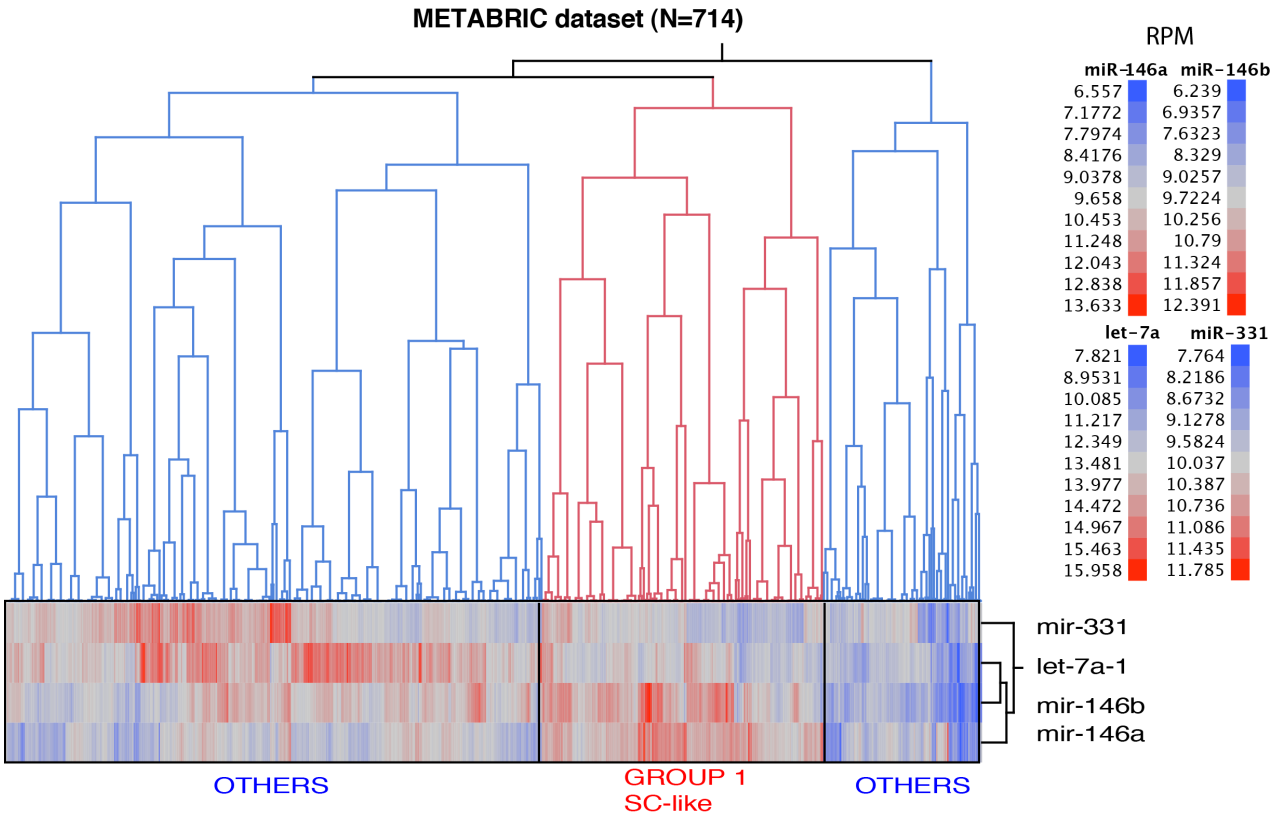


Figure 16- Breast cancer stratification according to "stem-cell signature"- Hierarchical clustering (Ward method, with standardized data) of 714 breast primary tumors according to expression of miR-146a/b, miR-331-3p and Let-7a (METABRIC dataset). The data are reported as reads per million mapped (RPM) in every tumor.

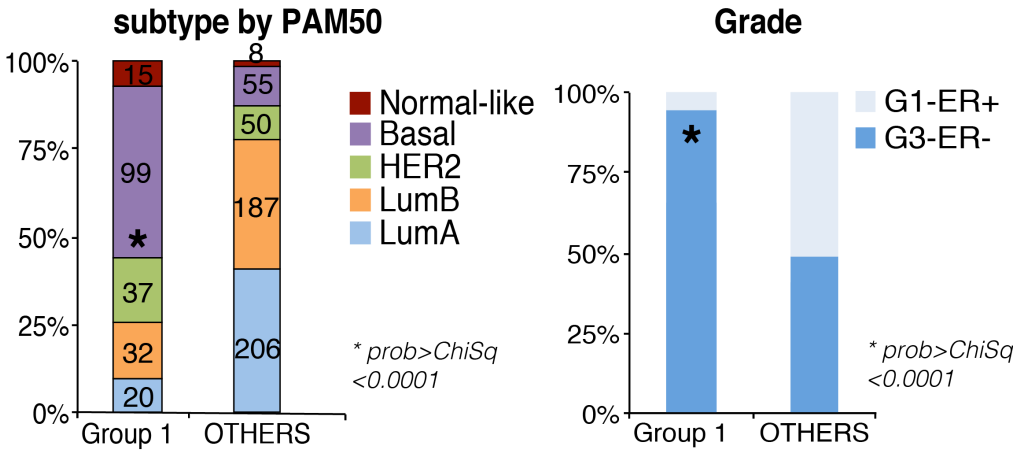


Figure 17- Contingency analysis on group composition- Subgroups identified by hierarchical clustering of METABRIC were analyzed according to PAM50 classification (number of samples are indicated) and grade. Data are reported as percentage over the total number of samples and enrichments observed within cluster A is statistically significant (Chi-square Pearson coefficient <0.001, marked with asterisk).

3.1.4. Validation of miR-146 as SC specific miRNA in human breast primary tumors

Shimono *et al*. demonstrated that normal breast stem cells and breast cancer stem cells (BCSCs) share common traits such as the regulation of stem cell properties and miRNA expression profiles. In support of this hypothesis, more recently, Pece et al. demonstrated that the transcriptomic profile obtained from human SCs could be used to predict the CSC content in breast tumors. Thus, we decided to test the hypothesis that the ”normal stem-cell signature” could be regulated in CSCs vs non-CSCs. We used the microRNAs expression data collected by Clarke lab from 11 human breast cancer patients. Clarke et al, isolated BCSCs from tumor specimens by using CD44 and CD24 surface markers, previously isolated by Al Hajj for BCSCs purification (Al-Hajj, 2003) and performed miRNAs expression profile of $CD44^{+}/CD24^{-}$ population, putatively enriched in BCSCs, as compared to non-tumorigenic $CD44^{-}/CD24^{+}$ counterpart (Shimono et al., 2009). As reported in Figure 18, miR-146 (both miR-146a and miR-146b) was consistently up regulated both in normal and cancer stem cell, strongly suggesting a role for this family in regulating stem cell biology.

To further corroborate the identification of miR-146 as *bona fide* SC marker, we isolated PKH^{pos} and PKH^{neg} from human primary breast biopsies. We collected 10 breast biopsy samples from patients that underwent breast surgery at IEO, divided in 3 normal-like and 7 tumors with different histology. Primary samples grew as spheroids *in vitro*, reflecting their content of SC/CSCs, but all had very low sphere-forming efficiency (SFE) (Pece et al., 2010), meaning that we had to deal with amount of cells ranging from 10 to 500 in total after one or two mammosphere generations. Figure 19 reports the workflow that we set up for SC profiling-grade purification from primary samples: after mammosphere dissociation, cells were cleaned up from human endothelial and hematopoietic

contaminants by MACS columns, and stained with DAPI to eliminate dead cells before FACS sorting.

We collected 100 and 50 cells from 1 normal and 1 tumor primary sample to set up a protocol for miRNAs detection from low-amount of starting material (low sample input, LSI protocol) that was compared with the standard TaqMan protocol from 10 ng of RNA coming from the same samples. Figure 20 shows the percentage of miRNAs detected with the standard protocol vs the LSI protocol, we reached almost the 80% of detection in LSI compared to standard protocol, even when starting from as little as 50 cells.

Having established a reliable protocol, we collected PKH^{pos} cells with their PKH^{neg} counterpart from the 10 primary samples and we analyzed the levels of miR-146a and miR-146b. As described in Figure 21, miR-146b is consistently up modulated in all the samples, with a 1.5 log_2fold difference between PKH^{pos} and PKH^{neg} cells, independently from the molecular subtype of breast tumors. This level of regulation is comparable with the one observed by Clarke in tumoral and normal samples. Instead, differently from the normal stem cell signature and from Clarke's data, we did not appreciate any significant regulation for miR-146a.

Together, our data point to a specific family of microRNAs, miR-146, being specifically accumulated in the stem compartment of both normal and cancer samples. Next, we focused on the biological validation of miR-146 as stem-specific miRNA and on its ability to modulate stem properties.

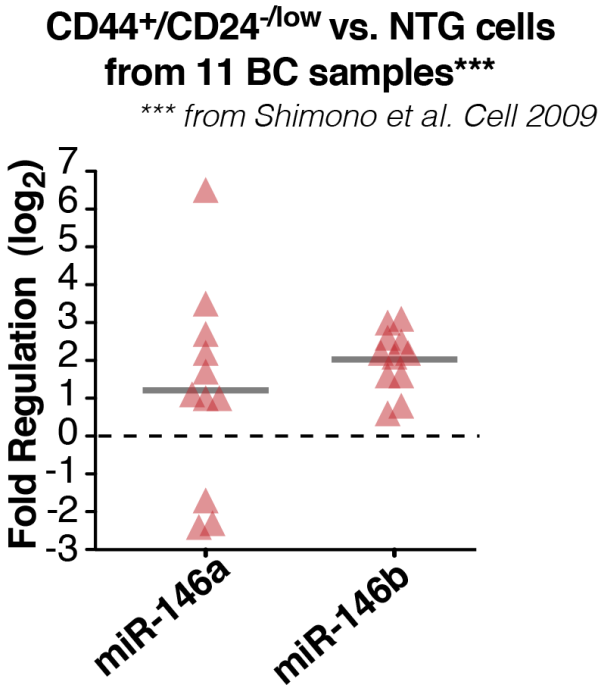


Figure 18- MiR-146 family regulation in BCSCs vs non tumorigenic CD44$^-$/CD24$^+$– The panel recapitulates the regulation of miR-146a and miR-146b measured in BCSCs isolated with surface markers by (Shimono et al., 2009). The data are reported as log2fold of CD44$^+$/CD24$^-$ over CD44$^-$/CD24$^+$ (NTG, non tumorigenic) counterpart, each triangle represents a different tumor and mean value of regulation is reported.

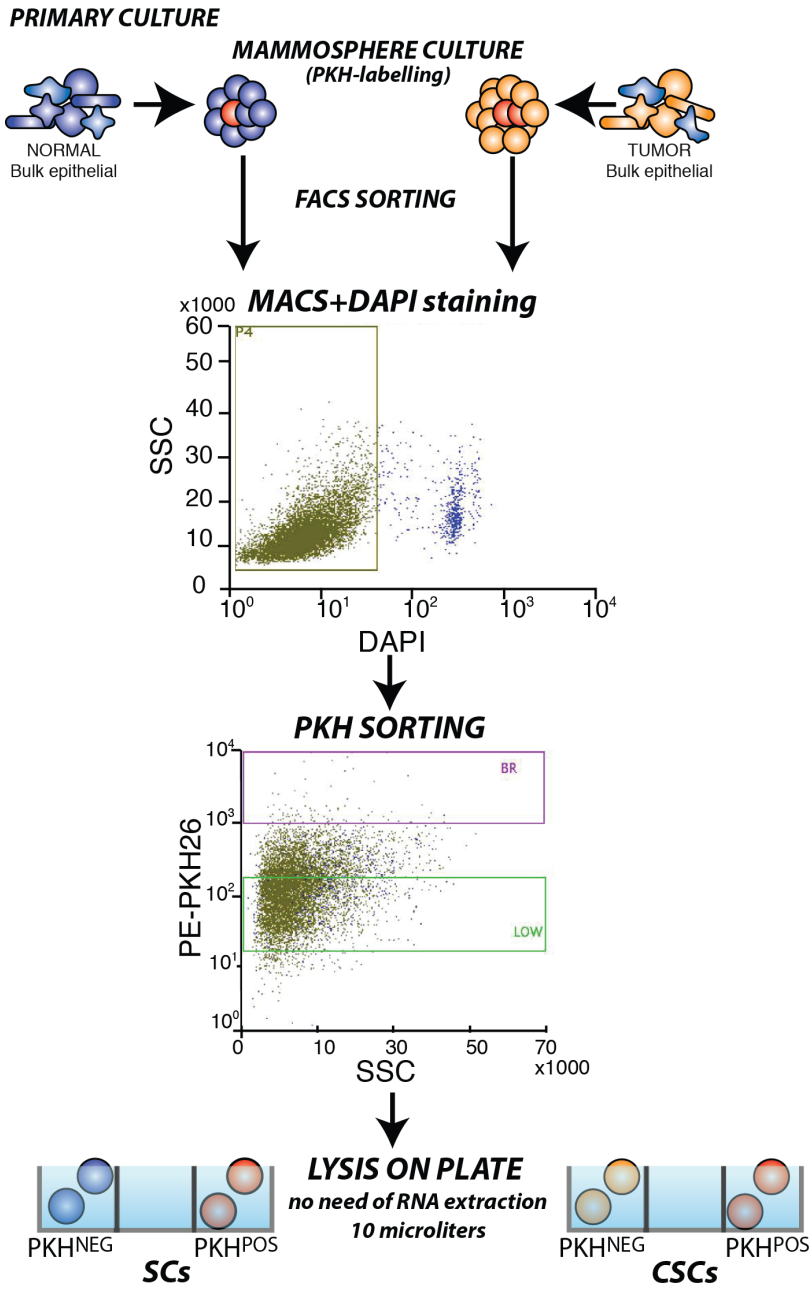
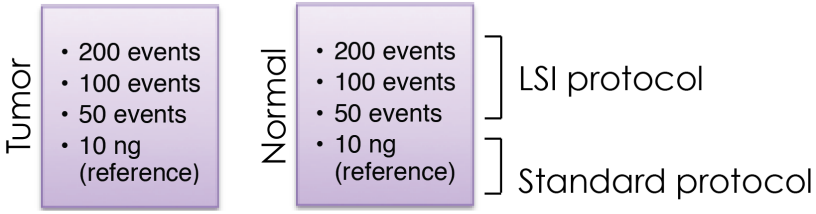
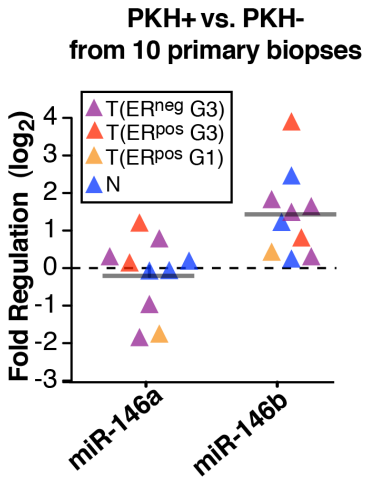


Figure 19– Scheme of microRNA HT profile from human primary samples- PKH-labeled mammospheres were dissociated to single cells and separated from non epithelial contaminants with CD31/CD45 MACS columns (Miltenyi). Before sorting, living cells were stained with DAPI for 1 minute then sorted according to PKH content in 96-well plates containing lysis buffer (Ambion). Total RNA was immediately reverse transcribed and preamplified for miRNAs detection, according to the manufacturer's instructions.



	REF	LSI	LSI	REF	LSI	LSI
Type	Tum	Tum	Tum	Norm	Norm	Norm
N. Cells (FACS)	10 ng	100	50	10 ng	100	50
Total miRNAs	48	48	48	48	48	48
Detected <30 Ct	39	40	35	35	34	34
% detected	81.3 %	83.3%	72.9%	72.9%	70.8%	70.8%

Figure 20- Set up of LSI protocol for HT miRNA profiling from human primary samples- we collected 100 and 50 FACS-sorted events in lysis buffer from 1 tumor and 1 normal sample. We used 10 ng of RNA from the same samples as reference. The reference was reverse transcribed and preamplified according to manufacturer's protocol; the LSI protocol introduced 16 cycles of preamplification and different dilutions of the RT and preamp material. The table reports the number of detected miRNAs under the threshold of 30 cycles out of a total 48 miRNAs analyzed.



NAME	ER (%)	PGR (%)	Her-2/Neu	Ki-67 (%)	Staging	Grading
12M68T	95	95	Neg.	28	pT3 pN3a	G3
12M82T	5	5	90	45	pT2 pN3a	G3
12M129T	95	95	Neg.	10	pT2(is) pN0(sn)	G1
12M136T	Neg.	Neg.	1+	70	pT3 pN3a	G3
12M154T	Neg.	Neg.	1+	27	pT3 pN1a	G3
12M182	90	90	1+	25	pT2(m)	G3
12M247T	Neg.	Neg.	2+	70	pT2 pN1a	G3
12M93N	/	/	/	/	/	N
12M247N	/	/	/	/	/	N
12MR2	/	/	/	/	/	N

Figure 21– miR-146 family regulation in PKH^{pos} and PKH^{neg} cells – Left panel- the plot summarizes the regulation of miR-146a and miR-146b measured in SCs isolated from 7 tumors and 3 normal samples. Data are reported as log_2fold difference between PKH^{pos} vs PKH^{neg} cells, each triangle represents different sample and mean value of regulation is reported. Right panel-the table reports the histological characteristics of the samples used for miR-146 analysis.

3.2. Validation of the biological properties of miR-146 as stem-specific miRNA

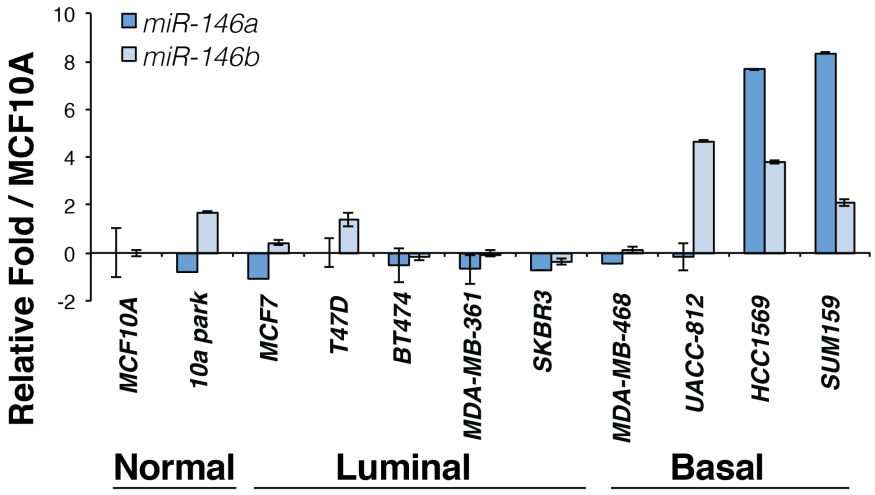
3.2.1. miR-146 expression correlates with basal breast tumors

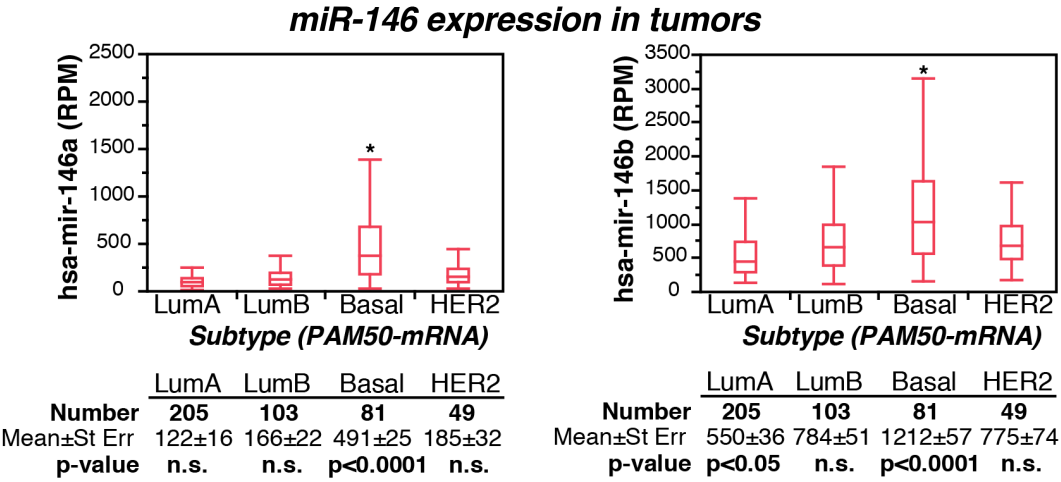
We analyzed the level of miR-146 expression in a panel of mammary epithelial normal and cancer cell lines to find a suitable model for further characterization of the biological role of miR-146. We collected total RNA from 1 normal cell line (MCF10A and its derivative clone MCF10A-Park), 5 luminal (MCF7, T47D, MDA-MB-361 and SKBR3) and 4 basal tumor cell lines (MDA-MB-468, UACC812, HCC1569 and SUM159) and then we measured the relative fold change of miR-146 in tumoral cell lines compared to normal. As reported in Figure 22A, while miR-146a and miR-146b were low/absent in normal and luminal subtypes, they were consistently overexpressed in basal tumor cell lines, with an average $\log_2$fold difference from 2 to 8 from normal cell lines.

To assess if miR-146 are overexpressed in basal tumors, we selected 438 primary tumors from the TCGA dataset for which PAM50 classification was available to classify tumor subtypes (Luminal A, Luminal B, Basal and Her2) according to a 50-gene expression pattern. We measured the difference in mean reads per million (RPM) between miR-146a and miR-146b in each subtype. Figure 22B shows that, similarly to the observation in cell lines, basal tumors are the most enriched for both miRNA-146a and -146b, in a statistically significant manner (p-value<0.0001, Each Pair, Student's t test), with an averaged number of RPM that is 2-3 times higher than in the other molecular subtypes. Both these observations suggest that basal tumors are associated with elevated levels of miR-146 family.

We thus chose SUM159 as a model system to manipulate miR-146a/b for several reasons: i) it expresses high endogenous levels of miR-146a and -146b ii) it resembles a triple negative breast cancer model that is *per se* the most aggressive subtype iii) it possesses a $CD44^+/CD24^-$ component that *in vitro* behaves as a breast cancer stem-like enriched

population (high efficiency of mammosphere formation in non-adherent conditions and high frequency of TICs assessed by *in vivo* transplantations) (Fillmore and Kuperwasser, 2008; Gupta et al., 2011).





	LumA	LumB	Basal	HER2
Number	**205**	**103**	**81**	**49**
Mean±St Err	122±16	166±22	491±25	185±32
p-value	**n.s.**	**n.s.**	**p<0.0001**	**n.s.**

	LumA	LumB	Basal	HER2
Number	**205**	**103**	**81**	**49**
Mean±St Err	550±36	784±51	1212±57	775±74
p-value	**p<0.05**	**n.s.**	**p<0.0001**	**n.s.**

Figure 22– miR-146 expression levels in breast cell lines and in primary breast cancers from the TCGA dataset – Upper panel: expression levels of miR-146a/b measured by RT-qPCR in normal, luminal and basal breast cell lines. Shown is the $\log_2$fold change of miR-146a and miR-146b normalized with respect to the expression of the normal cell line MCF10A. As housekeeping miRNA, we used snora73 (Qiagen, miScript system). Lower panel: from TCGA dataset, we selected 438 primary tumors for which PAM50 classification was available (LuminalA, LuminalB, Basal and Her2). Expression level of miR-146a and miR-146b is indicated as reads per million (RPM) and shown as box plots, mean level and *= p-value <0.0001 (Each Pair, Student's t test 0.05) are also reported in the table.

3.2.2. Endogenous high levels of miR-146a/b correlate with stem properties

To investigate if different endogenous levels of miR-146a/b were associated with populations with diverse stem cell properties, we generated a miR-146a/b GFP sensor containing four perfect complementary sequences for miR-146 at the 3'UTR of a destabilized GFP protein (Brown et al., 2007) in a lentiviral backbone. When miR-146a/b are expressed in the cell, they bind the 3'UTR of GFP mRNA, causing its degradation. Thus, miR-146a/b levels inversely correlate with GFP expression: the GFP^{high} population expresses low levels of miR-146a/b as opposed to GFP^{neg} population with high miR-146a/b levels. To normalize the number of lentiviral integrations, the vector expresses a truncated form of the nerve growth factor (ΔNGFR, generally not expressed in non neuronal cells), useful to select $\Delta NGFR^{+}$ population by FACS sorting Figure 23A.

First, we tested the responsiveness of the miR-sensor to exogenous changes in miRNA levels, by over-expressing miR-146 in MCF7 cells, which express low endogenous levels of miR-146a and miR-146b. As controls, we used a miR-sensor generated for miR-34a in response to miR-34 overexpression and scrambled (SCR)-sensor, with a non-targetable sequence at the 3'UTR of GFP. As shown in Figure 23B, the miR-146a/b sensors responded very efficiently to miR-146 over-expression, reducing the GFP positive population by more than the 50%, similarly to miR-34 sensor in response to miR-34a overexpression. As expected, the SCR-sensor did not alter its GFP expression in response to miR-146 overexpression.

Then, to assess how miR-146a/b levels affect SUM159 cells, we infected SUM159 with a SCR sensor or a miR-34a, miR-146a or miR-146b sensor. As shown in Figure 24, we selected a population $\Delta NGFR^{+}$ expressing homogeneous levels of the receptor, then we stratified the population according to GFP content ($\Delta NGFR^{+}$ GFP^{high} and $\Delta NGFR^{+}$ GFP^{neg}): both the SCR-sensor, as well as miR-34a, showed 90-100% of GFP positivity, in fact miR-34a is not expressed in SUM159. Instead, miR-146 sensors responded

efficiently to high endogenous levels of miR-146 in SUM159, in fact the GFP^{high} population was decreased to 10-15% compared to the controls.

After sorting, the cells were then tested for their ability to form mammospheres *in vitro*. As illustrated in Figure 25, cells that endogenously express low levels of mir-146a/b showed impaired mammosphere formation capacity compared to their miR-146a/b high counterpart, which was similar to the control vectors (p-value< 0.001, two biological replicas).

We repeated the same experiment in primary mouse MEC to see if miR-146 expression could stratify normal breast stem cells as well. Although we performed only one experiment (the second experiment is ongoing), we confirmed that miR-146 expressing cells retained the ability to form mammosphere *in vitro*. Conversely, in absence of miR-146 expression, cells failed to form mammospheres (Figure 25).

Taken together, our results underline a possible role for miR-146a/b as miRNAs specifying stem cell traits.

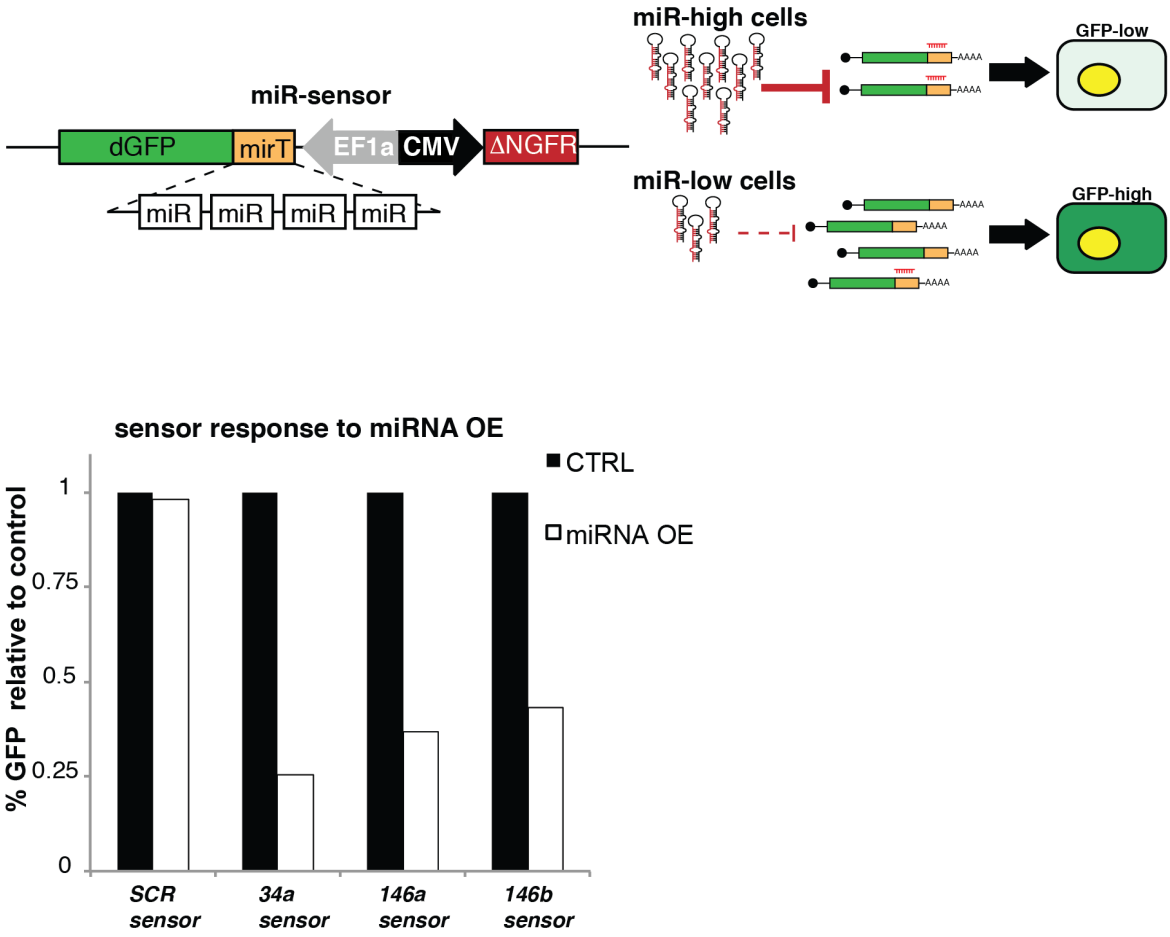


Figure 23– Generation of miR-146 sensor – upper panel: Schematic representation of the lentiviral vector miR-146a/b GFP sensor (courtesy of L. Naldini, (Brown et al., 2007)). Lower panel: Response of the miR-sensor to miRNA overexpression: MCF7 cells were transfected with 50 nM of miR-146a, miR-146b or miR-34a oligos, SCR oligo was used as control (CTRL, Qiagen). After 48h cells were analyzed for GFP expression. Data are represented as % of GFP positive cells relative to control.

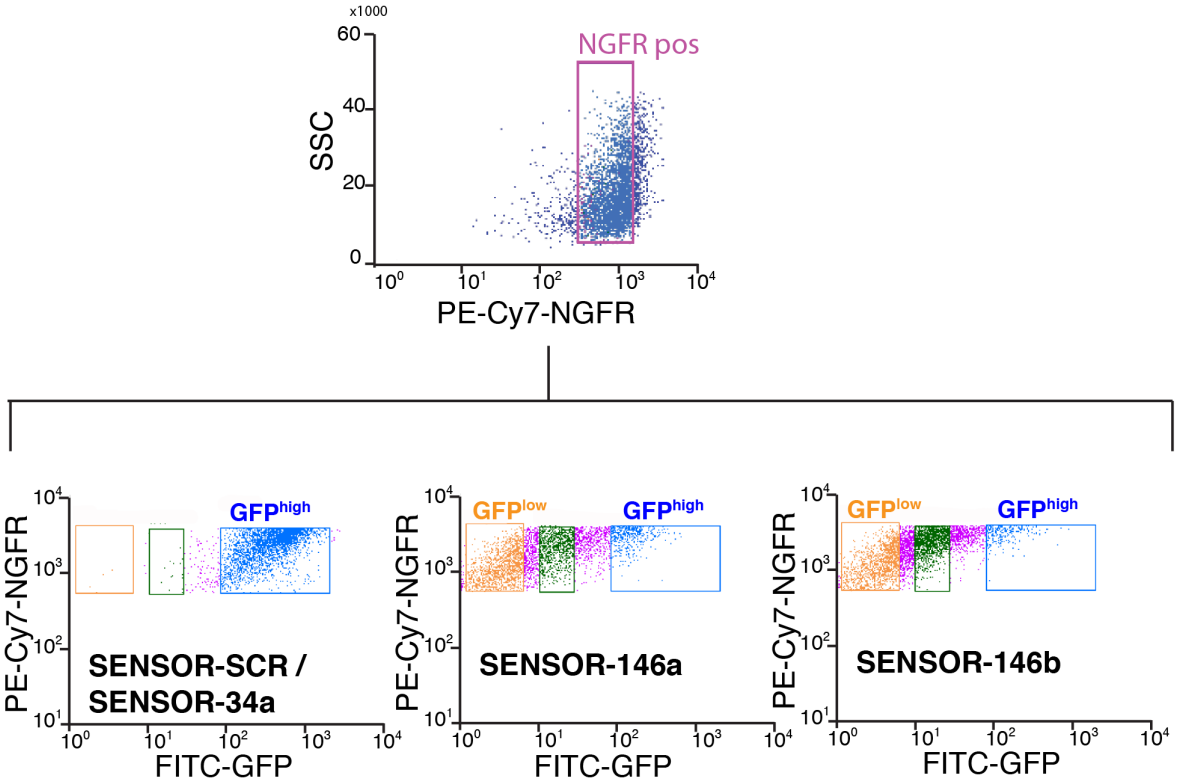


Figure 24- FACS sorting of miR-sensors- Upper panel: SUM159 cells were infected with SCR sensor or a miR-34a, miR-146a or miR-146b sensor. After infection, we FACS sorted the cells, according to ΔNGFR levels and then ΔNGFR^{+} cells were stratified according to its GFP content. Lower panel: FACS plot of ΔNGFR^{+} GFP^{high} and ΔNGFR^{+} GFP^{neg} populations of SUM159 infected with miR-146 sensor compared to CTR vectors (SCR and/or miR-34a, used as reference).

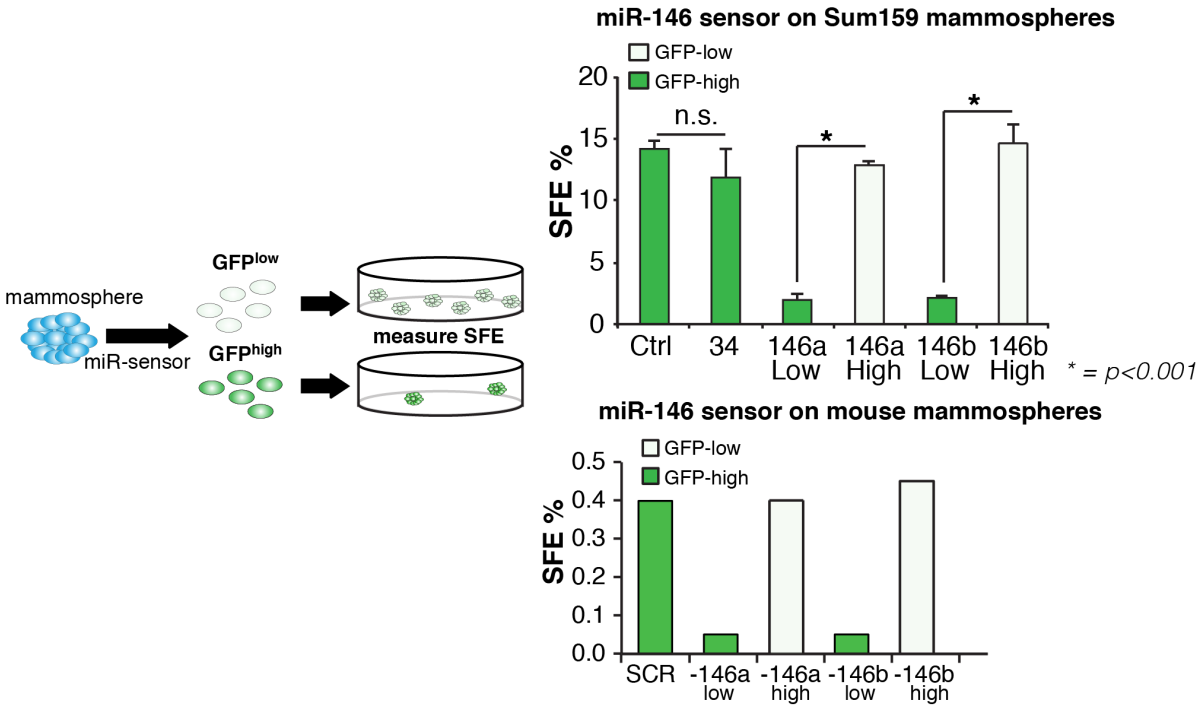


Figure 25– High levels of miR-146a/b correlate with stem properties – Top: mammosphere assay on miR-146 low versus miR-146 high SUM159 cells: sphere-forming efficiency is reported as the percentage of spheres with respect to the total number of cells plated. Error bars denote the standard deviation of two independent biological replicas. Bottom: mammosphere assay on miR-146 low versus miR-146 high in mouse primary cells: sphere-forming efficiency is reported as the percentage of spheres with respect to the total number of cells plated

3.2.3. Loss of miR-146 reduces stem cell self-renewal *in vitro* and *in vivo*

Having demonstrated that diverse endogenous levels of miR-146 can discriminate cells with different stem properties, we decided to perform high-resolution studies manipulating miR-146a/b levels to determine their relevance to stem cell phenotype. We used two model systems: SUM159 as model of BCSCs and mouse primary MEC as source of normal stem cells. To knockdown miR-146 expression, we infected SUM159 cells with a lentiviral vector expressing an "antimiR" for endogenous miR-146a/b and co-expressing turbo-GFP reporter protein and puromycin selection (miRzip, SBI system bioscience). Briefly, the antimiR is an exogenous molecule perfectly complementary to a miRNA and sequesters mature miRNAs within the cytoplasm, thus preventing them from binding to their targets (Ebert et al., 2007). We used a non-targeting antimiR (CTRL) as control. RT-qPCR was used to check the level of miR-146-KD achieved using the miR-146 antimiR: as reported in Figure 26A, the level of KD reached 50% for miR-146a and 70% for miR-146b with respect to the control levels. Nevertheless, measuring the KD levels by RT-qPCR is not the optimal method, due to the fact that the dimer miRNA: antimiR is not always degraded, leaving the mature miRNA levels unaltered. Thus, we performed an Ago2-RIP experiment to verify whether the mature miR-146a/b were less loaded on Ago2, the protein of miRISC complex necessary for the binding of mature miRNA to its targets. We immuno-precipitated RNAs in complex with Ago2 from cells not infected vs miR-146 KD. As control for specificity we used negative control IgGs and we compared the enrichment of mature miRNA from both the samples (Ago2 and IgG). Figure 26 shows that: i) as expected, miRNAs, but not small nuclear RNAs such as snoRNAs (snord72 or snord6) are enriched into the Ago2 in the two conditions, compared to IgG ii) there was no unspecific binding of miRNAs on IgG in both the samples iii) the percentage of miR-146 loading on Ago2 was strongly impaired in miR-146 KD compared to WT, and this effect was not observed for two unrelated microRNAs as Let-7a and miR-26a upon miR-146 KD.

We then analyzed the impact of miR-146 KD on cell proliferation by growth curve analysis; Figure 26B reports that there was a 20-25% of cell growth reduction on SUM159 in the miR-146 KD compared to controls and that this effect was consistently observed in three different biological replicas. Therefore, miR-146 KD does translate in reduction of miR-146 loading on the miRISC and correlates with SUM159 reduced growth.

We then used the mammosphere assay to measure the effect of miR-146 KD on stem cell properties. We infected SUM159 cells and after puromycin selection, GFP+ cells were plated in non-adherent conditions with methylcellulose to prevent cell aggregation. We measured mammosphere propagation for at least 4 generations. At each generation, we acquired the images of all the spheroids (in 12-well plates, in duplicate) and we counted GFP^+ mammospheres greater than 70 μm of diameter. We used these data to calculate the sphere forming efficiency (SFE). Figure 27A summarizes the results: KD of miR-146 reproducibly reduced by 30-50% and in a statistically significant way the sphere-forming capacity of SUM159 (p-value<0.001, 3 independent experiments). Importantly, this effect was maintained for at least 3 generations of mammospheres, suggesting that miR-146 expression is necessary to regulate the self-renewal of cancer stem cells. We performed the mammosphere propagation assay using normal stem cells from MECs of FVB mice, and as shown in Figure 27A, the loss of miR-146 expression strongly impaired stem cell self-renewal, almost reducing the SFE to zero. Finally, we performed a tumorigenicity assay *in vivo*, injecting orthotopically 1000 GFP+ SUM159 cells, CTRL vs miR-146 KD, in the mouse mammary fat pad of immunocompromised mice (NOD SCID ILR2 gamma null). After tumors became palpable, animals were sacrificed and number and volume of each tumor was measured. As reported Figure 27B, 8/8 animals injected with CTRL vector developed tumor, with a mean volume of 485.8 mm^3, compared with 7/8 animals injected with 146 KD with a mean tumor volume of 46.6 mm^3, indicating approximately a 10-fold reduction of the mean tumor dimension. To further verify if self-renewal capacity is

affected *in vivo*, we are actually retransplanting these tumors, to measure the TIC ability in the 2nd generation of tumors.

Taking our *in vitro* and *in vivo* experiments together, we can conclude that the loss of miR-146a/b affects stem cell self-renewal both in normal and cancer stem cells, suggesting a potential role as a target to inhibit the expansion of CSCs in tumors.

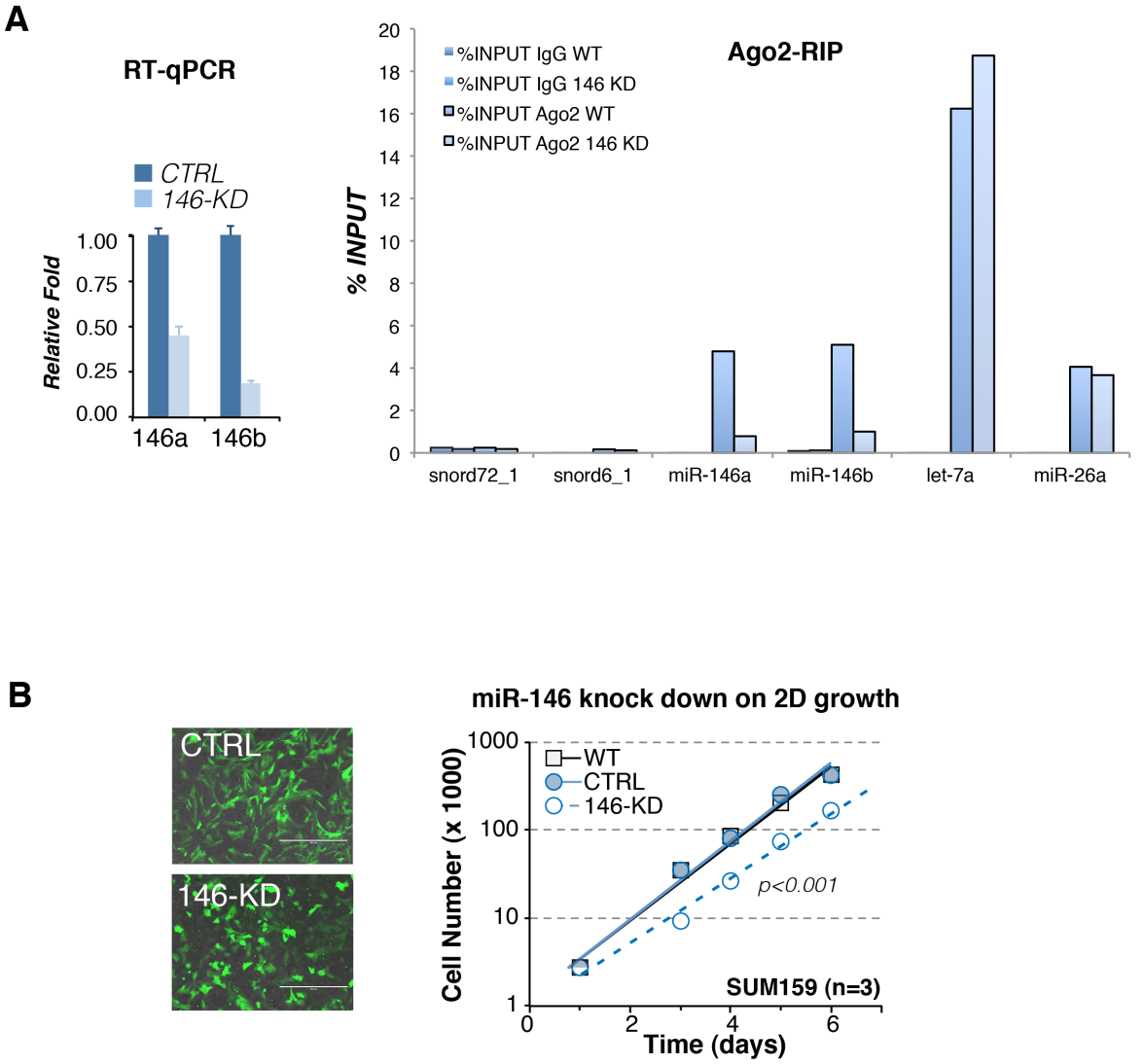


Figure 26- Testing the effects of miR-146 KD - A) Left panel: expression level of miR-146a/b measured by qRT-PCR in control cells (CTRL) vs miR-146 KD cells (146-KD). miR-146a and miR-146b were normalized on housekeeping snord6_1. Right panel: Ago2 RIP experiment was performed on lysates from 10^7 cells (WT or miR-146 KD) with 2.5 µg Ago2 antibodies or 2.5 µg of IgGs. The plot reports the % of miRNAs loaded on Ago2 measured as (copies in $RIP_{Ago2\ or\ IgG}$/ total number of copies)*100; as reference we used 5% of Input. miR-146a/b showed a reduction of 80% of Ago2 loading in miR-146 KD compared to WT, differently from Let-7a and miR-26a that are not significantly affected. The plot is an average between two independent biological replicas. B) Left panel: representative images of cells infected with CTRL or miR-146 KD lentiviral vector, 10X magnification, scale bar 400 µm. Right panel: the growth curve was performed by plating 2800 cells infected with CTRL or miR-146 KD in 12-well plates and counting cells every 24h, the graph reports the total number of cells at each time point as average of three biological replicas.

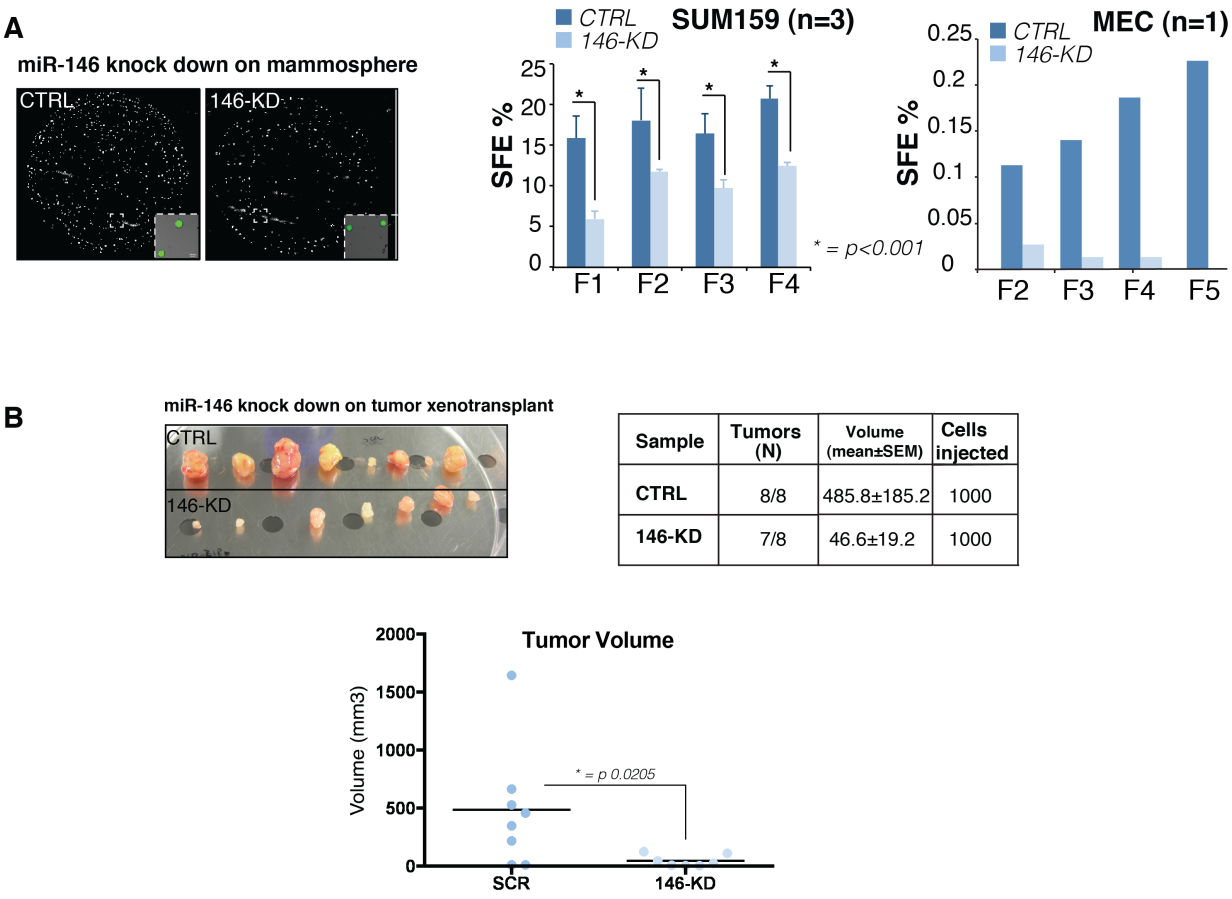


Figure 27– Biological effects of miR-146 KD- A) Fluorescent image acquisition of GFP+ mammospheres in CTRL and 146-KD in 12-well format, inset panel is a magnification of white boxed areas. Serial mammosphere passage upon miR-146a/b KD: sphere-forming efficiency is reported as % (total number of spheres/total number of cells plated). For SUM159 data are plotted as mean of three independent biological replicas and the error bars denote the standard error of the mean, p-value< 0.001 compared to CTRL. For MECs (mouse epithelial cells) data are plotted as result of one experiment. B) 1000 GFP+ cells expressing CTRL or miR-146 KD vector were injected orthotopically in mouse mammary fat pad in 8 mice per condition. Tumor growth was measured every week until they became palpable and the volume was measured by caliper. The right panel showed the number and mean volume of tumors as mean± standard error of the mean (SEM). In the plot, each dot represents the tumor volume in mm^3 from one mouse as measured with the formula $V= [(width)^2 * length]/2$; p-value was measured with two-sample t Test to be < 0.0205.

3.2.4. Generation of patient-derived xenografts (PDXs) for *in vivo* manipulation of miR-146

SUM159 are a powerful model for the dissection of miR-146a/b functions using both *in vitro* and *in vivo* approaches. However, they are strongly limited by being a tumor cell line that departs significantly from real human tumors. For example, when injected in immunocompromised mice, tumor cell lines gave rise to tumors that are generally homogenous, not resembling at all the heterogeneity that exists in primary tumors from patients. Thus, patient–derived xenografts are being used as a more powerful tool that better maintains tumor heterogeneity *in vivo* and promote a better understanding of tumor biology in a preclinical setting (Williams et al., 2013). Thanks to the infrastructure provided by the IEO biobank (IBBRI) we had access to tumor tissues isolated from patients who underwent breast surgery at IEO hospital. We collected five primary tumors, for which histological parameters were analysed via immunohistochemistry by the IEO Tissue bank as reported in Table 2. After orthotopic transplantation, all the tumors engrafted in NOD Scid Gamma (NSG) mice, with latency time ranged from 3-5 months for basal/triple negative tumors to 10 months for the Luminal B tumor, that is generally less aggressive. We propagated the solid tumor tissue *in vivo* for 2-3 serial transplantations and examined whether the masses originated were of human epithelial origin and whether they preserved the histological characteristics of the original tumor. We analyzed a complete panel of markers on the first two tumors and its derived xenografts, including: i) H&E staining for evaluating tissue histology; ii) human Ki67 staining as index of proliferation; iii) human receptor staining (estrogen receptor ER, progesterone receptor PgR and HER2 receptor) to classify tumor subtype; iv) expression of human cytokeratins, as a readout of basal (CK5+) or luminal (CK8+) cancer cells and lastly v) E-cadherin and vimentin to evaluate a possible process of epithelial to mesenchymal transition (EMT) *in vivo*.

Figure 28 and Figure 29 summarized the results of the analyses performed on one of the PDX and its original tumor, namely 12-B1-00339p; Figure 28 showed that the architecture of infiltrating ductal carcinoma, arranged in nests and solid narrows of cells poorly differentiated, is still maintained in the xeno as reported from HE staining. Proliferation index (Ki67) was further increased from 30% of the original tumor to 70% of the PDX, indicating a highly proliferative carcinoma. As for tissue architecture, the receptors status and markers expression were maintained in the xenograft. Tumor cells had very low positivity to ER (less than 2%), negativity for PgR and a score 1+ to HER2 (according to FDA scoring); expressed high levels of basal cytokeratins (CK5) and diffused low levels of luminal cytokeratins (CK8) and several regions showed high levels of both E-cadherin and vimentin (Figure 29).

The same characterization was performed on the second PDX, named 12-B1-00197p. The original tumor possessed the characteristic of an infiltrating ductal carcinoma, poorly differentiated and constituted of solid cell nests with high proliferative index (Ki67 around 35%) and no positivity to human receptors. The xenografts maintained exactly the same histological characteristic, apart for an increased Ki67 index around 80% (Figure 30). Cancer cells were highly positive to CK5 (basal marker) and negative for CK8 (luminal marker) in the original tumor, with large areas of positive and mainly cytoplasmic E-cadherin and vimentin (Figure 31). In the xenograft, cancer cells were almost totally positive for the basal cytokeratin (CK5), negative for the expression of the luminal cytokeratin (CK8), and showed an increased expression of vimentin and reduced expression of E-cadherin, likely suggesting an ongoing epithelial-to-mesenchymal transition (EMT) (Figure 31).

By these characterizations, we concluded that the PDXs resembled quite accurately the histology and molecular characteristics of their primary tumors, even though the PDXs acquired features of increased aggressiveness. Hence, we limited our analysis of the

remaining three samples (original tumors and their PDXs) to a smaller set of markers, which included H&E, Ki67 and receptor status. As reported in Figure 32, Figure 33 and Figure 34, in all cases the histo-pathological features were maintained, apart for the increased proliferative index.

We, next, decided to manipulate the levels of miR-146 in the PDXs, to analyze whether miR-146 inhibition could affect tumor formation as observed for SUM159 cells. We set up a protocol for miR-146 manipulation in human PDXs. The scheme in Figure 35 summarizes all the steps: xenografts are collected at different passages, kept in non-adherent conditions for 2-3 days to avoid any adaptation to cell culture and/or modification of the original tumor, purified of mouse non-epithelial contaminants and infected with SCR and miR-146 KD lentiviruses. Then, Lin^-/GFP^+ cells were injected in mouse mammary fat pad of immunocompromised mice at different doses (from 5000 to 50.000 per injection). After 20 weeks, we compared frequencies of tumor onset and mean dimensions of the masses. This set of experiments is long-lasting and requires multiple round of injections to reach a number of independent observations to clearly establish the effect (if any) of miR-146 inhibition. As reported in Figure 36, in the first PDX analyzed, 12-B1-00197p, we obtained tumors from 7/11 animals injected with 50.000 control cells (SCR) and 1/5 animal injected with 5.000 control cells. Conversely, 146 KD cells gave rise to 3/13 (50.000 injected cells) and 0/5 tumors (5.000 injected cells), indicating impairment in the frequency of tumor formation. The boxplot in Figure 36 showed that the mean volume of the tumors was reduced 30 fold between SCR and miR-146 KD. The 10-fold serial dilution of injected cells allowed us to calculate the CSCs frequency, using the Extreme Limiting Dilution Analysis (ELDA) software (Hu and Smyth, 2009). Briefly, ELDA calculates the SC/CSC frequency, keeping into account the dose of cells injected, the numbers of animal tested and the number of responses obtained; in this case the response is represented by the development of tumors in the mouse. Then, it fits the data

according to a single-hit model explained by the fact that in each population injected there must be at least one SC/ CSC that give rise to the biological response, i.e the tumor. Thus, we calculated the CSC frequency of 12-B1-00197p sample, which was of 1:45777 cells for control treated cells (CTRL) and of 1:198954 for miR-146-KD cells, meaning that the CSC frequency was diminished by miR-146 inhibition.

The same experiment was repeated for another PDX, named 12-B1-00339p. Due to a lower number of tumor cells obtained, we only had sufficient material to inject cells at lower dilution (5k). As shown in Figure 36, in the 12-B1-00339p-146 KD there was a decrease in the number (4 tumors compared to 7 of SCR) but not a significant difference in the mean volume of the tumors, due to high biological variability (336 mm^3 vs 415 mm^3 of SCR) compared to control (p-value=0.7768).

As reported in Figure 36, the last PDX, named 13-B2-00430p, exhibited at 50.000 cells injected both a decrease in the number (2 tumors compared to 4 of SCR) and in the mean volume of the tumors (260 mm^3 vs 1336 mm^3 of SCR) compared to control (p-value=0.2472). Clearly, the p-values for these complex experiments have not achieved statistical significance, something that could require a greater number of experiments to be performed. Nonetheless, our preliminary results are in line with a putative role for miR-146 as a regulator of CSCs expansion.

ALIQUOT ID	DIAGNOSIS (>M-SNOMED)	GRADE	pT	pN	ER (%)	PGR (%)	HER-2	KI67 (%)	Subtype
12-B1-00339-01	Infiltrating duct carcinoma	/	y4b(m)	x	/	/	/	/	post-chemio
12-B1-00197-01	Infiltrating duct carcinoma	3	2	0(sn)	Neg	Neg	Neg	35	TN
13-B1-00037-01	Infiltrating duct carcinoma	3	3(m)	3a	Neg	Neg	Neg	35	TN
13-B1-00351-01	Infiltrating duct carcinoma	3	2(m)	1a	90	60	2+	35	LumB
13-B2-00430-01	Infiltrating duct carcinoma	3	1c	1a	<1	<1	Neg	95	TN

Table 2-Histology of human primary samples- the table summarizes all the clinical parameters measured by immunohistochemistry (IHC) on formalin-fixed paraffin embedded (FFPE) sections of primary tumors resected from patients. Listed here are grade, pT (primary tumor staging, as mean size of tumors), pN (nodal status); ER (estrogen receptor) status, PGR (progesterone receptor) status, HER-2 (Herceptin) status and Ki67 are reported as percentage of expressing cells with respect to the total number of cells analyzed.

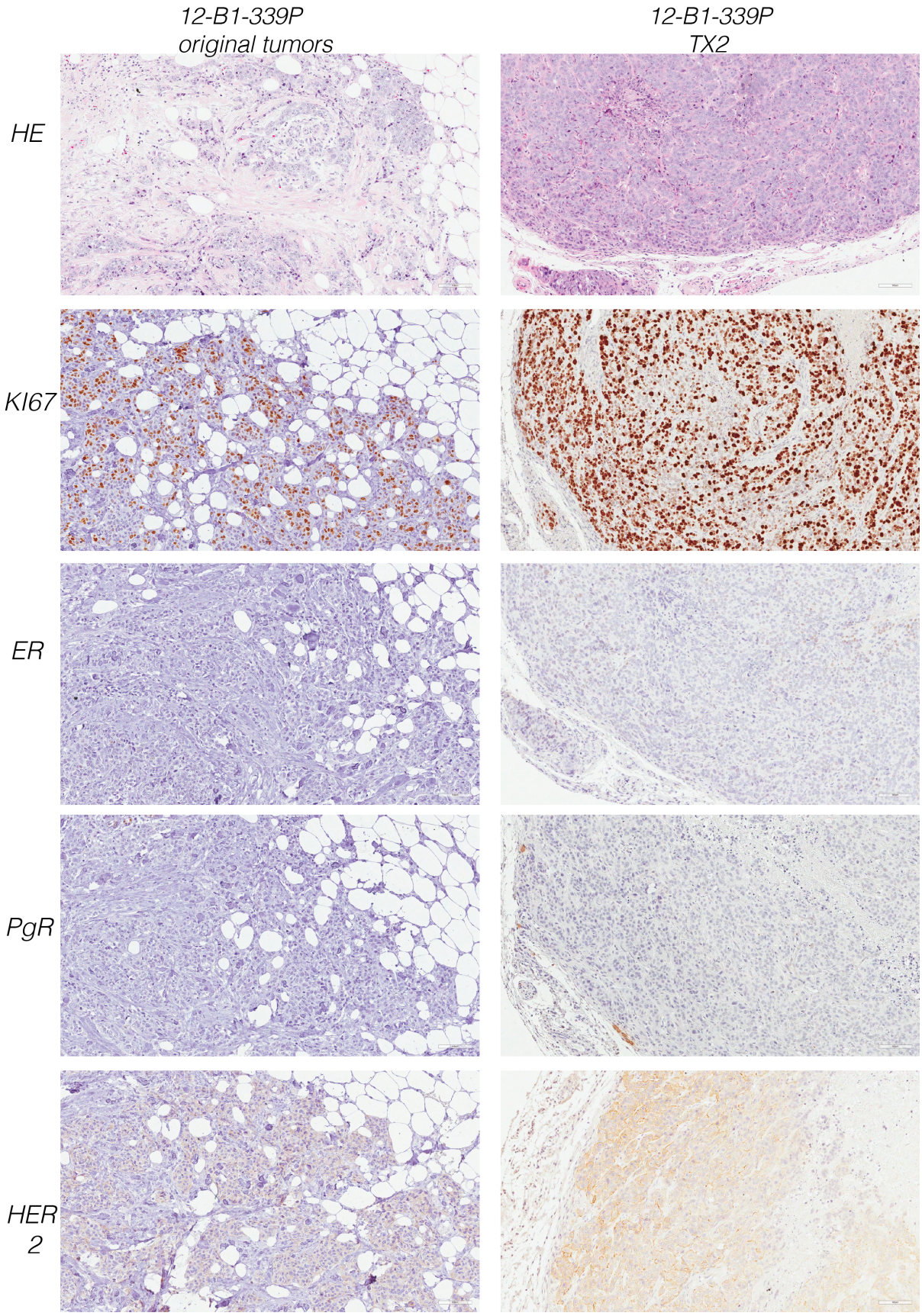


Figure 28- Characterization of human PDXs vs primitive tumor- representative images of IHC analysis of the original tumor 12-B1-339p and its corresponding PDX, with the indicated antibodies, magnification 20X.

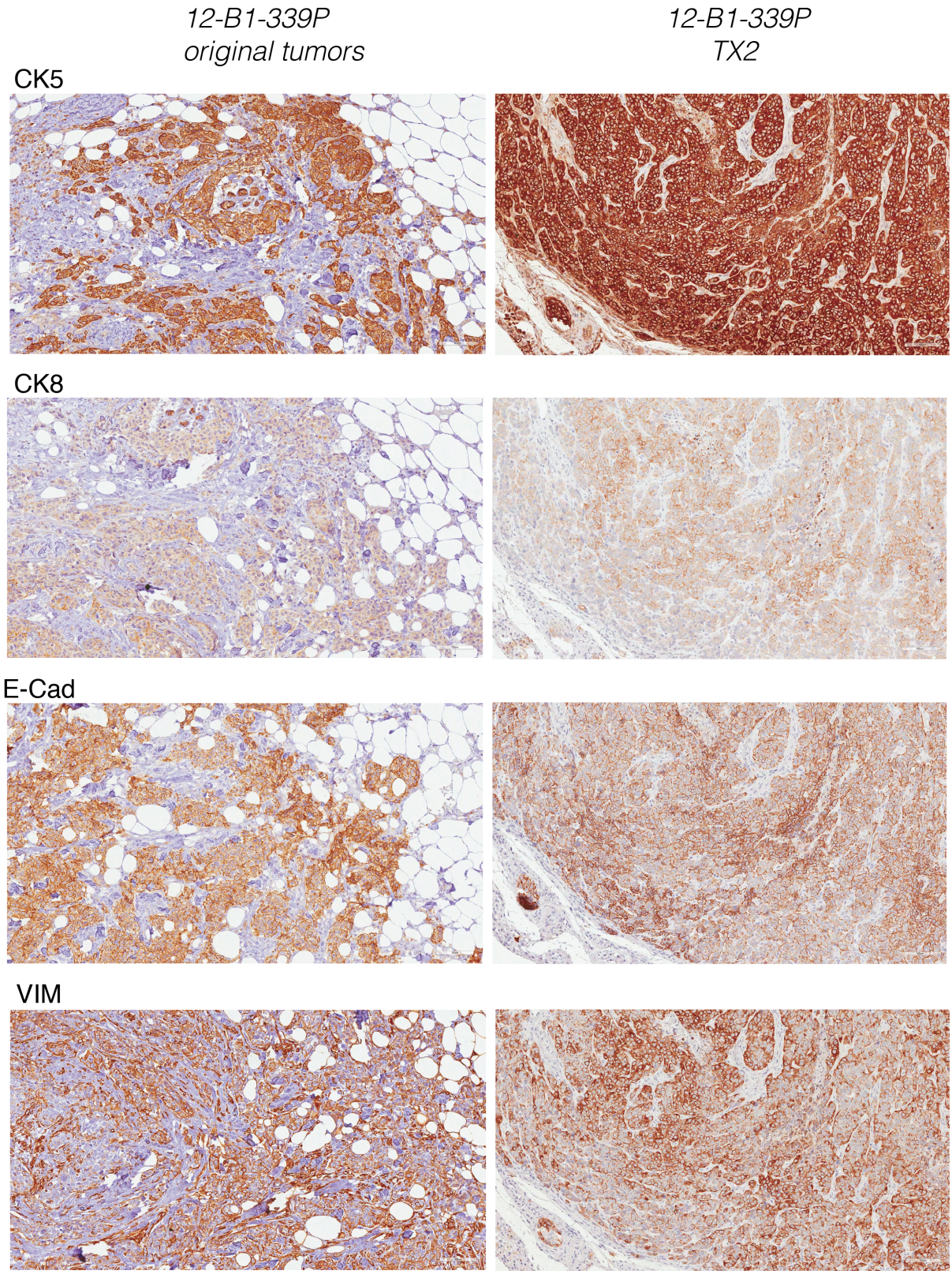


Figure 29- Characterization of human PDXs vs primitive tumor- representative images of IHC analysis of the original tumor 12-B1-339p and its corresponding PDX, stained as indicated, magnification 20X.

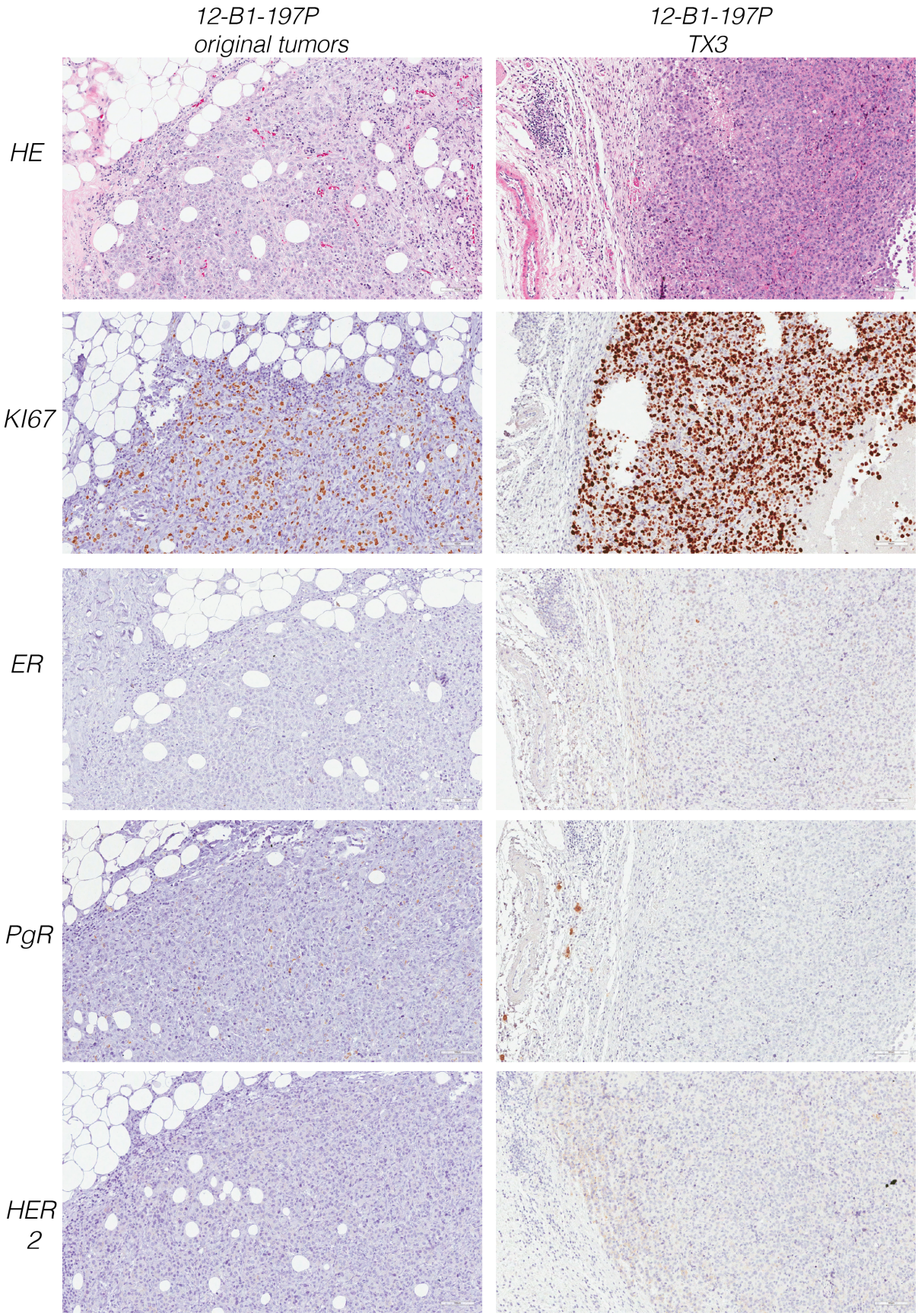


Figure 30- Characterization of human PDXs vs primitive tumor- representative images of IHC analysis of the original tumor 12-B1-197p and its corresponding PDX, stained as indicated, magnification 20X.

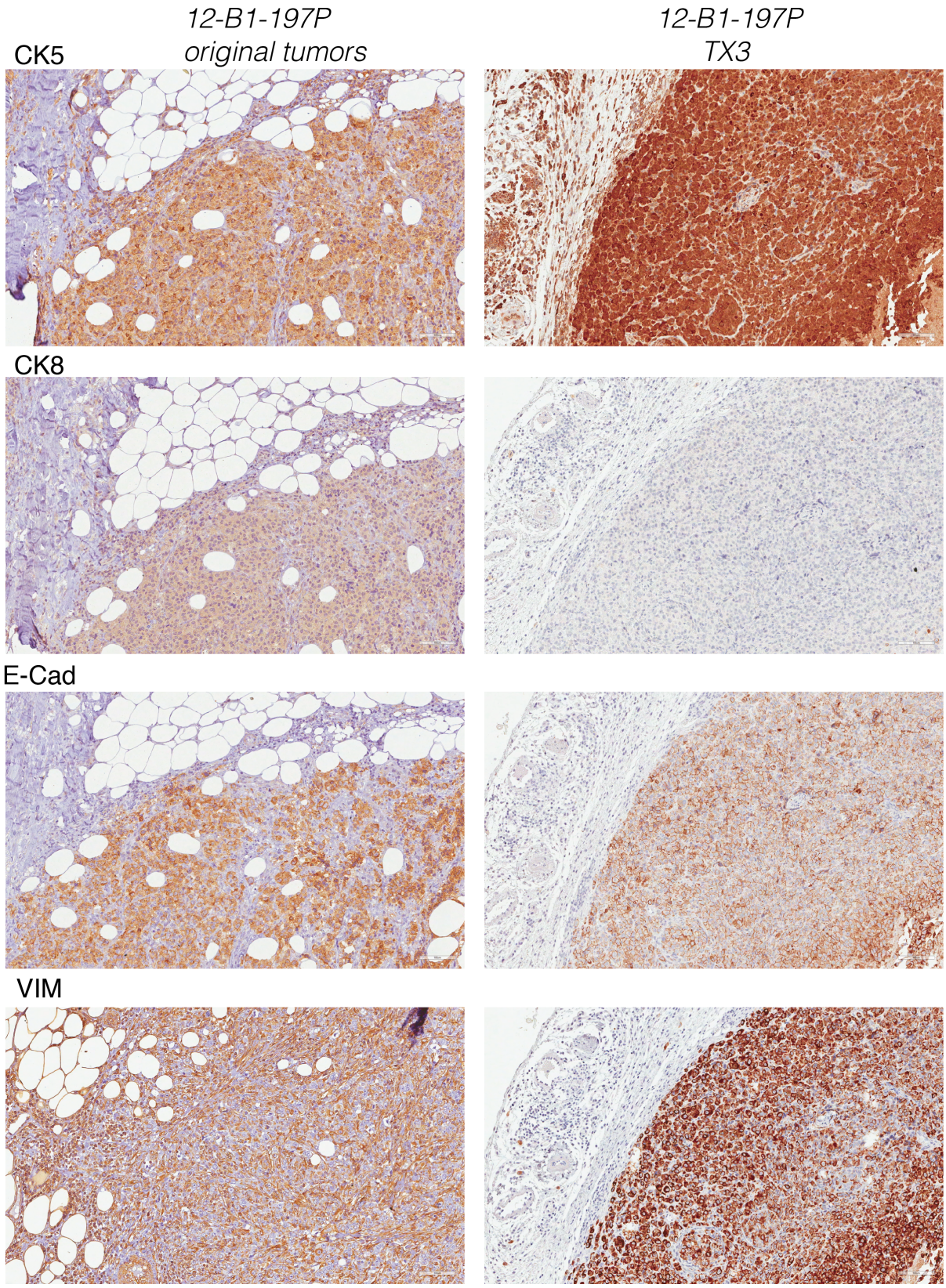


Figure 31- Characterization of human PDXs vs primitive tumor- representative images of IHC analysis of the original tumor 12-B1-197p and its corresponding PDX, stained as indicated, magnification 20X.

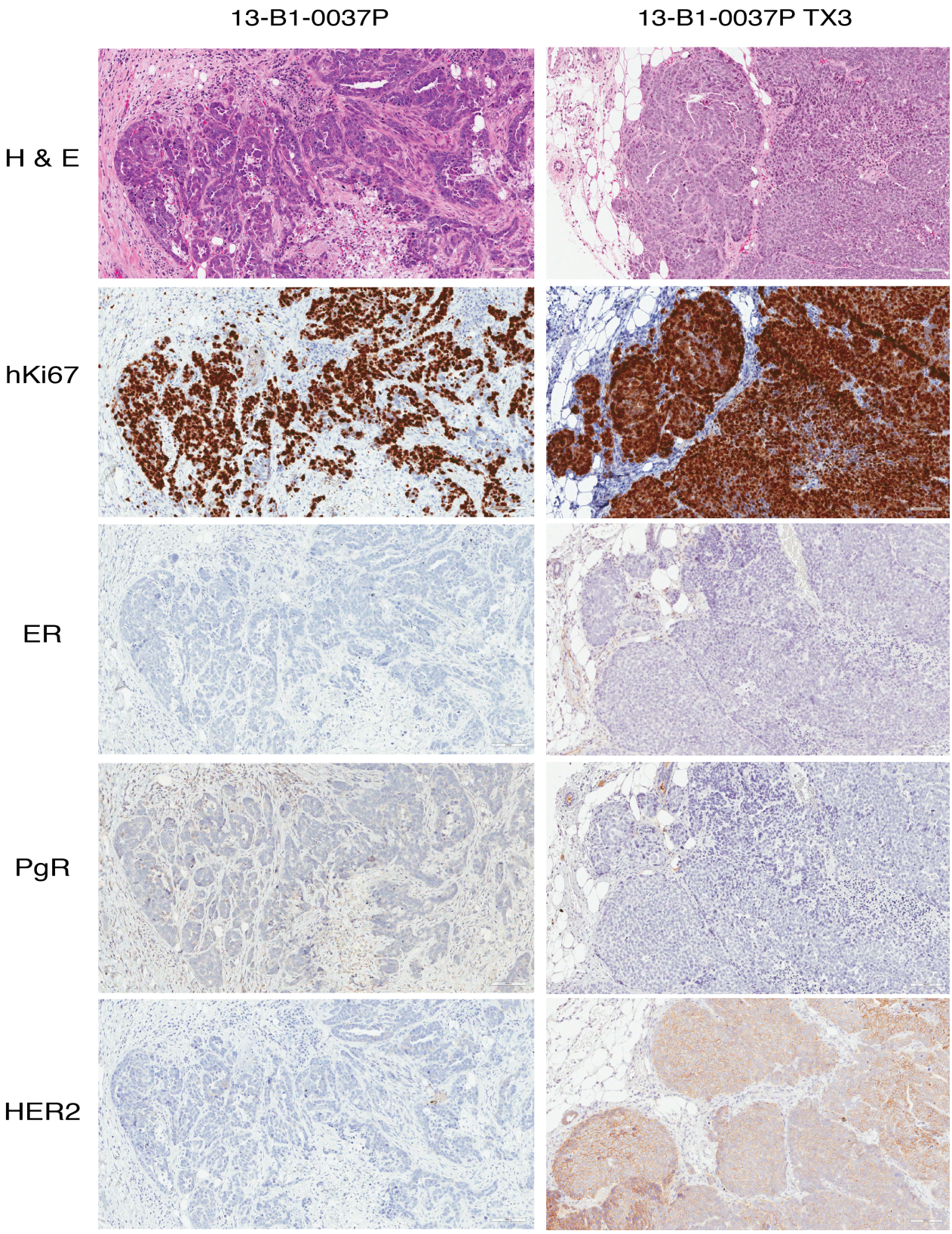


Figure 32- Characterization of human PDXs vs primitive tumor- representative images of IHC analysis of the original tumor 13-B1-0037p and its corresponding PDX, stained as indicated, magnification 20X.

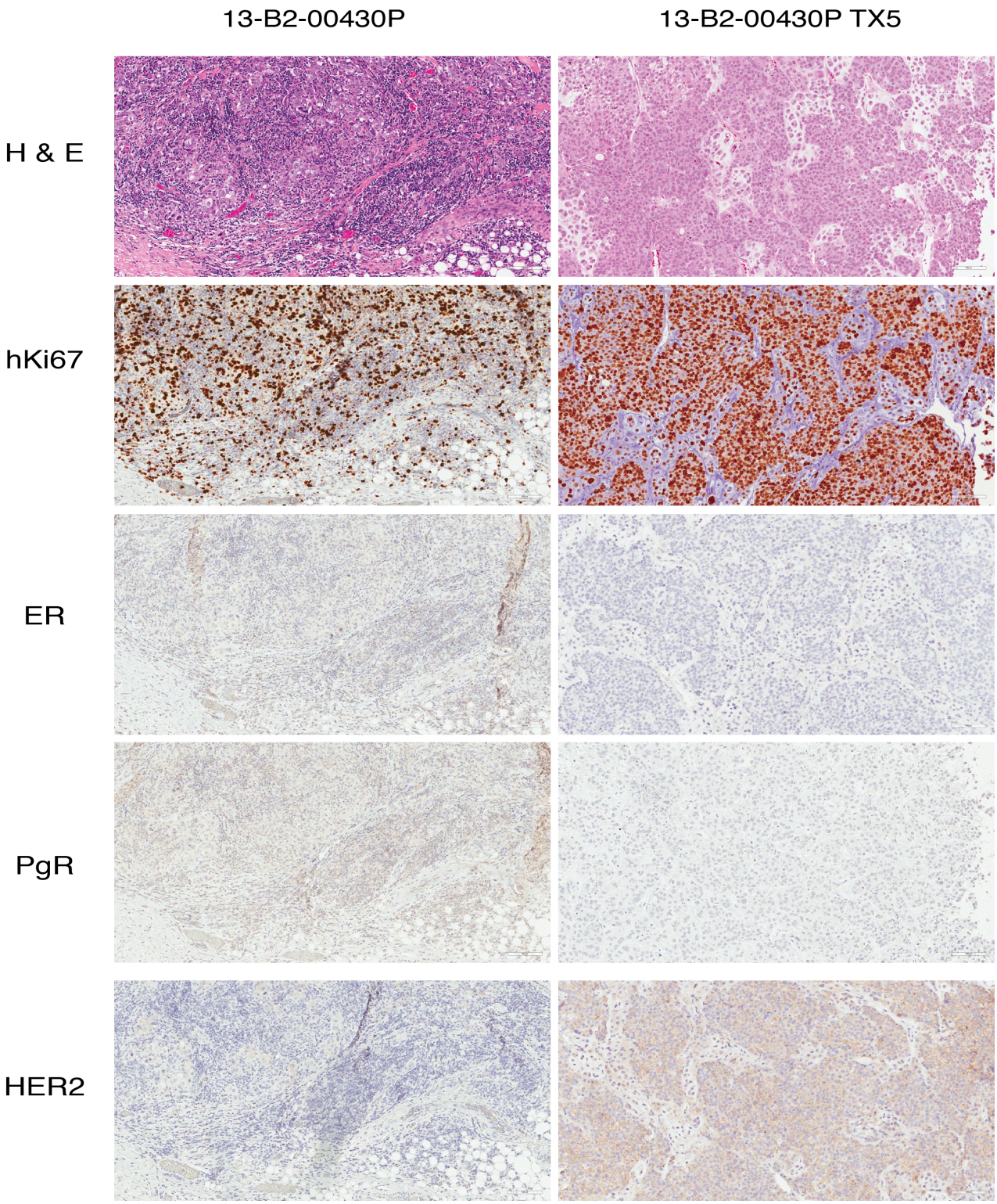


Figure 33- Characterization of human PDXs vs primitive tumor- representative images of IHC analysis of the original tumor 13-B2-00430p and its corresponding PDX, stained as indicated, magnification 20X.

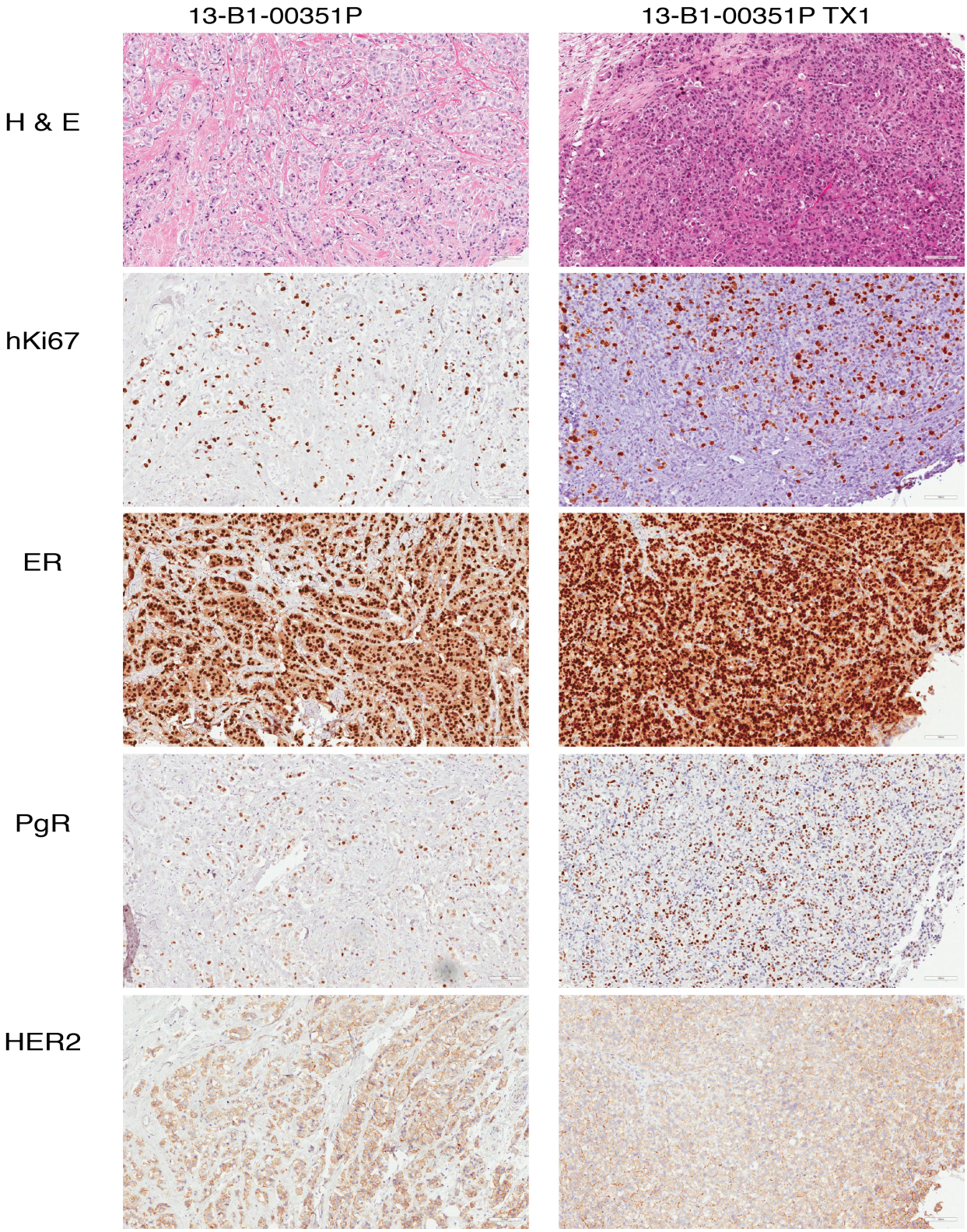


Figure 34- Characterization of human PDXs vs primitive tumor- representative images of IHC analysis of the original tumor 13-B1-00351p and its corresponding PDX, stained as indicated, magnification 20X.

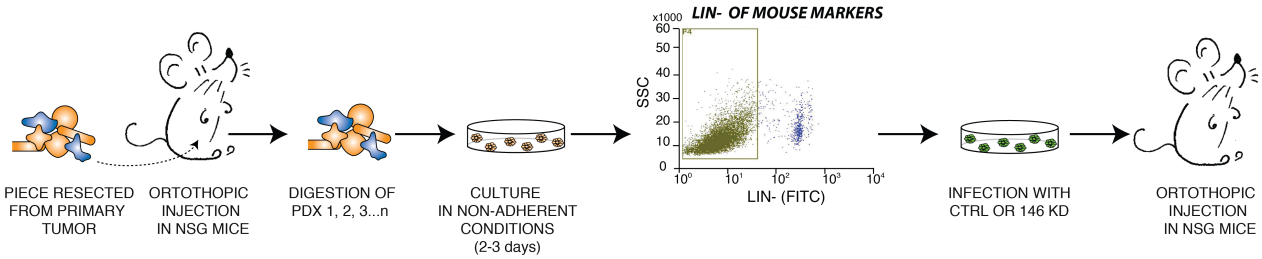
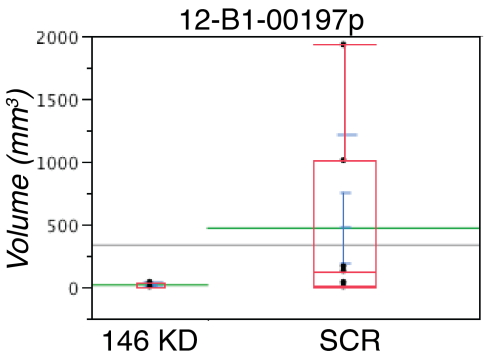
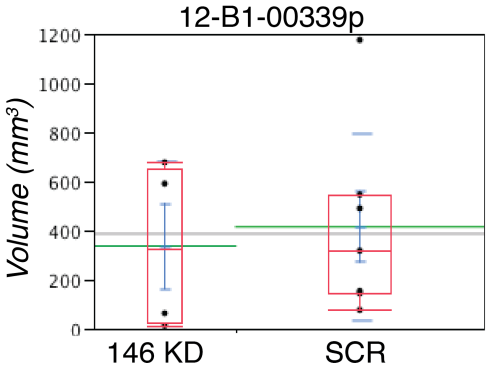


Figure 35– Scheme of the procedure used to manipulate miR-146 expression in human PDXs- Human PDXs were collected at different passages *in vivo*, digested at single cell level and put in culture for 2-3 days in non-adherent conditions. Before miRzip infection, cells were cleaned up using an epithelial cell enrichment kit to remove mouse cell contaminants (mouse lineage-, Easysep) and then infected with lentivirus CTRL or miR-146 KD. After 48h cells were selected for puromycin resistance and GFP positivity, then re-injected in mouse mammary fat pad of NSG mice at different doses (from 5.000 to 50.000 cells per injection).



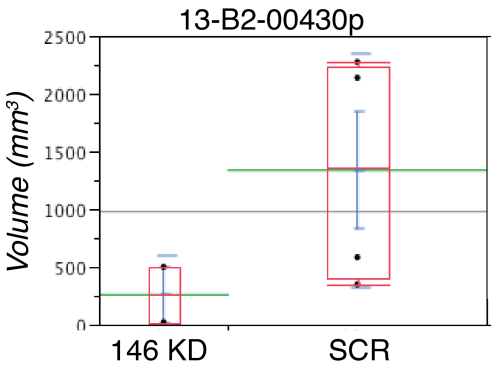
Sample	Mean Volume (mm3)	SEM
SCR	470.2	278.5
146 KD	15.6	12.5

p-value: 0.1689



Sample	Mean Volume (mm3)	SEM
SCR	415.4	143.6
146 KD	336.4	173.2

p-value: 0.7768



Sample	Mean Volume (mm3)	SEM
SCR	1336.7	505.3
146 KD	260	240

p-value: 0.2472

Outgrowths/injections

Sample	50k	5k	CSCs frequency
12-B1-00197p SCR	7/11	1/5	1:45777 (1:21985-1:95316)
12-B1-00197p 146 KD	3/13	0/5	1:198954 (1:617125-1:64140)
12-B1-00339p SCR	NA	7/8	NA
12-B1-00339p 146 KD	NA	4/8	NA
13-B2-00430p SCR	4/4	NA	NA
13-B2-00430p 146 KD	2/4	NA	NA

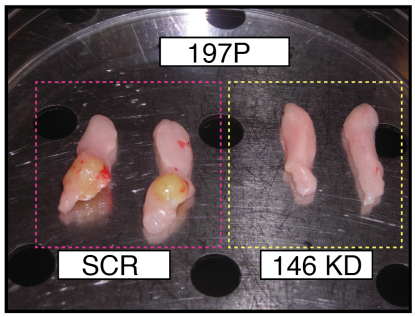


Figure 36- Effect of miR-146 KD in tumor transplantation assay – 50.000 or 5.000 Lin-/GFP^{+} cells expressing SCR or miR-146 KD vector were injected orthotopically in mouse mammary fat pads in different number of mice per condition. Tumors growth was measured every week until they became palpable, after 5 months tumors from 12-B1-00197p, 12-B1-00339p and 13-B2-00430p were surgically excised and the volume was measured by caliper. A representative picture of tumors obtained from 12-B1-00197p SCR vs 146 KD cells (50.000 cells) is reported. The table summarizes the result of tumor transplantations for each condition and, where possible, CSC frequency was calculated using ELDA (extreme limiting dilution analysis) software (Hu and Smyth, 2009), upper and lower limits are reported and p-value was calculated according to the fitting of the data to single-hit model. The upper panel summarizes the two-sample t-test performed on all the tumors, 12-B1-00197p, 12-B1-00339p and 13-B2-00430p, tumor volume in mm^3 was calculated with the formula V= [(width)^2 * length]/2; the table reports mean volume ± standard error of the mean (SEM) and p-value was measured with two-sample t Test.

3.3. Dissecting the molecular mechanism of miR-146 family

3.3.1. Identification of miR-146 KD regulated genes

To understand what are the molecular mechanisms through which miR-146 exerts its biological function on stem cell pathways, we performed a transcriptomic expression profile. We used the Affymetrix Human Gene 2.1 platform (≈ 22.000 genes) and we compared SUM159 in three biological conditions: i) cells grown in adherent conditions (WT 2D) ii) cells grown as mammospheres (3D) not infected (WT 3D) and iii) cells grown as mammospheres infected with a control vector (SCR 3D) or 146 KD vector (146 KD 3D) Figure 37A. Three independent biological replicates were analyzed for each condition, generating two signatures: i) genes regulated upon mammosphere formation between WT_2D vs WT_3D (referred as SUM-3D signature) ii) genes regulated in mammospheres upon miR-146 loss between SCR vs miR-146 KD (SCR_3D vs 146 KD, referred as 146 KD signature).

We performed a principal component analysis (PCA) to look globally at the transcriptional differences between samples. As shown in Figure 37B, the biological triplicates were clustered toghether, meaning that the transcriptional profiles were highly reproducible in each different condition. The gene expression profile between WT and SCR mammospheres was comparable, so from here to now we will refer to these sample as controls (CTRLs) to compare the gene regulation upon miR-146 KD in mammosphere formation (WT+SCR_3D vs 146 KD). PCA revealed that loss of miR-146 in mammospheres strongly changed the gene expression of mammospheres, in part moving towards the gene expression pattern of adherent cells (WT_2D, component 1), suggesting a reversion of mammosphere gene expression, and in part due to miR-146 specific effects (component 2).

The miR-146 KD signature consists of 1975 genes regulated >|0.25| $\log_2$fold with a p-value < 0.05 by student's T-test in 146 KD vs CTRLs_3D, of which 1127 are

dowregulated and 848 are up-regulated genes. Hierarchical clustering of 146 KD regulated genes, shown in Figure 38, revealed two biological effects: i) activation/repression of genes specifically responsive to miR-146 KD ii) reversion of the gene program activated during mammosphere formation, highlighted in the second heatmap composed of 785 genes with an opposite trend of regulation (WT_2D vs WT_3D, common genes).

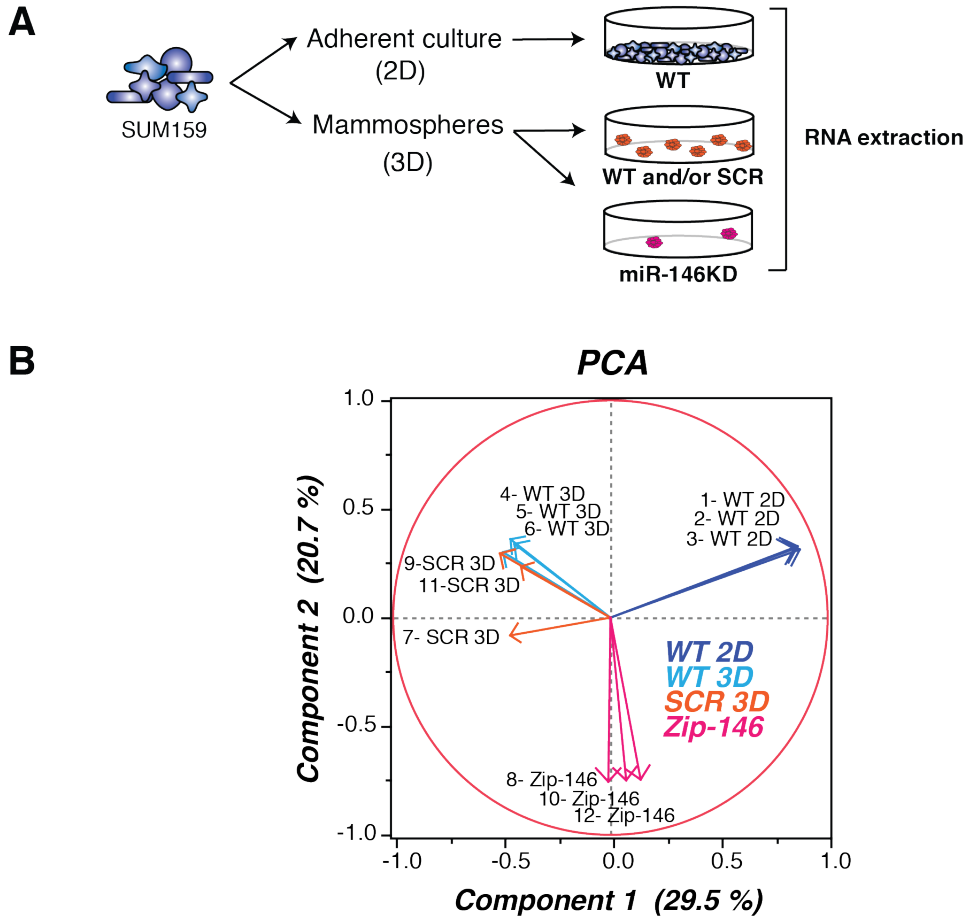


Figure 37- Summary and quality control of the samples used on gene expression data- A) the scheme summirizes the samples used for gene expression analysis. B) PCA analysis on Affymetrix samples revealed great concordance between biological replicas. The distance between miR-146 KD samples and the controls in both the components reveals the magnitude of the transcriptional changes induced by loss of miR-146.

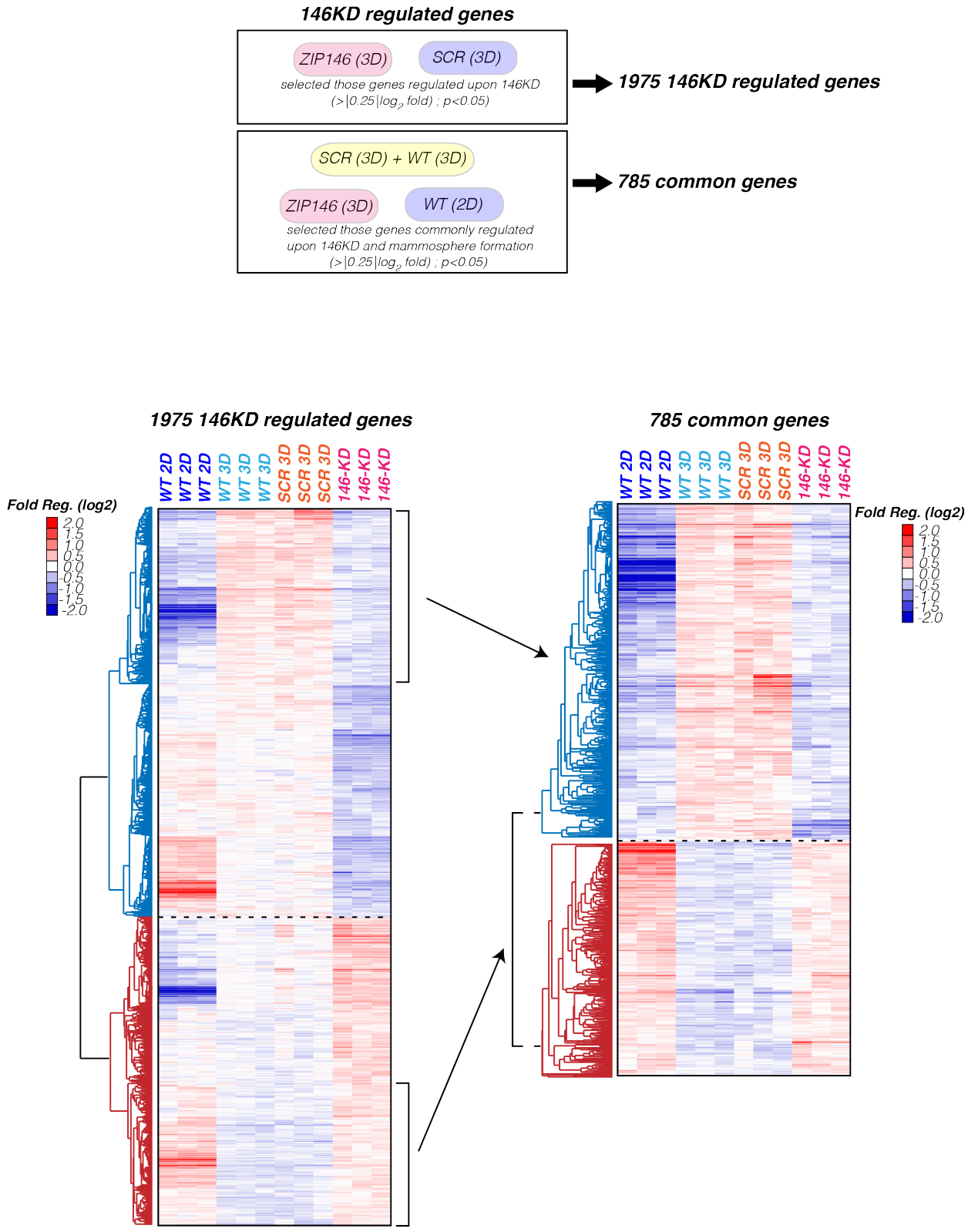


Figure 38- Identification of miR-146 KD regulated genes- Transcriptional profiling of genes regulated in mammospheres upon miR-146 KD. The 1975 genes regulated in the triplicates with a log_2fold >|0.25| (p-value < 0.05 by student's T-test) were used for unsupervised hierarchical clustering analysis. 785 genes had an opposite trend of regulation upon mammosphere formation (2D vs 3D) and upon miR-146 KD and are reported as a separate heatmap. The color bar indicates the value of the regulation.

3.3.2. Gene set enrichment analysis on stem cells signatures

Our hypothesis is that the gene expression changes induced by loss of miR-146 expression in mammospheres are somehow related to stem cells. We retrieved from public literature available gene signatures related to breast stem-enriched populations (SC-signatures) isolated with different approaches. These included: i) genes regulated in human normal and breast cancer SC isolated as lin-/procr^{+}/CD44^{+}/CD24^{-} cells, here referred to as Polyak (Shipitsin et al., 2007), ii) genes regulated in mammary repopulating units (MRU) from murine normal mammary tissue isolated with Sca-1^{neg}/CD49f ^{high} / CD24 ^{med} markers, here referred to as Stingl (Stingl et al., 2006), iii) genes regulated in mammary stem cells (MaSCs) isolated from murine normal mammary tissues with Sca1-/CD61^{hi}/CD29^{hi} /CD24^{+} cells, here called "Visvader" (Lim et al., 2010) and lastly iv) quiescent label-retaining PKH^{pos} cells isolated from human normal primary tissues, here referred to as Pece (Pece et al., 2010). For the analysis, in the case of mouse datasets (Stingl and Visvader), we considered only the genes conserved between human and mouse.

We used a computational tool known as Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) to correlate the SC-signatures with the transcriptional changes that occurred upon miR-146 KD. Briefly, each gene set is ordered according to the regulation of a query dataset (in our case, the miR-146 KD dataset) from the most upregulated to the most downregulated genes. If the gene set is not correlated with the query dataset, its genes will be distributed randomly in the query dataset. Conversely, if the two datasets are correlated (positively or negatively), the distribution will be asymmetric. GSEA computes the overlap between the gene sets mathematically, returning a normalized enrichment score (NES), which expresses the overlap within the two signatures, and a False Discovery Rate (FDR, q-value) as measure of statistical significance.

We performed GSEA analysis distinguishing each SC signature in UP and DOWN regulated genes that were ranked according to their regulation in the miR-146 KD dataset.

Figure 39 shows that the major enrichment scores in miR-146 KD dataset were measured in the Up-regulated genes of the SC-signatures, meaning that the core of genes downregulated by miR-146 was strongly enriched for genes that are generally activated in the stem compartment. The FDR q-values were statistically significant in three out of four stem UP signatures (Polyak, Visvader and Stingl UP < 0.01). On the other hand, no significant enrichment scores were observed for SC-signatures DOWN (Figure 39).

We concluded therefore that the expression of miR-146 is required to sustain the activation of genes (UP genes) specifying or maintaining stem cell functions. Of note, the enrichment analysis was significant for SC-signatures obtained for both human and mouse samples, meaning that miR-146 is likely affecting key conserved hubs.

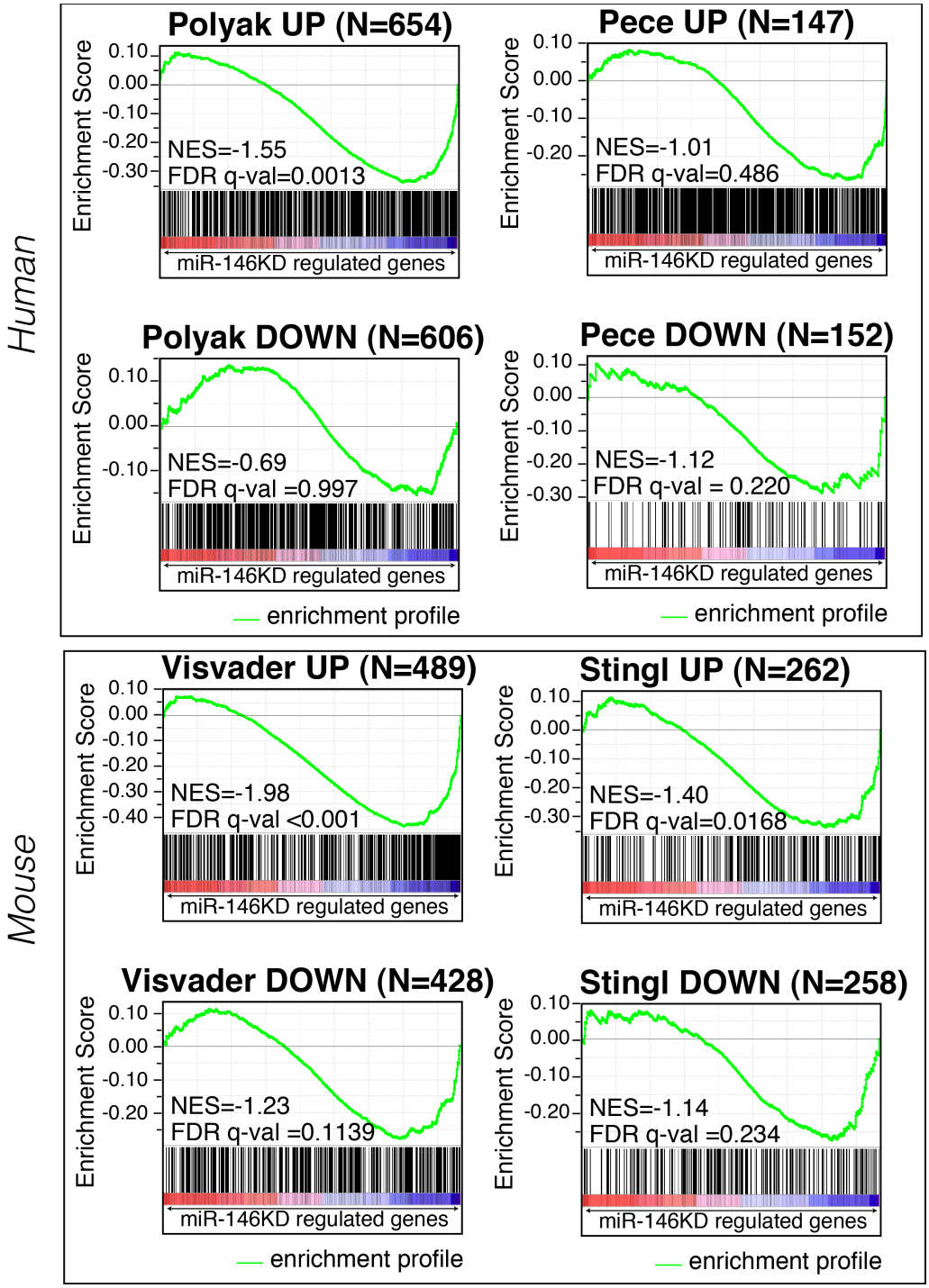


Figure 39- miR-146 down-regulated genes are enriched for stem UP signatures - GSEA was used to correlate stem cell signatures with the gene expression changes observed upon miR-146 KD. The enrichment plot together with the score (NES) and significance (FDR q-val) for each gene set analyzed are shown.

3.3.3. Identification of stem-cell genes regulated by miR-146 KD

To uncover the critical hubs affected by miR-146 and the signaling pathways to which they belong, we performed an upstream regulator analysis by Ingenuity Pathways software (IPA) on the 1975 genes regulated by miR-146 KD. Briefly, IPA searches within a given gene list (i.e. the miR-146 KD gene set) for enrichment of gene sets that are associated to upstream regulators, such as signalling hubs or transcriptional factors. The "upstream regulator analysis" computes a Z-score, which predicts if the upstream regulator is activated (Z-score >0) or inhibited (Z-score <0). The higher the value of the Z-score, the more regulated the pathway (activated or inhibited). The tool also performs statistical test (p-value), measuring if the overlap between the selected gene set and each upstream regulator gene set is significant.

We selected those upstream regulators with a p-value of overlap $< 10^{-3}$ and a z-score > |2|, which means pathways consistently regulated. With these selection criteria, we identified 62 upstream regulators for the miR-146 KD signature. Next, we narrowed the selection to those relevant to stem cell biology, exploiting the stem cell signatures (Polyak, Stingl, Visvader and Pece) and SUM159 mammosphere gene signature. To do this, we performed IPA analysis of each stem list to obtain a number of upstream regulators with a p-value of overlap $< 10^{-3}$; then, we merged these results with those coming from the analysis of miR-146 regulated upstream signaling pathways, leading to the identification of 24 common upstream regulators. Figure 40 describes the workflow for selecting the stem regulators; as shown in the heatmap, the cluster identified two separate groups. One group contains pathways that are inhibited in stem cells signatures and upon mammosphere formation (SUM 3D_2D) but strongly activated upon miR-146 KD , here called "DOWN in SCs" and containing i.e. Myc and MycN.

The second group contains pathways that are activated in stem cells and upon mammosphere formation but strongly inhibited by miR-146 KD, here called "UP in SCs".

This group includes some regulators known to be involved in epithelial to mesenchymal transition and/or acquisition of stem cells traits, such as TGF-β, Stat3 or Nf-κB.

The results clearly suggest that miR-146 affects the activation of critical pathways involved in the regulation or in the mainteinance of stem cell traits. This could explain why, in absence of miR-146 expression, the stemness genetic program cannot be maintained any longer, resulting in a depletion of SCs/CSCs as observed both in *in vitro* and *in vivo* experiments.

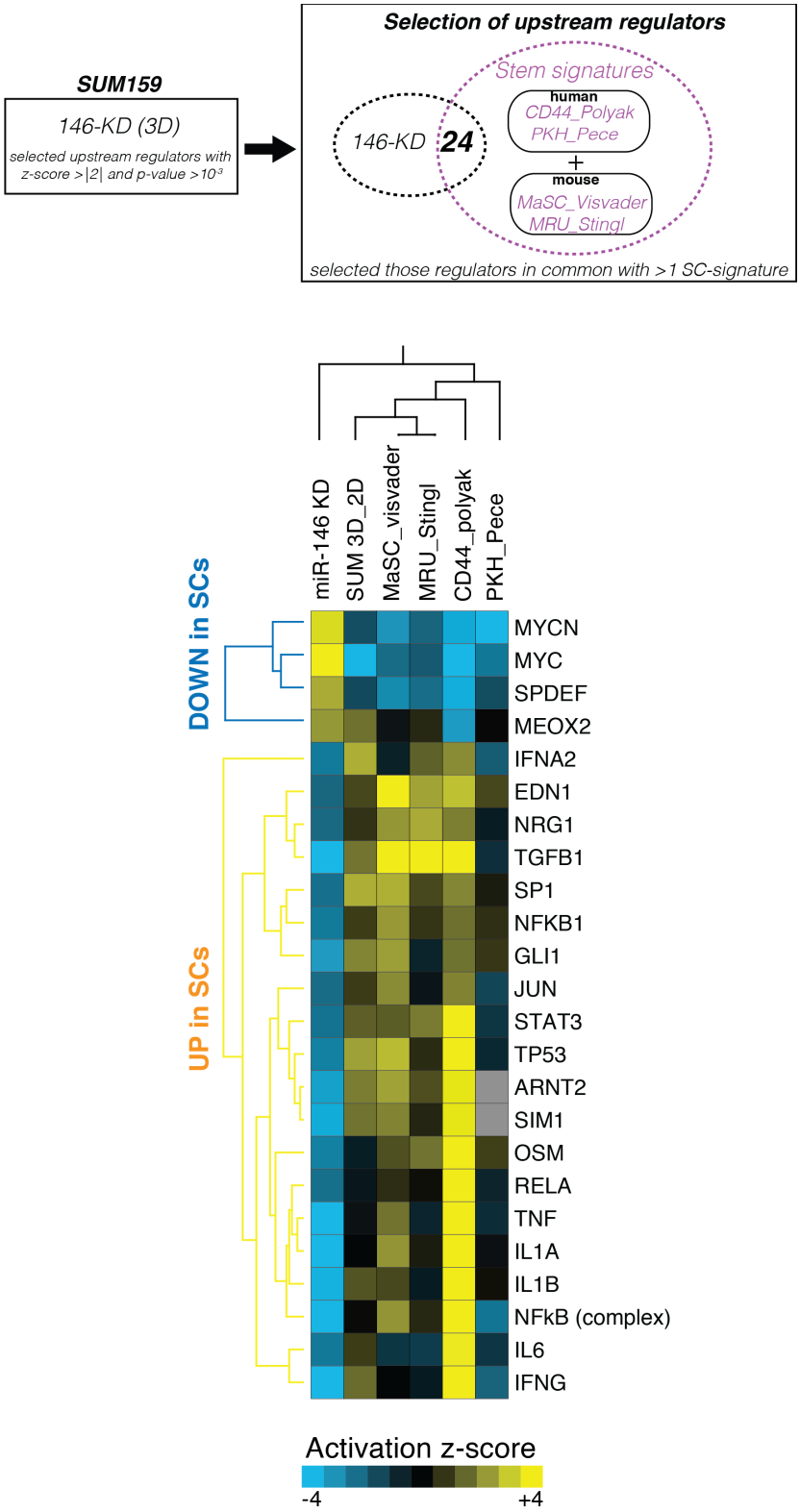


Figure 40- Identification of stem cell upstream regulators affected by miR-146 kD- Top: the scheme summarizes the pipeline for the selection of upstream regulators involved in stem cell biology. Bottom: the heatmap shows the unsupervised hierarchical clustering of the 24 upstream regulators regulated in miR-146 KD with a z-score >|2| and with a p-value < 10^{-3}. A color bar representative of the magnitude of activation is also shown (-4 to +4 of z-score).

3.3.4. Identification of putative stem cell targets of miR-146 in SC pathways

Based on the evidence that mammalian miRNAs predominantly act to decrease target mRNA levels (Guo et al., 2010a), we speculated that miR-146 targets might be transcriptionally regulated upon miR-146 manipulation (overexpression or knockdown). We took advantage of two miR-146 related datasets that were generated in our lab. As previously described, one dataset was generated by measuring gene expression in SUM159 mammospheres upon miR-146 KD (146 KD dataset). SUM159 cells resemble highly aggressive cancer cells and express high endogenous levels of miR-146a/b. Therefore, miR-146 targets should be upregulated upon miR-146 loss. Another dataset was generated by measuring the transcriptional changes upon transient overexpression of miR-146b (50 nM) in adherent MCF10A cells, a normal breast epithelial cell line that normally expresses low levels of miR-146a/b. In this case, miR-146 targets should be downregulated upon miR-146 overexpression. MCF10A regulated trancripts were obtained by measuring transcriptional changes with respect to a control oligo (scrambled, SCR) occuring at short time points (16h and 24h post transfection), to ensure the enrichment for primary miR-146 effects. At first, we search for the enrichment of miR-146 target in the two datasets. We exploited a computational tool, known as *Sylamer* (van Dongen et al., 2008), which systematically searches for over-represented words (i.e. "seed" matching words, as 7mers or 8mers) in the sequence of genes coming from a dataset that has been ordered from the most downregulated to the most upregulated. If the distribution of any word is asymmetric, it means that the gene expression is affected by a miRNA, with the target enriched among the upregulated genes (in case of downregulated miRNAs) or downregulated genes (in case of upregulated miRNAs).

The analysis by *Sylamer* was performed with sequences of the 3' untranslated region (3'UTR), which usually contains the miRNA responsive elements (MREs), as well as with sequences of the coding (CDS) or the 5' untranslated region (5'UTR) as control. In the

MCF10A OE dataset the most represented words (7mers and 8-mers) in 3'UTR sequences exactly matched with the canonical seed of the miR-146 family, which is a common sequence shared by miR-146a and miR-146b. The maximum enrichment score was observed within the threshold that we used to define "downregulated" genes (-0.2 log_2fold ratio), while a progressive depletion was found over the "not regulated" and "upregulated" gene classes, as shown in Figure 41. Conversely, in the 146 KD dataset we did not score any significant enrichment for either the 8-mer or the 7-mer of miR-146a/b in 3'-UTR, 5'UTR or CDS of regulated genes, suggesting that the transcriptional changes observed are mainly long-term effects on targets due to miRNA loss.

We set out to identify those genes belonging to SC pathways that are targeted by miR-146. We subselected the strongest 14 upstream regulators, regulated in at least two or more stem cells datasets, consisting in 1997 genes in total. In this list we searched for the potential targets of miR-146 using a multi-step approach, which exploits the transcriptional changes induced upon miRNA manipulation (overexpression MCF10A-OE dataset; knockdown SUM159-KD dataset) and bioinformatics, using algorithms that predict miRNA targets.

A complete list of predicted miR-146 targets was generated by using both Targetscan 6.2 and miRanda databases: Targetscan predicts biological targets of miRNAs by looking at the presence of conserved 8-mer (and 7-mer) sites that bind with perfect complementarity to miRNA responsive elements (MRE) in the 3'UTR of mRNAs and calculates a "context score" as measurement of the likelihood of a mRNA to be repressed by the selected miRNA. Conversely, miRanda does not require perfect complementary binding to the 3'UTR of the target gene and keeps into account also free energy binding contribution. miRanda expresses the strengh of the miRNA:mRNA predicted interaction as a score, named a miRsvr score. The lower the score (context or miRsvr score) the stronger the predicted interaction.

We cross-compared the list of predicted targets (TargetScan+ miRanda) with the 1977 stem cell gene list derived from upstream regulator analysis, to highlight those pathways/genes targeted directly by miR-146. We identified 336 genes that were predicted from miRanda database, 58 of these possess a miRsvr score < -0.75. On the other way, Targetscan identified 326 targets, 55 of which had a context score < -0.20. From the union of these two lists, we obtained 91 top stem genes putatively targeted by miR-146 (shown in Figure 42). Eighteen (of the 91) were upregulated upon miR-146 KD with a $\log_2$fold > +0.20 and nineteen were downregulated by miR-146 OE at 16/24h with a $\log_2$fold < -0.20. Four genes were found to be regulated in common between the two systems analyzed: IRAK1, B3GNT5, UHRF1, KRT6B.

The 91 stem genes putatively targeted by miR-146 contained some interesting candidates: IRAK1 and TRAF6 reported as the most potent miR-146 targets, belonging to the Nf-κB network; Jagged1, the ligand of Notch, that activates an important signaling pathway in stem cell manteinance; and NUMB known as fate determinant that antagonizes and represses Notch1.

Looking at the number of miR-146 targets that belong to a single upstream regulator, we found that pathways such as TGF-β, p53, TNF or Nf-κB are strongly enriched for predicted targets of miR-146 (Figure 42), suggesting that miR-146 may act cooperatively on multiple targets belonging to pathways that are generally activated in stem compartement. So, once identified the critical pathway/s modulated by miR-146, we have already available a list of target/s ready to be validated in biological assays.

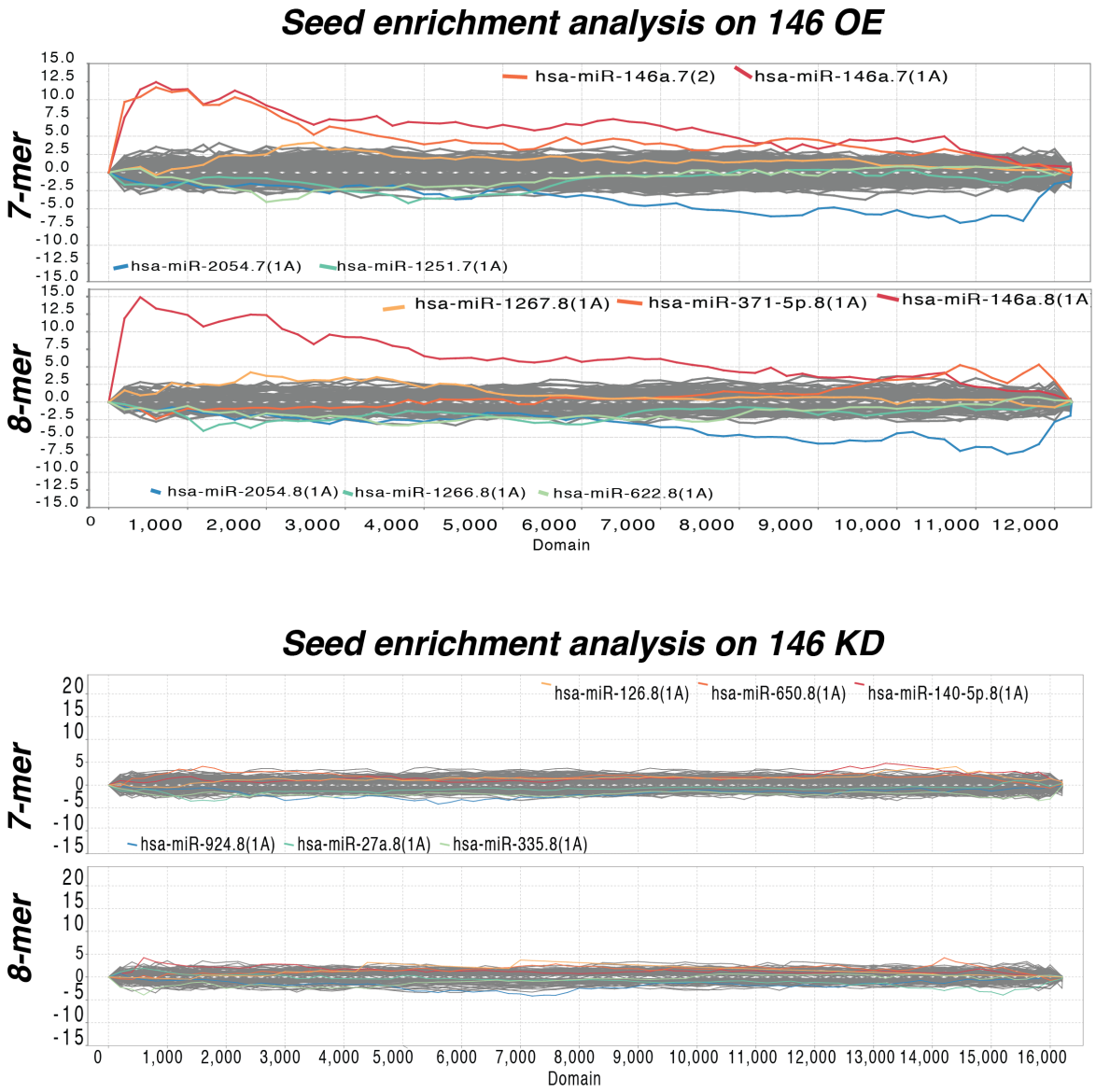


Figure 41- Sylamer analysis on 146 KD vs 146 OE- Seed enrichment analysis was performed using Sylamer to search for miRNAs binding sites in 3'-UTR, 5'UTR and CDS of all the genes regulated by miR-146 OE in MCF10A cells (upper panel) or by miR-146 KD in SUM159 cells (lower panel). Shown are the 8-mer or the 7-mer of miR-146 significantly enriched in miR-146 OE but not in 146 KD.

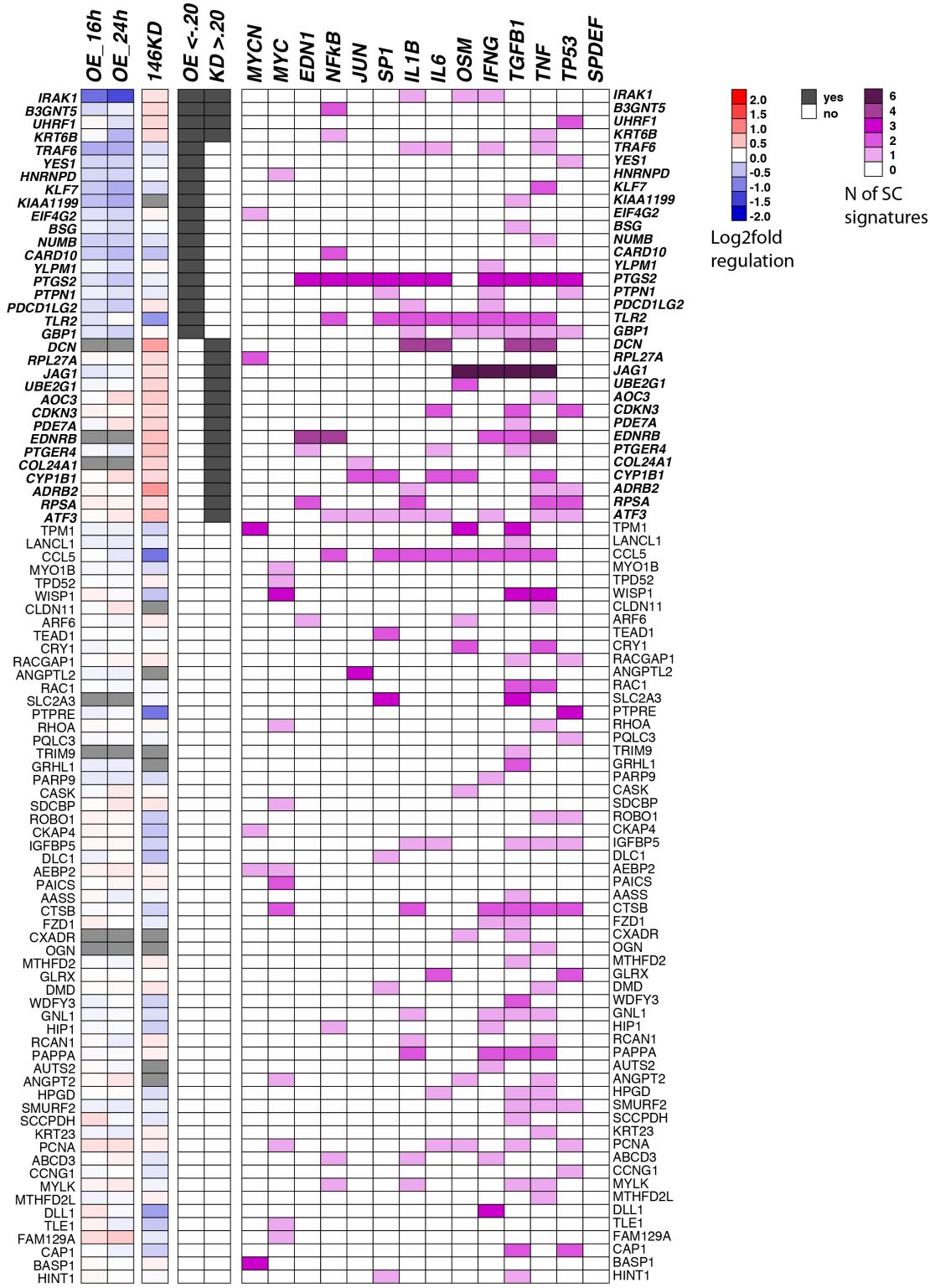


Figure 42 - Identification of stem cell targets of miR-146- the heatmap shows the 91 stem cell genes predicted using Targetscan and miRanda algorithms as putative targets of miR-146. For all these genes we report the regulation in i) SUM159 upon miR-146 KD (reported the log2fold regulation vs CTRLs) ii) miR-146 OE in MCF10A (16h and 24h), (reported the log2fold regulation vs SCR oligos). The right panel reports the 14 top regulated upstream regulators, and reflects how often the gene is regulated in the SC-signatures in a statistically significant manner (color code from pale to dark pink).

3.3.5. Candidate approach on SC targets: NUMB

We have identified 91 target genes putatively regulated by miR-146 that could potentially explain how miR-146 depletion regulates stem cell self-renewal. One of these is the cell fate determinant NUMB, a likely candidate in the regulation of "stemness". NUMB plays an important role in determining cell fate decisions by antagonizing and repressing Notch1 and by stabilizing the tumor-suppressor p53. Our lab, and others, has demonstrated that the NUMB/Notch and NUMB/p53 axes are relevant to breast tumorigenesis (Pece, 2004) and regulate SC self-renewal (Tosoni *et al*., unpublished). Additionally, a recent paper has reported the miR-146/ NUMB interaction as a mechanism to control self-renewal of CSCs in colon cancer (Hwang et al., 2014). In this paper, the authors found that high levels of miR-146a in CSCs keep NUMB repressed, and in turn enhance Wnt/β-catenin signaling, thus fostering CSC symmetric cell division. Conversely, restoring NUMB levels in CSCs by knocking down miR-146a leads to a switch from symmetric to asymmetric cell division thus reducing the CSC pool both *in vitro* and *in vivo* (Hwang et al., 2014).

Our data show NUMB to be regulated in MCF10A cells upon miR-146 OE. Additionally, both the algorithms miRanda and Targetscan predict NUMB as direct target of miR-146 with a miRsvr score of -1.33 and context score of -0.29. Hence, we speculated that the miR-146 dependent effects we observed in breast SC and CSC self-renewal might be dependent at least in part on the miR-146:NUMB interaction.

We first verified the direct binding of miR-146 to NUMB 3'UTR. We generated a NUMB luciferase reporter assay by cloning the NUMB 3'UTR downstream of a firefly luciferase reporter gene (pmiR-NUMB). This construct also expresses a Renilla Luciferase, used to normalize the efficiency of transfection between samples. In the absence of miRNA binding to pmiR-NUMB, the luciferase gene is expressed at high levels. However, if a miRNA binds to the 3'UTR, this causes the degradation of the luciferase mRNA and, in turn, a reduction in the luciferase signal. As a specificity control, we generated a mutant

NUMB 3'UTR reporter, deleted for the MRE of miR-146 (pmiR-NUMB-del-146-MRE). As shown in Figure 43 the overexpression of either miR-146a or miR-146b reduced the luciferase signal by 20-35% with respect to control cells. When the miR-146 MRE was deleted, the luciferase signal of pmiR-NUMB-del-146-MRE was restored to levels comparable with the SCR control, demonstrating that miR-146a/b were specifically able to interact with NUMB 3'UTR.

Then, we tested the involvement of NUMB in the miR-146 dependent effects on CSC self-renewal, using SUM159 cells as a model system. We exploited a doxycycline- inducible lentiviral vector to express an shRNA targeted against NUMB, under the control of a tetracycline responsive promoter (Tet-on system), along with a turbo-RFP as a reporter. By FACS-sorting, we selected a homogenous RFP^+ population and then we infected the NUMB-silenced cells (shNUMB) with a lentiviral vector for miR-146 KD (zip-146 KD) or with SCR control (zip-SCR), both of which expressed a GFP reporter gene. We reasoned that, if miR-146-dependent regulation of CSC self-renewal is mediated by NUMB repression, then i) the inhibition of miR-146 expression (miR-146 KD) should induce NUMB protein accumulation, and, more importantly, ii) preventing NUMB accumulation (by silencing of NUMB) should also prevent the effects of miR-146 KD (i.e. the inhibition of CSC self-renewal in the mammosphere assay).

As shown in Figure 44, cells that expressed both the vectors exhibited an intense yellow color, useful to distinguish them from the cells infected only with one vector (RFP, shNUMB only or GFP, miRNA sponge only). We verified the levels of NUMB protein by immunoblot and observed that NUMB protein was maintained silenced both in single- and in double-infected SUM159 cells (shNUMB+zip-SCR and shNUMB+zip-146 KD).

The results of a mammosphere assay performed on these cells are reported in Figure 44:

i) shNUMB alone increased the sphere forming efficiency of SUM159 by 30-35% with respect to control cells; this is in agreement with a previous work in which loss of NUMB

regulation over Notch seemed to have a positive impact on mammosphere formation (Dontu et al., 2004) ii) when miR-146 is knocked down alone, the SFE of SUM159 was reduced almost to an half, as previously observed iii) the miR-146 KD effect was unaltered in the presence of NUMB shRNA, since we again observed a SFE reduction by 50% with respect to the control (Figure 44).

Even though NUMB is a direct target of miR-146, this result demonstrates that in the breast context miR-146 is able to regulate CSC self-renewal independently of NUMB.

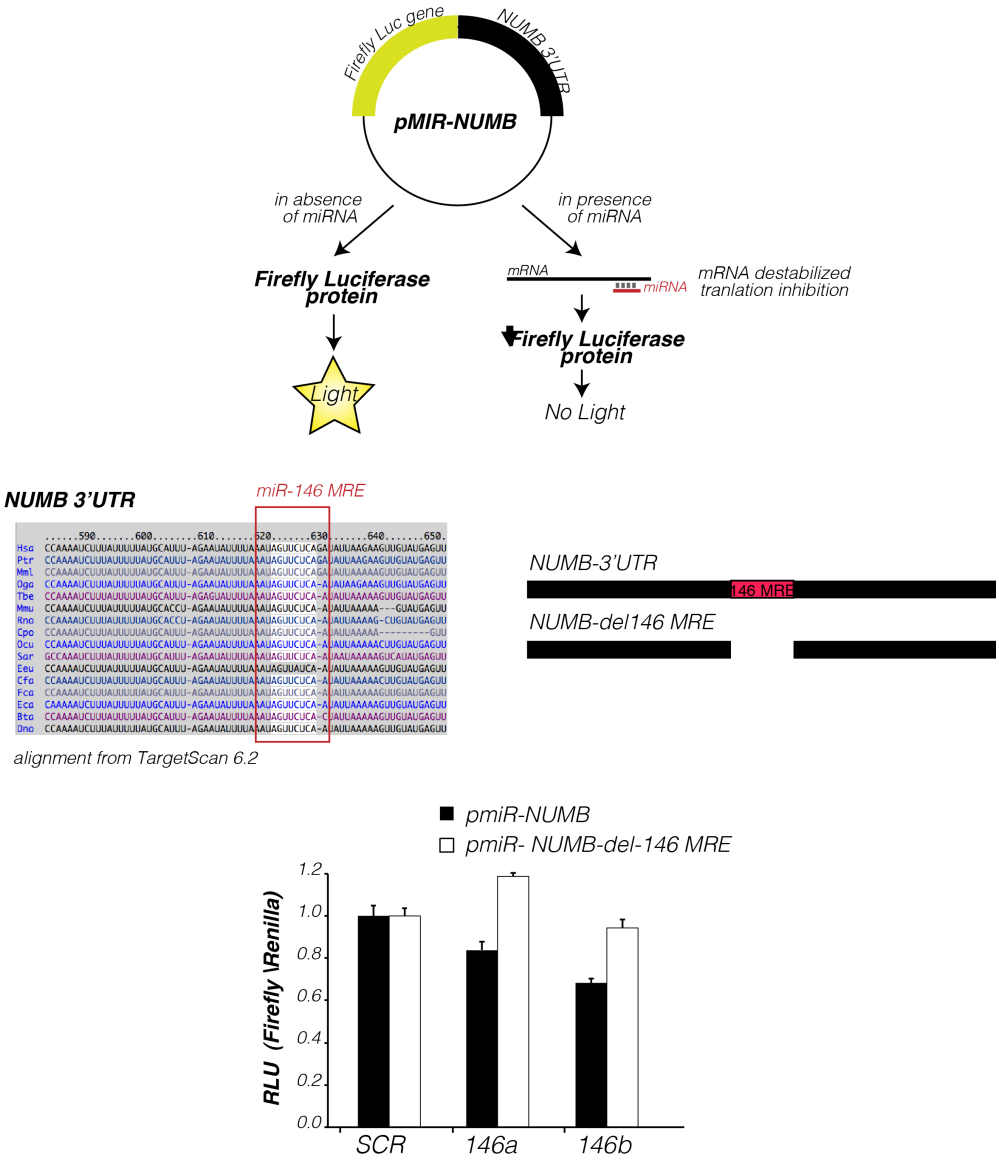


Figure 43- Luciferase assay to test miR-146 binding to NUMB 3'-UTR- 293t cells were transfected with pmiR-NUMB (black bar) or pmiR-NUMB-del-146-MRE (white bar) vectors in presence of 50 nM of mimic control (SCR) or miR-146a and mir-146b. After 24h Firefly luciferase signal was measured through Dual Glo Luciferase Assay system. The plot reports the relative luminescence units (RLU) as ratio of Luciferase over Renilla signal. The data are normalized with respect to SCR and the error bars report the standard deviation of two biological replicas.

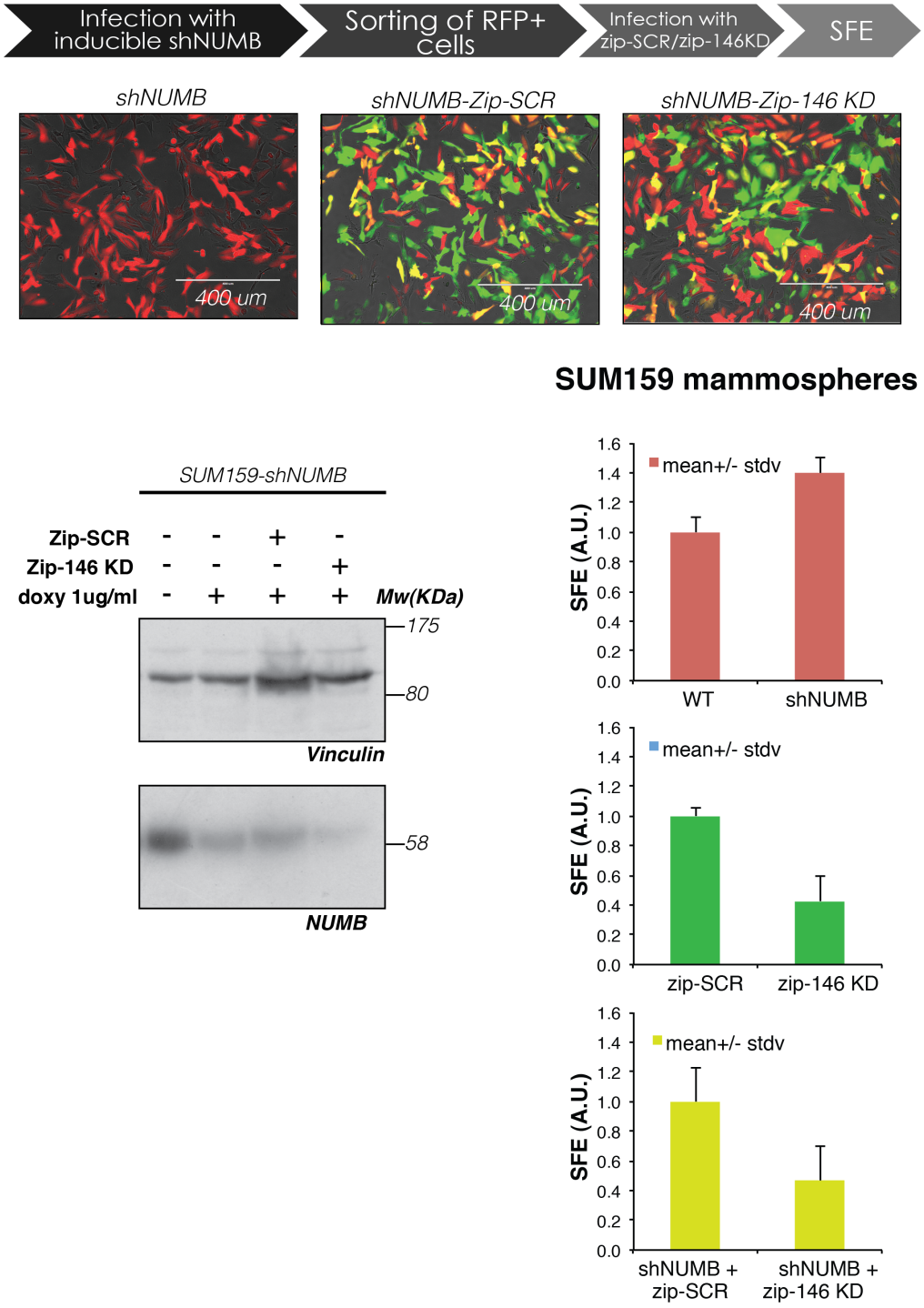


Figure 44- Numb regulation is dispensable for miR-146 KD phenotype- Top: representative images of cells infected with shNUMB alone, or in combination with zip-SCR or miR-146 KD lentiviral vector, 10X magnification, scale bar 400 μm. Bottom: endogenous levels of NUMB protein were examined by western blotting on the indicated samples; vinculin was used as loading control. The plot reports the sphere forming efficiency as arbitrary units of mammosphere assay for SUM159 cells targeted with shNUMB alone, zip-SCR and zip-146 KD alone or shNUMB-zip-SCR and shNUMB-zip-146-KD.

3.3.6. Analysis of stem cell pathway modulation upon miR-146 KD using the "Cignal cancer 10-pathways reporter assay"

Genome wide analysis of transcriptome revealed that a number of pathways involved in SC biology were modulated upon miR-146 KD. To assess the miR-146 dependent regulation of these pathways and to distinguish which pathways are primarily affected by miR-146 expression, we performed a multiple reporter assay ("Cignal cancer 10-pathways reporter assay"). This assay consists of 10 dual-luciferase reporters, each constituted by a Renilla luciferase construct, to normalize the efficiency of transfection between samples, and a transcription factor-responsive luciferase. Briefly, the promoter of each transcription factor-responsive Firefly luciferase is under the control of several repeats of Transcriptional Response Element (TRE) for 10 different transcriptional factors: Wnt, Notch, p53, TGF-β, Nf-κB, Rb-E2F, Myc/Max, Hypoxia, MAPK/ERK and MAPK/JNK. In this way, each construct can monitor the activation or repression of each key transcription factor, by increasing or decreasing the luciferase signal upon miR-modulation.

First of all, we set up all the conditions for the efficient cotransfection of plasmid and RNA oligos (miRNA mimics or inhibitors). We compared two different reagents, Lipofectamine®2000 and Lipofectamine®3000 (Invitrogen), both for efficacy of transfection and for cell viability. As model system, we used SUM159 cells transfected with one reporter plasmid alone or in combination with 50 nM of miR-146 oligos; since the reporter plasmid expresses both GFP and a mix of constitutively active luciferase plus constitutively active Renilla luciferase, we measured the efficiency of transfection both by luciferase assay and by visual inspection of GFP positivity. As shown in Figure 45, Lipofectamine®3000, in comparison with l Lipofectamine®2000, resulted in an increased efficiency of cotrasfection without affecting cell proliferation.

Even though the level of overexpression of miR-146a/b was quite comparable between the two transfection reagents, the luciferase assay revealed that: i) in presence of plasmid alone both the reagents performed efficiently in a comparable way ii) in presence of plasmid+miRNAs, the transfection efficiency of Lipofectamine®2000 dropped severely compared to Lipofectamine®3000, therefore we chose Lipofectamine®3000 as optimal reagent for cotransfection experiments.

Next, we transfected the panel of 10 reporter pathways in serum-depleted SUM159 cells, and evaluated their regulation at basal levels, without providing any specific ligand to promote pathway activation. This first experiment served to monitor the effect of miR-146 transient inhibition in a steady state condition. We achieved miR-146 inhibition by transfecting at 100nM the miR-146 LNA power family inhibitor (that acts by simultaneously blocking all the members of miR-146 family in a stable and highly specific manner). As negative control, cells were transfected with a non-targeting control (SCR) at 100 nM for 24h. Figure 46 summarizes the results of a single preliminary experiment. We noticed that at the basal level not all the pathways were constitutively active; in fact, Wnt, TGF-β and Notch reporters expressed 10-fold less Firefly luciferase than constitutively active pathways such as Nf-κB or MAPK. In addition, several pathways were not modulated in 146 KD vs ctrl, such as myc/max, p53, hypoxia and Mapk/JNK, meaning that there was no direct transcriptional effect on these pathways. Of note, Rb-E2f, Wnt, Notch and TGF-β were slightly transcriptionally activated upon miR-146 KD, in spite of their very low basal level of activation.

This preliminary experiment suggests that miR-146 could effectively modulate the transcriptional response related to specific pathways; of course, we will also monitor the effect of miR-146 on the 10 pathways in the presence of ligand stimulation, especially for those that are not constitutively active at the basal levels (Wnt, TGF-β and/or Notch) and in culture condition that are related to stem cell properties (such as mammosphere culture).

We expect that, as for other miRNAs, miR-146 may also display significant biological activity upon acute activation of its target genes, especially if it plays a major role in a negative feedback response (as already observed in the NFkB circuitry, (Taganov et al., 2006).

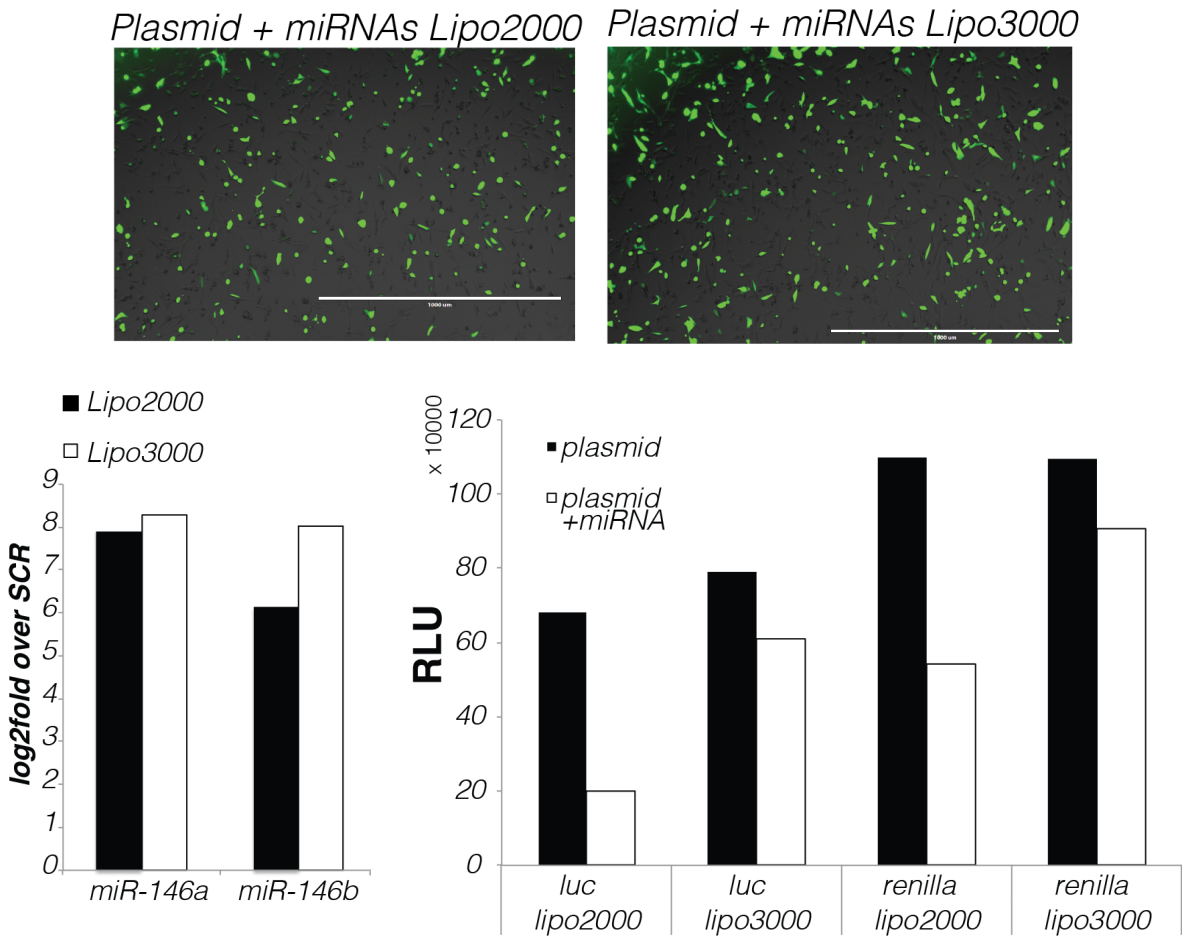


Figure 45- Protocol set up for Cignal 10-pathways reporter assay on SUM159- Representative images of 10.000 SUM159 cells transfected with a combination of GFP-reporter plasmid with 50 nM of miR-146 mimic, to compare transfection efficiency of Lipofectamine®2000 vs Lipofectamine®3000 (Invitrogen) 4X magnification, scale bar 1000 µm. The plot on the left reports the log_2fold of miR-146 overexpression (relative to SCR oligos) obtained from a comparison of Lipofectamine®2000 and Lipofectamine®3000. The plot on the right reports the relative luminescence units (RLU) of Firefly Luciferase and Renilla luciferase signals, transfected alone (plasmid, black bar) or in combination with 50 nM of miR-146 mimic (plasmid+miRNAs, white bar), from a comparison of Lipofectamine®2000 and Lipofectamine®3000.

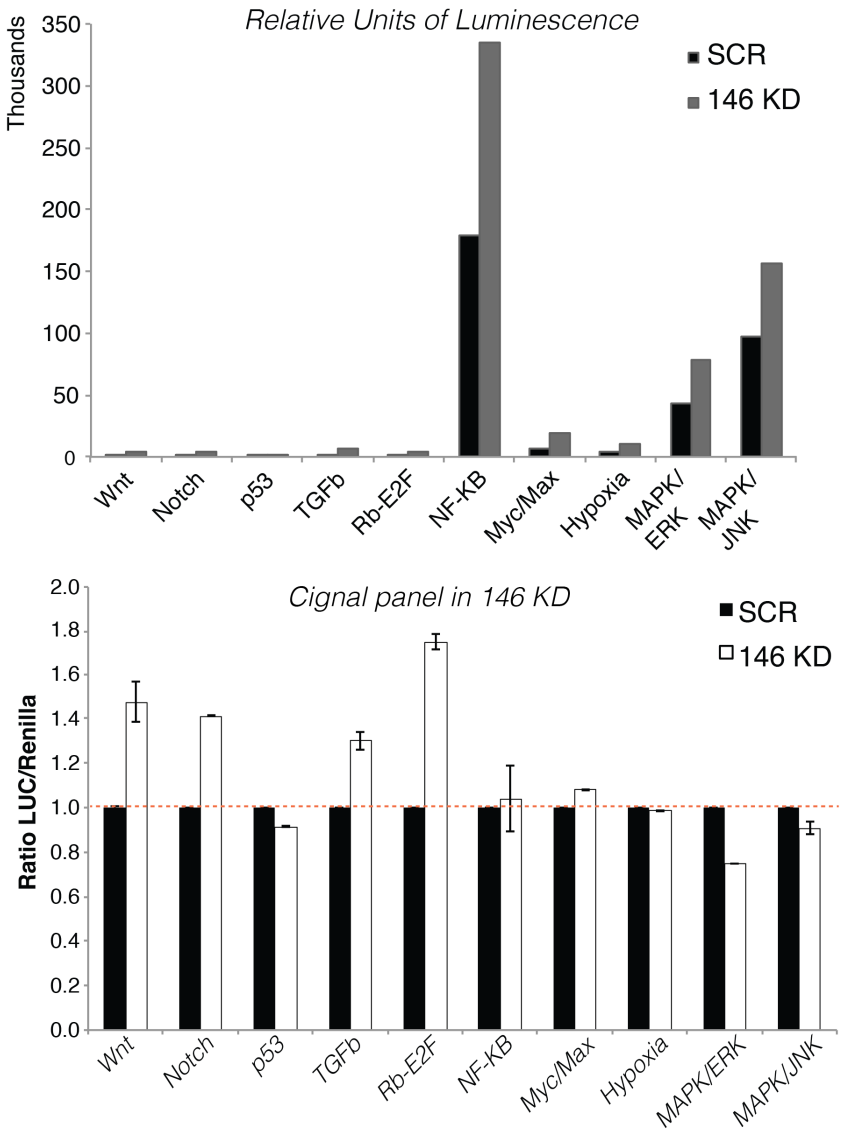


Figure 46- 10-cancer pathways modulation upon miR-146 KD in SUM159- 10.000 SUM159 cells, in low-serum conditions, were transfected with a panel of 10 reporters in the presence of 100 nM of control LNA or miR-146 LNA. After 24h Firefly luciferase and Renilla luciferase signals were measured through Dual Glo Luciferase Assay system. In the upper plot we report the relative units of firefly luminescence, indicating basal level of pathway activation. The lower plot reports the relative luminescence units (RLU) as ratio of Luciferase over Renilla signal (SCR, black bar, and miR-146 KD, white bar). The data are normalized over SCR and the error bars measured the standard deviation of three technical replicas.

4 Discussion

4.1 Identification of a "normal stem cell" miRNA signature

The past decade has seen the accumulation of a large body of evidence that miRNAs are critical regulators of stem cell self-renewal and important determinants for the balance between cell proliferation and differentiation. Normal SC regulation is often usurped in CSCs, and the relevance of this process to cancer biology is, by now, undisputed. It is not surprising, therefore, that we are discovering frequent instances of miRNA-dependent regulation of cancer-relevant pathways. However, we have only scratched the surface, and much remains to be understood of the exquisitely complex modes of regulation exerted by miRNAs. Our interest is breast cancer, and the role that SCs/CSCs play in its development. Therefore, out first step was to identify miRNAs that could specify the stem cell state in a physiological context.

We used the PKH26 dye-retaining assay to isolate quasi-pure populations of SCs/CSCs from mammosphere cultures, a strategy that has been validated in previous studies from our lab (Pece et al., 2010) and from others (Cicalese et al., 2009). To maximize our chances of isolating conserved, and thus important, miRNA regulators, we profiled the miRNAs from two independent normal breast models: mouse primary mammary epithelial cells (MECs) and a human normal cell line, MCF10A. The two models provide advantages and limitations. MECs are primary cells, directly isolated from the mammary gland of young mice and placed in short-term mammosphere culture, as a source of SCs. Although this model closely resembles physiological conditions, primary cultures can contain contaminants, such as stromal cells or endothelial/hematopoietic cells, which might adversely impact the miRNAs profiling. Conversely, the MCF10A model is a normal human established cell line, with no contaminating non-epithelial cells. This particular cell line is one of the few normal *in vitro* models that displays stem cell properties such as the ability to differentiate *in vitro* (Debnath et al., 2003), and to withstand anoikis while

generating clonal mammospheres in suspension culture. We hypothesized that miRNAs involved in the transcriptional circuitry that specifies SC identity and determines SC properties would be conserved in human and mouse. Thus, we searched for common miRNAs between the two models. We observed that the mouse MEC and human MCF10A profiles of SC miRNAs overlapped partially, but significantly ($p<0.05$), leading us to define a "normal stem cell miRNA signature", consisting of three miRNAs: miR-146a and miR-331-3p which were both upregulated in PKH^{pos} cells (phenotypically identical to SCs) vs. PKH^{neg} cells, and let-7a which was lost in SCs in PKH^{pos} cells vs. PKH^{neg} cells,

Let-7 is a known developmental regulator and general tumour suppressor. It has been shown to inhibit the self-renewal of SCs from various tissue compartments, including breast (Yu et al., 2007). In this latter case, the authors described that overexpression of let-7 in BCSCs dramatically impaired proliferation, mammosphere-forming ability, tumor formation and metastatic spread *in vivo.* On the contrary, reducing let-7 levels restored the self-renewing capacity of BCSCs (Yu et al., 2007). The expression of let-7 can also reduce self-renewal and promote the differentiation of ES cells (Melton et al., 2010), and the Lin28/let-7/HMGA2 axis strictly controls the self-renewal potential of fetal haematopoietic stem cells (Copley et al., 2013).

The other two miRNAs of our SC-signature, miR-146a and miR-331-3p, were not previously known to be expressed in the breast stem cell compartment. In human cancers, miR-331-3p is downregulated in prostate cancers with respect to normal tissues and appears to be a negative regulator of HER2 (Epis et al., 2009). It is also downregulated in gastric cancers where it can regulate E2F1 causing a cell-cycle block (Guo et al., 2010b). Recently, miR-331-3p was found to participate in a regulatory circuit with the long non-coding RNA HOTAIR, a very well known player in cancer progression and metastasis, that acts as a competing endogenous (ce)RNA for miR-331-3p thus relieving miR-331-3p-dependent repression of HER2 (Liu et al., 2014). Hence, the high levels of miR-331-3p in

SCs would be required to keep HER2 signalling inhibited. Indeed, HER2 overexpression has been shown to induce SC expansion in mice by increasing proliferation and inducing a symmetric mode of cell division (Cicalese et al., 2009). Thus, we can envisage that, through the inhibition of HER2 signalling, high levels of miR-331-3p might help maintain the proliferative equilibrium of SCs on the side of quiescence.

The role of miR-146a/b is very well characterized in the hematopoietic system, where it acts as fundamental component of the immune-cell regulation and as a critical tumor suppressor, keeping levels of Nf-κB signaling inder control, through a negative feedback loop involving the downmodulation of TRAF6 and IRAK1 (Taganov et al., 2006). In human cancer, miR-146a/b are described as either tumor suppressors or oncogenes, depending on the context. These miRNAs are upregulated in thyroid carcinomas (Geraldo et al., 2012), melanomas (Forloni et al., 2014) and triple-negative basal breast cancer (Garcia et al., 2011) and involved in the regulation of different signalling pathways, respectively the TGF-β pathway via SMAD4 inhibition (Geraldo et al., 2012), the Notch signalling pathway through Numb downmodulation (Forloni et al., 2014), and BRCA1 regulation in triple-negative breast tumors (Garcia et al., 2011). Considering that we found miR-146a to be upregulated in SCs, it is not surprising that these three pathways are all involved in SC biology (see Introduction, Section 3). Furthermore, very recent work highlights the existence of a positive feedback loop between Snail/miR-146a/ β-catenin/ Numb that is crucial for the expansion of the SC population and for controlling symmetric cell division of colorectal CSCs (Hwang et al., 2014).

Having identified a credible miRNA SC signature, we investigated whether the expression of miRNAs from the “normal stem cells signature” could stratify human breast cancers according to their phenotypic aggressiveness. Our rationale was based on the fact that human tumours, including breast cancers, contain a sub-population of CSCs with properties reminiscent of normal stem cells (Al-Hajj, 2003; Visvader and Lindeman,

2012). Further support to our rationale come from the fact that a normal SC gene transcriptional profile could be used to predict the CSCs content of breast tumours (Pece et al., 2010; Visvader and Lindeman, 2012). By using two independent datasets coming from The Cancer Genome Atlas network and the METABRIC consortium, we analysed 1152 breast cancer samples in total for expression of our SC-miRNA profile. We identified one group of tumors, here referred to as Group1/SC-like, that was significantly correlated with histological markers of aggressiveness (basal tumor subtype, ER and PR negative status), and with genetic alterations, such as p53 mutations and Myc amplification.

Furthermore, the Group1/SC-like tumors were enriched for high grade and estrogen receptor negative (G3-ER-) samples, a class that typically displays a higher frequency of CSCs, compared to low-grade ER+ tumours (G1-ER+) and normal breast epithelia (Pece et al., 2010). Hence, the miRNA normal SC signature seems to possess diagnostic/prognostic features related to stem cell biology in real human cancers.

4.2. miR-146 is a SC/CSCs specific miRNA

Although the normal miRNA SC signature that we identified contains three miRNA species, we focused on miR-146a/b for several reasons: i) miR-146a expression appeared to be specific for SCs compared to progenitors/differentiated cells (upregulated in PKH^{pos}); ii) miR-146a/b were also found by an independent group to be upregulated in both breast normal and cancer SCs, isolated from human biopsies through the use of surface markers (Shimono et al., 2009); and iii) our expression data obtained from SCs/CSCs isolated from three normal-like and seven primary breast tumors by using the PKH26 dye-retaining assay confirmed that upregulation of miR-146b was a common trait for breast SCs and CSCs, independently of the tumor molecular subtypes. Additionally, Hwang *et al.* performed deep sequencing analysis and found miR-146a to be the top upregulated miRNA in colorectal CSCs with respect to parental cells (Hwang et al., 2014). In bulk breast tissues, miR-146a/b expression was low/absent in normal epithelia and luminal

breast cancers, while it appeared to be specifically expressed in the basal-like tumors and in claudin-low tumor subtypes (our data from TGCA dataset and human cell lines), the most aggressive subtypes of breast cancers, closely associated to SCs (Pece et al., 2010). This finding is in agreement with previous work from another group, showing that miR-146a/b were overexpressed in triple-negative/high-grade breast primary tumors of an independent cohort (Garcia et al., 2011).

The expression pattern of miR-146a/b suggested it could behave as a marker of SCs/CSCs. We tested this hypothesis directly with the use of two sensors for miR-146a/b, which were capable of stratifying subpopulations with different endogenous levels of miR-146. Using primary MECs or SUM159 cancer cells, we found that not all the cells within the population expressed the same levels of miR-146a/b and, more importantly, cells with high endogenous levels of miR-146a/b could form mammospheres *in vitro* more efficiently than their miR-146a/b^{low} counterpart, confirming a role for miR-146a/b as marker for SCs and CSCs.

In the future, it will be interesting to evaluate if miR-146^{high} cells can retain additional properties associated with SCs beyond their mere ability to form mammosphere *in vitro*. For example, by using 2D clonogenic assay and/or 3D organotypic assay, we still need to determine if miR-146^{high} cells have an increased ability to differentiate *in vitro* or to reconstitute an entire mammary gland in *in vivo*, compared to their miR-146^{low} counterpart. This would be in line with the hypothesis that a miR-146^{high} population is enriched in SCs, and would bring the miR-146a/b family one step close to being considered a *bona fide* SC marker.

4.3. Regulation of expression of miR-146a/b in normal and cancer stem cells

Given their important role in the regulation of the SC phenotype, we needed to understand the transcriptional regulation mechanisms of the two members of miR-146 family in the

normal and tumoral breast tissues. As previously discussed, (Introduction, Section 1.6) miR-146a has been extensively studied in the immune and haematopoietic system, where its expression regulation is NF-κB dependent (Taganov et al., 2006). Taganov *et al.* showed that the genomic locus of pri-miR-146a possesses three NF-κB responsive elements exactly 1.5 kb upstream of the pri-miRNA, but similar sites were not found upstream to pri-miR-146b (Taganov et al., 2006).

Conversely, the miR-146b gene possesses two binding sites for C/EBPβ (CCAAT/enhancer binding protein β), located 2 kb upstream from its transcription starting site (TSS) and able to induce miR-146b transcription (Li et al., 2014). Since the TGF-β response in breast epithelial cancer cells has been linked to C/EBPβ (Gomis et al., 2006) and C/EBPβ was found to regulate stem cells activity in the mammary gland (LaMarca et al., 2010), it is tempting to speculate that miR-146b expression in breast could be dependent on the TGF-β-C/EBPβ axis. Therefore, although miR-146a and -146b biological functions appear to be redundant, their expression in the breast compartment could be dependent on different signalling pathways (NF-κB and C/EBPβ). Within this scenario, transcription of the two miR-146 members in the normal and in the tumor compartment could either be regulated separately (with miR-146b more frequently induced in breast CSCs than miR-146a), or be dependent, at different extent, on a common signaling mechanism involved in self-renewal and/or transformation (such as IL-6 pathway) (Hartman et al., 2013; Hinohara and Gotoh, 2010).

4.4. miR-146 as a target for inhibition of CSC self-renewal

As frequently occurs in miRNA biology, the specific expression pattern of a miRNA correlates with its role in regulating the biological features and/or the behaviour of a given cell type. Our results suggest that miR-146a/b is not simply a marker for breast SCs/CSCs, but it is also required for their biological functions. Strong evidence came from our

experiments inhibiting self-renewal in both human SUM159 mammospheres and primary MEC mammospheres upon loss of expression of miR-146a/b. This was achieved with the use of a lentiviral vector expressing an "antimiR" for miR-146a/b, able to reduce the endogenous levels of miR-146a/b in the cells, impairing their loading on the miRISC complex and preventing their interaction with natural targets mRNAs.

A tumorigenicity assay *in vivo* showed that miR-146 inhibition also had a marked effect in reducing the efficiency of tumor formation of the human SUM159 cancer cell line. In addition, miR-146 loss also affected the efficiency of patient-derived xenografts (PDXs). At the time of writing, we have collected five different PDXs, whose complete characterization upon miR-146 loss is still ongoing. Our preliminary results showed that miR-146 KD impairs tumor formation, reduces tumor mean volume and decreases CSC frequency in at least two different PDXs. These results suggest that miR-146 targeting might be exploited to inhibit human tumours and to weaken CSCs self-renewal and proliferation.

Nowadays, a number of so-called "stem-inhibitor" miRNAs have been identified in the breast context. This includes let-7, miR-200c and miR-93, whose levels are low in SCs with respect to progenitor/differentiated cells and are typically increased upon differentiation. The mechanism of SC/CSC inhibition exerted by these miRNAs relies on the targeting of critical cell fate determinants, such as BMI regulated by miR-200c, (Shimono et al., 2009), HMGA2 targeted by Let-7 (Yu et al., 2007), or multiple SCs-related targets by miR-93, (Liu et al., 2012). In all these cases, it remains unclear whether the regulation of the stem cell fate is the physiological function of these miRNAs, since the inhibition of self-renewal and the induction of differentiation has so far been obtained upon the forced overexpression of miRNAs in compartment (such as SCs or BCSCs) where they are normally little or barely expressed. Differently to what has been uncovered to date, we have identified a family of miRNAs, miR-146a/b, whose expression is specific for the stem

cell compartment, and similarly is required for SC functions in murine primary and human cells, highlighting a potential conserved mechanism controlling SC properties. To our knowledge, this is the first time a miRNA has been discovered that is specifically expressed in the SCs/CSCs compartment of the mammary gland, and that is specifically required for stem cell function as demonstrated by a loss-of-function (miRNA inhibition) approach.

4.5. miR-146 is required to sustain multiple stem cell pathways

The function of miRNAs is often exerted through various epigenetic and transcriptional mechanisms, thus generating complex regulatory networks of gene expression that can robustly and sharply define cell fate and behaviour (Gangaraju and Lin, 2009). We identified 1975 genes regulated in mammospheres upon loss of miR-146, a condition in which self-renewal of BCSCs was impaired. A part of this transcriptional response was due to the reversion of the gene expression program that is activated upon mammosphere formation, suggesting that the loss of miR-146 expression impairs the gene expression program needed to specify stem properties. Indeed, comparison with four independent gene signatures of mammary SCs isolated usinf different approaches (SC signatures – (Stingl et al., 2006; Shipitsin et al., 2007; Lim et al., 2010; Pece et al., 2010) revealed that the core of genes downregulated by miR-146 KD significantly overlaps with genes activated in the stem compartment. Hence, miR-146 expression appears to be required to sustain the stem program. In particular, we identified 20 upstream signalling pathways activated in multiple SC signatures, whose regulation is mirrored by loss of miR-146, including TGF-β signalling, pro-inflammatory cytokines and the Nf-κB pathway. As mammary stem cell is a complex system in which individual pathways interact extensively to maintain the delicate balance between self-renewal, proliferation and differentiation (Liu et al., 2005; Izrailit and Reedijk, 2012), it is reasonable to suggest that the miR-146 family

might function as a rheostat coordinating multiple signalling pathways. Importantly, these signalling pathways are significantly enriched in stem compartment of both normal and cancer SCs, as previously shown by (Stingl et al., 2006; Shipitsin et al., 2007; Lim et al., 2010; Pece et al., 2010).

We hypothesize two main mechanisms through which miR-146 exerts its biological activity: i) miR-146 might repress a negative regulator that controls a single or multiple stem-pathways ("hub" mechanism); ii) miR-146 might act on multiple genes, "buffering" their variation to allow optimal signalling ("fine-tuner" mechanisms).

The first mechanism might apply to the control exerted by miR-146 on TRAF6/IRAK1 for the Nf-κB pathway or on Numb for the Notch pathway. According to this model, the loss of miR-146 in SCs/CSCs causes the upregulation of these negative regulators that limit the SC/CSC pool; moreover, according to the model the SC phenotype can be rescued by inhibiting the negative regulator (such as Numb or TRAF6). This mechanism has been recently shown to occur in colorectal CSCs, where the induction of miR-146a causes the downregulation of Numb, thus increasing Notch signalling activation and β-catenin stabilization, and resulting in a positive feedback loop that is critical for the expansion of the cancer stem cell pool (Hwang et al., 2014). Although a role for Numb and Notch signalling in the control of self-renewal of breast SCs/CSCs has been extensively suggested (Bouras et al., 2008; Dontu et al., 2004), our preliminary experiments seem to exclude the involvement of Numb as the main effector of the miR-146 mediated SC-phenotype. Indeed, circumventing Numb accumulation (through preventive silencing of Numb) in miR-146 KD cells, was not sufficient to rescue the impaired self-renewal of SUM159 mammospheres, suggesting that other mechanisms (that not impinge on Numb) are at work. A likely candidate might be the Nf-κB pathway, which is regulated by at least two miR-146 targets, TRAF6 and IRAK1 (Taganov et al., 2006). Indeed, recent works has highlighted the role of NF-κB in controlling breast TICs/CSCs, particularly in the human

basal-phenotype, through the activation of Jag1 and the Notch-dependent expansion of CSCs (Yamamoto M, 2013); alternatively, NF-κB can promote self-renewal in claudin-low breast cancers through the induction of EMT and proinflammatory cytokines (such as interleukin-1β and Il-6) (Kendellen et al., 2014). This last hypothesis is particular appealing, since NF-κB signalling is very active in the claudin-low SUM159 cell line, which also shows very high levels of miR-146. It is tempting to speculate that, in absence of miR-146, NF-κB signalling is impaired, leading to inhibition of EMT program (TGF-β signalling) and to proinflammatory cytokine signalling (such as IL-6), thus affecting self-renewal.

In a different perspective, the fine-tuner mechanism suggests that miR-146 could keep the activation of stem pathways under strict surveillance by "buffering" the levels of multiple endogenous targets and limiting undesired signalling fluctuations. We performed bioinformatic predictions to derive a list of 91 stem cell genes (from the above-mentioned SC-signatures), that might be putatively targeted by miR-146. Of these, we found 37 genes (approximately the 40% of the total) that were modulated at the transcriptional level by miR-146 overexpression or by its knockdown. This list includes several examples of targets that belong to the same pathway but act in opposite directions, such as CARD10 and IRAK1/TRAF6, respectively an activator and two repressors of Nf-κB signalling (Taganov et al., 2006; Crone et al., 2012); or JAG1 and NUMB, which act in opposite ways to modulate Notch signalling (JAG1 is a ligand/activator and NUMB a repressor).

As previously discussed, in physiological conditions the stem cell activates several pathways necessary to maintain tightly controlled self-renewal and the balance between asymmetric/symmetric cell division. These include the induction of: the Notch pathway (as demonstrated for miR-34 and for miR-146a in colorectal CSCs), the TGF-β pathway (Shipitsin *et al*. reported the TGF-β pathway as one of the most potent pathways activated in SCs and CSCs (Shipitsin et al., 2007), or the inflammatory responses converging on Nf-

κB pathway, mediated by TNF and interleukins stimulation and extensively characterized as regulators of BCSCs (Iliopoulos et al., 2009; Kendellen et al., 2014; Kim et al., 2013).

According to the "buffering" hypothesis, in absence of miR-146 these signalling that balance stem cell functions become unstable, with too much or too little activity, resulting in impaired SC/CSC self-renewal ability. Importantly, compared to the "hub-mechanism" hypothesis in which the self renewal of SCs/CSCs could be restored by acting directly on the hub targeted by the miRNA, in the "buffering" hypothesis none of the miR-146 targets idividually should rescue the effect of miR-146 loss-of-function, although restoring miR-146 signalling to physiological levels should.

In conclusion, we propose that the miR-146 family members act as key controllers of multiple signaling pathways, that are tightly interconnected to modulate SC/CSC self-renewal. Because of the ability of miR-146a/b to interact at multiple layers of gene regulation, a promising strategy to limit CSC expansion might be the inhibition of these miRNAs, with low or limited chances for the tumor to escape from miR-146 loss-of function effects. To date, the therapeutic inhibition of miRNA function in vivo has been effectively approached for specific diseases through the use of "antagomiRs" or "LNA-based inhibitors" therapies, as reviewed in (Li and Rana, 2014). Whether this approach will be exploitable for the treatment of breast cancers remains to be seen, but the promise of a therapy that can effectively target the BCSCs should be a sufficiently tantalizing reward to motivate further investigations.

5 References

1. Al-Hajj, M. (2003). Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences 100, 3983-3988.
2. Ayyanan, A., Civenni, G., Ciarloni, L., Morel, C., Mueller, N., Lefort, K., Mandinova, A., Raffoul, W., Fiche, M., Dotto, G.P., *et al.* (2006). Increased Wnt signaling triggers oncogenic conversion of human breast epithelial cells by a Notch-dependent mechanism. Proceedings of the National Academy of Sciences of the United States of America *103*, 3799-3804.
3. Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.
4. Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215-233.
5. Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nat Genet *35*, 215-217.
6. Borchert, G.M., Lanier, W., and Davidson, B.L. (2006). RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol *13*, 1097-1101.
7. Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M.L., Oakes, S.R., Lindeman, G.J., and Visvader, J.E. (2008). Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell Stem Cell *3*, 429-441.
8. Brantley, D., Chih-Li Chen, Rebecca S. Muraoka,, Paul B. Bushdid, J.L.B., Frances Kittrell, Daniel Medina,, and Lynn M. Matrisian, L.D.K., and Fiona E. Yull (2001). Nuclear Factor-kB (NF-kB) Regulates Proliferation and Branching in Mouse Mammary Epithelium
9. Brown, B.D., Gentner, B., Cantore, A., Colleoni, S., Amendola, M., Zingale, A., Baccarini, A., Lazzari, G., Galli, C., and Naldini, L. (2007). Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. Nature biotechnology *25*, 1457-1467.
10. Bu, P., Chen, K.-Y., Chen, Joyce H., Wang, L., Walters, J., Shin, Yong J., Goerger, Julian P., Sun, J., Witherspoon, M., Rakhilin, N., et al. (2013). A microRNA miR-34a-Regulated Bimodal Switch Targets Notch in Colon Cancer Stem Cells. Cell Stem Cell 12, 602-615.
11. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., *et al.* (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A *101*, 2999-3004.
12. Cao, Y.G.B., Tiffany N. Seagroves, Florian R. Greten, Randall Johnson,Emmett V. Schmidt, and Michael Karin (2001). IKK Provides an Essential Link between RANK Signaling and Cyclin D1 Expression during Mammary Gland Development.
13. Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.
14. Chao, C.H., Chang, C.C., Wu, M.J., Ko, H.W., Wang, D., Hung, M.C., Yang, J.Y., and Chang, C.J. (2014). MicroRNA-205 signaling regulates mammary stem cell fate and tumorigenesis. The Journal of clinical investigation J Clin Invest. doi:10.1172/JCI73351.
15. Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83-86.

16. Cheng, L.C., Pastrana, E., Tavazoie, M., and Doetsch, F. (2009). miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. Nature neuroscience *12*, 399-408.
17. Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.
18. Chua, H.L., Bhat-Nakshatri, P., Clare, S.E., Morimiya, A., Badve, S., and Nakshatri, H. (2007). NF-kappaB represses E-cadherin expression and enhances epithelial to mesenchymal transition of mammary epithelial cells: potential involvement of ZEB-1 and ZEB-2. Oncogene *26*, 711-724.
19. Cicalese, A., Bonizzi, G., Pasi, C.E., Faretta, M., Ronzoni, S., Giulini, B., Brisken, C., Minucci, S., Di Fiore, P.P., and Pelicci, P.G. (2009). The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. Cell *138*, 1083-1095.
20. Copley, M.R., Babovic, S., Benz, C., Knapp, D.J., Beer, P.A., Kent, D.G., Wohrer, S., Treloar, D.Q., Day, C., Rowe, K.*, et al.* (2013). The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. Nature cell biology *15*, 916-925.
21. Croce, C.M. (2008). Oncogenes and cancer. N Engl J Med *358*, 502-511.
22. Crone, S.G., Jacobsen, A., Federspiel, B., Bardram, L., Krogh, A., Lund, A.H., and Friis-Hansen, L. (2012). microRNA-146a inhibits G protein-coupled receptor-mediated activation of NF-kappaB by targeting CARD10 and COPS8 in gastric cancer. Molecular cancer *11*, 71.
23. Debnath, J., Muthuswamy, S.K., and Brugge, J.S. (2003). Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods *30*, 256-268.
24. Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes & development *17*, 1253-1270.
25. Dontu, G., Jackson, K.W., McNicholas, E., Kawamura, M.J., Abdallah, W.M., and Wicha, M.S. (2004). Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. Breast cancer research : BCR *6*, R605-615.
26. Eirew, P., Stingl, J., Raouf, A., Turashvili, G., Aparicio, S., Emerman, J.T., and Eaves, C.J. (2008). A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. Nature medicine *14*, 1384-1389.
27. Epis, M.R., Giles, K.M., Barker, A., Kendrick, T.S., and Leedman, P.J. (2009). miR-331-3p regulates ERBB-2 expression and androgen receptor signaling in prostate cancer. The Journal of biological chemistry *284*, 24696-24704.
28. Fernandez, A., Huggins, I.J., Perna, L., Brafman, D., Lu, D., Yao, S., Gaasterland, T., Carson, D.A., and Willert, K. (2014). The WNT receptor FZD7 is required for maintenance of the pluripotent state in human embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *111*, 1409-1414.
29. Fillmore, C.M., and Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast cancer research : BCR *10*, R25.
30. Forloni, M., Shaillay Kumar Dogra, Yuying Dong, Darryl Conte Jr,, Jianhong Ou, L.J.Z., April Deng, Meera Mahalingam,, and Michael R Green, N.W. (2014). miR-146a promotes the initiation and progression of melanoma by activating Notch signaling. eLife *DOI: 10.7554/eLife.01460*.

31. Fre, S., Huyghe, M., Mourikis, P., Robine, S., Louvard, D., and Artavanis-Tsakonas, S. (2005). Notch signals control the fate of immature progenitor cells in the intestine. Nature *435*, 964-968.
32. Gangaraju, V.K., and Lin, H. (2009). MicroRNAs: key regulators of stem cells. Nature reviews Molecular cell biology 10, 116-125.
33. Garcia, A.I., Buisson, M., Bertrand, P., Rimokh, R., Rouleau, E., Lopez, B.S., Lidereau, R., Mikaelian, I., and Mazoyer, S. (2011). Down-regulation of BRCA1 expression by miR-146a and miR-146b-5p in triple negative sporadic breast cancers. EMBO molecular medicine *3*, 279-290.
34. Garzon, R., Calin, G.A., and Croce, C.M. (2009). MicroRNAs in Cancer. Annu Rev Med *60*, 167-179.
35. Geraldo, M.V., Yamashita, A.S., and Kimura, E.T. (2012). MicroRNA miR-146b-5p regulates signal transduction of TGF-beta by repressing SMAD4 in thyroid cancer. Oncogene *31*, 1910-1922.
36. Ghani, S., Riemke, P., Schonheit, J., Lenze, D., Stumm, J., Hoogenkamp, M., Lagendijk, A., Heinz, S., Bonifer, C., Bakkers, J.*, et al.* (2011). Macrophage development from HSCs requires PU.1-coordinated microRNA expression. Blood *118*, 2275-2284.
37. Gomis, R.R., Alarcon, C., Nadal, C., Van Poznak, C., and Massague, J. (2006). C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells. Cancer cell *10*, 203-214.
38. Gjorevski, N., and Nelson, C.M. (2011). Integrated morphodynamic signalling of the mammary gland. Nature reviews Molecular cell biology *12*, 581-593.
39. Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.
40. Guo, X., Guo, L., Ji, J., Zhang, J., Chen, X., Cai, Q., Li, J., Gu, Q., Liu, B., Zhu, Z.*, et al.* (2010). miRNA-331-3p directly targets E2F1 and induces growth arrest in human gastric cancer. Biochemical and Biophysical Research Communications *398*, 1-6.
41. Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633-644.
42. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M.*, et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.
43. Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.
44. Harrison, H., Farnie, G., Howell, S.J., Rock, R.E., Stylianou, S., Brennan, K.R., Bundred, N.J., and Clarke, R.B. (2010). Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Research *70*, 709-718.
45. Hartman, Z.C., Poage, G.M., den Hollander, P., Tsimelzon, A., Hill, J., Panupinthu, N., Zhang, Y., Mazumdar, A., Hilsenbeck, S.G., Mills, G.B.*, et al.* (2013). Growth of triple-negative breast cancer cells relies upon coordinate autocrine expression of the proinflammatory cytokines IL-6 and IL-8. Cancer Research *73*, 3470-3480.
46. Hinohara, K., and Gotoh, N. (2010). Inflammatory signaling pathways in self-renewing breast cancer stem cells. Current opinion in pharmacology *10*, 650-654.

47. Houbaviy, H.B., Murray, M.F., and Sharp, P.A. (2003). Embryonic Stem Cell-Specific MicroRNAs. Developmental Cell *5*, 351-358.
48. Howard, B.a.B.A.G. (2000). Human Breast Development. Journal of mammary gland biology and neoplasia *Vol. 5* 1083-3021/1000/0400-0119.
49. Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L.*, et al.* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC genomics *7*, 96.
50. Hu, Y., and Smyth, G.K. (2009). ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. Journal of immunological methods *347*, 70-78.
51. Hwang, W.-L., Jiang, J.-K., Yang, S.-H., Huang, T.-S., Lan, H.-Y., Teng, H.-W., Yang, C.-Y., Tsai, Y.-P., Lin, C.-H., Wang, H.-W.*, et al.* (2014). MicroRNA-146a directs the symmetric division of Snail-dominant colorectal cancer stem cells. Nature Cell Biology *16*, 268-280.
52. Iliopoulos, D., Hirsch, H.A., and Struhl, K. (2009). An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. Cell *139*, 693-706.
53. Imamura, T., Masao Takase*, Ayako Nishihara, Eiichi Oeda, Jun-ichi Hanai, Masahiro Kawabata & Kohei Miyazono (1997). Smad6 inhibits signalling by the TGF-b superfamily. Nature *VOL 389 | 9 OCTOBER 1997*.
54. Iso, T., Kedes, L., and Hamamori, Y. (2003). HES and HERP families: multiple effectors of the Notch signaling pathway. Journal of Cellular Physiology *194*, 237-255
55. Iwatsubo, T. (2004). The gamma-secretase complex: machinery for intramembrane proteolysis. Current opinion in neurobiology *14*, 379-383.
56. Izrailit, J., and Reedijk, M. (2012). Developmental pathways in breast cancer and breast tumor-initiating cells: therapeutic implications. Cancer letters *317*, 115-126.
57. Jolicoeur, F. (2005). Intrauterine breast development and the mammary myoepithelial lineage. Journal of mammary gland biology and neoplasia *10*, 199-210.
58. Joshi, P.A., Di Grappa, M.A., and Khokha, R. (2012). Active allies: hormones, stem cells and the niche in adult mammopoiesis. Trends in endocrinology and metabolism: TEM *23*, 299-309.
59. Karamboulas, C., and Ailles, L. (2013). Developmental signaling pathways in cancer stem cells of solid tumors. Biochimica et biophysica acta *1830*, 2481-2495.
60. Karginov, F.V., Conaco, C., Xuan, Z., Schmidt, B.H., Parker, J.S., Mandel, G., and Hannon, G.J. (2007). A biochemical approach to identifying microRNA targets. Proceedings of the National Academy of Sciences of the United States of America *104*, 19291-19296.
61. Keller PJ, A.L., Skibinski A, Logvinenko T, Klebba I, Dong S, Smith AE, Prat A, Perou CM, Gilmore H, Schnitt S, Naber SP, Garlick JA, Kuperwasser C. (2012). Defining the cellular precursors to human breast cancer. PNAS *vol. 109 | no. 8*, 2772–2777.
62. Kendellen, M.F., Bradford, J.W., Lawrence, C.L., Clark, K.S., and Baldwin, A.S. (2014). Canonical and non-canonical NF-kappaB signaling promotes breast cancer tumor-initiating cells. Oncogene *33*, 1297-1305.
63. Kendellen, M.F., Bradford, J.W., Lawrence, C.L., Clark, K.S., and Baldwin, A.S. (2014). Canonical and non-canonical NF-kappaB signaling promotes breast cancer tumor-initiating cells. Oncogene *33*, 1297-1305

64. Kogo, R., Mimori, K., Tanaka, F., Komune, S., and Mori, M. (2011). Clinical significance of miR-146a in gastric cancer cases. Clinical cancer research : an official journal of the American Association for Cancer Research *17*, 4277-4284.
65. Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research *42*, D68-73.
66. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.
67. Lal, A., Navarro, F., Maher, C.A., Maliszewski, L.E., Yan, N., O'Day, E., Chowdhury, D., Dykxhoorn, D.M., Tsai, P., Hofmann, O.*, et al.* (2009). miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. Molecular cell *35*, 610-625.
68. Lamb R, M.P.A., Katherine Spence, Goran Landberg, Andrew H. Sims,, and Clarke, R.B. (2013). Wnt pathway activity in breast cancer sub-types and stem-like cells. Plos one *Volume 8 | Issue 7 | e67811*.
69. Lamouille, S., Connolly, E., Smyth, J.W., Akhurst, R.J., and Derynck, R. (2012). TGF-beta-induced activation of mTOR complex 2 drives epithelial-mesenchymal transition and cell invasion. Journal of cell science *125*, 1259-1273.
70. Lee, D., and Shin, C. (2012). MicroRNA-target interactions: new insights from genome-wide approaches. Annals of the New York Academy of Sciences *1271*, 118-128.
71. Lee P. Lim, N.C.L., Philip Garrett-Engele, Andrew Grimson, Janell M. Schelter, John Castle, David P. Bartel, Peter S. Linsley & Jason M. Johnson (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *VOL 433*.
72. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.
73. Kim, S.Y., Kang, J.W., Song, X., Kim, B.K., Yoo, Y.D., Kwon, Y.T., and Lee, Y.J. (2013). Role of the IL-6-JAK1-STAT3-Oct-4 pathway in the conversion of non-stem cancer cells into cancer stem-like cells. Cellular signalling *25*, 961-969.
74. LaMarca, H.L., Visbal, A.P., Creighton, C.J., Liu, H., Zhang, Y., Behbod, F., and Rosen, J.M. (2010). CCAAT/enhancer binding protein beta regulates stem cell activity and specifies luminal cell fate in the mammary gland. Stem cells *28*, 535-544.
75. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S.*, et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.
76. Lena, A.M., Shalom-Feuerstein, R., Rivetti di Val Cervo, P., Aberdam, D., Knight, R.A., Melino, G., and Candi, E. (2008). miR-203 represses 'stemness' by repressing DeltaNp63. Cell death and differentiation *15*, 1187-1195.
77. Leung, A.K., Young, A.G., Bhutkar, A., Zheng, G.X., Bosson, A.D., Nielsen, C.B., and Sharp, P.A. (2011). Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. Nature structural & molecular biology *18*, 237-244.
78. Li, J., Shan, F., Xiong, G., Wang, J.-M., Wang, W.-L., Xu, X., and Bai, Y. (2014). Transcriptional regulation of miR-146b by C/EBPβ LAP2 in esophageal cancer cells. Biochemical and Biophysical Research Communications *446*, 267-271.

79. Li, Z., and Rana, T.M. (2014). Therapeutic targeting of microRNAs: current status and future challenges. Nature reviews Drug discovery *13*, 622-638.
80. Lim, E., Wu, D., Pal, B., Bouras, T., Asselin-Labat, M.-L., Vaillant, F., Yagita, H., Lindeman, G.J., Smyth, G.K., and Visvader, J.E. (2010). Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. Breast Cancer Research *12*, R21.
81. Lindvall, C., Evans, N.C., Zylstra, C.R., Li, Y., Alexander, C.M., and Williams, B.O. (2006). The Wnt signaling receptor Lrp5 is required for mammary ductal stem cell activity and Wnt1-induced tumorigenesis. The Journal of biological chemistry *281*, 35081-35087.
82. Liu, S., Dontu, G., and Wicha, M.S. (2005). Mammary stem cells, self-renewal pathways, and carcinogenesis. Breast Cancer Research *7*, 86.
83. Liu, S., Patel, S.H., Ginestier, C., Ibarra, I., Martin-Trevino, R., Bai, S., McDermott, S.P., Shang, L., Ke, J., Ou, S.J.*, et al.* (2012). MicroRNA93 regulates proliferation and differentiation of normal and malignant breast stem cells. PLoS genetics *8*, e1002751.
84. Liu, M., Sakamaki, T., Casimiro, M.C., Willmarth, N.E., Quong, A.A., Ju, X., Ojeifo, J., Jiao, X., Yeow, W.S., Katiyar, S.*, et al.* (2010). The Canonical NF- B Pathway Governs Mammary Tumorigenesis in Transgenic Mice and Tumor Stem Cell Expansion. Cancer Research *70*, 10464-10473.
85. Liu, X.H., Sun, M., Nie, F.Q., Ge, Y.B., Zhang, E.B., Yin, D.D., Kong, R., Xia, R., Lu, K.H., Li, J.H.*, et al.* (2014). Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. Molecular cancer *13*, 92.
86. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.
87. Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M.*, et al.* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715.
88. Marzi, M.J., Puggioni, E.M.R., Dall'Olio, V., Bucci, G., Bernard, L., Bianchi, F., Crescenzi, M., Di Fiore, P.P., and Nicassio, F. (2012). Differentiation-associated microRNAs antagonize the Rb-E2F pathway to restrict proliferation. The Journal of Cell Biology *199*, 77-95.
89. Massague, J. (2008). TGFbeta in Cancer. Cell *134*, 215-230.
90. Massague, J. (2012). TGFbeta signalling in context. Nature reviews Molecular cell biology *13*, 616-630.
91. Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes & development *19*, 2783-2810.
92. Maziere, P., and Enright, A.J. (2007). Prediction of microRNA targets. Drug discovery today *12*, 452-458.
93. Mei, J., Bachoo, R., and Zhang, C.L. (2011). MicroRNA-146a inhibits glioma development by targeting Notch1. Molecular and cellular biology *31*, 3584-3592.
94. Melton, C., Judson, R.L., and Blelloch, R. (2010). Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature *463*, 621-626.
95. Pece, S. (2004). Loss of negative regulation by Numb over Notch is relevant to human breast carcinogenesis. The Journal of Cell Biology *167*, 215-221.
96. Pece, S., Tosoni, D., Confalonieri, S., Mazzarol, G., Vecchi, M., Ronzoni, S., Bernard, L., Viale, G., Pelicci, P.G., and Di Fiore, P.P. (2010). Biological and

molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell *140*, 62-73.

97. Perou CM, T.S., Michael B. Eisen, Matt van de Rijn, Stefanie S. Jeffreyk, Christian A. Rees,Jonathan R. Pollack Douglas T. Ross, Hilde Johnsen,Lars A. Akslen, éystein FlugeI, Alexander Pergamenschikov, Cheryl Williams, Shirley X. Zhu, Per E. Lùnning, Anne-Lise Bùrresen-Dale, Patrick O. Brown & David Botstein (2000). Molecular portraits of human breast tumours. Nature *VOL 406*.
98. Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., Iscove, N., Jones, S., McKinney, S., Emerman, J.*, et al.* (2008). Transcriptome analysis of the normal human mammary cell commitment and differentiation process. Cell Stem Cell *3*, 109-118.
99. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.
100. Rios, A.C., Fu, N.Y., Lindeman, G.J., and Visvader, J.E. (2014). In situ identification of bipotent stem cells in the mammary gland. Nature *506*, 322-327.
101. Sakaki-Yumoto, M., Katsuno, Y., and Derynck, R. (2013). TGF-beta family signaling in stem cells. Biochim Biophys Acta *1830*, 2280-2296.
102. Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.
103. Sevignani, C., Calin, G.A., Siracusa, L.D., and Croce, C.M. (2006). Mammalian microRNAs: a small world for fine-tuning gene expression. Mammalian genome : official journal of the International Mammalian Genome Society *17*, 189-202.
104. Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84-88.
105. Shimono, Y., Zabala, M., Cho, R.W., Lobo, N., Dalerba, P., Qian, D., Diehn, M., Liu, H., Panula, S.P., Chiao, E.*, et al.* (2009). Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. Cell *138*, 592-603.
106. Shipitsin, M., Campbell, L.L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., Nikolskaya, T., Serebryiskaya, T., Beroukhim, R., Hu, M.*, et al.* (2007). Molecular definition of breast tumor heterogeneity. Cancer cell *11*, 259-273.
107. Shostak, K.a.A.C. (2011). NF-κB, stem cells and breast cancer: the links get stronger. *Breast Cancer Research 2011, 13:214*.
108. Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S.*, et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America *98*, 10869-10874.
109. Sorrentino, A., Thakur, N., Grimsby, S., Marcusson, A., von Bulow, V., Schuster, N., Zhang, S., Heldin, C.H., and Landstrom, M. (2008). The type I TGF-beta receptor engages TRAF6 to activate TAK1 in a receptor kinase-independent manner. Nature cell biology *10*, 1199-1207.
110. Speiser, J., Foreman, K., Drinka, E., Godellas, C., Perez, C., Salhadar, A., Ersahin, C., and Rajan, P. (2012). Notch-1 and Notch-4 biomarker expression in triple-negative breast cancer. International journal of surgical pathology *20*, 139-145.
111. Starczynowski, D.T., Kuchenbauer, F., Wegrzyn, J., Rouhi, A., Petriv, O., Hansen, C.L., Humphries, R.K., and Karsan, A. (2011). MicroRNA-146a disrupts

hematopoietic differentiation and survival. Experimental hematology *39*, 167-178 e164.

112. Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993-997.
113. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.
114. Taganov, K.D., Boldin, M.P., Chang, K.J., and Baltimore, D. (2006). NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. Proceedings of the National Academy of Sciences of the United States of America *103*, 12481-12486.
115. Tang, B., Bottinger E, Jakowlew S, Bagnall K, Mariano J, Anver M, Letterio J & Wakefield L (1998). Transforming growth factor B1 is a new form of tumor suppressor with true haploid insufficiency. Nature Medicine *Volume 4, Number 7*.
116. TCGA (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61-70.
117. Uyttendaele H, J.V.S., Roberto Montesano, and Jan Kitajewski (1998). Notch4 and Wnt-1 Proteins Function to Regulate Branching Morphogenesis of Mammary Epithelial Cells in an Opposing Fashion. Developmental Biology *196*, 204-217.
118. van Amerongen, R., Bowman, A.N., and Nusse, R. (2012). Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland. Cell Stem Cell *11*, 387-400.
119. van Dongen, S., Abreu-Goodger, C., and Enright, A.J. (2008). Detecting microRNA binding and siRNA off-target effects from expression data. Nature methods *5*, 1023-1025.
120. Van Keymeulen, A., Rocha, A.S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. Nature *479*, 189-193.
121. Ventura, A., and Jacks, T. (2009). MicroRNAs and cancer: short RNAs go a long way. Cell *136*, 586-591.
122. Visvader, J.E., and Lindeman, G.J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nature reviews Cancer *8*, 755-768.
123. Visvader, J.E., and Lindeman, G.J. (2012). Cancer stem cells: current status and evolving complexities. Cell stem cell *10*, 717-728.
124. Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes & development *28*, 1143-1158.
125. Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M.*, et al.* (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. Proceedings of the National Academy of Sciences of the United States of America *103*, 2257-2261.
126. Wang, D., Cai, C., Dong, X., Yu, Q.C., Zhang, X.O., Yang, L., and Zeng, Y.A. (2014). Identification of multipotent mammary stem cells by protein C receptor expression. Nature.

127. Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nature genetics *39*, 380-385.
128. Wellner, U., Schubert, J., Burk, U.C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., zur Hausen, A.*, et al.* (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. Nature cell biology *11*, 1487-1495.
129. Williams, S.A., Anderson, W.C., Santaguida, M.T., and Dylla, S.J. (2013). Patient-derived xenografts, the cancer stem cell paradigm, and cancer pathobiology in the 21st century. Laboratory investigation; a journal of technical methods and pathology *93*, 970-982.
130. Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol *11*, 228-234.
131. Yamaguchi, N., Ito, T., Azuma, S., Ito, E., Honma, R., Yanagisawa, Y., Nishikawa, A., Kawamura, M., Imai, J.-i., Watanabe, S.*, et al.* (2009). Constitutive activation of nuclear factor-κB is preferentially involved in the proliferation of basal-like subtype breast cancer cell lines. Cancer Science *100*, 1668-1674.
132. Yamamoto, M., Taguchi, Y., Ito-Kureha, T., Semba, K., Yamaguchi, N., and Inoue, J. (2013). NF-kappaB non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype. Nature communications *4*, 2299.
133. Yi, R., Poy, M.N., Stoffel, M., and Fuchs, E. (2008). A skin microRNA promotes differentiation by repressing 'stemness'. Nature *452*, 225-229.
134. Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J.*, et al.* (2007). let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell *131*, 1109-1123.
135. Yu, J., Li, A., Hong, S.M., Hruban, R.H., and Goggins, M. (2012). MicroRNA alterations of pancreatic intraepithelial neoplasias. Clinical cancer research : an official journal of the American Association for Cancer Research *18*, 981-992.
136. Zeng, Y.A., and Nusse, R. (2010). Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. Cell Stem Cell *6*, 568-577.
137. Zhang, W., Tan, W., Wu, X., Poustovoitov, M., Strasner, A., Li, W., Borcherding, N., Ghassemian, M., and Karin, M. (2013). A NIK-IKKalpha module expands ErbB2-induced tumor-initiating cells by stimulating nuclear export of p27/Kip1. Cancer cell *23*, 647-659.
138. Zhao, J., Dinesh S. Rao, Mark P. Boldinc, Konstantin D. Taganov, Ryan M. O'Connell, and David Baltimore (2011). NF-κB dysregulation in microRNA-146a–deficient micedrives the development of myeloid malignancies.