PhD degree in Molecular Medicine

European School of Molecular Medicine (SEMM)

University of Milan and University of Naples "Federico II"

Faculty of Medicine

A whole-genome approach to identify microRNA "modifiers" of breast cancer stem cell self-renewal

Maria Elena Bicchieri

IFOM-Firc Institute of Molecular Oncology, Milan

14-01-2015

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: Prof. Pier Giuseppe Pelicci, M.D. PhD.

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

List of Abbreviations

Nomenclature

ACD: Asymmetric Cell Division
BCSCs: Breast Cancer Stem Cell
CD29: Integrin β1
CD24: Cluster of Differentiation 24
CD44: Cluster of Differentiation 44
CD49f: Integrin subunit α6
CSC: Cancer Stem Cell
CTRL: Control
DMEM: Dulbecco's Modified Eagle Medium
E-Cad: E-cadherin
EGF: Epidermal Growth Factor
EMT: Epithelial to Mesenchimal Transition
ER: Estrogen Receptor
FACS: Fluorescence-activated cell sorting
FBS: Fetal Bovine Serum
FGF: Fibroblast Growth Factor
gDNA: Genomic DNA
GFP: Green Fluorescent Protein
GSEA: Gene-Set Enrichment Analysis
HER2: Human Epidermal growth factor Receptor 2
IEO: European Institute of Oncology
IN: Integration number
IPA: Ingenuity Pathway Analysis
K5: Cytokeratin 5
K8: Cytokeratin 8
Lin-: Lineage negative
MaSCs: Mammary Stem Cells
MEBM: Mammary Epithelial Basal Medium
miRNA: microRNA
MOI: Multiplicity of infection
MRE: microRNA responsive element
MRU: Mammary Repopulating Unit
OE: Over Expression
PBS: Phosphate buffered saline
PCA: Principal Component Analysis
PDX: Patient-Derived Xenografts
PgR: Progesterone Receptor
pre-miRNA: precursor microRNA
pri-miRNA: primary microRNA
RFP Red Fluorescent Protein
RISC: RNA-induced silencing complex
RNA-SEQ: RNA sequencing
SBI: System biosciences
SC: Stem Cell
Sca-1: Stem cell antigen
SCD: Symmetric Cell Division

SCR: Scrambled
TCGA: The Cancer Genome Atlas
TIC: Tumor Initiating Cell
TLDUs: Terminal Lobular-Ductal structures
UTR: Untraslated region
WT: Wild Type

ABSTRACT

A whole-genome approach to identify microRNA "modifiers" of breast cancer stem cell self-renewal

An emerging notion in the breast cancer field is that a rare subpopulation of cells within the tumor bearing stem cell (SC)-like properties, cancer stem cells (CSCs), are responsible for the degree of aggressiveness of the tumor, as well as the emergence of therapeutic resistance and disease relapse. Therefore, the development of novel therapeutic strategies that specifically target the CSCs population within a tumor could be the key to achieve an effective cure for breast cancer.

One strategy for targeting CSCs could be to inhibit their altered self-renewal mechanism and induce a quasi-normal differentiation process in tumor cells. However, the mechanisms that control the replicative mode of SC division and the degree of "stemness" of tumours are poorly characterized.

Recent research has highlighted the role of microRNAs (miRNAs), a class of small non-coding RNAs, as key regulators of gene expression in a variety of cellular processes, including SC self-renewal and differentiation. miRNAs negatively regulate gene expression at a post-transcriptional level and their expression is often deregulated in disease, making them ideal candidates as tumour biomarkers.

Despite recent studies uncovered new microRNA molecules linked to stem cell biology, we definitively miss a defined picture of which microRNAs are involved in the regulation of breast cancer stem cell self-renewal and their contribution to tumorigenesis.

The overall goal of this project was to identify key miRNA "modifiers" of breast cancer SC self-renewal that could either inhibit or enhance the self-renewal potential of cancer stem cells, with the purpose of identifying key molecules involved in the

acquisition/regulation of stem-cell traits and *bona fide* novel therapeutic targets.

We used a lentiviral microRNA library, composed of approximately 650 precursor microRNAs, to perform a functional whole-genome screening based on phenotypic competition assays on a very aggressive breast cancer cell line with stem-like properties (SUM159). Infected cells were challenged in an *in-vitro* 3D competition assay based on self-renewal ability of CSCs. In the competition assay, miRNAs that supported stem cells expansion were positively selected during passages, while microRNAs that inhibited self-renewal were depleted overtime. In parallel we also performed a 2D assay based on cell proliferation to gain insights into the ability of miRNAs to alter the proliferation of SUM159 cancer cells in adherent conditions. For each screening the positive and the negative selected miRNAs were identified by means of Next Generation Sequencing analysis.

The screening yielded to 20 candidate microRNAs selected as potential modifiers of self-renewal: 18 presumably acting by decreasing self-renewal (and hypothetically with tumor suppressing functions) and 2 by increasing self-renewal (and potentially with oncogenic functions). A proof-of-principle validation revealed that 6 out of 10 tested clones, confirmed their effects even when analyzed as single clones, underlining the potentiality of the whole-genome phenotype screening.

We focused our attention on the most promising candidates, namely miR-124 (which inhibited self-renewal) and miR-657 (which increased self-renewal) and, in order to search for the mechanism through which these microRNAs exert their function, we performed an RNA-seq analysis of transcriptional changes induced upon the overexpression of miR-124 and miR-657. We revealed that miR-124 in particular was able to regulate hundreds of genes and control independently different pathway related to self-renewal, migration and proliferation, suggesting that this miRNA could

effectively act at multiple levels to silence the self-renewal potential of cancer stem cells and, likely, inhibit the proliferation and migratory ability of the tumor, too.

As further experiments, we will definitively need to completely understand the role of miR-124 with the potential of being effective even on the most aggressive breast cancer disease.

1 INTRODUCTION

1.1 Normal mammary stem cells and breast cancer stem cells

1.1.1 Mammary gland development and mammary stem cells

The mammary gland is a glandular epithelium composed mainly of ducts that terminate in lobular structures, termed terminal lobular-ductal units (TLDUs). TLDUs are generally composed of two cell types: secretory or inner luminal cells and basal myoepithelial cells (Watson and Khaled, 2008).

As the names suggest, luminal cells are positioned around the central lumen, whereas basal myoepithelial cells are positioned in direct contact with the basement membrane (Figure 1). Immunohistochemically, these cells can be distinguished by the expression of different cytokeratin proteins: cytokeratins 8 and 18 (CK8 and CK18) identify luminal cells, cytokeratins 5 and 14 (CK5 and CK14) instead are typically expressed by basal-myoepithelial cells. Besides, the mammary epithelium is surrounded by a stroma, which is composed of adipose tissue, fibroblasts, blood vessels and nerve terminals. Stroma is also commonly referred to as the "mammary fat pat" (Figure 1).

The capability of the mammary gland to undergo different cycles of remodeling throughout the life of the organism renders this gland a unique organ. Indeed, it is possible to distinguish different stages in the maturation of the human mammary gland (schematized in Figure 2). The first stage is in the embryo, when the human fetal breast is formed from an epithelial structure called the "primordium", which, at 21 weeks, starts migrating towards the mesenchyme (Jolicoeur, 2005). In the neonatal stage, the terminal lobular units and the end buds appear and remain unchanged until puberty, when sex hormone signaling leads to duct elongation, side-branching and maturation of the breast. During pregnancy, other modifications occur. In particular,

luminal cells are stimulated to proliferate and differentiate into the milk-producing cells, called “acini” or “alveoli”. Simultaneously, basal myoepithelial cells, which surround the milk-producing cells, assist and favor the ejection of milk by contracting (lactation stage) (Gjorevski and Nelson, 2011).

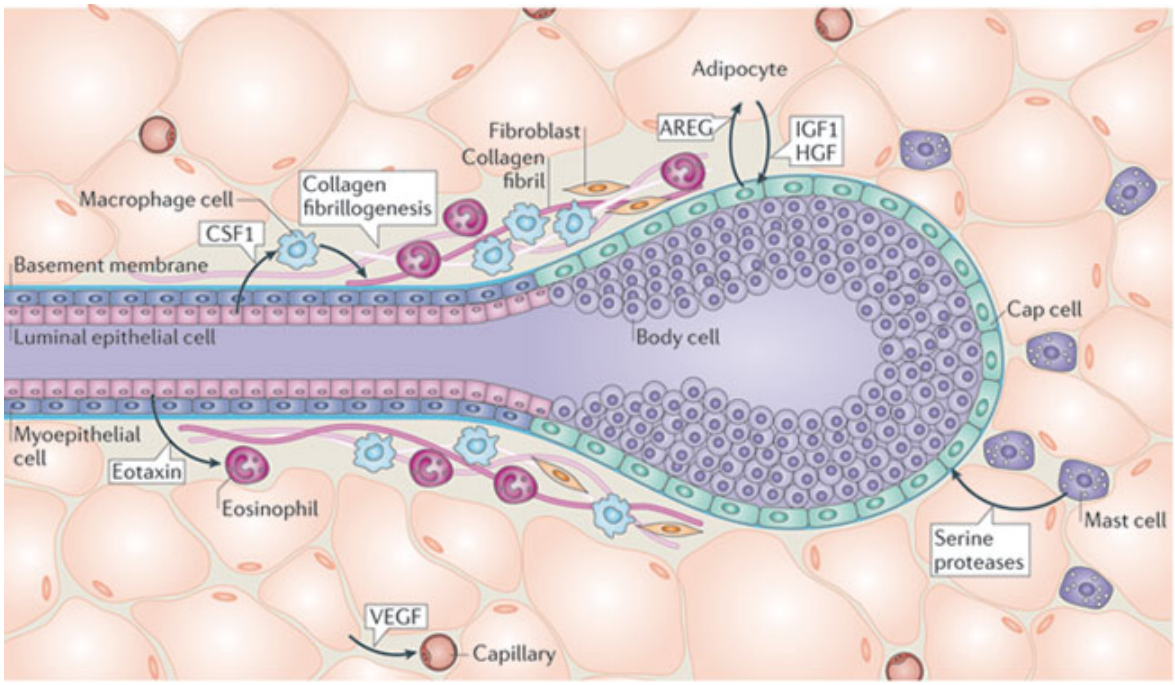


Figure 1 – Representation of the mammary gland and its components. The figure shows a terminal lobular-ductal unit composed of luminal epithelial cells and myoepithelial cells. Adipocytes, fibroblasts, and all the stroma components, which surround the duct, are also shown. Figure taken from (Gjorevski and Nelson, 2011).

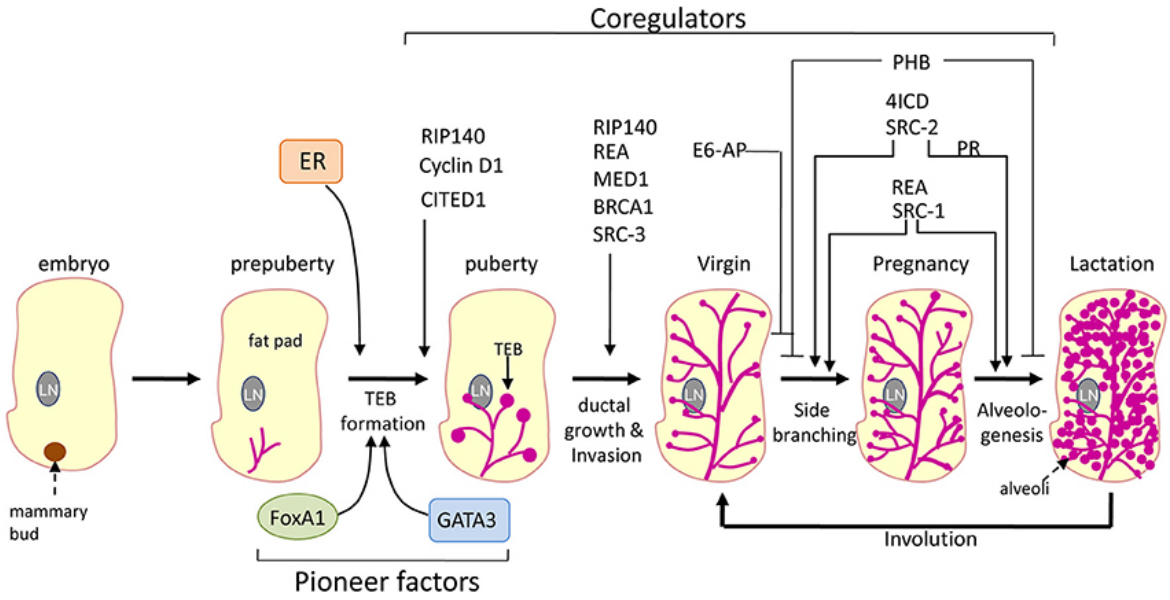


Figure 2 – Schematic representation of the phases of mammary gland development. 6 stage of mammary gland development are shown: embryonic, prepuberty, puberty, virgin, pregnancy, lactation. Critical molelcular regulators of the different stages are also indicated. Figure taken from (Manavathi et al., 2014).

Notably, at the end of lactation the mammary gland returns almost, but not completely, to its state before pregnancy (involution). Overall, the mammary gland is able to support multiple cycles of remodeling (pregnancy, lactation, involution) for several cycles. This observation suggests the existence of a pool of cells, termed "mammary stem cells" (MaSCs), which can i) generate all the cellular lineages of the mammary gland and ii) sustain the plasticity that is required whenever lactation is needed. The evidence that such a cell population is, indeed, likely to exist came definitively from Stingl's lab in 2006 (Stingl et al., 2006).

They exploited *in vivo* transplantation analysis and injected single rare epithelial FACS-sorted adult mouse mammary cells into cleared mammary fat pads of mice. Under these conditions, they were able to regenerate an entire mammary gland within 6 weeks upon injection, thus demonstrating that a multipotent population of rare cells,

MaSCs, exists (Stingl et al., 2006).

MaSCs share the typical characteristics of adult tissue stem cells, including the ability to self-renew and support multilineage differentiation (Van Keymeulen et al., 2011). Under physiological conditions, self-renewal is achieved by a unique pattern of asymmetric cell division (ACD): each ACD generates a daughter stem cell that retains stem identity and properties, and a daughter progenitor cell, which is committed to differentiation. This type of cell division preserves the stem cell pool over time, while allowing the generation of differentiating progenitors. Under particular conditions, for example, when tissue regeneration is required (such as upon injuries), a stem cell could divide symmetrically (SCD), generating two daughter cells with stem cell features. The daughter SCs would then undergo asymmetric cell division (ACD) to generate the differentiating progenitors and promote rapid regeneration of the tissue. Another stem cell peculiarity is its quiescent or slowly dividing state. Conversely, progenitor cells (such as those of the transit-amplifying compartment) sustain tissue proliferation during mammary gland formation or pregnancy. Lastly, stem cells are able to withstand anoikis, surviving in anchorage-independent growth. This characteristic allowed the development of a typical MaSC assay by Dontu and co-workers, called the "mammosphere assay" (Dontu et al., 2003). This assay consists of growing epithelial cells *in vitro* under anchorage-independent conditions (by the use of tissue culture plates that impair the adhesion of cells) in a growing medium, which is depleted of serum components to preserve the undifferentiated state of cells. Under these conditions, only MaSCs are able to survive and grow as clonal spheroids, called mammospheres, composed of quiescent stem cells and differentiating progenitor cells. Thus, the number of clonal spheroids formed is a proportional to the number of stem cells in the initial population. Mammospheres can be dissociated into single cell

suspensions, in which only stem cells can support a new round of mammosphere formation and create the second generation of spheres by self-renewing. Hence, serial propagation of mammospheres is considered as an *in vitro* assay to measure self-renewal of MaSCs (Dontu et al., 2003). Importantly, under mammosphere culture MaSCs retain also their ability to differentiate, and when grown under differentiating conditions can form both luminal-like and myoepithelial-like colonies (Dontu, 2003). Additionally, they can generate β-casein^{+} acini structures in the presence of prolactin (Dontu et al., 2003).

After the discovery of a rare population of adult stem cells residing within the mammary gland, scientists have exploited multiple approaches to specifically isolate these MaSCs. The typical approach relies on the use of cell surface markers that are associated with MaSCs. Stingl's lab, for example, demonstrated that MaSCs express low levels of Sca-1, medium levels of 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). These murine MaSCs are capable of serial transplantion *in vivo* and, thus, were defined as mammary repopulating units (MRUs). Similarly, Eirew's lab showed that MaSCs could be isolated as CD49f^{hi}/ Epcam$^{low/neg}$ basal cells (Eirew et al., 2008).

Our laboratory recently developed another approach for MaSC isolation from mammosphere cultures that does not rely on surface markers. This approach exploits the ability of MaSCs to slowly and asymmetrically divide in mammosphere cultures and, hence, asymmetrically dilute the lipophilic fluorescent dye known as PKH26. The dye is retained by the stem cells (which are PKH26 positive – PKH^{pos}), and is, instead, diluted in the progenitors and differentiated cells (PKH26 negative– PKH^{neg}). Using this approach, it is possible to isolate a quasi-pure population of MaSCs by

FACS-sorting the PKH^{pos} cells from dissociated mammospheres. Notably, the PKH^{pos} cells are fully functional, able to form mammospheres *in-vitro* and regenerate the entire mammary gland upon fat pad transplantation experiments *in-vivo* (Pece et al., 2010) (Cicalese et al., 2009).

1.1.2 Breast cancer and breast cancer stem cells

Human breast tumors are considered a heterogeneous disease, mainly classified histologically on the basis of their hormonal receptor status. In particular, pathologists look mainly at 1) the immunoreactivity for the estrogen receptor (ER) and the progesterone receptor (PR), 2) overexpression of the human epidermal growth factor receptor 2 (HER2), and 3) the proliferative index (Ki67), to stratify breast cancers. Tumors that express the ER are named luminal ER^{+} breast cancers and are associated with a better prognosis. In contrast, tumors that lack expression of the 3 receptors (i.e., ER^{-}, PR^{-}, $HER2^{-}$) are named "triple negative", and are associated with a more aggressive phenotype and a worse prognosis.

Recently, a new type of classification of breast cancers has been developed, the so-called "molecular classification", which takes into account the molecular characteristics of tumors to distinguish five major subtypes: Luminal A, luminal B, HER2 amplified, Basal and Claudin-low (Prat et al., 2010). Of note, the different molecular subtypes better define breast tumor heterogeneity, patient outcome and therapeutic responses (Hu et al., 2006) (Sabatier et al., 2014). The subtype classification can be derived using an entire transcriptome analysis or by a simple 50-gene set (PAM50), extensively validated on formalin-fixed, paraffin-embedded tissues and acting as independent predictor of survival in breast cancer (Guiu et al., 2012). The molecular classification suggests also that the different tumor subtypes

resemble different gene expression profiles of the breast hierarchy. For example, the transcriptome of Luminal A and B breast cancers is similar to that of mature luminal cells, whereas the expression profile of the basal-like subtype (and claudin-low subtype in particular) is similar to the profile of MaSCs (Liu et al., 2014) (Figure 3).

The parallelism between physiological development of a normal mammary gland and the aberrant development of breast cancer also extends to the apex of the cellular hierarchy, the stem cells. Hence, the development and maintenance of the breast cancer appears to be fueled by certain cells, known as tumor-initiating cells (TICs) or cancer stem cells (CSCs), which are able to form tumors when transplanted in mice (Visvader and Stingl, 2014). Breast cancer was one of the first malignancies for which a CSC hypothesis was postulated (breast cancer stem cells or BCSCs). Although this hypothesis struggled initially to gain acceptance, the experimental proof that CSCs reside within the tumor mass came from the seminal work by Al Hajj and colleagues, who isolated, using two surface markers CD44 and CD24, a population of cells enriched for SC traits from the tumor mass (Al-Hajj, 2003). CSCs defined as $Lin^-/CD44^+/CD24^-$ cells were able to generate tumors when transplanted in immunocompromised mice, even when as few as 100 cells were injected. Conversely, no tumor formation was detected when thousands of non-tumorigenic $CD44^-/CD24^+$ counterpart (NTG) cells were injected (Al-Hajj, 2003). Importantly, the same heterogeneity of the initial tumors could be reproduced upon TICs/CSCs transplantation, and the process could be repeated serially (serial transplantation), concluding that the $CD44^+/CD24^-$ population shares many features with normal SCs, including self-renewal and differentiation.

CSCs are now believed to be responsible for the most dangerous features of cancer, such as recurrence, therapeutic resistance and metastasis (Visvader and Lindeman,

2008). In particular, the self-renewal ability of CSCs appears to be more pronounced than that of normal MaSCs and aberrant outgrowth of cancer cells with SC traits is frequently observed (Liu and Wicha, 2010; Visvader and Lindeman, 2012) Recent observations suggest that BCSCs undergo SCD (rather than ACD) more frequently than their normal counterparts, thereby ensuring that the CSC pool never extinguishes and acquires unlimited self-renewal potential (Stingl, 2007) (Cicalese et al., 2009). Despite many efforts to reveal how MaSCs control their mode of division, the mechanisms that mediate the acquisition of stem cell traits, as well as the control ACD or SCD are mostly unknown.

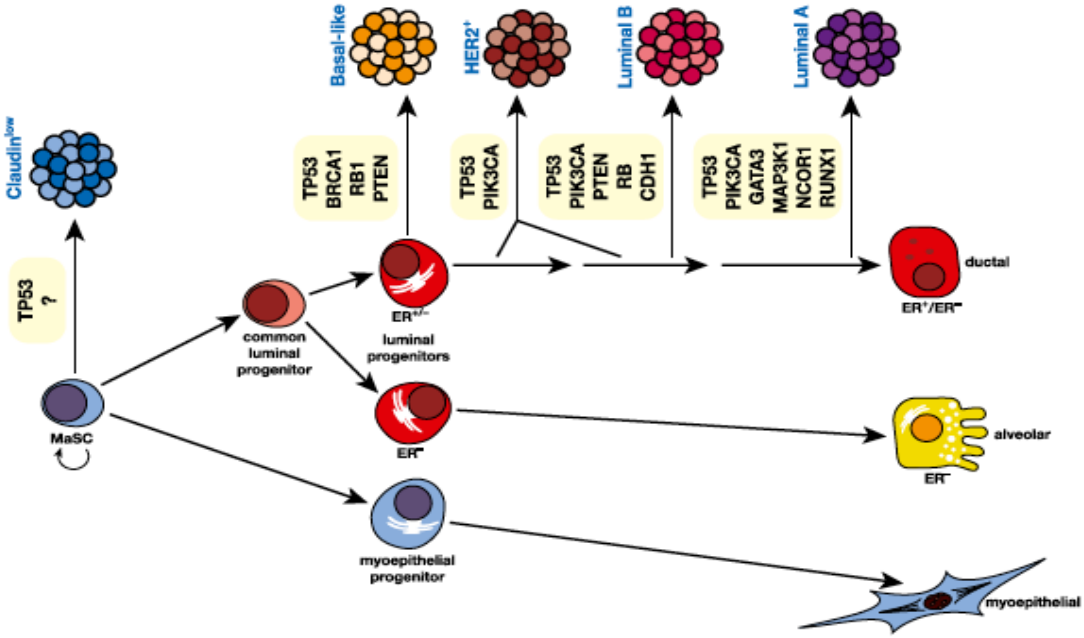


Figure 3 – Possible relationships between cell of origin and breast cancer subtypes. The figure shows a self-renewing mammary stem cell (MaSC) that gives rise to different progenitors (myoepithelial and luminal), which in turn differentiate into ductal, alveolar or myooepithelial cells. Different molecular subtypes of breast cancer are shown along the top, originating from different cells in the mammary gland cellular hierarchy. The critical molecular players involved in cancer development such as PTEN, TP53, BRCA1, RB, are indicated– Figure taken from (Visvader and Stingl, 2014).

1.2 microRNAs

microRNAs (miRNAs) have recently emerged as appealing molecules able to control several important cellular processes, including stem cell biology.

miRNAs are small RNA molecules, typically 20-22 nucleotides in length, which act by negatively regulating messenger RNAs (mRNAs). Their ability to affect gene expression is accomplished through base pairing of the miRNA to the 3' untranslated region (UTR) of a target mRNA (Bartel, 2009).

The discovery of the first miRNA molecule goes back to 1993, when Lee's laboratory identified *Lin-4* (Lee et al., 1993), a small RNA molecule fundamental in the control of the timing of *C. elegans* larval development. A few years later, in 2000, another small RNA molecule, named *let-7*, was discovered with a similar function in an advanced stage of *C. elegans* development (Reinhart et al., 2000). Of note, *let-7* was also found to exist in other species, including human and mouse, suggesting that miRNAs existed in higher eukaryotes too. Since then, a great number of miRNAs have been identified in different organisms, such us plants, animals, flies and viruses (Lagos-Quintana et al., 2001). A recent analysis in humans revealed that 2588 unique sequences of mature miRNAs have now been identified (according to last miRbase database, release 21) (Kozomara and Griffiths-Jones, 2014).

miRNAs are able to regulate many genes at the same time and, conversely, the expression of one gene might be controlled by different miRNAs. Bioinformatics analyses predict that miRNAs might control the activity of at least 30% of all the protein-coding genes. This feature, together with the fact that many miRNAs are conserved among different species, confers to these molecules the unique ability to be involved in the regulation of many cellular processes. Indeed, functional studies indicate that miRNAs participate in the control of almost all the physiological

processes within the cell, such as development, cell cycle, differentiation, and apoptosis (Filipowicz et al., 2008) (Bushati and Cohen, 2007) (Chekulaeva and Filipowicz, 2009). Moreover, a role for miRNAs as regulators of pathological processes has been pointed out, with particular emphasis on cancer disease, where miRNA deregulation is a frequent hallmark (Calin et al., 2004) (Ventura and Jacks, 2009) (Esquela-Kerscher and Slack, 2006).

With these observations in mind, it is not surprising that there has been great interest in the study of mechanisms controlling miRNA biogenesis and of their mechanisms of action.

1.2.1 miRNA nomenomenclature and occupancy in the genome

The nomenclature of miRNAs is based on their order of discovery, meaning that they are progressively numbered. A "letter suffix" is added when two or more miRNAs show a very high sequence homology. Conversely, a "number suffix" identifies miRNAs that have the same mature sequence, but derive from different transcripts. Lastly, the suffix 5p or 3p is added when mature miRNAs are derived from the 5'-3' or 3'-5' arm of the precursor miRNA, respectively.

miRNA genes occupy different genomic regions. About half of them are located within protein coding genes, such as in exons or introns, and are therefore known as "intragenic" miRNAs (Hinske et al., 2010). These intragenic miRNA genes usually have the same orientation as their "host" gene, signifying that they can be co-transcribed with the host. Alternatively, miRNA genes are located in regions of the genome quite distant from any protein-coding gene (gene deserts), thus, they are transcribed as independent transcriptional units by their own promoters. Such miRNAs are called "intergenic" miRNAs (Bartel, 2004) (Lee and Ambros, 2001)

(Garzon et al., 2009).

Sometimes, multiple miRNA genes are located in the genome in close proximity to each other (< 10 kb). These miRNA clusters are typically transcribed as a unique multi-cistronic primary transcript (Lagos-Quintana et al., 2001) (Lau, 2001). A well-known cluster-miRNA is the 17-92, which is composed of 6 different miRNA genes (miR-17, 18a, 19a, 20a, 19b-1 and 92a-1), which share a common promoter and processing mechanism (Olive et al., 2010).

1.2.2 miRNA biogenesis

The main steps of miRNA biogenesis are illustrated in Figure 4. miRNA biogenesis begins in the nucleus, where these molecules are transcribed by RNA polymerase II (Pol II) to generate a primary miRNA (pri-miRNA) with a characteristic stem loop structure (Lee et al., 2002). Pri-miRNAs can be quite long, from hundreds of nucleotides to tens of kb and are often capped and polyadenylated. In some cases, miRNA transcription can be mediated by RNA polymerase III (Pol III), in particular, for miRNAs located in genomic regions with a high content of Alu repeats (Borchert et al., 2006). For most intragenic miRNAs, transcription occurs by exploiting thei same regulatory machinery of the "host" gene and miRNA maturation is associated with mRNA splicing (Bartel, 2004).

Hereafter, the biogenesis of miRNAs requires two maturation steps. During the first step, the primary miRNA is cleaved in the nucleus to create a ≈ 70-100 nucleotide stem loop intermediate, known as the precursor miRNA (pre-miRNA). This process is executed by Drosha, a ribonuclease (RNase) III enzyme, that cuts both strands of the pri-miRNA, leaving a phosphate at its 5' end and an ≈ 2 nucleotide overhang at its 3' end (Basyuk et al., 2003). Drosha forms part of a multi-complex nuclear protein,

together with DGCR8 (DiGeorge syndrome critical region gene 8), known as the Microprocessor complex. DGCR8 protein has two ds-RNA-binding domains (ds-RBDs) and functions as a Drosha partner, since it guides the latter in the pri-miRNA cleavage by recognizing the ssRNA-dsRNA junction at the base of the hairpin (Han et al., 2006). Of note, some miRNA molecules, called "mirtrons", can bypass this cleavage step and originate the pre-miRNAs from intron splicing events (Filipowicz et al., 2008). The pre-miRNA is, then, exported from the nucleus to the cytoplasm by the Exportin-5 enzyme, which recognizes its peculiar arm (≈ 2-nucleotide 3' overhang), and together with RAN-GTPase directs the export event (Lund et al., 2004).

In the cytoplasm, the second maturation step takes place through another RNase III enzyme, similar to Drosha, known as Dicer. Dicer acts by cutting both strands of the duplex at about two helical turns away from the base of the stem loop, generating another 5' phosphate end and a 2-nucleotide 3' overhang, thus defining the second arm of the mature miRNA molecule (Bartel, 2004). Dicer functions with its partner molecule TRBP (TAR RNA Binding Protein), which binds to the N-terminal part of Dicer and facilitates the recruitment of one of the Argonaute proteins (Ago). In higher eukaryotes, four different isoforms of Ago exist (Ago1 – Ago4). Once the mature miRNA is loaded onto the Ago, the RNA-induced silencing complex (RISC) is formed, which executes silencing functions. The loading of the miRNA onto the Ago requires Ago-mediated unwinding of the ds mature miRNA and the selection of just one of the two strands, which is retained in the RISC complex. Ideally, both of the strands could be loaded into the complex, however, it has been observed that the strand that has the less stable 5' end is preferentially incorporated. This strand is called the "guide strand" as it guides the silencing function. Conversely, the discarded

strand, called “passenger strand”, is typically immediately degraded (Khvorova et al., 2003).

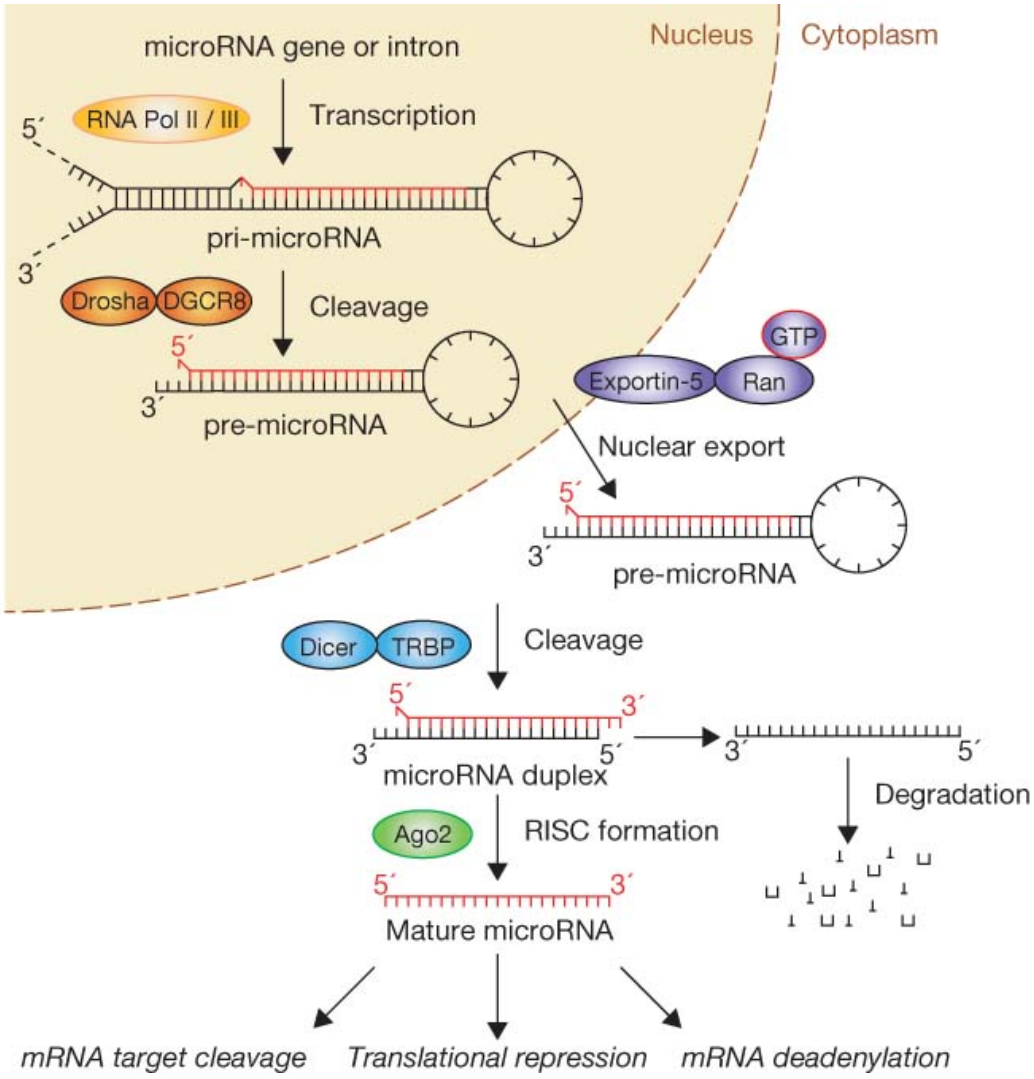


Figure 4 - miRNA biogenesis. The figure illustrates all the steps involved in miRNA biogenesis, including transcription, nuclear cleavage and export, cytoplasmic cleavage and the final loading of the mature miRNA onto the RISC complex [taken from (Winter et al., 2009)].

1.2.3 miRNA target recognition

The critical region involved in miRNA target recognition lies at the 5' end of the miRNA, positioned between the 2nd and 8th nucleotide and called the "seed sequence". This region is able to base pair, with a certain degree of complementarity, to the miRNA responsive element (MRE) usually located in the 3' untranslated region (UTR) of the target mRNA. The number of complementary bases between the miRNA and its target mRNA determines the strength of interaction and is used to classify the MREs as an 8-mer, 7-mer or 6-mer (Fang and Rajewsky, 2011) (Figure 5). In the case of a perfect match, the miRNA guides the RISC complex to cleave the mRNA through the AGO2 protein. The latter represents the only human AGO protein having an intrinsic and specific "slicing" activity (Bartel, 2004). However, in the majority of the cases a perfect complementarity does not occur, and target downregulation occurs through mRNA destabilization or translational repression by multiple mechanisms, briefly described here and summarized in Figure 6. For example, miRNAs may induce target destabilization by mRNA deadenylation, followed by de-capping and degradation. Alternatively, miRNAs may induce translational repression by interfering with protein synthesis at different levels. miRNAs can interfere with the initiation phase of translation by competing with eIF4E (eukaryotic translation initiation factor 4E), an initiation factor that mediates mRNA circularization for the binding of the polyA-tail to the cap-5' terminus of the mRNA. Alternatively, miRNAs may prevent the access or the formation of the 60S ribosomal subunit though a mechanism that is not yet completely understood. Translational repression can also occur after the initiation phase (mainly during elongation), either by decreasing the elongation rate or by inducing ribosome drop-off, which renders ribosomes prone to terminate their job prematurely. Finally, an extreme and not totally accepted

possibility, is that miRNAs may somehow impair the maturation of the nascent peptide by inducing proteolysis.

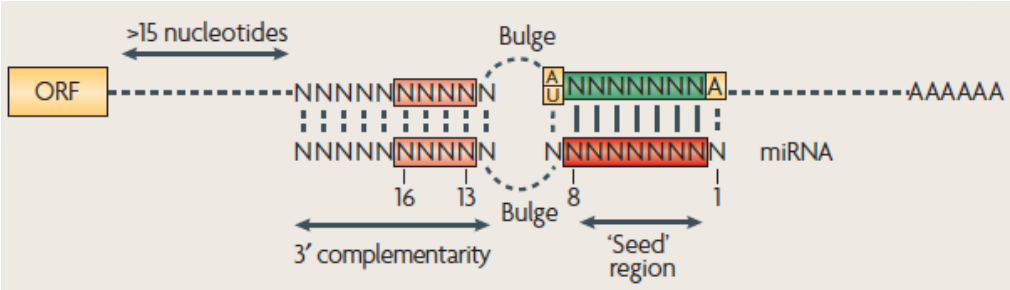


Figure 5 – Representation of the miRNA – mRNA binding site. The miRNA "seed sequence" and mRNA "MRE" are represented in red and green, respectevely. Different MREs are possible: i) a 6-mer site lacks the match in the 8th position (complementary at positions 2-7); ii) a 7-mer is complementary at positions 2-8, as represented in green; iii) an 8-mer site includes a further base pair in the 9th position plus an A residue in the first position. Generally, an A residue at position 1 of the mRNA and an A/U residue at position 9 (shown in yellow) improve the site efficiency even without base pairing. Good base pairing at positions 13-16 is important when the complementarity of the seed region is suboptimal [taken from (Filipowicz et al., 2008).

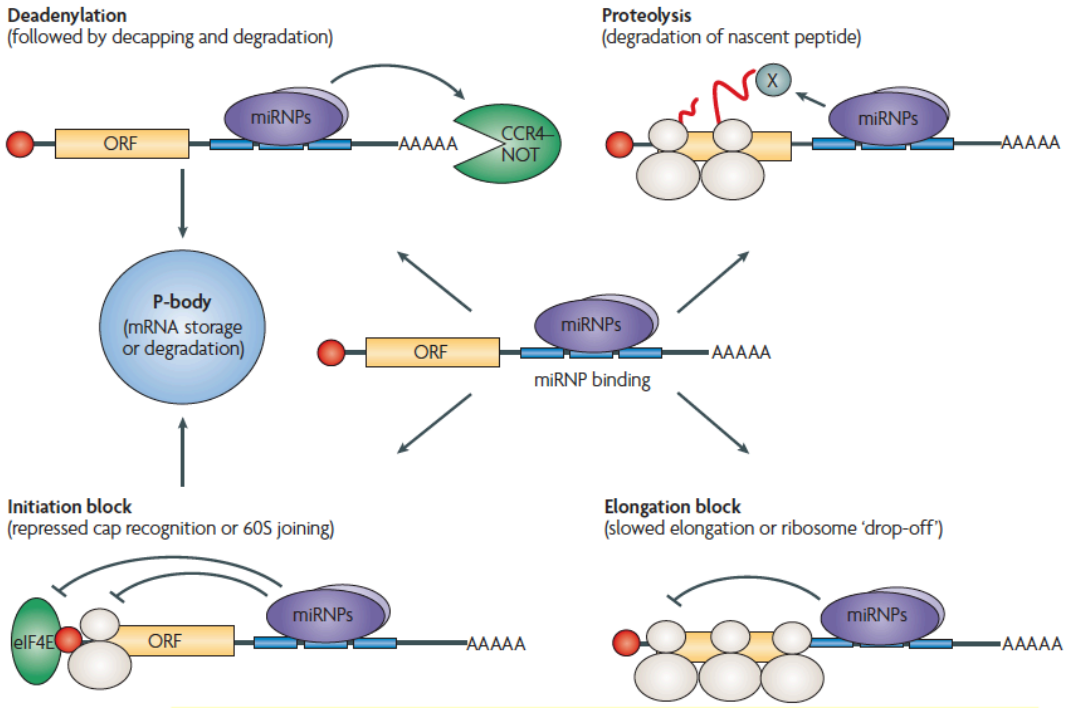


Figure 6 - Possible mechanisms of target repression. miRNAs repress mRNA targets by: 1) destabilizing the mRNA by polyadenylation; 2) inhibiting the initiation phase of translation; 3) inhibiting the elongation phase of translation; 4) possibly degrading nascent polypeptides miRNAs (Filipowicz et al., 2008).

1.2.4 miRNA: mode of action

The result of mRNA repression by miRNAs might be the reduction in the amount of the target protein to a level that is below that required for a protein to function (meaningless biological activity). This mode of regulation is called "binary off-switch" and is mainly observed for genes that should not be expressed in a particular cell type or stage (Bartel, 2009). In this case, the miRNA:mRNA interaction is fundamental in the regulation of a binary decision that is critical for cell fate establishment. One relevant example is the interaction between ZEB1 and miR-200s. ZEB1 is a transcriptional factor gene involved in the "epithelial-to-mesenchymal transition (EMT)" program, physiologically active during embryonic development

(Kalluri and Weinberg, 2009). It has been demonstrated that this gene is able to suppress the miRNAs of the miR-200 family by inhibiting the transcription of the two miR-200 genomic clusters (Bracken et al., 2008). In turn, miRNAs can extensively target ZEB1 to suppress its function (more than 6 different MREs are located in the 3' UTR of ZEB1 transcripts), thus generating a negative feedback loop (Park et al., 2008). The epithelial and mesenchymal states are regulated simply by the equilibrium between ZEB1 and miR-200s. The double negative feedback loop ensures that, once the fate is established, it is maintained over time unless a shift in the abundance of one of the two players (ZEB1 and miR-200s) is induced (Wellner et al., 2009).

Alternatively, miRNAs can function by refining protein levels to one that is optimal for the functional role exerted in the cells. In this case, miRNAs function by "tuning" one or multiple targets, in order to achieve smooth fluctuations in protein expression and ensure a more robust biological effect (Nazarov et al., 2013).

1.2.5 miRNA targets: computational predictions and tools to study miRNA-mRNA interactions

The need for scientists to know, with a certain degree of reliability, the identity of target genes regulated by the miRNA under investigation has become more and more imperative. Indeed, the discovery of target mRNAs can lead to the discovery of new pathways regulating critical biological processes under physiological and pathological conditions. Since miRNAs are capable of base pairing with target mRNAs displaying limited complementarity, they can potentially downregulate several mRNA targets, simultaneously. For this reason, many genome-wide approaches have been developed with the aim of identifying miRNA targets. These include: i) computational tools and ii) gene/protein expression profiling upon miRNA modulation. A schematic

representation of these methods is depicted in Figure 7.

Computational tools are based on prediction algorithms. Five major algorithms are available: miRanda, Targetscan, Pictar, DIANA-microT and RNAhybrid. All of them share common rules, such as the presence of seed-matched elements, the degree of MRE conservation among different species, and the stability of miRNA:target interaction (Maziere and Enright, 2007). The main two algorithms used are miRanda and Targetscan. 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 strengh of the miRNA:mRNA predicted interaction measured as a score (miRsvr score). The lower the score, the stronger the predicted interaction. Typically, miRanda yields more predicted targets than other algorithms, however, it has a high rate of false positives (30 – 40%). Targetscan predictions are instead more stringent since it 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 (Creighton et al., 2008).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

An alternative or complementary strategy to identifying miRNA targets is to perturb the expression of a particular miRNA and identify any genes or proteins that change expression as a consequence. A seminal study applying this strategy came from Lim and colleagues, who transiently transfected the muscle specific miR-1 and miR-124

into HeLa cells and exploited microarray analysis to determine transcriptional changes (Lee P. Lim, 2005). This studied revealed that miRNA effects could be scored by transcriptome analysis. Indeed, they found that ~100 mRNA were downregulated upon miR-1 and miR-124 overexpression and that the vast majority of these (>75%) contained an MRE in their 3' UTR corresponding to the overexpressed miRNAs (Lee P. Lim, 2005). An alternative strategy is to determine the consequences of knockdown or deletion of endogenous miRNAs; however, changes in mRNA levels after knockdown were not as pronounced as those after overexpression (Lee and Shin, 2012).

A more recently established approach to miRNA target identification involves the use of high-throughput RNA sequencing (RNA-seq) to detect transcriptional changes (Wang et al., 2009). Compared with microarray technology, RNA-seq possesses a higher sensitivity together with a broader dynamic range and does not require sequence information. With this approach, Burow *et al.* evaluated the global effects of a well-studied oncogenic miRNA, miR-155, on the transcriptome of an estrogen receptor-positive (ER^+), luminal, breast cancer cell line, MCF7. By performing RNA-seq analysis, the authors studied changes in the transcriptome mediated by miR-155 modulation in MCF7, and demonstrated that this miRNA is able to alter many pathways, with the MAPK signaling cascade being the main one (Martin et al., 2014).

Another strategy to decipher miRNA targets is to use proteomic approaches to observe protein changes in response to miRNA overexpression or downregulation (Li et al., 2012). In this case, the stable isotope labeling by amino acids in cell culture (SILAC) technique, or its variants, have been exploited to analyze protein abundance by mass spectrometry (Li et al., 2012).

Lastly, biochemical approaches to study miRNA-mRNA interactions have been used.

These methods rely on the physical interaction between the miRISC (miRNA loaded RISC complex) and the target mRNAs, and are considered as the most accurate way to define the miRNA targetome, although technically very challenging. Briefly, these methods require the immunopurification of the miRISC complex, through one of its subunits (such as Ago2), together with associated mRNAs (Ago-RNA immunoprecipitation or Ago-RIP). Ago-associated mRNAs are revealed through Affymetrix (Ago-RIP-Chip) or via deep sequencing (Ago-RIP-seq) (Karginov et al., 2007). Crosslinking treatments are usually employed to avoid contaminants and to precisely map MREs. Two main techiniques are used: High-throughput Sequencing of RNA isolated by crosslinking immunoprecipitation and Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (HITS-CLIP and PAR-CLIP).

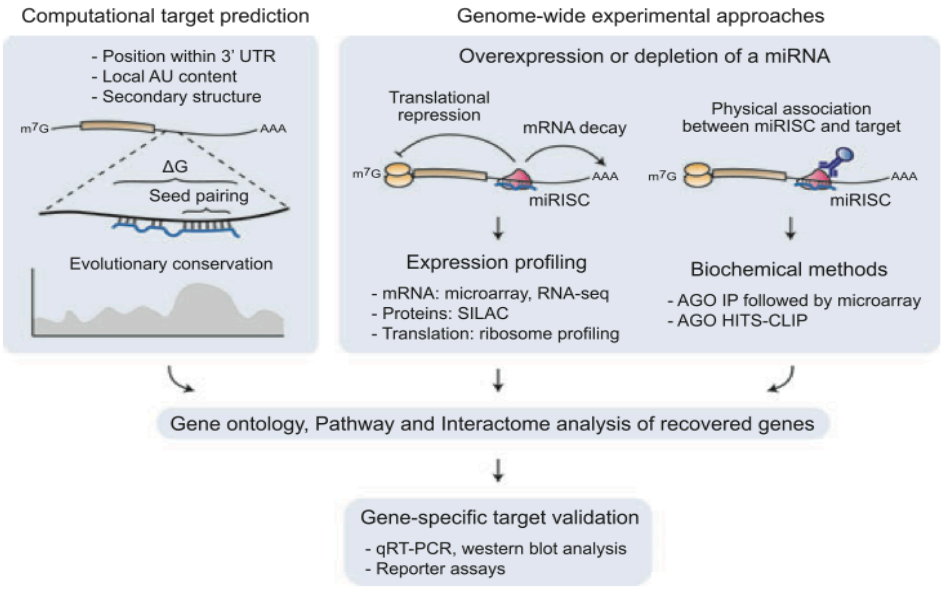


Figure 7 - Summary of genome-wide approaches for miRNA target identification. Computational methods and experimental approaches, including expression profiling and biochemical methods, have been applied to miRNA target identification. Figure taken from (Lee and Shin, 2012).

1.3 miRNAs and stemness

Given the central role of miRNAs in controlling gene expression programs that regulate cell identity and behavior, it is not surprising that miRNAs have been linked to stem cells (both embryonic and adult) and to the control of their peculiar characteristics, such as self-renewal and differentiation (Guo et al., 2011). Indeed, the stem cell and miRNA fields of research have converged following the identification of stem cell-specific miRNAs. These miRNAs target genes required for stem cell function, thereby, critically affecting stem cell fate and behavior.

1.3.1 miRNAs and embryonic stem cell differentiation

The first evidence that miRNAs can control mammalian stem cells came from experiments on embryonic stem cells (ESCs). These cells can be cultured *in vitro* as pluripotent stem cells and pushed to differentiate into different cell lineages by specific differentiation media. Analysis of miRNA expression profiles in undifferentiated ESCs revealed that these cells possess a unique pattern of miRNAs, with the miR-290 species alone comprising more than 60% of all miRNA molecules (Tiscornia and Izpisua Belmonte, 2010). By analyzing the expression profiles of mouse embryonic fibroblasts (MEFs) and neuronal precursors (NPC), it was immediately clear that the miRNAs of this family (plus those of the miR-302 family) are uniquely expressed in ESCs (ESC-specific miRNAs) (Marson et al., 2008).

Of note, genetic depletion of Dicer, one of the key enzymes involved in miRNA biogenesis, in mice caused severe impairment of expression of most, if not all, the miRNAs, defects in differentiation, and embryonic lethality (Bernstein et al., 2003). This strong phenotype was immediately associated with a possible role for miRNAs in promoting differentiation (Bernstein et al., 2003). Similarly, DGCR8-deficient

ESCs displayed impairment of the differentiation process (Wang et al., 2007). Both Dicer-null and DGCR8-deficient ESCs also exhibited a cell cycle delay together with a failure to shut down some of the transcription factors (OCT4, Nanog, Sox2) associated with self-renewal (Wu et al., 2014). These defects could be rescued by the expression of the ESC-specific miRNAs (miR-290s and miR-302s) suggesting that they are, indeed, necessary for self-renewal of this cell type (Wang et al., 2008). Furthermore, the expression of ESC-specific miRNAs enhanced the production of murine induced pluripotent stem (iPS) cells, indicating a role of these miRNAs in specifying ESC identity (Judson et al., 2009).

Conversely, other miRNAs, such as miR-145, let-7 and miR-21, are downregulated in self-renewing ESCs, but become upregulated during differentiation (Niu et al., 2011). These miRNAs act by silencing pluripotency genes, such as OCT4, Nanog, Sox2 and Klf4. Thus, they suppress the self-renewal program and promote the differentiation of ESCs towards a specific cell lineage (Xu et al., 2009) (Fre et al., 2005).

These findings unequivocally established a pivotal role of miRNAs in dictating the transition from an embryonic state to a more differentiated one by promoting cell cycle progression or silencing of self-renewal genes.

1.3.2 miRNAs and adult stem cell differentiation

In addition to studies in ESCs, the function of miRNAs has also been explored in adult stem cells by many research groups. The number of miRNAs that were found to participate in the control of adult stem cells is quite high. Here, I will briefly discuss some of these miRNAs to illustrate how they were identified and their modes of action.

One of the first adult stem cell models to be investigated was that of the

hematopoietic system, for which it is possible to isolate stem cells and progenitors, and to follow their differentiation processes. In this system, several miRNAs were found to be involved in the regulation of specific steps of hematopoiesis. For example, miR-181, miR-223 and miR-142 are all upregulated when bone marrow progenitor cells begin to differentiate towards the lymphoid or the myeloid lineages (Chen et al., 2004). Importantly, miR-155, expressed both in mature myeloid and lymphoid cells and upregulated in several types of lymphoma, is a required for the correct differentiation process and affects immune response *in vivo*. Indeed, transcriptome analysis in miR-155-deficient T lymphocytes revealed that cytokines and chemokines are among the key target genes regulated by this miRNA (Rodriguez et al., 2007).

In non-hematopoietic tissue, miR-203, one of the most abundant miRNA in mammalian skin, was among the first miRNAs to be shown to control cell fate decision. In a seminal study by Yi and co-workers, miR-203 was described to have a well-defined spatio-temporal expression regulation, through *in-situ* hybridization analysis of skin primary mouse keratinocytes. In particular, this miRNA was poorly expressed in epidermal stem cells and its expression started to increase as soon as these cells differentiated. Moreover, the authors showed that activation of miR-203 depletes the stem cell pool, while the inhibition of endogenous miR-203 resulted in the normal pattern of differentiation of epidermal stem cells. Mechanistically, miR-203 targets ΔNp63α, a transcription factor essential for maintenance of epidermal stem cells (Yi et al., 2008).

In another study by Green *et al.*, a role for miR-205 in the control of differentiation of mammary epithelial cell progenitors has been described (Greene et al., 2010). The authors initially isolated a subpopulation of cells, enriched in progenitors, from a

mouse mammary epithelial cell line (COMMA-DbetaGeo) by FACS-sorting for stem cell antigen 1 expression (Sca-1^{+}). The authors then compared the expression profiles of this Sca-1^{+} subpopulation with the Sca-1^{-} counterpart. By microarray analysis, they found that miR-205 was highly expressed in the Sca-1^{+} subpopulation. Moreover, overexpression of this miRNA *in-vitro (*COMMA-DbetaGeo) caused an expansion of the progenitor cell population and increased cell proliferation. They also identified the tumor suppressor gene PTEN as miR-205 target. Taken together, these results indicate that miR-205 is involved in the maintenance of self-renewal, at least in mammary epithelial stem cells (Greene et al., 2010).

1.3.3 miRNAs and self-renewal maintenance

A role for miRNAs in the maintenance of stem cell traits has also emerged. For example, it has been shown that the miR-290-295 cluster is ESC-specific and is directly involved in maintaining pluripotency; indeed, its levels decrease as the ESC differentiates. This miRNA family exerts its role by targeting the repressor of DNA methyl transferase (DMNTs), called RBL2, which in turn is no longer able to inactivate OCT4, maintaining active a critical effector of self-renewal (Houbaviy, 2003).

1.4 miRNAs and cancer biology

The role for miRNAs in controlling cell transformation and tumor progression has been also explored. Although this field has only recently emerged, several miRNAs have been implicated in human cancer and shown to influence key cancer-relevant pathways and cancer genes (Jansson and Lund, 2012). One of the initial observations made by Carlo Croce and co-workers was that miRNA loci frequently map to

genomic regions commonly amplified or deleted in various human cancers (fragile sites), suggesting a possible link between tumors and miRNAs (Calin et al., 2004). This hypothesis was corroborated when expression profile analysis revealed that tumors presented a widespread deregulation of miRNA expression (Volinia et al., 2006). Notably, miRNAs may not act only on master cancer regulatory genes, but may influence the properties of the network to which these gene belong, thereby stabilizing or de-stabilizing the whole network (Kwak et al., 2010).

microRNAs may act as either oncogenic or tumor suppressor molecules. Hence, in tumor formation we assist to a strongly reduction or even to a complete depletion of tumor suppressor microRNAs (Figure 8B) or to an amplification/overexpression of the oncogenic ones (Figure 8C).

A miRNA functioning as a tumor suppressor might inhibit the expression of an oncogene under normal conditions; when expression of this miRNA is lost, oncogene expression increases thus promoting tumor formation (Figure 8B). In contrast, a miRNA functioning as an oncogene might promote tumor formation by inhibiting the expression of a tumor suppressor gene (Figure 8C). For example, the miR-17-92 cluster acts primarily as an oncogene and cooperates with c-Myc to accelerate tumorigenesis in a mouse model of B cell lymphoma (He et al., 2005). Indeed, it maps to a chromosomal region that is frequently amplified in a subset of human B cell lymphomas (Ota et al., 2004). Similarly, miR-155 has been found to be upregulated in several hematopoietic malignancies and tumors of the breast, lung and pancreas (Kluiver et al., 2006). Experiments performed on genetically engineered mice with gain- and loss-of-function alleles of miR-155 demonstrated that this miRNA is able to generate tumors. Indeed, ectopic expression of miR-155 in B cells of transgenic mice is sufficient to induce proliferation of pre-B cells and consequently leukemia

(Costinean et al., 2006).

Conversely, miR-15 and miR-16 are located in a genomic region that is deleted in the majority of chronic lymphocyte leukemias (CLLs) and in prostate cancer, suggesting that these miRNAs may act as tumor suppressors (Calin et al., 2002). Importantly, a significant regression of prostate tumor xenografts was observed upon intratumoral delivery of miR-15a and miR-16-1 (Bonci et al., 2008). Another well-studied tumor suppressor miRNA is let-7. In humans, let-7 refers to a family of miRNAs composed of a dozen members organized in eight different genomic loci. Some of these loci are deleted in a variety of human cancers and overexpression of multiple let-7 miRNAs can suppress tumor development *in vivo*, in xenograft or mouse models of lung cancer (Esquela-Kerscher et al., 2008). Besides let-7, many other miRNAs were found to have a tumor suppressor function. Many miRNAs are associated with cell differentiation, which is by definition a tumor suppressive mechanism. Indeed, a global decrease in miRNA expression has been observed in tumors as a consequence of genetic (deletion or mutations in miRNA genes), epigenetic (methylation-induced silencing of miRNA promoters), or molecular events (regulation of transcription factors that control miRNA expression)(Chang et al., 2007) (Iorio et al., 2005) (Ozen et al., 2007). Accordingly, mutations or alterations of the genes involved in the miRNA biogenesis pathway (such as Dicer or Drosha) are associated with increased tumor malignancy and poor prognosis (Lin et al., 2010; Merritt et al., 2008). Indeed, Tyler Jack's lab demonstrated that silencing of three different components of the miRNA processing machinery (Dicer, Drosha and DGCR8) resulted in enhanced oncogene-induced transformation and accelerated tumor formation *in-vivo*. These findings suggest that global loss of miRNAs favors tumorigenesis (Kumar et al., 2007).

Finally, miR-22 has been identified as a potent oncogene in leukemia. Transgenic mice conditionally expressing miR-22 showed an increase in hematopoietic stem cell self-renewal and defects in differentiation. Moreover, miR-22 inhibition was able to block proliferation of leukemic cells. In the same study, they demonstrated that miR-22 exerts its role by targeting the tumor suppressor protein TET2 (Song et al., 2013).

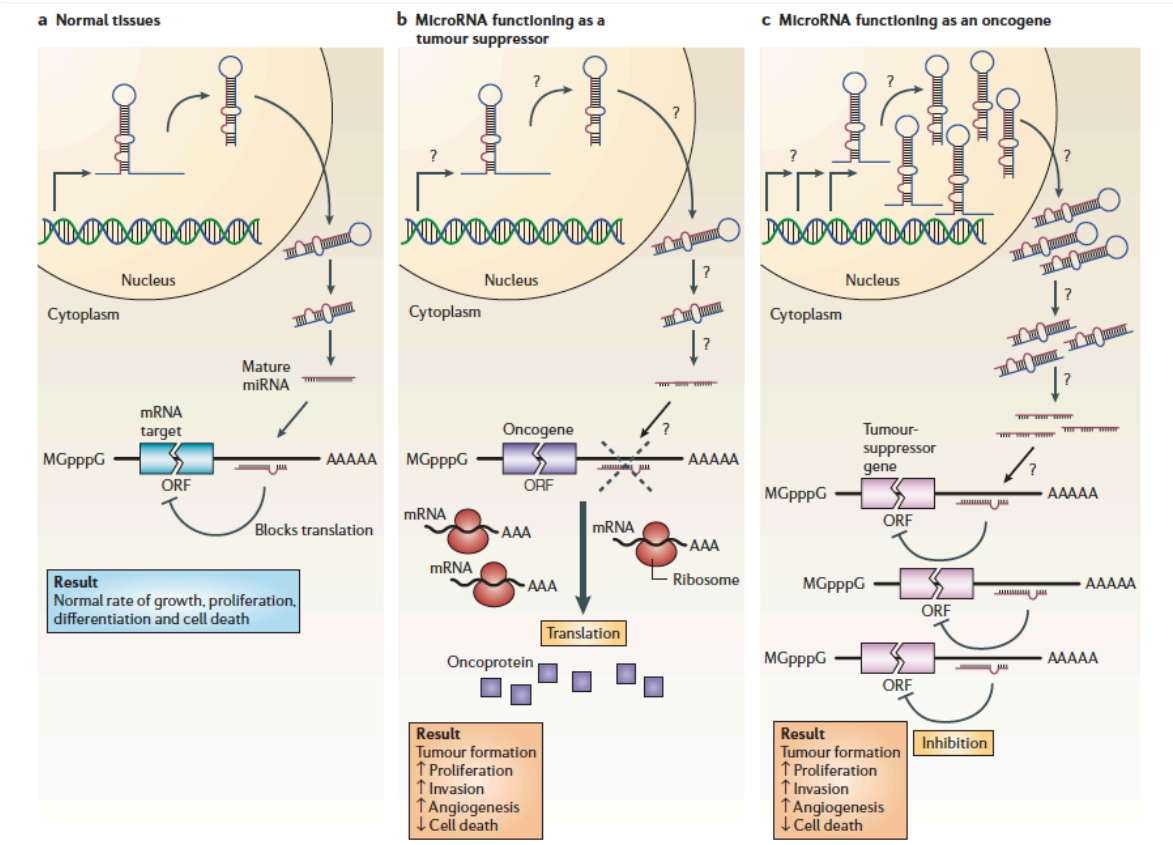


Figure 8 – Possible roles of miRNAs acting as tumor suppressor or oncogenic molecules. A) In normal tissues, miRNA molecules are regularly processed and cells proliferate and differentiate normally. B) miRNAs acting as tumor suppressor molecules are often lost in cancer and might promote the over-expression of oncogenes. C) miRNAs functioning as oncogenes are often overexpressed in cancer and might promote tumor formation by inhibiting the expression of a tumor suppressor gene. Figure taken from (Esquela-Kerscher and Slack, 2006).

1.4.1 miRNAs and breast cancer stem cells

In 2007, Yu *et al.* first linked the let-7 miRNA family to BCSCs. In their work, they compared the expression of selected miRNAs in $lin^-/CD44^+/CD24^-$ tumor initiating cells (TICs) to that in the $CD44^-/CD24^+$ non-tumorigenic counterpart (NTG) and found that levels of the let-7 family were strongly reduced in TICs (Yu et al., 2007). As chemotherapy increases the proportion of CSCs within breast tumors, they used as a model system a derivative of the SK-BR-3 breast cancer cells, called SK-3rd, generated by serial passaging in NOD/SCID mice treated with chemotherapy. They demonstrated that let-7 levels were decreased in SK-3rd cells, compared to the parental clone; however, upon SK-3rd cell differentiation *in vitro,* let-7 levels increased (Yu et al., 2007). Overexpressing let-7 in TICs decreased mammosphere-forming ability, and tumor formation and metastasis *in-vivo*. The authors dissected the mechanism of action of let-7 in SK-BR-3 and SK-3rd cells and found that this miRNA directly targets H-Ras and HMGA2, which control self-renewal and differentiation (Yu et al., 2007).

A few years later, Shimono *et al.* used a similar approach and compared the miRNA expression profiles of 460 miRNAs, by RT-qPCR, in human breast cancer cells (isolated by surface markers $lin^-/CD44^+/CD24^-$) and in their non-tumorigenic counterparts. Using this approach, the authors were able to derive a signature of 37 miRNAs differentially expressed in BCSCs compared to their non-tumorigenic counterparts (Shimono et al., 2009). Multiple members of miR-200 were downregulated in the stem cell compartment and were also found to be downregulated in MaSCs isolated from normal human and murine mammary gland tissue. High-resolution experiments on the miR-200 family, revealed that ectopic expression of miR-200c suppressed the ability of normal MaSCs to form mammary outgrowths *in-*

vivo and, more importantly, inhibited *in vivo* tumor formation by human BCSCs. At the molecular level, miR-200c inhibited repression of BMI, a member of the Polycomb repressing complex 1 (PRC1) and a regulator of self-renewal and differentiation of adult stem cells (Shimono et al., 2009).

In a recent study, a role for miR-205 in the regulation of MaSC fate and tumorigenesis was elucidated. In particular, the repression of miR-205 by the Notch ligand Jagged 1 (JAG1) in breast epithelial cells, promoted EMT and loss of epithelial polarization (Chao et al., 2014). These effects were accompanied by an expansion of the stem cell population by SCD and the formation of spontaneous mammary lesions *in vivo*. Mechanistically, miR-205 targets both ZEB1/2 (similarly to the miR-200s) and Notch2, generating another feedback mechanism involved in the control of epithelial *vs.* mesenchymal states (Chao et al., 2014).

Thus, mounting evidence indicates that miRNAs are key regulators of SC (including MASCs), self-renewal and that their deregulation can lead to aberrant regulation of the SC compartment and tumor formation.

1.5 Genome-wide screenings

Genetic screenings represent a reverse-genetic approach, in which multiple genomic elements (i.e. genes) are analyzed at the same time to search for those that are relevant to the phenotype of interest, by a totally unbiased approach. When applied at the genome-wide level, all (or most of) the genes in the genome are analyzed at the same time, to systematically explore their contribution to a certain biological process. The advantage of using such an approach is that it allows new functions to be associated with specific genes, and explores interactions between genes and their biological pathways. The main problem with genome-wide screenings is that they are extremely

technologically challenging due to the high number of genes being screened at the same time, and the necessity to keep confounding variables under control, which could influence the phenotype and thus alter the screening results.

Biologists generally exploit RNA interference (RNAi) methods to perturb the function of genes of interest in cells using short-interfering RNA (siRNA) molecules or short hairpin RNA (shRNA) molecules to exert the silencing (loss-of-function studies). Once the phenotype of interest has been observed, the sequence responsible for that phenotype is tracked back and, hence, genes capable of influencing the phenotype are identified.

Transfection using oligonucleotide siRNAs has greatly facilitated functional analysis of the human transcriptome, however one of the main downsides of such an approach is the need of delivering siRNAs into recipient cells by transfection, which makes it hardly feasible in non-dividing and/or primary cells. Moreover, siRNAs have transient effects and, thus, are not suitable to perform long-term assay or *in-vivo* experiments.

To overcome these limitations, shRNA vectors are frequently used. In this case a retroviral or lentiviral vector can be stably integrated into a target cell genome ensuing into the stable expression of a specific shRNA, resulting in the permanent reduction of the target gene product. Viral integration ensures a broader range of infectable target cells, including non-dividing and primary cells. However, to achieve robust biological effects, several conditions have to be satisfied, such as: i) efficient delivery of DNA-based plasmid into the host cell; ii) stable genomic integration; iii) transcription of the pri-shRNA hairpin; and iv)precise processing of pre-shRNA hairpin to yield a functional silencing duplex.

In large-scale applications, shRNA vector libraries directed toward entire genomes

could be used. In this case, silecing is accomplished by individual transfection of cells with a single shRNA from the library. This approach is very time-consuming, expensive and suffers of many sources of variability, for instance due to different reagent batches, , different confluence of plated cells. Alternatively, pooled shRNA library screenings could be exploited. Briefly, from hundreds to thousands of hairpins are combined (pooled) together and used to interrogate simultaneously a large population of cells, thus minimizing the experimental variability. In this case, cells are infected in a way that allows just a single shRNA expression vector to be integrated in the genome. Hence, different cells in the population have different silenced genes and could behave differently according to the effect of the silencing. Transduced cells could be screened for a phenotype of interest, to select those shRNAs that confer a specific phenotypic advantage/disadvantage. In particular, those shRNAs that inhibit the functions under the selection process are depleted from the initial pool (negative-selected/ drop-out screen), whereas constructs that enhance the functions will be enriched, as cells harboring these shRNAs will survive/expand in the selection process (positive-selected/ drop-in screen).

The main advantage of pooled shRNA libraries is that all the cells are cultivated under the same conditions, minimizing variability, and simplifying the experiments. Furthermore, the stronger is the phenotypic effect of the shRNA, the stronger and the quicker is the depletion/enrichment from the initial pool. However, clone devonvolution is not trivial and usually requires the exploitation of ad-hoc high-throughput approaches such as microarrays or next generation sequencing. One of the main drawbacks is, for instance, that RNAi-based-loss-of-function studies require that a strong shRNA expression should be maintained throughout the experiment. Indeed, if the shRNA expression is prevented because of unfavorable integration sites or other

inhibitory effects, than the effect of the shRNA is hidden. To overcome this problem, Scott Lowe developed a tool to track and isolate cells with strong target knockdown. The tool consists of an inducible shRNA expression system, the TRMPV vector, containing two fluorescent reporters: 1) the constitutive active yellow-green protein Venus and 2) the inducible dsRed fluorescent protein, whose expression is coupled to that of the shRNA of interest.

To test this vector, they transduced an established mouse lymphoma cell line (Eu-myc; Trp53 $^{-/-)}$ and performed a proof-of-principle experiment. By doxycycline treatment, they obtained a strong induction of dsRed expression and demonstrated that the double positive cells ($VENUS^{+}$, $dsRred^{+}$) provided a more sensitive assessment of deleterious shRNA effects than Venus alone. To this end, they propose the use of the TRMPV vector to facilitate the selection by FACS analysis of pure clonal populations capable of strong shRNA induction (Zuber et al., 2010).

Genome-wide shRNA libraries are extremely powerful and could be employed to get insights into many biological processes, including the study of genes related to cancer, stem cell biology or both (cancer stem cells).

For instance, in the work by Luo *et al,* an shRNA library was exploited to uncover putative therapeutic targets by identifying genes that were selectively required for the proliferation or survival of cancer cells with certain mutations. They developed a library containing 74,905 retroviral shRNAs targeting 32,293 unique human transcripts and applied it (6 pools of ~ 13,000 shRNAs per pool) to the colorectal cancer cell line DLD-1, carrying a K-Ras active point mutation that ensures a malignant phenotype (Luo et al., 2009). With such a loss-of-function experiment, they were able to detect shRNAs that displayed an anti-proliferative effect and were, thus, depleted from the population, by calculating the change in relative abundance of each

shRNA over time using microarray hybridization (Luo et al., 2009). Among the negatively selected genes, they discovered strong enrichment in genes regulating mitotic functions, meaning that the depletion of such genes impaired the viability of Ras mutant cells. They were able to describe a pathway involving the mitotic kinase PLK1, the anaphase-promoting complex/cyclosome (APC/C), and the proteasome that is essential for Ras mutant cells. Hence, they proposed to employ pharmacological inhibition of such pathway to treat cancers harboring Ras mutations (Luo et al., 2009).

To reveal new transcriptional regulators that govern the self-renewal capacity and regenerative potential of stem cells, Fuchs and co-workers performed an RNA-interference-based loss-of-function screening on hair follicle stem cells (HF-SCs) purified from their native niche in-vivo (Chen et al., 2012). Primary HF-SCs were transfected with a lentiviral pool encoding about 2000 candidate shRNAs (400 genes in total) that were found to be enriched in stem cells compared to progenitors. Cells were cultured for different passages and shRNAs were amplified and sequenced at each passage to reveal shRNAs depleted or enriched as a consequence of their ability to influence long-term proliferative potentials. They focused on the strongest shRNAs able to suppress long-term self-renewal and selected TBX1 gene for *in-vivo* analyses. TBX1-null stem cells, indeed, failed to replenish the niche during tissue regeneration, thus uncovering a new key regulator of skin stemness (Chen et al., 2012).

Similarly, in the work of Ali et al., a pooled lentiviral shRNA library was used to infect primary human hematopoietic stem cells (HSCs) searching for modifiers of HSC self-renewal (Ali et al., 2009). HSCs isolated from umbilical cord blood were transfected with a pooled library targeting 1300 human genes and passaged in liquid cultures for 10 weeks followed by colony-forming-cell assays to select for clones that

acquired enhanced self-renewal potential. DNA was extracted from each colony and sequenced. Three hit candidates were pointed out and further validated, demonstrating the potential of RNAi screening on human primary stem cell populations (Ali et al., 2009).

In the elegant work by Gupta and co-workers, an RNAi screen was performed to uncover agents that specifically kill breast epithelial cancer stem cells (CSCS) (Gupta et al., 2009). About 16000 compounds were screened for their effects on a normal breast cell line that was previously induced to an epithelial-to-menchymal transition in order to favor the acquisition of stem cell traits ($HMLE^{shEcad}$). Cells were individually tested for viability 3 days upon treatment with chemicals by a luminescence assay. 32 compounds were found to specifically reduce the viability of $HMLE^{shEcad}$ as compared to control HMLE. Eight candidates were chosen and four of them were individually validated. Of note, the strongest compound, Salinomycin, was also tested on the breast cancer cell line HMLER and found to inhibit cancer stem cells, as i) it specifically decreased the cancer stem cell content by FACS ($CD44^{high}$ $CD24^{low}$); ii) it reduced the number of tumorspheres by 10-fold; 3) it decreased the tumor-seeding efficiency by 100 fold and tumor volume and metastasis *in-vivo* (Gupta et al., 2009).

1.5.1 Functional microRNA screenings

A new emerging challenge is elucidating the biological processes in which miRNAs play a role. As for protein coding genes, the function of a miRNA might be also revealed by knocking it down or out from the genome of interest. However, this method requires prior knowledge where a miRNA is expressed and sufficient knockdown to demonstrate an observable effect is not always guaranteed. Adding to

this challenge is the possibility that many miRNAs may elicit only subtle changes or are redundant with other family members.

A more convenient approach, which overcomes these problems, is to overexpress a given miRNA in the system of interest. As miRNAs closely resemble siRNAs/shRNAs, it is possible to exploit the same kind of approaches, such as i) siRNA oligos with the sequence of miRNAs (miRNA mimics); ii) viral vectors expressing miRNA precurosors.

MicroRNA mimics are chemically synthesized miRNA duplexes, which express miRNAs in their mature form, so that the canonical miRNA biogenesis pathway is not required. Bypassing this physiological processing step has several implications. For instance, it is possible that transfection of miRNA mimic achieves cellular concentrations beyond physiological relevance. Conversely, the lenti- or retro-viruses are integrated in to the host genome and swiftly and stably express miRNA genes, allowing long term assays or the use of primary cells hard to transfect, such as adult stem cells.

To identify specific miRNAs that promote the G1/S transition in embryonic stem (ES) cells, a screening using 266 mouse microRNA mimics individually transfected into the Dgcr8 knockout embryonic stem cells was performed (Wang et al., 2008). As Dgcr8-mutant ES cells lack endogenous miRNAs and display a proliferation defect, transfected cells were evaluated for changes in their rate of cell proliferation. MiRNAs that rescued the proliferation defect of Dgcr8-/- cells were identified, including members of the miR-290 cluster (miR-291a-3p, miR-291b-3p, miR-294, and mir-295) and the miR-302a-d. Authors revealed that these miRNAs control the expression of key regulators of the G1/S transition such as *Cdkn1a* (Wang et al., 2008) and could promote the induction of pluripotency (iPS) when introduced into

mouse embryonic fibroblasts along with reprogramming transcriptional factors (*Oct4*, *Sox2* and *Klf4* (Judson et al., 2009).

By combining miRNA expression pattern and phenotype screening, Yu and collaborators identified key miRNAs that promote ES cells differentiation (Ma et al., 2014). In particular, they focused on 40 miRNA individually transfected as mimics in Dgcr8 knockout embryonic stem cells to search for those responsible of ES differentiation. They pointed out miR-27a and miR-24, able to directly target pluripotency factors and signal transducers of ESC self-renewal networks such us Foxo1, Oct4 and Smad4 (Ma et al., 2014).

By performing a high-throughput functional screening, the group of Giacca uncovered a set of microRNAs that stimulates cardiomyocytes (CM) proliferation and promotes cardiac repair (Eulalio et al., 2012). A library of 875 miRNA mimics was used to infect neonatal rat cardiomyocytes and two microRNAs in particular (miR-590 and miR-199a) significantly induced proliferation of CM. By injecting these synthetic microRNAs *in-vivo* into the heart of neonatal rats, they further prove that these molecules not only are able to control CM proliferation but also they boost the normally ineffective myocardial repair that takes place after myocardial infarction (Eulalio et al., 2012). Notably, the findings that exogenously administered miRNAs have the potential to restore cardiac function to almost normal levels after myocardial infarction have evident translational value.

1.5.2 Functional screening assays to reveal cancer-associated miRNAs

Given the great biological impact of miRNAs on cancer and considering that the number of miRNAs is considerably lower than the total human genes, it is not surprising that a lot of effort is now being directed to finding new strategies to study

miRNAs involved in cancer. Currently, two main approaches are used: expression profiling studies and functional screenings. The first approach is mainly based on the frequently observed differences in miRNA expression profiles between tumoral and not-tumoral tissues, due to aberrant expression of miRNA genes in cancer. This approach is a good strategy for identifying novel miRNA tumor biomarkers; however, it does not usually allow the detection of "driver" tumor miRNAs. In contrast, functional screenings allow the identification of miRNAs with cancer-associated phenotypes, without considering their level of expression. This type of screening involves two steps: 1) a systematic introduction of miRNAs into cells; 2) detection of miRNAs with a cancer-associated phenotype by microarray analysis or next generation sequencing.

Recently, vector-based (either retroviral or lentiviral) miRNA libraries have been exploited for this purpose. Such pooled libraries are used to stably express precursor miRNAs that are challenged in phenotype competition assays. As example, in their work Agami and co-workers (Voorhoeve et al., 2006) introduced a homemade miRNA pooled library of 197 miRNAs into immortalized human foreskin fibroblasts (BJ|ET) and subsequently transduced the cells with $RAS_{V12.}$ The difference in abundance of all miR-vectors was then detected by a homemade miR-microarray after 3 weeks by growing cells in 2D adherent conditions. They found that miR-372 and miR-373 were positively selected in three independent experiments, capable of conferring growing advantage. They further demonstrated that cells overexpressing this cluster were capable of growth in soft agar conditions, and demonstrated that miR-372 and miR-373 collaborate with RAS in the transformation process in a manner that resembles p53 inactivation. Finally, they found that these miRNAs were overexpressed in 4 out of 7 testicular germ cell tumors tested, all of which were WT

for p53 and presented chemotherapy resistance (Voorhoeve et al., 2006). This study is an example of a positive screening assay in which a cell population with a phenotype of interest increases during the selection process.

To examine the potential of miRNAs for treatment of head and neck squamous cell carcinomas (HNSCC), a retroviral human miRNA expression library containing 370 miRNAs was used to infect the cancer cell line population VU-SCC-120 and the normal oropharyngeal keratinocytes ciOKCs as control (Lindenbergh-van der Plas et al., 2013). The majority of the clones were not able to influence the survival of either the tumor cell or the control, or the lethal effect was similar in both models. Conversely, 19 microRNA candidates specifically affected the head and cancer cell line VU-SCC-120. By individual transfection, they validated 6 out of 19 microRNA candidates, as they were able to decrease proliferation specifically in HNSCC cells. Moreover, they demonstrated that 3 of these microRNAs target the nuclear protein kinase ATM, that activates DNA damage through several downstream pathways (Lindenbergh-van der Plas et al., 2013).

In the work of Nakagama *et al.* (Izumiya et al., 2010), authors performed a high-throughput functional screening using a pooled miRNA library containing about 450 miRNA clones. The pancreatic cancer cell line MIA PaCa-2 was infected with this library to detect miRNAs able to suppress cell proliferation in adherent 2D condition. They deconvoluted miRNA clones by the use of a custom-made microarray and determined changes in the relative abundance of a miRNA, by comparing its depletion or increase over time relative to the initial population. They thus identified 5 negatively-selected miRNAs, putatively acting as anti-proliferative miRNAs. By single validation, they showed that 4 out of 5 miRNAs were able to cause a cell cycle block, confirming the feasibility of the drop-out screening in which a microRNA is

negatively-selected and thus decreases from the initial population (drop-out).

The analogous group then applied the same library to detect tumor suppressor miRNAs in colon carcinogenesis (Tsuchiya et al., 2011). They infected the HCT116 cell line and identified 24 putative anti-proliferative miRNAs by microarray analysis. To narrow the selection, they integrated the functional analysis to expression profiling and identified miR-22 as miRNA candidate for tumor suppressor. They then moved on to study the mechanism of action of miR-22 and identified p21 as its direct target. In turn, miR-22 is a transcriptional target of p53. Briefly, they propose the p53–miR-22–p21 axis is a crucial regulatory component involved in the determination of p53-dependent apoptosis.

Overall, genome-wide screening approaches for identifying miRNAs involved in particular phenotypes appear to be a successful strategy.

2 MATERIAL AND METHODS

2.1 Cellular biology

2.1.1 Cell lines

MCF10A (human normal mammary gland cell line) cell line was provided by ATCC and cultured in DMEM/F12 (1:1) + Glutamax, 5% horse serum, EGF (20 ng/mL), Cholera Toxin (100 ng/ml), Human insulin (10 µg/ ml), Hydrocortisone (500 ng/ml), Pen/strep 1%. [CO_2] = 5%

MCF7 (luminal human primary breast cancer cell lines) cell line was provided by ATCC and was cultured in DMEM supplemented with 10% fetal bovine North American serum (Hyclone, TermoScientific), 2mM Glutammine under an atmosphere of 5% CO_2 at 37°C.

SUM159PT (triple negative human primary breast cancer cell lines) cells were commercially available (Asterand, Detroit, MI) and were cultured in Ham's F12 medium (Invitrogen Life Technologies) supplemented with 5% fetal bovine North American serum (Hyclone, TermoScientific), 2mM Glutammine, 5ug/mL insulin, 1 µg/mL hydrocortisone and 10mM HEPES and were cultured under an atmosphere of 10% CO_2 at 37°C.

2.1.2 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 hs 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.

2.1.3 microRNA lentiviral vectors

MicroRNA overexpression and non-targeting controls (SCR and PCDH) was obtained by lentiviral infection with commercially available vectors from System Biosciences (SBI). List of clones: Scramble control CD511B-1; PCDH control CD511B-1 pCDH-CMV-MCS-EF1; Lenti-miR-34a MI0000268; Lenti-miR-191 MI0000465; Lenti-miR-483 MI0002467; Lenti-miR-96 + miR-183 MI0000273, MI0000098; Lenti-miR-200c MI0000650; Lenti-miR-146a MI0000477; Lenti-miR-342 MI0000805; Lenti-miR-24-1 MI0000080; let-7a-2 MI0000061; Lenti-miR-125a MI0000469; Lenti-miR-124a-1 MI0000443; Lenti-miR-1-1, MI0000651; Lenti-miR-153-2 MI0000464; Lenti-miR-379 MI0000787; Lenti-miR-204 MI0000284; Lenti-miR-889 MI0005540; Lenti-miR-664 MI0006442; Lenti-miR-657 MI0003681; Lenti-miR-144+miR-451a+miR-4732 MI0001729, MI0017369, MI0000460.

microRNA lentiviral vectors enabled to set up stable and heritable microRNA overexpression. microRNAs were all expressed as precursor transcripts (pre-miRNAs) in their native sequence context to ensure a natural interaction with endogenous processing machinery and leading to authentic mature microRNAs.

Transduced cells were monitored thanks to the co-expression of copGFP fluorescent marker by FACS analysis.

2.1.4 Colony assay

5000 transduced MCF10A cells, (overexpressing one of the following constructs: PCDH, miR-34, miR-483, miR-200c, miR-191) were plated in 2D adherent conditions onto three different cell culture dishes and incubated for 10 days before visualizing colonies by crystal violet staining. Crystal violet staining solution was prepared by dissolving 1g of crystal violet powder in 100 mL of 20% EtOH and filtered at 0.45 micron. For the staining procedure, plates were initially washed with 1X DPBS and after, the staining crystal violet solution was gently and homogeneously distributed with a sufficient volume to cover the surface along the plates. The staining solution was left in contact with cells for few minutes and then removed. Plates were finally washed with water at least twice and were left to dry inverted before visualizing colonies.

2.1.5 Calculation of integration number (IN)

We needed to control the number of integration upon infection, thus we introduced two parameters:

- Molteplicity of infection (MOI)

$$\frac{\text{N° of virus particles}}{\text{N° of total cells}}$$

- Integration number (IN)

$$\frac{\text{N° of integrated particles}}{\text{genome}}$$

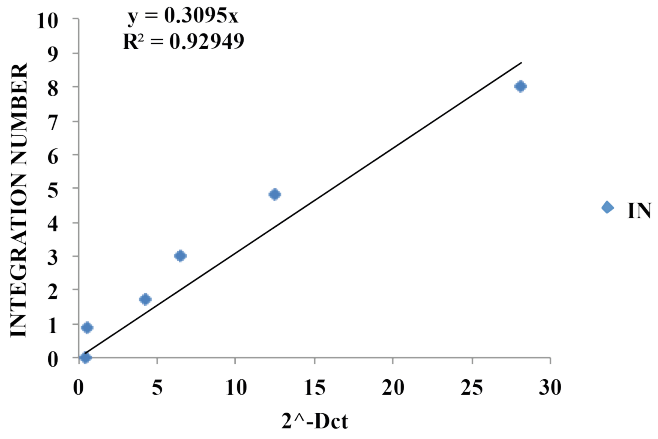
The MOI can be calculated prior to the infection, since it depends on the virus titer and on the number of cells to be infected. On the contrary, IN must be determined upon infection.

To measure the IN, a calibration curve was generated by means of q-PCR using the Global UltraRapid Lentiviral Titer Kit containing six DNA samples with known integration numbers (INs) as it is here described:

The q-PCR was performed using two primers for each standard sample 1) the Upstream conserved region 1 (UCR1), a common region of the human genome, and 2) the woodchuck hepatitis post-transcriptional regulatory element (WPRE), a specific region for the integrated viral cassette.

UCR1 Forward: CAGGCTTTGCTGAGCAGTTT
UCR1 Reverse: CCTCTCCCAGTCAGGGATTT
WPRE Forward: TCCTTCTGCTACGTCCCTTC
WPRE Reverse: GAGATCCGACTCGTCTGAGG

Since IN is proportional to $2^{-[Ct\ (WPRE)\ -\ Ct\ (UCR1)]}$, we obtained the liner equation y=0.3095X,



where X is $2^{-[Ct\ (WPRE)\ -\ Ct\ (UCR1)]}$ and can be calculated experimentally for any new condition, 0.3095 is the slope value and Y is the IN. We used this linear equation to determine the number of integration of any unknown sample infected at different

MOI. In particular low MOI (1-2) corresponded to IN=1 and higher MOI (>3) corresponded to multiple number of integrations. To keep IN number equal to one, we used a low MOI in every experiments discussed throughout this text.

Moreover we determined a correlation between the MOI and the GFP% upon infection, so that the IN could be easily deduced from the GFP content evaluated by FACS analysis.

MOI	% GFP
10	70
5	38
2	14
1	11

2.1.6 Library infection

Human miRNAs lentiviral library was provided by System Bioscience. SUM 159 cells (650000) were seeded in a 15 ∅ plate one day prior to infection. We infected at low multiplicity of infection (MOI = 2) with a viral titer of 7,2 x 10^8 (IFU/mL). At 72h post transduction, infection efficiency was checked with fluorescence microscopy. GFP expressing cells were selected by flow cytometry using the Influx cell sorter (BD with a 488 nm laser).

Sorting gates were maintained constant for all the experiments throughout this thesis: only cells with a bright GFP intensity (>10^2) were isolated. The gate for negative cells was selected according to the basal fluorescence of not infected cells.

Selected cells (GFP^+) were challenged in the 3D and 2D competition assays. The same protocol with the same library stock was used to perform the biological replica experiment.

2.1.7 *In-vitro* 3D-assay – Mammosphere assay

SUM159 cells transduced with the lentiviral library (GFP^+ cells) were cultured in non-adherent conditions using free serum MEMB (Lonza) basal medium supplemented with glutamine (1%), penicillin-streptomycin antibiotics (1%), insulin (5 ug/mL), hydrocortisone (0,5 ug/mL) and (eparine 1 U/mL) with fresh growing factors EGF (40 ng/mL), FGF (40 ng/mL) and B27 (Gibco, 2X). A total of 3 x 10^5 GFP^+ sorted cells were plated for the competition assay in non-adherent condition at 2000 cells/mL final concentration (15 ∅ cm plate) in stem medium + methylcellulose (ratio 1:1). 10 days after, mammospheres were nicely formed and dissociated (both enzymatically and mechanically using trypsin-EDTA diluted 1:10 in to PBS, 50mg/L trypsin final concentration). Single cells were then re-seed in non-adherent conditions for a new round of the assay. At each passage ≈ 3 x 10^5 cells were kept for gDNA extraction. This procedure was repeated for six passages, keeping GFP^+ cells by FACS-sorting every two passages. Cells were cultured for a total of 6 passages (F1-F6).

In parallel the sphere forming efficiency was checked plating cells at 1000 cells/mL in a 12-well plus methylcellulose (2% w/v, Sigma Aldrich) in triplicates (diluted 1:1 with stem medium with 2x of growth factors).

Each generation of mammosphere, that expressed the miRNA 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, we collected the images of each generation of all the12-well plate and the macro counted only the rounded (≥ 0.8) fluorescent spheres with a diameter > 70 µm.

The sphere forming efficiency was calculated using the formula:

$$\%SFE = [n° \text{ of spheres}/n° \text{ of plated cells}] X100)$$

All the data (as number of spheres, diameter in micron, circularity and mean dimension area) were collected in an excel file for further analysis.

2.1.8 *In-vitro* 2D-assay – Proliferation assay

SUM159 cells transduced with the lentiviral library (P0 = 90000 GFP^+ cells) were cultured onto 2D adherent condition as growing in cell culture plates (at least 250k cells). Cells were regularly split every two days in growing condition. Each week, GFP^+ cells were selected by sorting to avoid transgene silencing and keeping ≈ 3 x 10^5 cells for gDNA isolation. Cells were cultured for a total of 6 weeks (P1-P6).

2.1.9 Validation Step

SUM159 were individually infected using 10 microRNA candidates (miR-1, miR-124, miR-889, miR-664, miR-657, miR-125a, miR-144, miR-204, miR-379 and miR-153-3 plus a scramble vector as control). All lentiviral vectors were purchased by System Biosciences.

At first, 293T cells were transfected using 20000 ng of lentiviral DNA vector per plate with the protocol previous reported (see section "lentiviral infections") to produce lentiviruses expressing each specific microRNA. Viruses were concentrated at 100X final concentration. Finally, SUM159 were infected at high MOI using 30 uL or 70 uL of concentrated virus and monitored for their viability. Cells infected with 30 uL of virus were vital and their GFP expression was evaluated by FACS sorting 48h post infection. Cells expressed high levels of GFP (>75%) and were all sorted to generate the first round of mammospheres in 15 ∅ cm plate for propagation (2000 cells/mL final concentration, for a total of 200k cells for each condition) and at 1000 cells/mL in a 12-well to evaluate SFE%. After 10 days cells were dissociated to single

cells and plated for the second round of mammospheres (F2). Cells were not sorted, but GFP expression was evaluated by FACS analysis. Dissociation was performed again after 10 days and cells were sorted to avoid clone counterselection and plated for the last round of mammosphere (F3). SFE% was calculated as previously indicated (see section "*in-vitro* 3D assay") soon after the sorting procedure (F1 and F3).

2.1.10 Oligo-mimic transfection

SUM159 cells were plated in 60 mm dishes and transfected according to fast-forward protocol with miRNA Mimic (for hsa-miR-124-3p OE we used the mimic MSY0000422, for hsa-miR-657 OE we used the mimic MSY0003335 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 dish)

N. cells plated	450.000
Volume	3 ml
miRNA Mimic 20 µM	7. 5 µl
HiPerfect	20 µl
Optimem	Up to 100 µl

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

2.1.11 In-vivo experiments and animal manipulation

The sample material to be injected (SUM159 cells) was collected into sterile Eppendorf tubes and gently pelleted in microfuge. Cell pellets (50k cells) were resuspended in a mix composed of: 14 µl of PBS and 6µl 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) and were injected intra-nipple.

Animals injected with tumorigenic cells were euthanized after 4/6 weeks (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.

2.2 Molecular biology technique

2.2.1 gDNA extraction and transgene selection by PCR

Starting from ≈ 3 x 10^5 cells, gDNA was extracted using the " Puregene Core Kit B" provided by Qiagen according to manufacturer's protocol. The DNA quality and quantity was assessed by spectrophotometry using a Nanodrop 2000 (Thermo Scientific). The gDNA was used to select and amplify transgenes (precursor miRNAs) by PCR (Hot Star Taq PCR - Qiagen) using two primers specific for the viral transgene and common for all the precursor clones.

LVL Forward: GCCTGGAGACGCCATCCACGCTG

LVL Reverse: GATGTGCGCTCTGCCCACTGAC

Here is reported the PCR conditions for a 50 µL reaction: 200ng of gDNA, 0,25 µL enzyme, dNTPs 10 mM 1µL, primers 2uM 5 uL, buffer 10X 5 µL, H_20. PCR was

performed using the following thermo profile: 5' 95°, 30" 94°, 1' 67°, 1' 72°, 10' 72°, 30 cycles.

PCR products were run in a 2% agarose gel, extracted and purified (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel). The PCR amplicon size was ≈ 500 to 700 bp. and used to prepare a library for high-throughput sequencing by NGS.

2.2.2 Clones deconvolution – Next generation sequencing gDNA

PCR products were used to prepare standard libraries for DNA sequencing.

Amplicons were shared by sonication and adapters were added to the fragmented sequences. Tagged PCR products were then sequenced at 50 base pair, single-read mode with the HiSeq 2000 (Illumina). Sequencing was performed by the Genomic Unit at the Center for Genomic Science of IIT@SEMM.

Raw reads were mapped on the human genome hg19 and, then, aligned over the reference genome consisting of all the miRNA precursor clones. For each condition, the number of matching reads were normalized over the total reads number per million (RPM) and imported into an excel worksheet for further analysis.

$$RPM = (\text{number of reads miR-(X)} / \text{number of total reads}) \times 1*10^{\wedge 6}$$

Since reads consisted of some repeated microRNA within clusters, we filtered according to redundant sequences in order not to introduce biases by counting the same microRNA twice.

We, first, analyzed the distribution of the clone of the initial population (F0) before the competition assay. Thus, we could select the clones that were represented in the initial population by more than 100 independent cells/CSCs (Coverage>100). This parameter depends on the number of transduced cells and the number of different constructs in the library and is also influenced by the phenotype of interest, since it

depends on the percentage of cells capable of displaying a phenotypic response.

A coverage mean value was calculated with the following formula:

$$\varpi = \text{Coverage mean} = \frac{\text{Total sorted cells at P0}}{\text{N. of detected miRNAs}} \quad (* \text{SFE})$$

where SFE = [n° of spheres/n° of plated cells] and is considered only in the 3D experiments.

Than, the specific coverage value of each construct was calculated:

C (miR-X) = ϖ * miR-ratio

where,

miR-ratio = Numb. of reads miR-X (RPM) ⁄ Average numb. of reads all detected microRNAs (RPM)

Next, we evaluated if microRNA precursor clones were increased or depleted over passage as consequence of the phenotype competition. We calculated a ratio score comparing the RPM (expressed in $\log_2$) of the late passage (F6 in the 3D assay; P8 in the 2D assay) against the original population (F0) for each of the microRNA constructs.

3D score = $\log_2 (RPM_{F6}/ RPM_{F0})$

2D score = $\log_2 (RPM_{P6}/ RPM_{P0})$

We measured the 3D scores for both the biological replica experiments and generated the 3D bubble plot. Similarly, the 2D scores of the biological replica experiments were used to create the 2D bubble plot. A 3D vs. 2D bubble plot was finally generated by plotting the 3D scores (average of the two experiments) in the y-axis against the 2D scores in the x-axis (average of the two experiments) to select for microRNA modifiers.

2.2.3 Total RNA 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.

2.2.4 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

Sample was incubated 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.
The kit Superscript VILO (QIAGEN) was used to reverse transcribe mRNA. The mix was prepared as following:

Enzyme mix 10X	4 µl
Vilo reaction mix 5X	2 µl
Template RNA	500 ng
RNase-free water	Up to 20 uL

Sample was incubated for 60' at 42° C, then 5' at 85 ° C. cDNA was diluted at a final concentration of 1 ng/µl for Real-Time PCR.

2.2.5 Quantitave RT-qPCR

After the RT, for microRNA expression we routinely 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 or triplicate):

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-124 (MS00006622) miR-657 (MS00010479), as controls we used SNORD61 (MS00033705) or SNORD72 (MS00033719).

For mRNA expression, we used the QuantiFast SYBR Green PCR (Qiagen)

Reagent	Volume/ reaction
2x quantifast SYBR Green PCR master mix	7,5 μl
10uM primer (F+R)	1,5 μl
Template cDNA	5 ng
RNase-free water	Up to 10 uL

qPCR detection of mature microRNAs was assayed with the following home-designed primers:

Primer sequences	**miR-124 targets**
ATGCTCTGTGGCCTTTCTTG	**FZD4 Forward**
TCTGAAAGTCCCCCAGATTG	**FZD4 Reverse**
ACGAATTTGGTCAGGCTCTC	**CLDN1 Forward**
AGGTTTTGGATAGGGCCTTG	**CLDN1 Reverse**
ATGGTTGGGAGGGAAAAGAC	**SOS1 Forward**
TCATGTTTGGCTCCTACACG	**SOS1 Reverse**
TTCTCCAGAAGGTGGTTTCG	**ITGB1 Forward**
ACATTCCTCCAGCCAATCAG	**ITGB1 Reverse**
TGCAGAAGACCTTGTGCTTG	**LAMC1 Forward**
TGGATAGGAATTGCCCTGAG	**LAMC1 Reverse**
AGGGGTGAATGGCTCTTTTC	**VAMP3 Forward**
GTTTCTGCTTCCAGCATTCC	**VAMP3 Reverse**
AGCAACCAAAGTGGGACTTC	**RAB27A Forward**
TCCCTGACCCTTCAATCAAC	**RAB27A Reverse**
TCCAGCTTTCCACACAGTTG	**RHOG Forward**
ATTCTGTGGCTCCCTCATTG	**RHOG Reverse**
CCCACACGGGACATGTGAAT	**MAGT1 Forward**
TCCATGTCAGAGGTAGCAGC	**MAGT1 Reverse**

Primer sequences	**miR-657 targets**
AGGCTTCCAGGGAGAAAAAG	**COL13A1 Forward**
TGGAACTCCAATTGGTCCTG	**COL13A1 Reverse**
TGATTTGGCCAGGCTCTATG	**CDKN2C Forward**
CCTCCCCACGTTTATTGAAG	**CDKN2C Reverse**

TGGACCCACAGAATTCCTTC	**PDCD1LG2 Forward**
TCCACACACCTCTGCTTTTG	**PDCD1LG2 Reverse**
GGTTGTTTTGCTACCTGCCG	**TSPAN6 Forward**
CGGTGACACCACAACAATGC	**TSPAN6 Reverse**
GCGATTTCAAACCTGGAGAC	**PSIP1 Forward**
TGGCTTTACAGCTCCATCAG	**PSIP1 Reverse**
TGTGTGGAAGAGCTGAATGC	**EIF4G3 Forward**
GTAATTGGCCCATGTGATCC	**EIF4G3 Reverse**
CAAAAGCCTGGGTGAGAATC	**SHC1 Forward**
ATCAGGTTGGCACAGGAAAG	**SHC1 Reverse**
CTGCAACTCGGGGACATACA	**CD83 Forward**
TGTAGCCGTGCAAAACAAGTG	**CD83 Reverse**
TGTAGCTCCGAGAGGCATGA	**CDH8 Forward**
CAGCTCTGCCACGTGTCTAT	**CDH8 Reverse**
GAAGCTGTCTCTGATCCGCA	**CCND2 Forward**
TGCTCCCACACTTCCAGTTG	**CCND2 Reverse**
GGAACTCACCCAACACCTCG	**KCNMA1 Forward**
AGGTCCCAGATACGGCATCC	**KCNMA1 Reverse**
GAGTCAGAGGGGTGACCAGA	**FRMD6 Forward**
GGCAATTCCTCAGGTTCCCA	**FRMD6 Reverse**
GCGCTTGAAGAGGCTCAGAA	**VPS53 Forward**
CCGTATTGTCTTCGCCTGGT	**VPS53 Reverse**

For normalization, we used the housekeeping genes GAPDH (QT00079247) provided by Qiagen.

2.2.6 RNA-SEQ

RNA extracted from cells transfected with miR-124 and miR-657 at 24h was used to create libraries for RNA-sequencing. According to the manufacture's manual, libraries were prepared with the use of "TruSeq RNA Sample Preparation LS (low sample)" protocol provided by Illumina.

Biefly, poly-A containing mRNAs molecules were purified using poly-T magnetic beads and fragmented. After cDNA synthesis, adpaters were ligated and fragments were amplified by PCR. A quality control analysis on the library was performed with the use of Bioanalyzer instrument. Finally libraries were run (3X plexity, 9 samples

in total) and sequenced at 50bp length and single read mode with a depth of 80 million reads.

Reads were analyzed by Benedetta Cerruti using an in-house pipeline, available in our campus IFOM-IEO The latter operates by i) filtering reads on the base of their quality ii) removing adapters and duplicate reads due to PCR amplification iii) aligning reads on the human reference genome.

At the end of the procedure, final data consisted in about 30 millions aligned reads for sample. Reads were mapped on the reference genome, which included 23368 genes, and normalized in FPKM (fragment per kilobase of exon per million reads mapped) in order to measure homogeneously all the genes within the libraries.

FPKM = numbers of reads / (genes dimension (kilobase)* number of total reads) *10^9

Genes with FPKM >2 were selected to get rid of the low expressed ones and eliminate genes not distinguishable from technical noise; FPKM > 2 = 10950 genes.

We obtained a total of 10950 expressed transcripts that were further analyzed.

2.2.7 mRNA gene expression and data analysis

We analyzed the mRNA gene expression of SUM159 OE miR-124, SUM159 OE miR-657 and SUM159 control SCR.

Principal Component Analysis was used to evaluate the variability among the different samples and performed on the 10950 gene list (with FPKM > 2) using JMP® software (SAS).

Seed enrichment analysis was performed by Sylarray software (http://www.ebi.ac.uk/enright-srv/sylarray/). All the ≈10000 genes were ordered from

the most down-regulated to the most up-regulated upon overexpression of miR-124 OR Mir-657 in SUM159 cells. This list was uploaded in Sylarray (Bartonicek and Enright, 2010) to compute word enrichment. The analysis was performed searching 7mers sequences within the 3' UTR, which usually contains the MREs. For SUM159 OE mir-124 dataset, the highest enrichment was observed in 3' UTR sequences within the "down-regulated" genes, and the most represented words (7mers) exactly matched with the canonical seed of the overexpressed miR-124. For SUM159 OE mir-657 dataset, the highest enrichment was observed in 3' UTR sequences within the "down-regulated" genes, and the most represented words (7mers) exactly matched with the canonical seed of the overexpressed miR-657.

We identified miRNA-regulated transcripts by comparing the expression regulation between groups (SCR vs. 124 and SCR vs. 657) with a statistical tool (DeSEQ2, (Anders and Huber, 2010)). We selected only genes that showed either up-regulation or down-regulation greater than |0.2| log2fold change with a false discovery rate (FDR) < 0.01. We identified a list of 769 genes regulated upon miR-124 overexpression (455 down-regulated genes and 314 up-regulated genes) and a set of 182 genes (130 down-regulated genes, 52 up-regulated genes) regulated upon miR-657 overexpression.

Ingenuity system pathway analysis was performed using IPA software (Qiagen). The 769 genes list of miR-124 OE 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}.

Gene Set Enrichment Analysis (GSEA, http://www.broadinstitute.org/gsea/index.jsp) was performed using the ≈10000 original gene list of miR-124 OE 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 genes labels to create ES null distribution.

The lists of miR-124 and miR-657 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).

2.2.8 Clustering analysis

Heat map was 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 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.

3 RESULTS

3.1 General strategy for whole-genome phenotype screening

The aim of the present study was to identify miRNA "modifiers" of mammary stem cell self-renewal by performing an unbiased genetic screening. A miRNA lentiviral library expressing the most comprehensive set of human miRNAs was used for this purpose. The library, Lenti-miR virus precursor library (System Biosciences), consists of a pool of 632 lentiviral human miRNA constructs (representative of approx. 80% of the breast miRNome) expressed in their native genomic context (pre-miRNA). This latter point is important as it preserves putative miRNA hairpin structures involved in biologically relevant interactions with the endogenous processing machinery and with regulatory partners. Each library construct is able to induce stable expression of a single miRNA upon lentiviral infection. Moreover, the constructs contain a human EF1α promoter that drives the expression of a GFP reporter transgene, which acts as fluorescent marker and helps in isolating miRNA expressing clones, as GFP-positive cells, by flow cytometry (Figure 9).

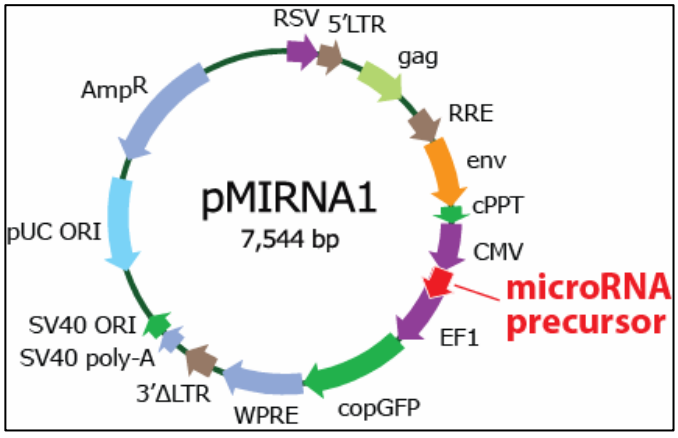


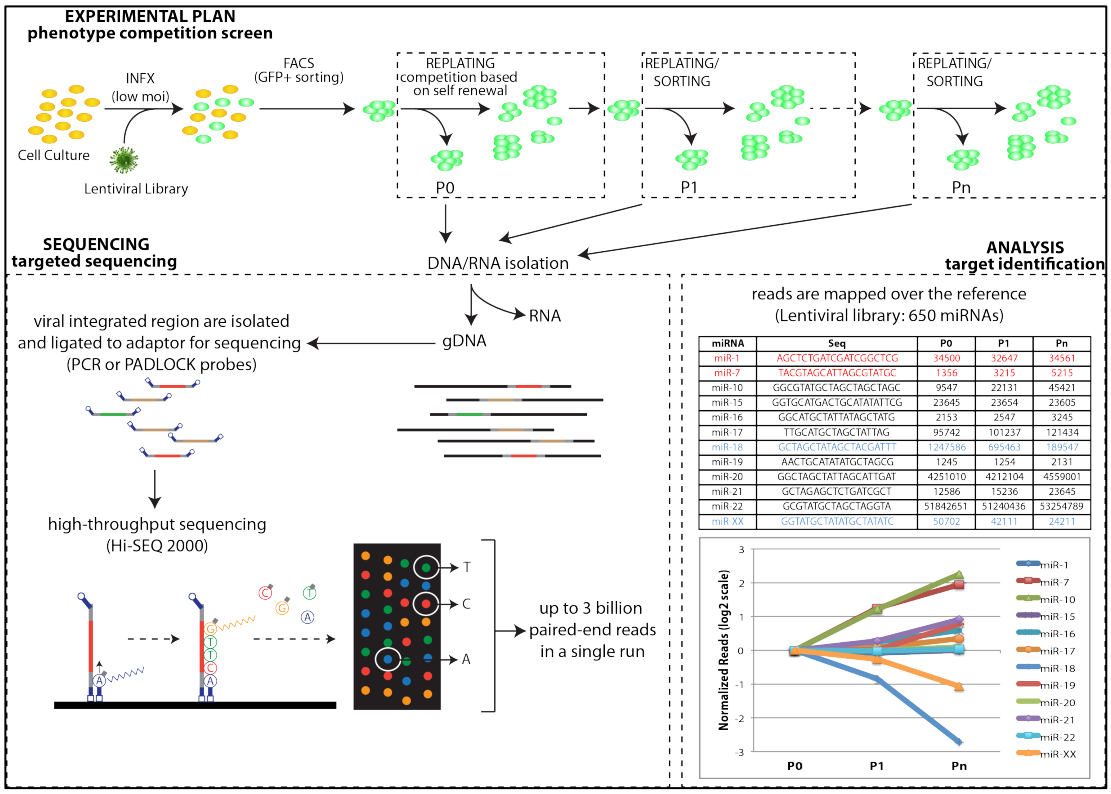
Figure 9 - Map of the HIV-based lentiviral plasmid expressing miRNA precursors from the System Biosciences (SBI) collection. The plasmid map shows the elements in the lentiviral vector pMIRNA1 that are required for different functions, including propagation in mammalian cells (SV40 Origin), selection (ampicillin resistance gene for selection in *E.Coli*), packaging and stable integration

(cPPT, GAG, LTRs). In particular, the CMV promoter guarantees high level of expression of the integrated miRNAs, while expression of the fluorescent reporter, copGFP, is driven by the Human EF1α promoter (image from www.systembio.com).

Our experimental plan consisted of expressing the miRNA lentiviral library in mammary epithelial cells (SUM 159 PT) that are able to self-renew *in vitro* (see Figure 10).

Upon infection, cells transduced with miRNAs were selected as GFP-positive (GFP^+) cells by fluorescence-activated cell sorting (FACS), and then challenged to the self-renewal phenotype by performing a 3D *in vitro* mammosphere assay. Through the serial propagation of mammosphere cultures, we selected for or against miRNAs that positively or negatively affect the self-renewal capability of mammary stem cells, respectively. In parallel, the same initial transduced cell population was subjected to a 2D proliferation assay, in order to identify miRNAs that could also affect the proliferation of mammary epithelial cells over several passages.

To identify miRNAs that were positively or negatively selected during serial propagation of mammosphere or 2D-proliferation cultures, genomic DNA (gDNA) was extracted from a small amount of GFP^+ cells at each passage and integrated viral regions were amplified by PCR. The PCR products were then analyzed by DNA sequencing (Figure 10). Using this approach, it was possible to determine the frequency of each miRNA transgene in the total cell population, at each passage. Positively or negatively selected miRNAs were then identified by comparing the frequency of miRNAs in transduced cells (GFP^+) at the baseline (P0, after transduction) and at different passages (Pn). These positively or negatively selected miRNAs represent candidate "modifier" miRNAs of the biological phenotypes being examined (i.e., stem cell self-renewal and survival/proliferation).



miRNA	Seq	P0	P1	Pn
miR-1	AGCTCTGATCGATCGGCTCG	34500	32647	34561
miR-7	TACGTAGCATTAGCGTATGC	1356	3215	5215
miR-10	GGCGTATGCTAGCTAGCTAGC	9547	22131	45421
miR-15	GGTGCATGACTGCATATATTCG	23645	23654	23605
miR-16	GGCATGCTATTATAGCTATG	2153	2547	3245
miR-17	TTGCATGCTAGCTATTAG	95742	101237	121434
miR-18	GCTAGCTATAGCTACGATTT	1247586	695463	189547
miR-19	AACTGCATATATGCTAGCG	1245	1254	2131
miR-20	GGCTAGCTATTAGCATTGAT	4251010	4212104	4559001
miR-21	GCTAGAGCTCTGATCGCT	12586	15236	23645
miR-22	GCGTATGCTAGCTAGGTA	51842651	51240436	53254789
miR-XX	GGTATGCTATATGCTATATC	50702	42111	24211

Figure 10 – Schematic diagram of experimental design. The figure summarizes the strategy used for the identification of miRNA modifiers of mammary stem cell self-renewal by phenotype competition screening. Mammary epithelial cells were transduced with miRNA-expressing viruses and replated for serial propagation of mammospheres. By isolating and sequencing the integrons, we physically counted how many times each miRNA was present in the initial mammosphere culture (passage 0, P0) and in successive passages (Pn). This enabled us to determine whether a particular miRNA was positively or negatively selected during the assay and, thus, represent a candidate modifier of self-renewal.

3.2 Competition assay

The approach used for selection of candidate miRNA "modifiers" of mammary stem cell self-renewal is based on a competition assay, in which different cell populations, each expressing a specific pre-miRNA, compete for a selected phenotype. In other words, each cell population is influenced by the miRNA that has been transduced, so that its ability to display the selected phenotype (i.e. self-renewal or survival/proliferation) is altered. In principle, there are three ways in which the

transduced miRNAs could affect the phenotype (summarized in Figure 11): 1) the miRNA could enhance the phenotype and, thus, the cells expressing that miRNA will be positively selected during the assay (positive effect, red line); 2) the miRNA could have little or no effect on the phenotype and, thus, cells expressing this type of miRNA will not be selected for or against during the assay (neutral effect, green line); 3) the miRNA could inhibit the phenotype; in this case the cells expressing this miRNA will be negatively selected or even lost over successive passages (negative effect, blue line).

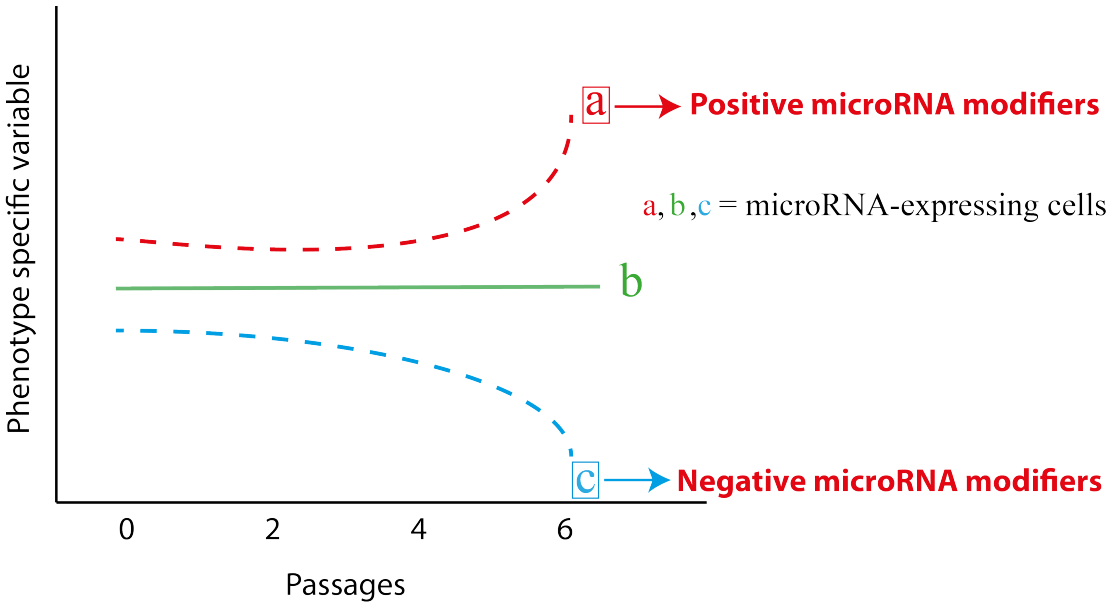


Figure 11 - Candidate selection. Diverse cell populations are generated upon miRNA library infection of the parental mammary epithelial cell culture. Each population expresses a miRNA that could influence the selected phenotype in different ways. We assumed that the vast majority of miRNAs would not influence, or only slightly influence, the phenotype (green line "b"), thus, this population will be neutral during the competition assay. Conversely, miRNAs able to enhance or inhibit the phenotype will be positively (red line "a") or negatively (blue line "c") selected over successive passages. In the case of the stem cell self-renewal phenotype, positively and negatively selected miRNAs might represent candidate oncogenic and tumor-suppressor miRNAs, respectively.

An enhancement or reduction in an individual miRNA representation would indicate a

selective advantage or disadvantage during the screening process. Thus, under selective pressure, resistant cells increase in the population and sensitized cells decrease. In our screening strategy, the selective pressure in the 3D mammosphere assay was represented by the ability of cells to survive in anchorage-independent conditions, so that only cells that have such ability, that is to say the stem cell population, could survive and generate spheres. Nevertheless, some confounding factors still exist and could impact on the screening influencing final results.

For instance, since we are using viral vectors to express miRNA precursors, which are integrated into the genome, there is the possibility that functional outcomes are associated with insertional mutagenesis events (activation of proto-oncogenes or silencing of a tumor suppressor) rather than the expression of the miRNA precursors. Therefore, we worked at low integration number, thus preventing multiple integration events, and considered only those miRNA precursors with high coverage, which are represented by multiple independent insertions. Under these conditions, we could associate a precise role to a specific microRNA avoiding data misinterpretation. Keeping low the integration number has also other advantages. For instance, if two or more miRNAs are integrated within a single cell, synergistic effects could take place, so that the observed phenotype depends on the combination of different miRNA molecules rather than a single one. All together these effects would lead to the impossibility to recognize i) miRNAs that do have an effect from those that are ineffective ii) miRNAs that act together, possibly amplifying the effect.

Besides, we reasoned that keeping high the coverage decreases the possibility to score false negative results, which typically occurs for transgenes that are a miRNA poorly represented in the initial population (P0). In a preliminary setting performed using *in-vivo* serial transplantation of tumor as cancer stem cell assay, we noticed that the

selection process was so fast that even at the first passage (formation of tumors after first round of injection) just few clones were found, while the vast majority was lost. However, the experiment was not reproduced when repeated, suggesting that the functional outcome was not dependent on the transgene but on the specific clones that were able to grow in the recipient mice. Likely with a very high number of independent experiments (injections), the sensitivity and specificity of the *in-vivo* condition could be enough to obtain reliable results. However, it would have required a number of initial cells and recipient mice so high to make the whole experiment unfeasible (too much expensive, too much cumbersome). Therefore, we decided to focus on the mammosphere assay for screening and use the *in-vivo* transplantation assay only for validation purposes.

However, keeping high the coverage has some drawbacks. Indeed, we discarded miRNA constructs that are toxic or not tolerated, including those which strongly affect cell proliferation or induce apoptosis. Indeed, in few days (3 days) necessary for infecting and stabilizing the initial population (P0), these miRNAs are already depleted and thus are not included in the competition screening. Besides, we discarded those constructs that had an insufficient number of virus particles within the library and could not be efficiently transduced. This event reflects the possible not-homogenous library production, so that some virus particles could be more represented than others.

Moreover, we aimed to control microRNAs representation also during the whole screening. Indeed, a decrease in library complexity, resulting in a reduction of miRNA clones, could be due to the loss of GFP. To overcome this bias, we introduced a sorting step throughout the screening to minimize loss of clones. With this approach, we selected only GFP^+ expressing miRNA cells that could compete in

the assay and could be able to affect the self-renewal of the mammosphere (3D screening) or influence cancer cell proliferation (2D screening).

Finally, we also aimed to distinguish between proliferation effects and self-renewal ones. Since our main aim was to identify miRNAs that specifically regulate self-renewal, we had to monitor those that simply interfere with proliferation.

Considering that the self-renewal mechanism is the way through stem cell proliferates and thus, the mammosphere contains both the stem cell and the early progenitors, which constitute the bulk of the sphere, we expect that if a microRNA impacts only on cell proliferation the sphere size should be reduced compared to the average sphere size of the specific cell line.

Based on this statement, we measured the sphere dimensions for our best candidates, so that if i) the miRNA candidate negatively acts only on the self-renewal the sphere forming efficiency should be reduced compared to the control, but the sphere size should remain equal; ii) the miRNA candidate acts on self-renewal and on the proliferation of progenitors within the sphere, we should observe both a decrease in sphere size as well as in the sphere number; iii) the miRNA candidate affects only the proliferation, we expect the same sphere forming efficiency but smaller spheres. Obviously, when performing the whole screening, it is impossible for us to discriminate between the two phenotypes: in this case, the gDNA will contain both smaller and larger spheres as well as all cells that are not stem cells and thus are not resistant to anoikis. Indeed, to better evaluate whether the microRNA is influencing specifically self-renewal rather than proliferation, we performed a validation step on our top hit candidates. All microRNA candidates were individually tested in the mammosphere assay and for each of them, we calculated the sphere forming efficiency as well as the sphere dimension of the validated clones.

3.3 Set-up of experimental conditions for the whole-genome phenotype screening

To perform the whole-genome phenotype screening, we first had to develop and optimize the different steps of the competition assay, from the lentiviral infection to the DNA sequencing. A detailed description of all the steps followed to carry out the screening is reported here.

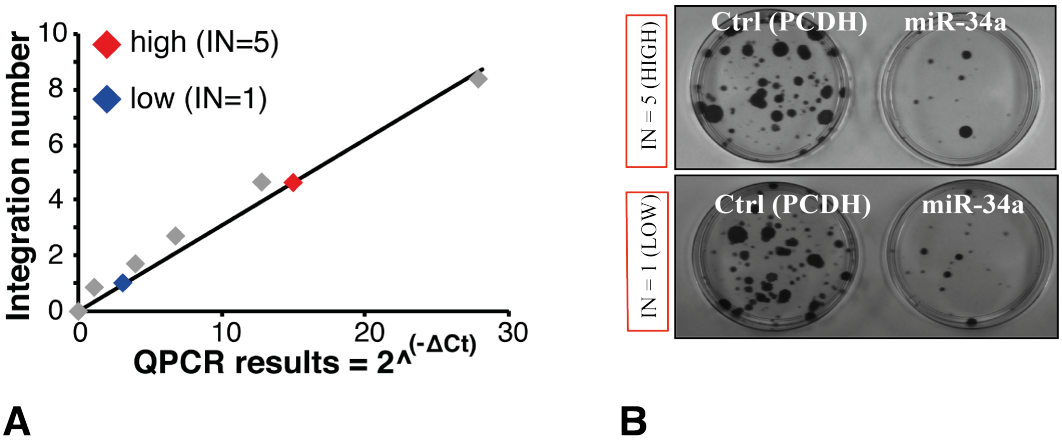
3.3.1 Lentiviral infection and competition assay

When setting up the conditions for the lentiviral infection, we aimed to have a transduction efficiency that would yield approximately one viral integration per cell. This was desirable in order to avoid multiple integrations per genome, which would increase the likelihood of random insertion of transgenes into critical genomic regions that could alter *per se* the phenotype under selection (such as, by insertion into a tumor suppressor gene). Furthermore, having multiple miRNAs within each population would increase the noise in the system, making it more difficult to associate a precise effect to a specific miRNA.

We set-up a quantitative method to count the number of viral cassettes integrated into the genome by means of q-PCR analysis of two genomic regions: i) the Upstream Conserved Region 1 (UCR1) that is a common region of the human genome; ii) the woodchuck hepatitis post-transcriptional regulatory element (WPRE), a region that is specific for the integrated viral cassette (see Figure 9). Both regions were amplified by q-PCR using the same gDNA as template. To obtain an absolute quantitative measurement, we built a calibration curve using DNA samples with known numbers of integration sites (see "Materials and Methods", Section 2.1.4 for details). We also

derived a formula to measure the number of integrations [i.e., the integration number (IN)] in each experiment, based on the fact that IN is proportional to $2^{-[Ct\ (WPRE) - Ct\ (UCR1)]}$. Following this approach, we measured the number of integration events following infection of the normal breast epithelial cell line MCFF10A with the PCDH lentiviral construct at different multiplicities of infection (MOI; Figure 12A). Specifically, at a MOI of 1 (low) we observed approximately one integration per genome (Figure 12A, blue diamond), while at a MOI of 5 (high) we observed 5 integration per genome (Figure 12A, red diamond).

We, then, verified that we could observe biological effects induced by the transduced miRNAs even when the IN was low. We infected normal human breast epithelial MCF10A cells with a miRNA, miR-34a, which is known to have an anti-proliferative effect on these cells, using MOI 1 and MOI 5, corresponding to IN=1 and IN = 5 respectively. We, then, checked if miR-34a still induced growth arrest in 2D adherent conditions at low IN as compared to the control vector (PCDH) and to high IN conditions by means of a colony assay. Results show that colonies expressing miR-34a both at high and low IN were smaller in size and reduced in number compared to the negative control, indicating that even with a low number of integrations (1), miRNA phenotypes are maintained (Figure 12B).



A B

Figure 12 – Set-up of the experimental conditions for the lentiviral infection. A) Calibration curve for the determination of the number of transgene integrations. The calibration curve was constructed by performing a qPCR analysis [using specific PCR primers designed for a common human ultraconserved region (UCR1) and for a lentiviral-specific integrated region (WPRE)] of different gDNA samples with known INs (standard samples, grey diamonds). Since the integration number (IN) is linearly correlated to q-PCR results [2^(-ΔCt), where ΔCt = Ct (WPRE) – Ct (UCR1)], we obtained a linear equation by which it is possible to measure the IN at different multiplicity of infections (MOIs). We next infected MCF10A cells with the PCDH lentiviral construct at high (5) and low (1) MOI (red and blue diamonds, respectively) and determined the IN using the standard curve. B) The effect of a known anti-proliferative miRNA (miR-34a) on normal human breast epithelial MCF10A cells was evaluated in the colony forming assay at different MOIs, 5 and 1, equivalent to an IN of 5 and 1, respectively. The PCDH control vector (Crtl) was used as a negative control. 5000 transduced MCF10A cells were plated in 2D adherent conditions onto three different cell culture dishes and incubated for 10 days before visualizing colonies.

Finally, we investigated whether we could decipher the effect of multiple miRNAs, infected at the same time, on a particular biological phenotype by analyzing the gDNA of the total cell population. For this purpose, we infected MCF10A cells with a "mini-pool" of four different miRNAs, namely miR-191, miR-34a, miR-483 and miR-200c, together with the control construct PCDH, and tested the ability of the miRNAs to regulate cell proliferation. Each construct had previously been assessed for their individual effects on MCF10A cell proliferation in the colony-forming assay in 2D adherent conditions; only miR-191 and miR-34a were shown to have an anti-proliferative effect (Figure 13A). In the "mini-pool" experiment, cells were initially infected with different amounts of the viral pool to select the condition corresponding to approx. one integration per genome. Infected cells were then grown for up to 4 passages. gDNA was collected at each passage for qPCR analysis to evaluate how each miRNA-expressing cassette varied over successive passages, using primers

specific for each of the miRNA constructs. We observed that the viral cassettes expressing miR-34a and miR-191 were promptly depleted over consecutive passages, while the other miRNAs were maintained (Figure 13B), in agreement with the results from the individual colony forming assays. Therefore, we concluded that, albeit in a preliminary setting with a limited number of miRNAs, the competition assay is feasible and that the genomic analysis is sensitive, since it allows us to score also clones that are negatively selected.

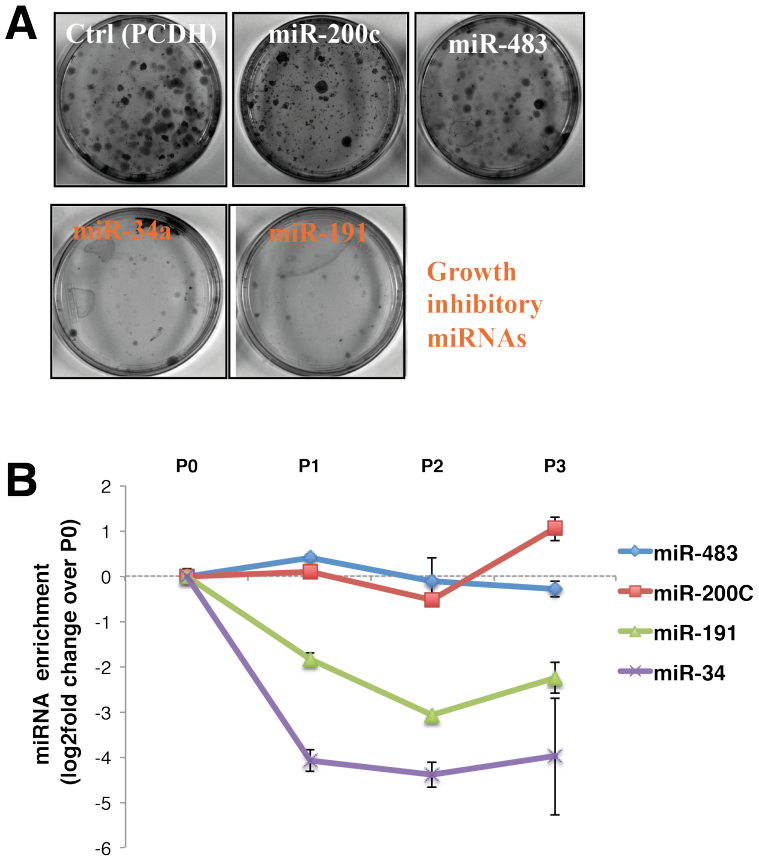


Figure 13 – Evaluation of the feasibility of the competition assay in the mini-pool experiment. A) The colony forming assay was performed on MCF10A cells separately infected with the indicated miRNAs and the control vector (Crtl PCDH) to evaluate their ability to inhibit proliferation. miR-34a and miR-191 (displayed in orange) showed an anti-proliferative effect on MCF10A cells, producing no or small colonies. B) The same miRNAs were used in a pooled experiment (mini-pool), mixing the different viruses together and infecting MCF10A cells at a low MOI. (IN=1). Cells were then sorted

for GFP expression and grown in 2D for 4 passages. At each passage, transduced cells were collected and gDNA was extracted and subjected to qPCR to measure how the different miRNAs were maintained in the cell population over time. The results are expressed as the enrichment of each miRNA in the population relative to P0 (as log_2fold-change). The mean ± ssdev of thee technical replicas are shown.

3.3.2 Evaluation of the advantages and disadvantages of selecting GFP$^+$ cells throughout the competition assay

We were aware of the fact that using a low MOI to prevent multiple integrations has the drawback that most cells of the initial population do not express any miRNA precursor. Furthermore, the transgene could lose expression during the assay due to an unfavorable integration event or heterochromatin formation induced by viral regions. Indeed, even few cells that fail to express the transgene will outcompete those in which expression is effective, thus altering the sensitivity of the assay. We reasoned that for the accomplishment of the phenotype competition screening, we should keep to a minimum the number of cells that do not express or that silence the transgene. To this aim, we could exploit the presence of the GFP marker within each miRNA construct, and preserve transduced cells that express the transgene by means of FACS-sorting. However, adding another step to the screening protocol would increase the complexity of the assay, and FACS-sorting could negatively influence the recovery and viability of cells at each passage. We, therefore, wanted to verify whether the selection of GFP$^+$ cells at each passage increases the sensitivity of the competition assay.

To this end, we performed another “mini-pool” experiment, which mimics the competition of different miRNA-expressing cell populations, using a pool of seven

viral constructs expressing miR-34a, miR-200c, miR-96, miR-146a, miR-342, miR-24 and let-7. MCF10A cells were infected with the mini-pool at a low MOI (IN = 1) and transduced cells were selected on the basis of GFP expression to generate the starting population (P0). The P0 populations were then divided into two groups and cultured in parallel in 2D using two different procedures: i) one group of cells was FACS-sorted for GFP expression at each passage, ii) the other group was simply replated without sorting. We extracted the gDNA from each of the samples (sorted and non-sorted) at every passage and evaluated the enrichment of miRNAs by qPCR analysis. This analysis revealed that the depletion of the anti-proliferative miRNAs, miR-34a and miR-96, was more evident in samples that were sorted for GFP expression at each passage compared to the non-sorted samples (Figure 14). Consequently, we decided to include FACS-sorting of GFP^+ cells in our competition screening protocol to increase the sensitivity of the assay. However, considering the requirements to perform the mammosphere assay, we decided to sort cells positive for GFP every two passages to avoid too stressful conditions and loose in this step too many cells.

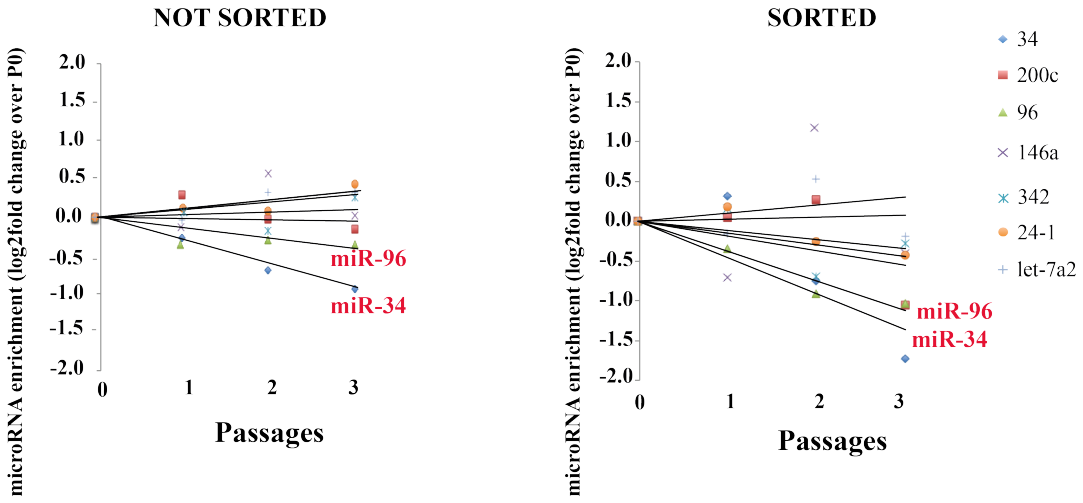


Figure 14 - Comparison of sorting *vs.* non-sorting procedure in the competition assay. A mini-pool of seven miRNAs (34a, -200c, -96, -146a, -342, -24 and -let-7) was used to infect MCFA10A

cells at low MOI. Upon infection, GFP^+ cells were selected by FACS-sorting, subdivided into two groups and cultured in 2D; in one group, cells were sorted for GFP expression at every passage (up to three, x-axis), in the other group, cells were not sorted. Q-PCR analysis was performed on extracted gDNA using specific primers for each of the miRNAs and fold-change of miRNA enrichment relative to the P0 population was calculated.

3.3.3 Deconvolution of miRNA integrons by deep sequencing

When using the entire miRNA library (approx. 650 clones) in the competition assay, it was not feasible to perform qPCR analysis on gDNA to measure all the miRNA integrons in the cell population at each passage. Instead, we decided to exploit high-throughput DNA sequencing that allows the identification of miRNA integrons in gDNA directly by sequencing. This approach has several advantages: it is quantitative (the number of sequences obtained for each miRNA cassette is proportional to its representation in the total cell population); it is unbiased (there is no need for specific PCR primers); it is cost-effective (a single sequencing lane is sufficient to analyze up to 12 samples at the same time, saving money and time). Thus, we needed to develop and optimize a protocol for the high-throughput sequencing of integrated regions using the Illumina HiSeq 2000 workstation, available in the lab. Briefly, this approach involves amplifying integrons from gDNA by PCR using primers that recognize a common sequence in the library constructs and then sequencing the PCR product by next-generation sequencing.

We reasoned that the most critical step in the detection of miRNA integrons is the selection and amplification of the integrons from the gDNA by PCR. Indeed, this PCR step is required to select from gDNA only lentiviral cassettes to be sequenced and, at the same time, to increase the amount of sequencing template. For this step,

we took advantage of the presence of specific primers (LVL) in the miRNA expressing constructs that flank the miRNA sequence.

A potential disadvantage of this amplification step is that it could alter the frequency with which miRNA integrons are detected. To test for this this possibility, we compared two different protocols for measuring miRNA integrons: i) the first method involved measuring integrons by qPCR analysis directly on gDNA using miRNA-specific primers; ii) the second approach involved first selecting and amplifying by PCR the integrons from gDNA using the LVL primers and then measuring the integrons by qPCR analysis. For this experiment, we used MCF10A cells infected with miR-191 and miR-34a viral constructs (Figure 15). We generated three samples: sample A, in which the quantity of gDNA of miR-34 was greater than that of miR-191; sample B, in which the quantity of gDNA of miR-191 was greater than that of miR-34; sample C, with equal quantity of gDNA of miR-191 and miR-34, to use as reference for normalization. We were able to score almost the same fold-change in the number of miRNA enrichment when performing the qPCR analysis directly on the gDNA or following the PCR selection step (Figure 15). This result indicates that selection and amplification of miRNA integrons from gDNA by PCR prior to their detection does not affect the original frequency of miRNA cassettes (Figure 15).

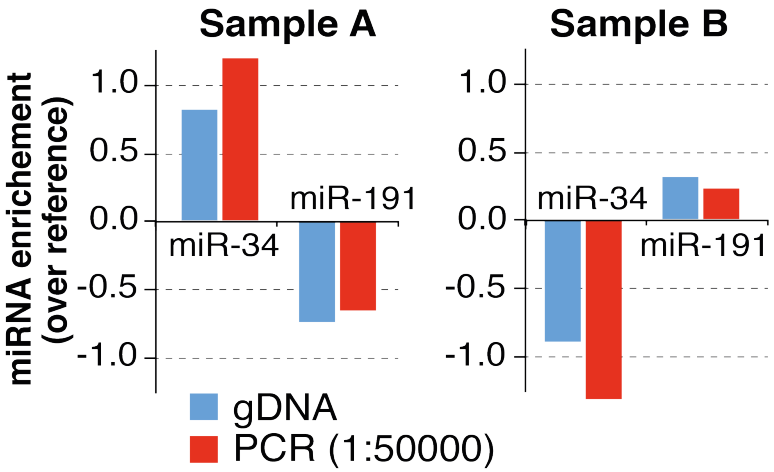


Figure 15 – Selection of viral integrons by PCR does not alter the quantification. Fold-change in miRNA enrichment (over the reference, which consisted in the mixture of miR-34 and miR-191 gDNA with a ratio 1:1) was calculated by q-PCR. MCF10A cells were infected with both miR-191 and miR-34 at varying the quantity of gDNA: sample A, gDNA of miR-34 > miR-191; sample B, gDNA of miR-191 > miR-34a. We measured miR-191 and miR-34 enrichment by performing a q-PCR analysis directly on gDNA (blue columns, gDNA). Alternatively, we selected the viral integrons by PCR and used the diluted PCR product, instead of the gDNA, to measure miR-191 and miR-34 abundance (red columns, PCR 1:50000).

3.4 Proof-of-principle experiment using MCF7 breast cancer cells: library detection and deconvolution of clones

Having defined the experimental strategy and optimized assay conditions (MOI, IN, gDNA purification, PCR selection and sequencing), we performed a proof-of-principle experiment using the entire miRNA library and the breast cancer cell line, MCF7. The aim of this experiment was: 1) to assess whether the experimental conditions defined in the assay set-up work well when using the whole miRNA library; 2) to determine whether we could detect integrons in gDNA by PCR amplification and next generation sequencing; 3) to verify whether we could detect a high percentage of miRNA clones with respect to the entire library.

MCF7 cells were infected with the lentiviral miRNA library at a low MOI (= 2), corresponding to an IN equal to one, and ~300,000 GFP^+ cells were recovered by FACS-sorting (equivalent to the P0 population). gDNA was extracted from these cells and used as a template for PCR amplification of the precursor miRNA transgenes using the LVL primers (Figure 16A). As the distribution of miRNA precursor lengths is centered on ≈ 600 bp (Figure 16A), we expected to generate PCR products of this size. Different protocols were used in the PCR selection step, varying the amount of gDNA template, enzyme type and annealing temperature to optimize the yield of the PCR, while keeping the number of cycles to a minimum (Figure 16B) to avoid biases in the amplification. The resulting PCR products were of the expected size and the sample with the highest yield was selected and sequenced at a 50 bp single-read mode with the HiSeq 2000 (Illumina) by the Genomic Unit of IIT@SEMM.

Raw sequencing reads were processed, filtered and aligned over the reference genome, which consisted of all the miRNA precursor clones (see "Material and Methods" Section 2.2.2 for details). Of note, we retrieved a total of 509 miRNAs precursors after sequencing. About 75% of the clones (381) were significantly represented (>10 reads) and the majority clones (342, 67%) had high-abundance reads (>10,000 reads, Figure 16C). We previously retrieved miRNA expression data from 466 human primary breast tumors generated by "The Cancer Genome Atlas (TCGA)" consortium (TGCA, 2012) Of the 421 miRNAs found to be expressed in breast cancer (> 10 RPM) in the TGCA dataset, approx. 80% (338) were represented in the miRNA library (Figure 16C).

Together, the results of the proof-of-principle experiment indicated that the assay conditions and the next generation sequencing approach for integrons detection worked successfully; we achieved a good level of detection of clones within the

library (509 out of 637 clones; 80%). Moreover, the breast cancer associated miRNAs were well represented in the library (80%), meaning that its use was a good starting point for the identification of miRNA modifiers of mammary stem cell self-renewal.

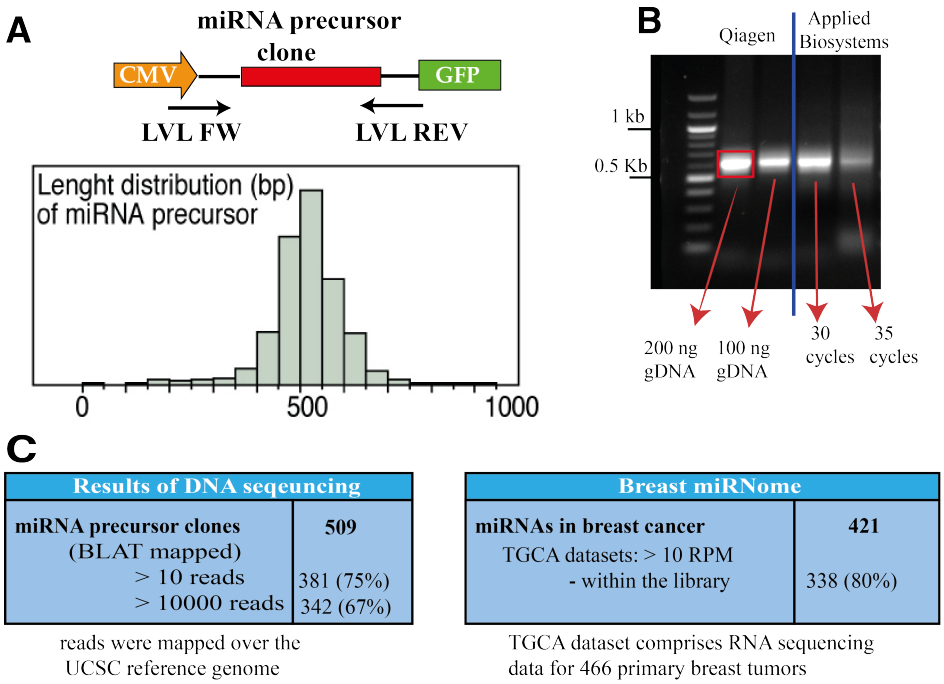


Figure 16 – Proof-of-principle experiment of miRNA library infection and integron detection in MCF7 breast cancer cells. A) Top: schematic diagram showing the position of PCR primers (LVL FW and LVL REV) used to select and amplify viral integrated regions (integrons) from gDNA. These primers were common to all miRNA precursors generated from the library. Bottom: histogram showing the distribution of the lengths (in bp) of miRNA precursors in the library. Lengths are centered on ≈ 600 bp. B) The entire miRNA library (approx. 650 constructs) was used to infect MCF7 cells at a low MOI (=2), corresponding to an IN = 1. After 72 h, GFP^+ cells were isolated by FACS-sorting and gDNA was extracted. Integrons were selected and amplified by PCR using the primers shown in A. The PCR protocol was optimized to maximize specificity and yield. The gel shows PCR products generated with two different enzymes (Qiagen and Applied Biosystems) and with two concentrations of template gDNA or with different numbers of PCR cycles. The most abundant PCR product (boxed in red) was purified from the agarose gel and used for DNA sequencing (Illumina). C) Left: sequencing results of the proof-of-principle experiment described in 'B'. miRNA precursors were identified by

mapping the sequencing data over the reference genome (UCSC database) using BLAT (BLAST Like Alignment Tool). A total of 509 precursor clones were identified, of which 75% were consistently identified in more than 10 reads, and 67% in > 10,000 reads. Right: to understand how many miRNAs related to human breast cancer are present in the library, we retrieved expression data from The Cancer Genome Atlas Consortium (TGCA). Of the 421 miRNAs expressed in breast cancer in the TGCA dataset (>10 reads per million, RPM), 338 (80%) miRNAs are present in the library.

3.5 Choice of cell line for the competition screening assay

A potential problem associated with genome-wide competition screening assays is the misinterpretation of the data due to stochastic events. To avoid such misinterpretation, a general rule that is applied in these kinds of assays is that only clones that are significantly represented in the initial population (P0), with at least 100 independent transduction events, can be considered to really compete each other during the assay. Hence, we defined a new parameter, ***Coverage*** (**C**), which indicates the number of cells in the initial population (P0) that bear the same transgene (precursor miRNA in our case), and which could effectively compete in the assay. Coverage is dependent on the number of transduced cells and the number of different constructs in the library. This parameter is also influenced by the phenotype of interest, since it depends on the percentage of cells capable of displaying the phenotypic response. In the case of the proliferation phenotype, all viable cells in the initial population (P0) have the potential to proliferate. However, in the case of the self-renewal phenotype, only a minor fraction of cells in the initial population (P0) possess stem cell traits and can therefore form mammospheres in 3D conditions. Hence, only this minor stem cell population can compete for the phenotype.

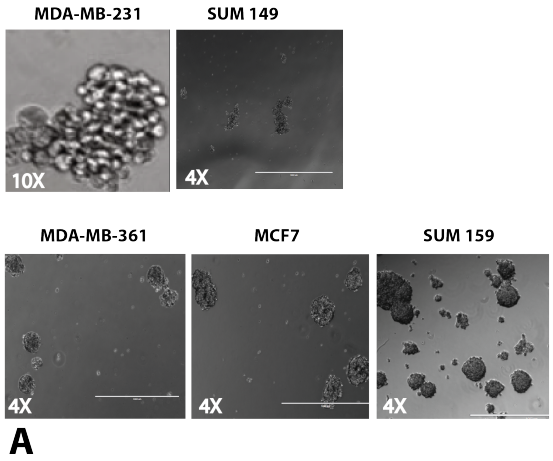
Considering this information, we screened different breast cell lines with the aim of identifying cells that can form mammospheres *in vitro* with a high SFE and thus

potentially achieve an optimal coverage value (100), which is needed for reliable results. Moreover, we tested the ability of the screened cell lines to be serially propagated as mammospheres, a prerequisite of the self-renewal phenotype competition screening. Many efforts were made to accomplish this goal, as the optimal mammosphere assay conditions varied from one cell line to another (cell concentration, medium, dissociation protocol) and were not always compatible with the sorting procedure. Indeed, the efficiency of FACS-sorting depends on the intrinsic cell shape and on cell vitality, which in this case depends on the conditions used for growth and dissociation of mammospheres.

The human breast cancer cell lines, MDA-MB-231, MDA-MB-361, MFC7, SUM 149PT, SUM 159PT, and the normal MCF10A breast cell line along with its malignant RAS-transformed derivatives (MII and MIV), were tested for their ability to form mammospheres. The cell lines MDA-MB 231, MII, MIV, and SUM 149PT were not able to form clonal mammospheres under the conditions used (Figure 17A). In contrast, MCF7, MCF10A and MDA-MB-361 cells were able to form mammospheres, albeit with a low sphere forming efficiency [SFE] (~0.1-0.15% at the first generation, F1), which dropped steadily upon serial propagation (i.e., they displayed a self-extinguishing "mortal" behavior; Figure 17A and B). The low SFE of these cell lines means that the number of cells needed to reach a good coverage in the competition screening assay is too high to be feasible. Instead, SUM 159PT cells (hereafter referred to as SUM159), a basal-like and claudin-low breast cancer cell line, displayed a very high SFE of ~10% at the first generation, which increased over successive generations (i.e., they displayed an "immortal" behavior, Figure 17A and B). Based on these results, the number of initial SUM159 cells needed to perform the phenotype competition assay (300,000) is feasible.

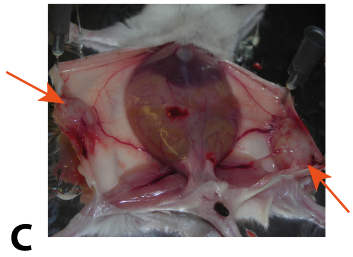
Next, we assessed whether SUM159 cells have the potential to generate solid tumors *in vivo*, a feature that could be useful in the future validation experiments. We injected 5 x 10^4 SUM159 cells into the mammary fat pads of recipient NOD/SCID mice. We observed that SUM159 are highly tumorigenic *in vivo* with 9 out 10 mice developing a visible and palpable tumor after a short latency period (4 – 6 weeks; Figure 17B and C). Moreover, it has been described in the literature that upon limiting dilution transplantation in mice, SUM159 cells are able to form solid tumors at 1000 cells⁄mL in 3 out of 8 mice (Fillmore and Kuperwasser, 2008).

Based on these *in vitro* and *in vivo* data, we chose the SUM159 cell line as the cell model system for performing the phenotype competition screening assay.



IN-VIVO TUMORIGENICITY

Weeks	Number of injected cells	Tumors formed
4\6	5 x 10^4	9\10



BREAST NORMAL AND CANCER CELL LINES

Name	Type	Subtype	Tumorigenic NOD\SCID	SFE % (F1)	Mammosph. assay
MCF10A	normal	basal	NO	0,15%	mortal
MCF7	tumoral	luminal	YES, low	0,1%	mortal
MDA-MB-361	tumoral	luminal	YES, high	0,1%	mortal
SUM 159PT	tumoral	basal	YES, high	10%	immortal

B

Figure 17 – Self-renewal capacity of different breast cell lines assessed by the mammosphere assay. A) Breast cell lines belonging to different histological subtypes were tested for their ability to form mammospheres *in vitro* in conditions compatible with the phenotype competition screening assay. MCF7 and 231 were plated at 5000 cells⁄mL; 361 and SUM 149 were plated at 20000 cells⁄mL

SUM159 cells were plated at 2000 cells/mL.Cells were cultured for eight days and colonies were visualized at light microscopy. Top panels: representative images of cells not able to form clonal mammospheres; MDA-MB-231 cells formed large clumps of cell while SUM149 PT cell line was not able to form spheres. Bottom panels: representative images of mammospheres formed by the MCF7, MDA-MB-361 and SUM159 cell lines. B) Table reporting: type (normal *vs.* tumor); breast cancer subtype; *in vivo* tumorigenicity (data taken from our laboratory and from the biological resource center ATCC); SFE at first generation of mammospheres (SFE%, F1); behavior (mortal *vs.* immortal) upon serial propagation of mammospheres under conditions compatible with the screening assay. C) Ability of SUM159 cells to form tumors upon xenograft transplantation *in vivo*. Cells (5×10^4) were injected into the mammary fat pads of 10 NOD/SCID mice and tumor growth was monitored 4 – 6 weeks after injection. Image of representative tumors (indicated by arrows) and table summarizing results is shown.

3.6 Verification of the clonal origin of SUM159 mammospheres

Our screening strategy for miRNA modifiers of self-renewal is based on the *in vitro* mammosphere assay. In this assay, cells with stem cell-like properties are able to grow in non-adherent conditions as clonally-derived cell spheroids composed of one (or a few) putative stem cells and a bulk of differentiating progenitor cells (Dontu et al., 2003). Hence, it was important to verify that, in the condition used for our screening assay, SUM159 cells give rise to spheroids that are clonally-derived rather than resulting from cell aggregation; a condition that would lead to misinterpretation of the screening results.

To assess the clonal origin of mammospheres generated by SUM159 cells, we used two different lentiviral vectors: one expressing a GFP reporter gene and the other carrying an RFP reporter gene. Since our phenotype competition screening assay involves the serial propagation of mammospheres, we decided to verify the clonal

origin of SUM159 mammospheres over several generations. Thus, we serially propagated FACS-sorted RFP-or GFP-expressing SUM159 cells as mammospheres, and at every generation we assessed the clonality of mammospheres (see Figure 18A for experimental design). To this end, we mixed together an equal number of FACS-sorted RFP- and GFP-SUM159 cells, derived from the starting population (F1) or dissociated mammospheres (Fn), and cultured cells as mammospheres (Figure 18A). With this approach, we could easily distinguish clonal spheroids, which are composed entirely of either red or green cells, from aggregates that contain both red and green cells. We observed that, at every generation (F1 – F3), the vast majority of SUM159 cell spheroids appeared either green or red, with just a few (<5%) appearing as mixed red and green aggregates (Figure 18B). This result confirms the clonal origin of the majority of SUM159 cell spheres, confirming their suitability as the model system for the screening assay.

In the same experiment, we also verified our previous observation that SUM159 cells display an "immortal" behavior in the mammosphere serial propagation assay (see Figure 17B). We assessed the SFE of RFP- and GFP-SUM159 cells over serial passages of mammosphere cultures (Figure 18A), and observed an increase in SFE over sequential passages: GFP-SUM159 cells, SFE at F1 ~10%, SFE at F3 ~20%; RFP-SUM159 cells, SFE at F1 ~4%, SFE at F3 ~28% (Figure 18C). This result confirms that SUM159 cells display an "immortal" behavior in the mammosphere serial propagation assay similar to that of high-grade breast tumors (Pece et al., 2010).

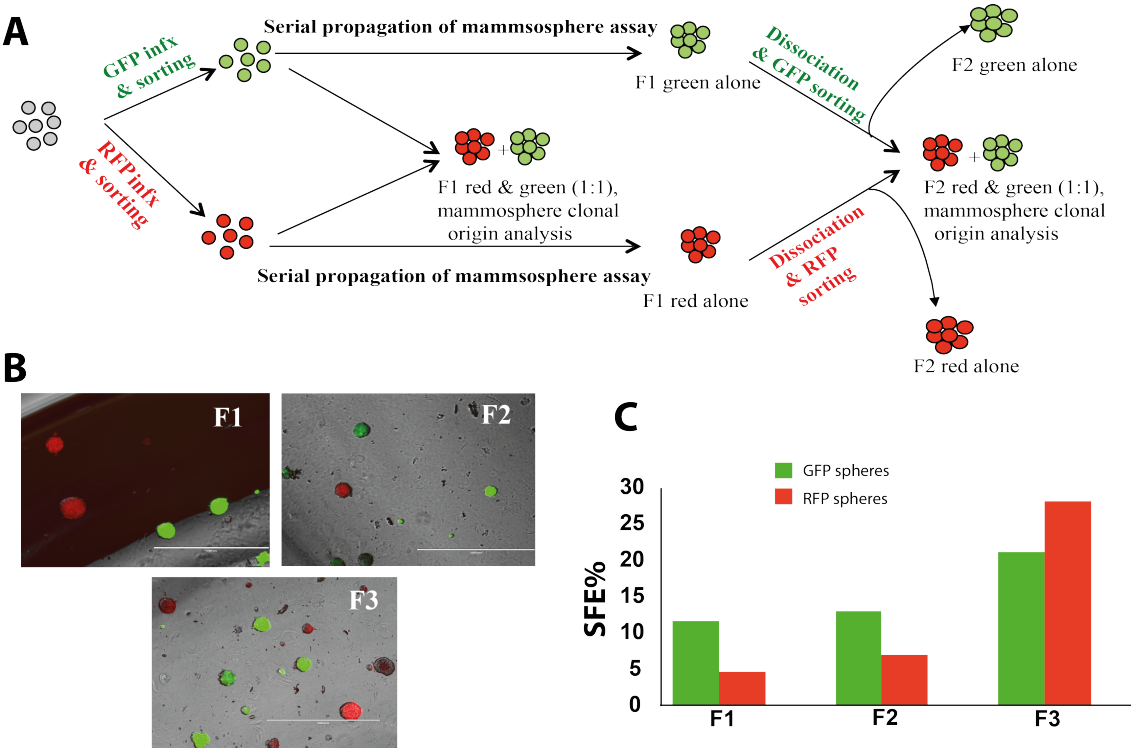


Figure 18 – Analysis of the ability of SUM159 cells to form mammospheres. A) Schematic diagram of experimental design. SUM159 cells were separately infected with lentiviral vectors carrying either a GFP or RFP reporter gene. Cells were sorted for reporter gene expression by FACS to obtain purified red and green cell populations. These cells were then analyzed separately for their ability to be serially propagated as mammospheres. In parallel, for every generation of mammospheres (i.e., F1, F2, F3) red and green cells were mixed together in a 1:1 ratio and cultured as mammospheres to check the clonal derivation of spheres *vs.* the formation of cell aggregates. B) Representative images of F1, F2 and F3 mammospheres visualized with an EVOS fluorescent microscope at 4X magnification. C) The efficiency of mammosphere formation (SFE) for the RFP and GFP expressing SUM159 cells over three different generations (F1, F2 and F3).

3.7 Identification of candidate miRNA "modifiers" of self-renewal using the SUM159 cell model system

3.7.1 *In-vitro* screening: 3D mammosphere assay and 2D proliferation assay

We performed an *in-vitro* phenotype competition assay with the miRNA library and SUM159 cells to identify candidate miRNA modifiers of self-renewal using serial

propagation of mammosphere as the phenotype read-out. To better evaluate effects that influence self-renewal or cell proliferation and survival, we used the same initial cell population transduced with the library (P0) into two parallel assays: a 2D proliferation assay and a 3D mammosphere assay (Figure 19A).

Cells were infected at MOI = 2, and those expressing the transgene were selected by FACS-sorting on the basis of GFP positivity to obtain the initial population (P0). This P0 population was divided into two groups that were challenged in parallel for the self-renewal (3D mammosphere serial propagation assay – 300,000 cells) and proliferation (2D proliferation assay – 90,000 cells) phenotypes (Figure 19A). Every second passage cells in from the 2D and 3D assays were sorted to ensure maintenance of transgene expression in the majority of cells during the screening (Figure 19B). In the 3D assay, the mammospheres were collected, dissociated and replated for a total of six generations (F1, F2 …F6). In parallel, the proliferation screen was carried out untill the 8^{th} passage (P1, P2 …P8). The entire procedure was repeated twice, with independent batches of cells and rounds of infection, thus generating two complete biological replicas of the screening.

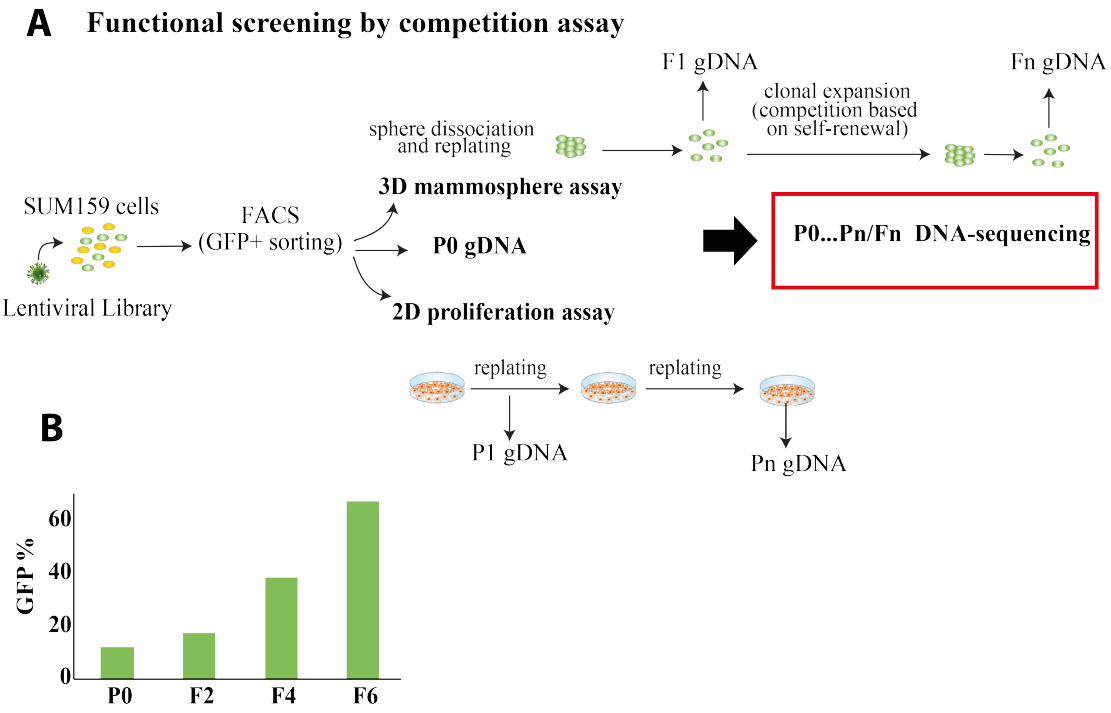


Figure 19 – Summary of the *in vitro* phenotype competition screening assay performed on SUM159 cells. A) SUM159 cells were infected with the lentiviral library at a MOI = 2 GFP$^+$ cells were FACS-sorted and the resulting P0 population was divided into two groups and assessed in two parallel competition assays based on the 3D serial propagation mammosphere assay (self-renewal) or a 2D serial proliferation assay. gDNA was extracted from each passage and integrons were identified by PCR and next generation DNA sequencing analysis. B) To enrich for transduced cells, samples were sorted for GFP$^+$ cells every second passage. The bar graph shows the fraction of GFP$^+$ cells in the starting population (F0) and in second (F2), fourth (F4) and sixth (F6) generations of the 3D mammosphere experiment.

3.7.2 Next generation sequencing for clone deconvolution and candidate identification

We collected gDNA from each passage of the 2D and 3D assays, and identified integrons by PCR amplification and next generation sequencing (Figure 20). Reads were mapped on the human genome hg19 and, then, aligned over the reference genome, which consisted of the entire human miRNA lentiviral library. Raw reads

obtained were then normalized by library size (RPM). At the end of this analysis (see "Materials and Methods" for details), we obtained 559 sequenced miRNA precursor clones.

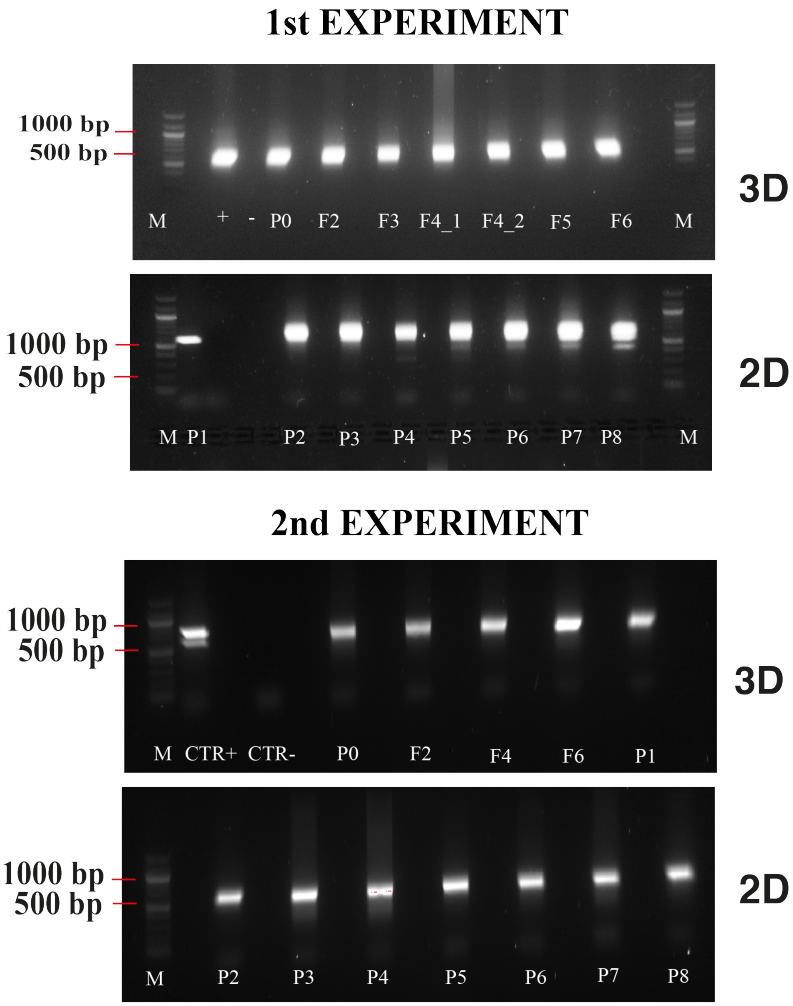


Figure 20 – miRNA screening in SUM159 cells: intergron selection by PCR. The phenotype competition experiment as described in Figure 11 was performed in SUM159 cells. gDNA was extracted at the indicated passages of the parallel 3D (Fn) and 2D (Pn) assays, and integrons were amplified by PCR. The PCR products were resolved on 2% agarose gels along with positive and negative controls (CTR+ and CTR-, respectively) and size markers (M). Shown are the results of the two independent biological replicas (1st and 2nd experiment).

The PCR amplification and sequencing step was repeated twice to generate a technical replicate for each sequenced sample. A bivariate analysis was used to test the reproducibility of the sequencing results among technical replicates (Figure 21). The sequencing results were extremely reproducible (R >0.95), in particular, for samples that were significantly represented (>100 reads). As the entire screening was performed twice, we also assessed the reproducibility of the two independent biological experiments; we compared the mean number of reads (derived from the two technical replicates) from the two independent experiments for the initial P0 cell population. This P0 population is composed of transduced GFP^+ FACS-sorted cells that have not yet undergone phenotype competition; hence, an analysis of this population will give an indication of the reproducibility of the library infection. Once again, we obtained a linear correlation (R= 0.76), which indicates good reproducibility of the two infections (Figure 22).

Similarly, we performed a bivariate analysis to evaluate the biological reproducibility at different passages (F2, F4, F6). As shown in Figure 22, the F2 reproducibility is almost identical to that of the initial population (0.75), indicating that at the early passages the vast majority of clones are still high represented and that the rate through which clones are lost, because of the selective pressure, is low.

Conversely, in the final passages (F4, F6) the biological replicas showed quite different behavior (Figure 22; R=0.1 at F4; R=0.01 at F6). This could be interpreted as poor consistence in the results or as the mere consequence of the variability in the speed and in the extent of loss/gain of clones between independent experiments. When we compared the final (F6) with the initial population (F0) in each of the two experiments, most clones were concordant in their trend of behavior (see Bubble Plot in Figure 25), suggesting similar biological effects.

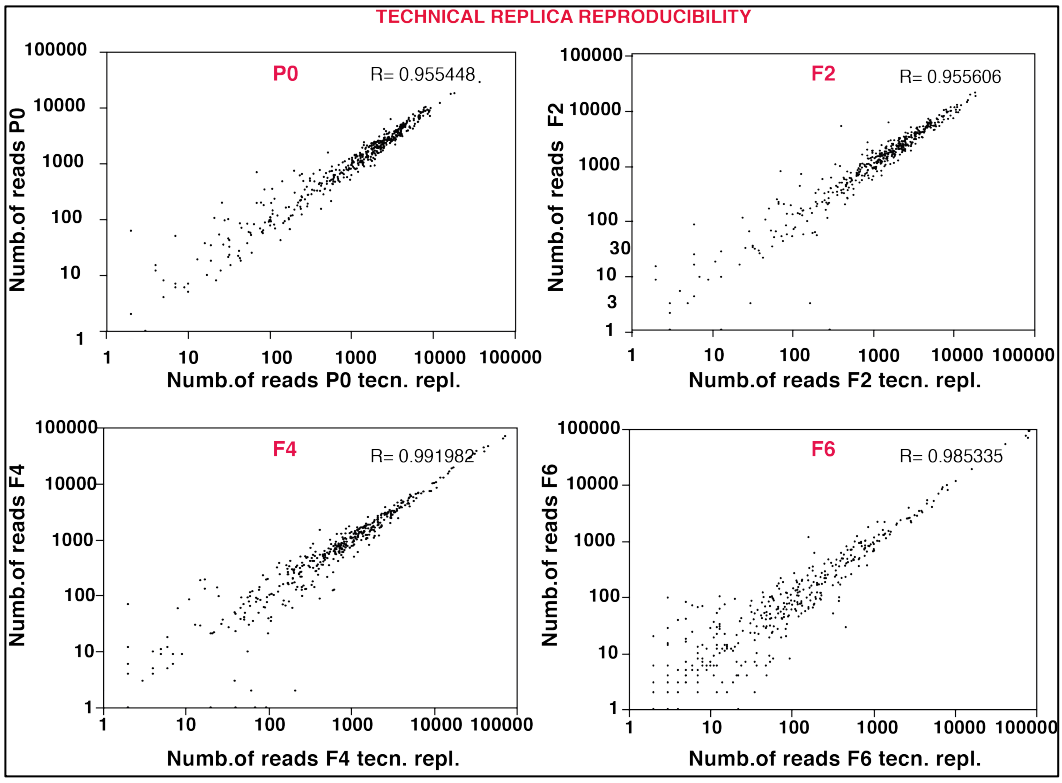


Figure 21 – Technical reproducibility of DNA sequencing analysis. Two independent PCR reactions were performed on each gDNA sample derived from different generations (P0, F2, F4, F6) of mammospheres in the 3D assay. The resulting PCR products were analyzed by next generation DNA sequencing. The graph shows a bivariate analysis of the number of reads obtained in the two technical replicates for each miRNA precursor (represented by each black dot), along with the linear correlation coefficient (R).

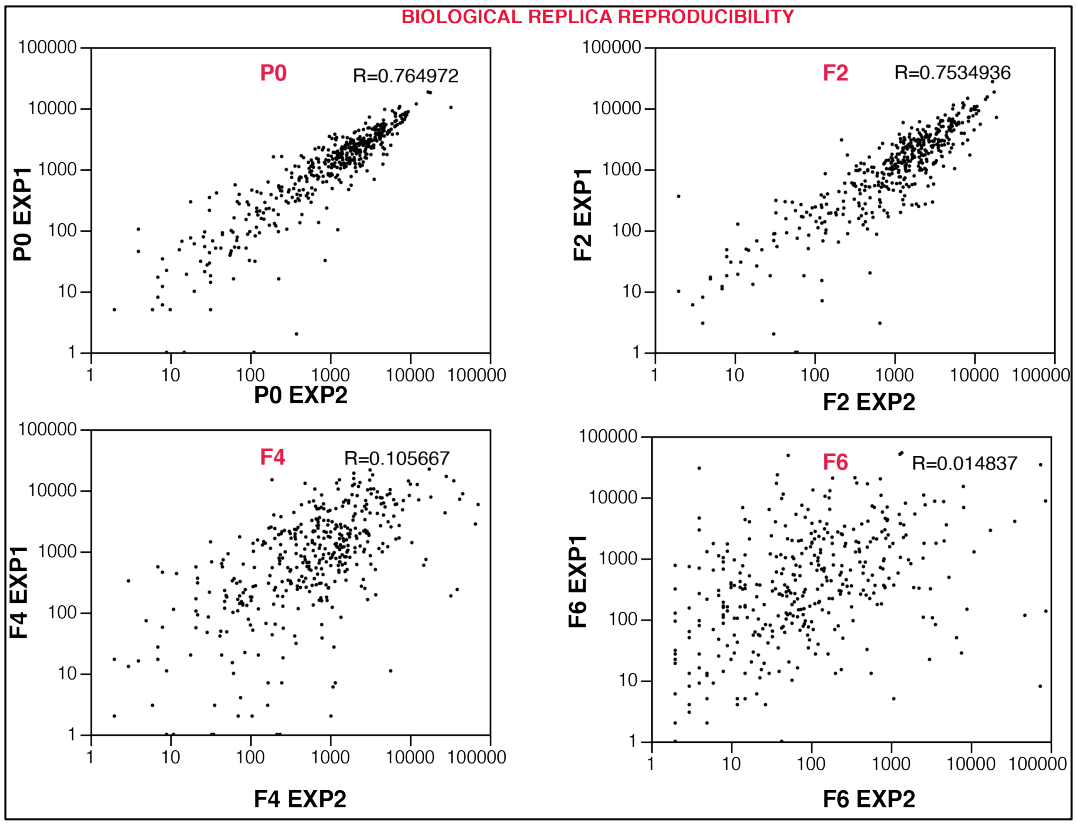


Figure 22 – Reproducibility of DNA sequencing data from the two biological replicates of the screening. Reproducibility of the DNA sequencing data across the biological replicates was assessed. The same human miRNA lentiviral library was used in two independent experiments on SUM159. gDNA samples from the initial P0 cell populations were sequenced twice (technical replicates) in each independent experiment and the mean number of reads calculated. The means from the two biological replicates at different passages (P0, F2, F4, F6) were then compared by bivariate analysis as shown in the graph where each dot represents a specific miRNA. The value of the linear correlation coefficient (R) is also shown.

We, then, focused on the initial population (P0) to select clones that were significantly represented and could therefore compete in the 2D and 3D assays. For this purpose, we calculated the coverage (C) of each construct in the 2D and 3D assays (Figure 23, see "Material and Methods" for details on the coverage calculation). Of note, all (100%) transduced SUM159 cells were able to compete for the proliferation phenotype, while only a minor fraction (about 10%) was able to grow as clonal

mammospheres in 3D conditions. Hence, the calculation of coverage had to take into account this difference. In particular, considering the clones that were represented by at least 100 independent transduction events (C>100), we found, in the first experiment, 375 and 179 different miRNA precursor clones in the 2D and 3D assays, respectively (Figure 23, 1st experiment). The same analysis was repeated with the second biological replica, yielding very similar numbers of miRNA precursor clones above the C>100 cut-off: 379 and 195 in the 2D and 3D assays, respectively (Figure 23, 2nd experiment).

Biologically the P0 coverage selection is determined by all those clones that are enough represented as virus particles within the commercial library and possess also a good and reproducible efficiency of transfection. Of course, with the coverage cut-off at the initial population, we excluded miRNAs that: i) were toxic or poorly tolerated by the cells and, thus were depleted immediately after infection, even before the competition started; ii) had an insufficient number of virus particles because of intrinsic library heterogeneity; iii) had a low transduction efficiency (see section 3.2 for details).

Since our first aim was to reveal microRNA modifiers of self-renewal, we selected the hits from the 3D experiments, irrespectively of their result in 2D. Conversely, we did not select those miRNAs that had an effect just on the proliferation screen.

For this reason, we filtered for the common miRNA clones from the 3D assays of the two biological repeats, considering only those with a coverage value ≥100 in both experiments. In total, 152 common miRNAs were identified (Figure 24) and selected for further analysis to establish whether they were positively or negatively selected (see below). Note that these 152 miRNAs were also in common with the 2D assay.

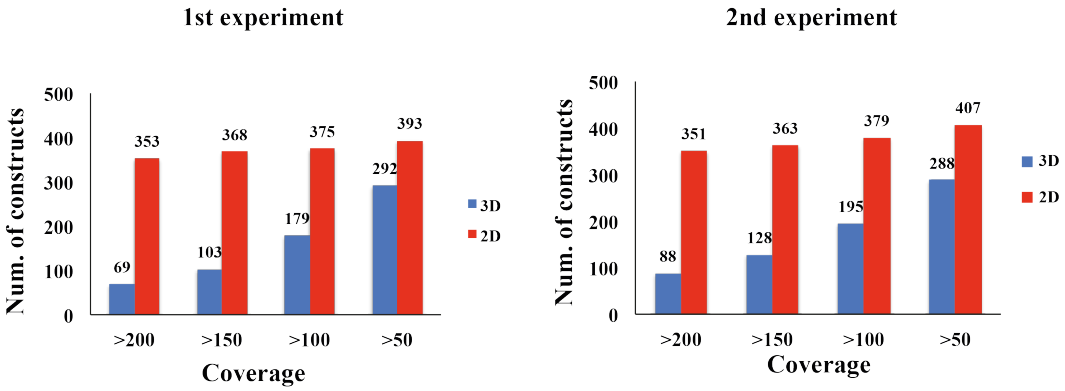


Figure 23 – Analysis of clone coverage. Based on the sequencing results, the coverage at P0 of miRNA precursor clones was calculated for the 2D and 3D assays. The coverage in the 3D assay was lower than in the 2D assay since only 10% of SUM159 cells are able to form mammospheres at the first passage. The bar graphs show the number of miRNA precursor clones (y-axis) above different coverage cut-off values (x-axis) for the 2D proliferation screening (red) and the 3D mammosphere assay (blue) in the two independent biological repeats. A high coverage value indicates a more robust result, but fewer constructs.

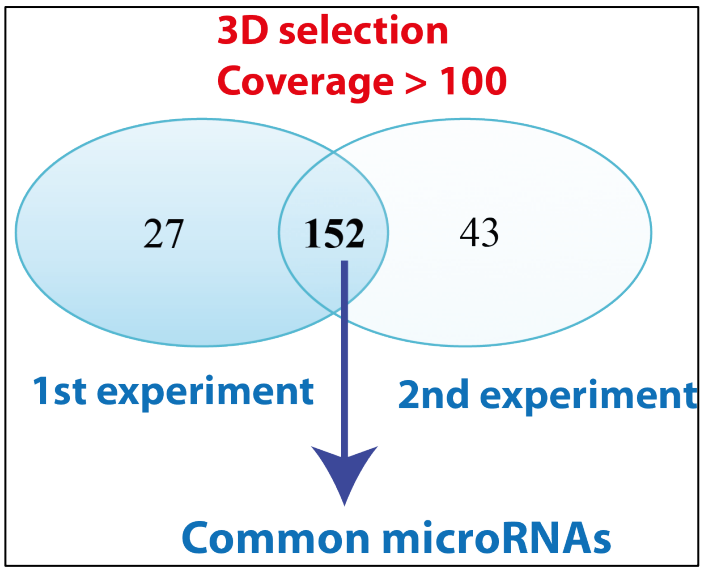


Figure 24 – Identification of common miRNA clones in the two biological repeats. Venn diagram showing the overlap of miRNA clones (C≥100) from the 3D assay, between the first and second screening experiment.

3D screening results

To reveal how each miRNA clone behaved in the mammosphere competition assay, we calculated the "3D enrichment scores" for the 152 common miRNAs using the equation:

3D enrichment score = $\log_2 (RPM_{F6}/ RPM_{P0})$

where RPM_{F6} is measure of clone abundance based on the number of reads, normalized by library size at the last passage (F6), and RPM_{F0} is the clonal abundance in the starting population (P0).

The 3D enrichment scores were calculated for both experiments and used to generate a graphical representation (the "bubble plot") of the screening results (Figure 25). Each bubble corresponds to a specific miRNA, with the size of the bubble being proportional to the miRNA coverage: the bigger the bubble, the higher the coverage. The coordinates of each bubble are relative to the 3D enrichment scores, with the values relating to the first experiment on the y-axis and those relating to the second experiment on the x-axis (Figure 25).

The bubble plot is divided into four regions: the top right and bottom left quadrants contain clones that are reproducibly enriched or depleted in the two experiments, respectively; the top left and bottom right quadrants contain clones with disjointed results, being enriched in one experiment and depleted in the other. Of note, most of the bubbles (miRNAs) were located in the bottom left quadrant close to the middle axes, suggesting that the vast majority of clones were not neutral (they should be located around zero), but tended to be depleted.

We claimed that this behavior of clone depletion could be mainly due to independent reasons. For instance, the stringent selective pressure of the mammosphere assay allows only stem cells to survive and generate spheres. Indeed, loss of clones is

generally observed when performing competition assays and the more severe is the selective pressure which is applied in the screening, the most rapidly the clones are depleted from the population. There is also the possibility that global loss is the result of the intrinsic ability of miRNA molecules to behave like tumor suppressor genes rather than oncogenes (Kumar et al., 2007). Accordingly, miRNAs that function by inhibiting self-renewal of CSCs are considered to have tumor suppressor potential. Finally, we still have to comment on the model used in the screening, which is a very aggressive cell line, with extremely high rate of sphere formation. Hence, using this cell line we could have a bias in identifying miRNAs that decrease rather than increase the efficiency of mammosphere formation. In these circumstances, miRNAs that decrease mammosphere formation tend to be depleted from the population over passages, so that the system ensures the propagation of the spheres. Even so, our first aim was to search for miRNA that potentially inhibit cancer stem cells, so this bias should have not marred the screening results.

To identify the most negatively selected clones from the bubble plot, we selected those with a ratio value of $\leq$ -4 (pink bubbles in Figure 25). Conversely, only two positively-selected clones were pinpointed (green bubbles in Figure 25).

In conclusion, from the 3D experiments, which consisted of growing cells in non-adherent conditions as mammospheres, 18 depleted and 2 enriched miRNAs were selected as top candidates for negative and positive regulators of self-renewal, respectively.

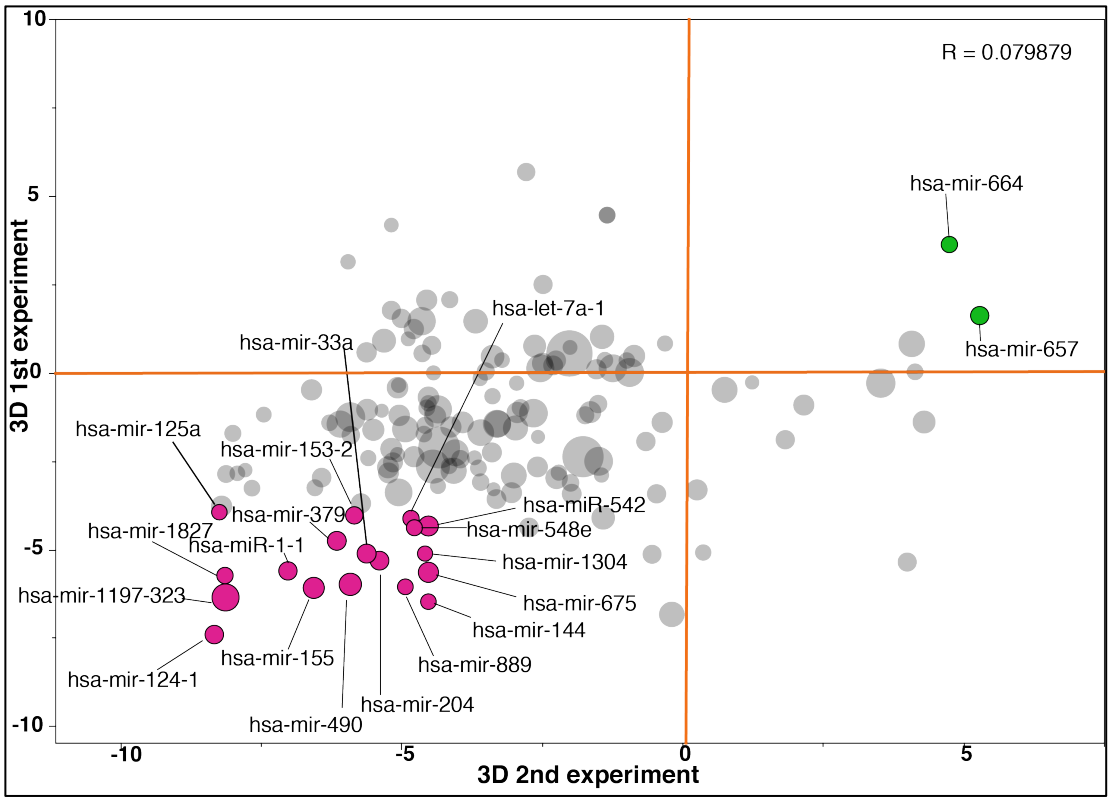


Figure 25 – Identification of candidate miRNA modifiers of stem cell self-renewal with C>100. The "Bubble-Plot" shows the behavior of the 152 common miRNAs (C > 100) during the mammosphere assay (3D) in the two independent experiments. Each bubble represents a different miRNA, where the size of the bubble is proportional to its coverage value. Based on sequencing results, an enrichment 3D score for each miRNA was calculated as the number of reads in the last passage (F6) divided by the number of reads in the starting population (P0), in log2 scale. The enrichment scores of the first experiment are plotted on the y-axis and those of the second experiment are plotted on the x-axis. The most downregulated (negatively-selected) and upregulated (positively-selected) miRNAs were selected for validation studies (shown as pink and green bubbles, respectively). The value of the linear correlation coefficient (R) is also shown.

To possibly broaden the selection of miRNA candidates and include known self-renewal inhibitor miRNAs (like miR-200 family), we reduced the coverage cut-off to 50 cells, which means starting with 269 miRNA clones present in the P0 in both biological replicas. Same as previous, we generated a bubble plot for those 269 regulated miRNAs with C>50.

Notably, we observed a similar plot, with most clones slightly decreased and with 22 additional miRNAs, in the negatively-selection area (Figure 26, blue bubbles). Among these clones, we found a member of the miR-200 (miR-200a), which has been previously described by several groups as inhibitor of breast Cancer Stem Cells (Shimono et al., 2009) (Wellner et al., 2009). No additional positively-selected clones were found. It could be interesting to test, some of the new candidates to i) reveal new other inhibitors of stem cell self-renewal, and ii) to understand whether the validation rate drops under this coverage, claiming for stochastic rather than functional effects in clone behavior.

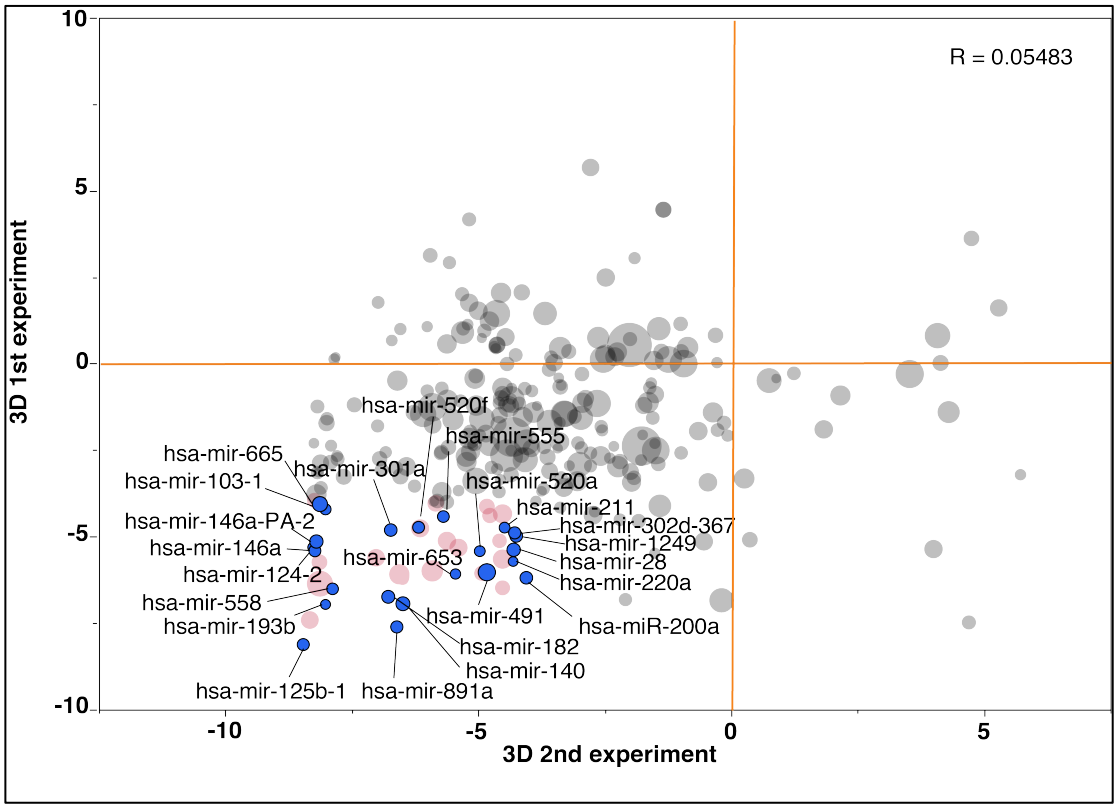


Figure 26 Identification of candidate miRNA modifiers of stem cell self-renewal with C>50. The "Bubble-Plot" shows the behavior of the 269 common miRNAs (C > 50) during the mammosphere assay (3D) in the two independent experiments. Each bubble represents a different miRNA, where the size of the bubble is proportional to its coverage value. Based on sequencing results, an enrichment 3D score for each miRNA was calculated as the number of reads in the last passage (F6) divided by the number of reads in the starting population (P0), in log2 scale. The enrichment scores of the first

experiment are plotted on the y-axis and those of the second experiment are plotted on the x-axis. The downregulated (negatively-selected) clones, that were not included in the previous analysis with C>100, are depicted in blue. The pink bubbles in the background represent the negatively-selected miRNAs regulated in the 3D experiment with C>100. The value of the linear correlation coefficient (R) is also shown.

2D screening results

We next assessed whether the 152 common miRNAs from the 3D assay were negatively or positively selected in the proliferation assay (2D). This approach would highlight if a microRNA might have a possible double role both in the self-renewal as well as in the proliferation of tumor cells. To this end, we calculated the 2D enrichment scores for the 152 miRNAs using the equation:

2D enrichment score = $\log_2 (RPM_{P6}/ RPM_{P0})$

where RPM_{P6} is the clone abundance (number of reads, normalized by library size – RPM) at the last passage (P6, corresponding to F6 in the 3D screen) and RPM_{P0} is the clone abundance in the initial population (P0).

As with the 3D assay, the 2D enrichment scores were calculated for both experiments and used to generate a "bubble plot" representative of the 2D screening (Figure 27). In this case, most of the clones had a neutral effect and were located around the center of the graph (around zero) with a tail in the bottom left quadrant, meaning that they have a slight negative effect on cell proliferation. We identified the location in the bubble plot of the 20 most depleted (in pink, Figure 27) and the most enriched miRNA clones (in green, Figure 27) from the 3D analysis (Figure 25) to determine how these clones were selected in the proliferation assay.

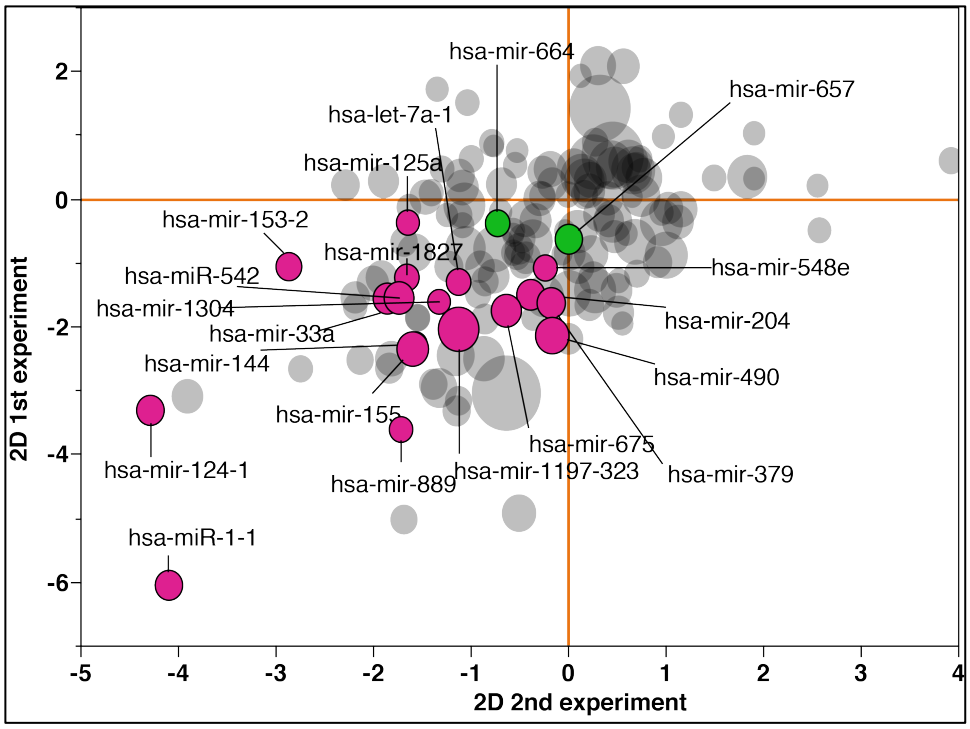


Figure 27 - Identification of candidate miRNA modifiers of proliferation and survival. The "Bubble-Plot" shows the behavior of the 152 miRNAs in the proliferation assay (2D) in the two independent experiments. Based on sequencing results, a 2D enrichment score was calculated as the number of reads in the last passage (F6) divided by the number of reads in the starting population (P0), in log2 scale. The enrichment scores of the first experiment are plotted on the y-axis and those of the second experiment are plotted on the x-axis. Each bubble represents a different miRNA with a size proportional to its coverage value. The pink and green bubbles represent most negatively and positively selected miRNAs regulated in the 3D experiment, respectively (see previous graph).

Next, we generated a 3D *vs.* 2D "bubble plot", in which we plotted the 3D enrichment score (average of the two experiments) on the y-axis against the 2D enrichment score on the x-axis (average of the two experiments), in order to identify hits that were unique to the 3D assay or might have also an effect on proliferation (Figure 28).

In general, we observed the tendency of clones to be neutral or slightly negatively-selected in the 2D screening and strongly depleted in the 3D assay, highlighting the fact that the selective pressure acting on the 3D assay is more severe compared to that

of the proliferation screening. This suggests that most of the miRNAs behaved as weak inhibitor of self-renewal of CSCs or more likely that the overexpression of any given miRNAs is less tolerated when cells are grown under 3D condition than in 2D conditions.

Self-renewal is, indeed, a particular mechanism of proliferation and, hence, the two mechanisms could share some kind of regulation. In other cases, proliferation and self-renewal could behave as distinct mechanisms. Indeed, we could distinguished these different behaviors in the 3D assay. We expect that the proliferation effect mainly impacts on the bulk population, thus determining a reduction of the sphere size, while acting on self-renewal mechanism means decreasing just the mammosphere number, whereas a combination of the two would indicated an effect on both stem cell and bulk sphere proliferation.

Accordingly, we distinguished three different categories of miRNAs within the bubble plot (Figure 28): i) 2 positively-selected miRNAs specific for the 3D assay (Figure 28 green bubbles); ii) 15 miRNAs stronger regulated in the 3D assay compared to the 2D one, suggesting that their effects could be specific for the self-renewal phenotype; iii) 3 miRNAs (Figure 28 red bubbles: miR-1-1, miR-124-1 and miR-889) strongly depleted in both 2D and 3D assays, suggesting that they might have a strong effect on both self-renewal and proliferation.

Based on these results, we selected the most promising candidates (based on the strongest regulation) for further analysis. Specifically, we selected as the “top hit candidates”: miR-1-1, miR-124-1 and miR-889 as negatively selected clones, and miR-664 and miR-657 as positively selected clones. These 5 “top hit candidate” miRNAs together with some other selected clones (miR-153-2, miR-144, mir-204,

miR-125a, miR-379 plus the scramble control) were individually validated for their effects on stem cell self-renewal.

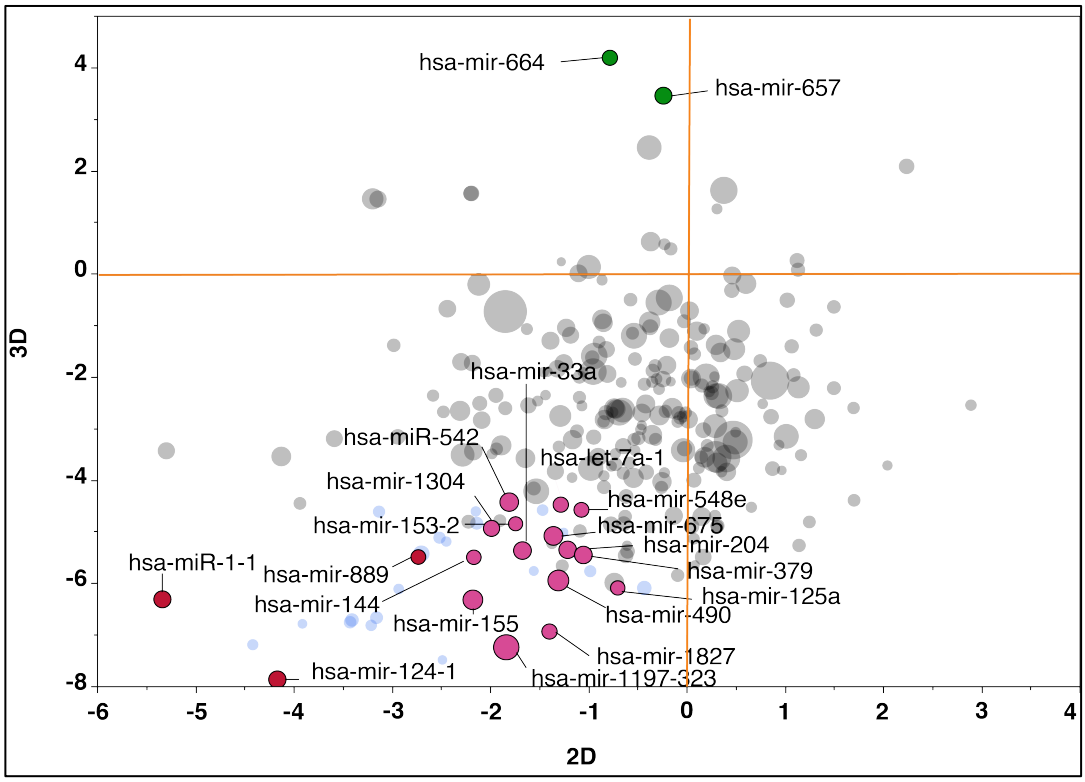


Figure 28 – Comparison of the enrichments scores in the 3D vs. 2D screening for the 152 miRNA. The bubble plot shows the enrichment scores (average of the two independent experiments, log2 scale) of the 152 miRNAs with C>100 in the 3D screening (y-axis) *vs.* their scores in the proliferation 2D screening (x-axis). Each bubble represents a different miRNA with a size proportional to its coverage value. Three miRNAs (red bubbles) were found to be depleted in both assays (2D and 3D) and thus are putative anti-proliferative and anti-self-renewal miRNAs. Fifteen miRNAs (pink bubbles) were found to be more markedly depleted in the 3D experiment compared with the 2D experiment, and thus represent candidate negative regulators of self-renewal. Two miRNAs (green bubbles) were found to be enriched only in the 3D mammosphere assay, thus representing candidate positive regulators of self-renewal.

3.8 Validation of microRNA modifiers

3.8.1 Single infection and sphere forming efficiency evaluation: looking for "real miRNA modifiers"

Through a genome-wide screening of the human miRNA lentiviral library, we identified 20 miRNA candidates that could potentially be involved in the regulation of mammary stem cell self-renewal. Three of these hits (miR-1, miR-124, miR-889) were also identified as putative negative regulators of breast cancer cell proliferation. To validate these potential miRNA "modifiers" of stem cell self-renewal, we evaluated their effects individually on the SUM159 breast cancer cell line. We first focused on the 5 "top hits", as these miRNAs showed the strongest effects on self-renewal and on cell proliferation, compared to the other hits. In this validation step, we also examined 5 other candidate miRNAs (miR-153-2, miR-144, mIr-204, miR-125a, miR-379), which appeared to be more markedly depleted in the mammosphere assay compared to the 2D proliferation screening.

We infected SUM159 cells with single lentiviral constructs (SBI) with high integration number (IN) to have a population that homogenously expresses the miRNAs with high transduction efficiency. As previously demonstrated (Figure 12), there is no difference in phenotype outcome using low or high MOI with single lentiviral constructs. Conversely, with low MOI the transduction efficiency is also very low (<10% GFP), which makes unfeasible the individual validation of miRNAs by serial propagation of mammosphere.

We optimized the virus titer to use to avoid toxicity and maximizing efficiency (see section 2.1.9 for details). Typically, we infected cells with different amount of concentrated virus (30uL and 70uL) and monitored cells for their viability and transduction efficiency (% of GFP+) up to 48h-72h post infection. We selected the

sample that showed a good efficiency (>50% GFP+) with the lowest amount of virus (30uL), nicely growing and with no signs of suffering. The samples were used to isolate the GFP^+ fraction by FACS-sorting.

With high MOI (MOI=10), we obtained more than 80% of GFP + cells at P0, meaning that we could easily perform a serial propagation of mammosphere experiments as assay to measure directly the effect of miRNAs on self-renewal (number of spheres at the first and in the next passages) and on proliferation (size of spheres and number of cells grown in 3D).

For the validation step, we performed a serial propagation of mammospheres over three generations. A control construct (SCR), which expresses a non-targeting sequence, was used in parallel to evaluate the background effects of lentiviral infection and transgene expression. As a control for the maintenance of transduced cells during the assay, we measured the percentage of GFP^+ positive cells at each generation and observed a general decrease in GFP levels over the passages, indicating that miRNA-expressing cells were being lost (Figure 29). This effect did not appear to be an effect of the lentiviral vector *per se,* since the SCR control construct did not show this behavior, but was common for all the miRNA expressing vector. We, thus, claimed that the decrease in GFP expression is the due to the fact the forced expression of any miRNA is poorly tolerated in 3D conditions. To limit the loss of GFP^+ cells, we FACS-sorted cells from the second passage (F2), in addition to the original FACS-sorting of the initial transduced population (P0), to enrich for GFP^+ clones and to allow the evaluation of SFE on a cell population that homogenously expresses the miRNA (SFE% calculated at F1 and F3). The GFP expression of cells from the F1 generation, that were not sorted, was simply monitored by FACS analysis.

Since the measure of SFE relies on the precise count of clonal spheroids, we further optimized the assay to increase its sensitivity and robustness by: i) seeding cells at a limited concentration (1000 cells\mL, in triplicate) in methylcellulose to avoid the formation of cell aggregates; ii) automatically counting spheres using a technique based on image processing (see "Material and Methods" for details); iii) discriminating between "green" (GFP^+) and "not-green" (GFP^-) spheres, to associate the phenotype only to miRNA-expressing cells (spheres that presented a heterogeneous GFP intensity were excluded) (Figure 30).

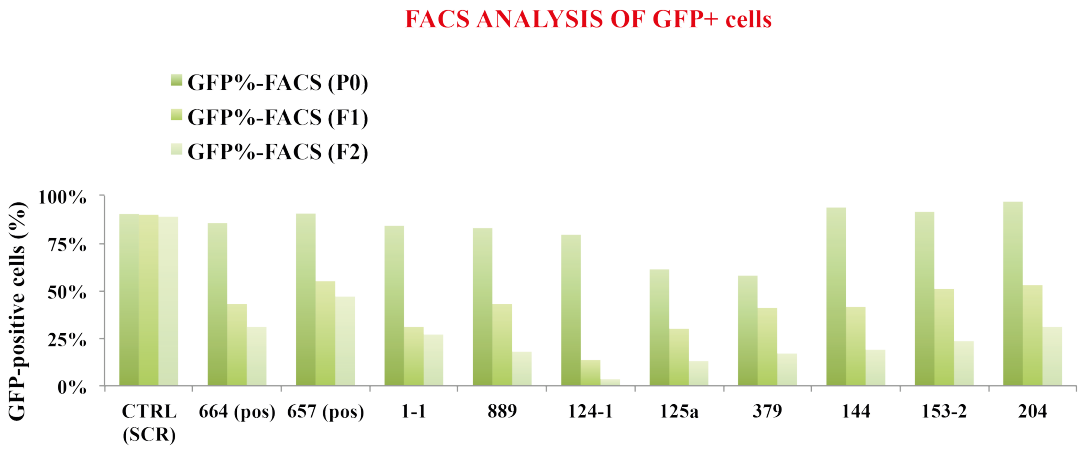


Figure 29 – Analysis of the maintenance of transduced cells upon serial propagation of mammospheres. SUM159 cells (250000 cells) were infected at high integration number with the System Bioscieneces lentiviral construct expressing the indicated miRNA hits or a scramble control sequence (CTRL SCR). Transduced cells were selected by FACS-sorting for GFP^+ cells and then plated in non-adherant conditions to allow mammosphere formation for 3 generations. The % of GFP^+ cells was assessed by FACS analysis of the initial, infected, FACS-sorted cell population (P0) or of dissociated first and second generation mammospheres (F1, F2).

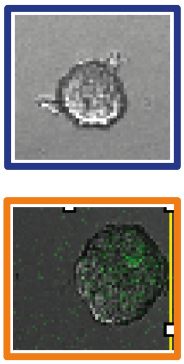


Figure 30 – Optimization of the mammosphere assay. To obtain a precise mammosphere count we: i) plated cells at a low concentration (1000 cells\mL) on methylcellulose, to avoid cell aggregation and preserve the clonality of mammospheres; ii) set-up the automatic counting of mammospheres based on a number of strict criteria that were established through a series of optimization experiments. The criteria included only counting spheres that were > 70 micron (top image), which is the average sphere size of SUM159, and that displayed a homogenous expression of GFP (bottom image of an excluded sphere displaying heterogeneous GFP expression).

Using the optimized assay, we measured the SFE of SUM159 cells transduced with the 10 selected miRNA hits, at the first (F1) and third generation (F3) of spheres, and evaluated the effects miRNA overexpression on SFE. Among the five "top candidate hits", Figure 31, two miRNAs (miR-124 and miR-657) markedly affected the SFE. In particular, miR-124, which was negatively selected in both the 2D and 3D competition screenings, significantly inhibited the SFE at F1 and F3 ($p<0.001$; Figure 31) In contrast, miR-657 expressing cells displayed an increase in SFE, which was particularly evident at the third passage ($p<0.001$ at F3; Figure 31). In addition, miR-1-1, another negatively selected miRNA in the 2D and 3D screening assays, also significantly decreased the SFE at the first and third passage ($p<0.05$ at F1 and $p<0.001$ at F3) compared to the control, although not to the same extent as miR-124 (Figure 31).

Regarding the other 5 candidate miRNA modifiers of self-renewal (miR-153-2, miR-144, mIr-204, miR-125a, miR-379) all of them were predicted by the screening to decrease self-renewal of SUM159 cells, while having a minimal effect on proliferation. They all decreased the SFE at the first generation, albeit not significantly (F1; Figure 32). However, at F3, most of the inhibitory effects were lost, with miR-144, miR-153 and miR-204 showing a significant albeit limited effect (Figure 32). We can conclude that out of 10 tested miRNAs, 6 of them had an effect on mammosphere formation, with two miRNAs, miR-124 and miR-657, having a particularly marked effect.

For these six validated miRNAs, we measured the sphere dimension over the control (SCR) to reveal whether a miRNA impacts exclusively on the self-renewal ability of stem cells, which is reflected by the decrease of the number of spheres, or may act also on the proliferation of the bulk of the sphere, determining a reduction on the sphere size.

As shown in Figure 33, most of clones formed spheres with a similar dimension to the control, whereas miR-124 formed less and smaller sphere (25%), possibly suggesting that this miRNA is able to act by negatively modulating both self-renewal, decreasing the sphere number, and tumor cell proliferation, diminishing the sizeof the sphere.

Natably, miR-124-1 and miR-657, which showed the strongest effect on self-renewal, were both in the “top five hit candidates”. These validation results indicate that our screening strategy is a robust and effective approach for identifying regulators of self-renewal. In addition, the results suggest that miR-124-1 and miR-657 could be important molecular tools for exploring the mechanisms regulating mammary stem cell properties, as well as potential molecular targets to control the self-renewal of breast cancer stem cells. Both possibilities are addressed in the next chapter.

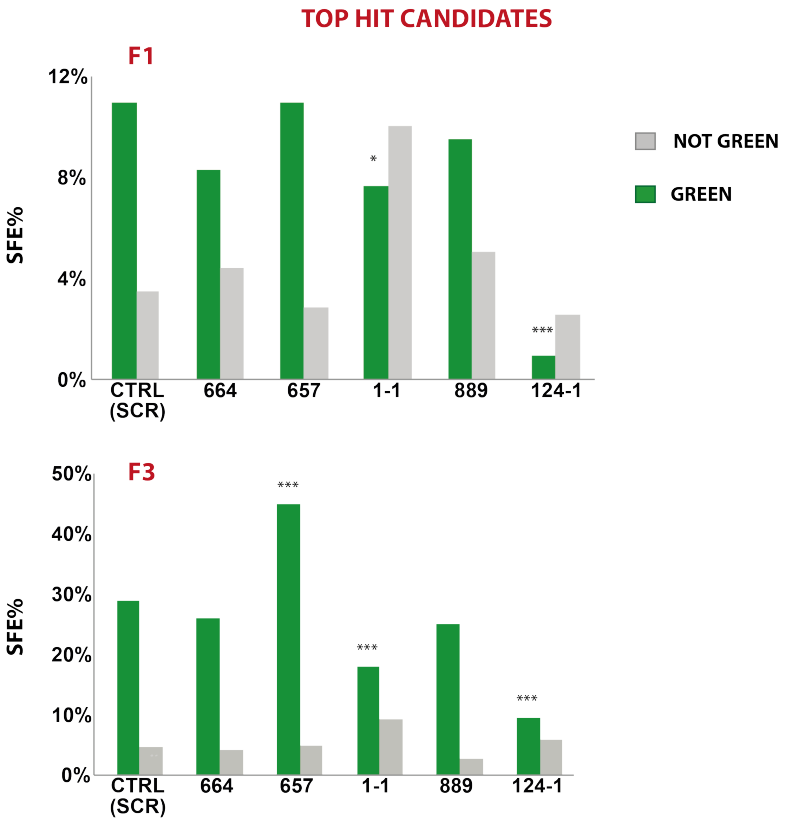


Figure 31 – Hit validation by the mammosphere assay. Evaluation of the individual effects of the top 5 miRNA hits on the sphere forming efficiency (SFE) of SUM159 cells. Cells were infected with lentiviral vectors expressing the indicated miRNA precursors or with a control vector (CTRL SCR). After 72 h, transduced GFP^+ cells were selected by FACS-sorting and plated in methylcellulose at 1000 cells\mL to allow mammosphere formation. Mammosphere cultures were then serially propagated for 3 generations. To preserve the expression of the transgene, cells were sorted after the 2^{nd} generation of spheres (F2>F3). The figure shows the SFE calculated for the first (F1, upper panel) and third (F3, lower panel) generations of spheres. Spheres were counted in triplicate per sample and the mean SFE ± sdev is shown. For each sample, the percentage of spheres expressing the GFP marker (GREEN), or not (NOT GREEN) is shown. Asterisks mark significant differences (Student's T-test, * = $p<0.05$; ***= $p<0.001$).

Figure 32 - Hit validation by mammosphere assay. Evaluation of the effects of the 5 additional selected hits on sphere forming efficiency (SFE) of SUM159 cells. Cells were infected with lentiviral vectors expressing the indicated miRNA precursors or with a control vector (CTRL SCR). After 72 h, transduced GFP^+ cells were selected by FACS-sorting and plated in methylcellulose at 1000 cells\mL to allow mammosphere formation. Mammosphere cultures were then serially propagated for 3 generations. To preserve the expression of the transgene, cells were sorted after the 2^{nd} generation of spheres (F2>F3). The figure shows the SFE calculated for the first (F1, upper panel) and third (F3, lower panel) generations of spheres. Spheres were counted in triplicate per sample and the mean SFE ± sdev is shown. For each sample, the percentage of spheres expressing the GFP marker (GREEN), or not (NOT GREEN) is shown. Asterisks mark significant differences (Student's T-test, ***= $p<0.001$).

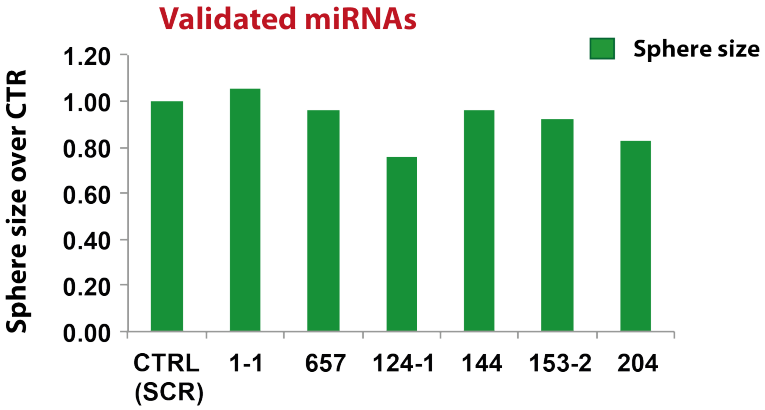


Figure 33 – Sphere dimension evaluation. Sphere size data were calculated using the automated macro-based in-house system (see section 2.1.7) and collected in an excel file. For each miRNA, the average sphere size between the triplicates was measured. The histograms shows, for each validated miRNA, its average sphere dimension over the average size of CTR spheres.

3.8.2 Transcriptional changes induced by miR-124 and miR-657

Since miR-124 and miR-657 were capable of influencing mammosphere formation in SUM159 cells, we decided to investigate the effects of these miRNAs at the molecular level. In particular, our aim was to identify their target genes, as well as to gain insights into the pathways involved in the control of self-renewal of breast cancer stem cells.

To this end, we transiently overexpressed miR-124 and miR-657 in SUM159 cells using miRNA oligo mimics and performed an in-depth analysis of the transcriptome by RNA-sequencing. As a negative control, we used an unrelated oligo (SCR) transfected at the same concentration [50 nM] in SUM159 cells. As we were mainly interested in early targets of the miRNAs, we collected RNA at 16 h and 24 h after transfection from three independent biological replicas. Initially, we verified the transfection efficiency of the two miRNAs by RT-PCR and observed a significant

increase in miR-124 and miR-657 levels over the control, comparable to that observed using lentiviral miRNA vectors (Figure 34).

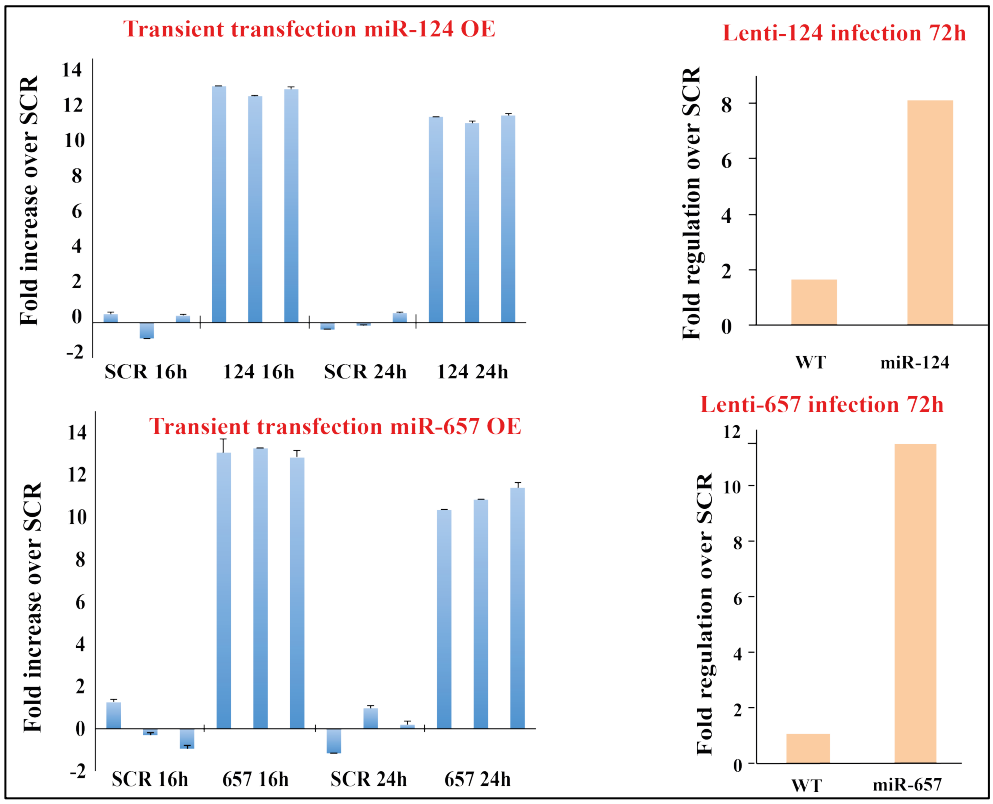
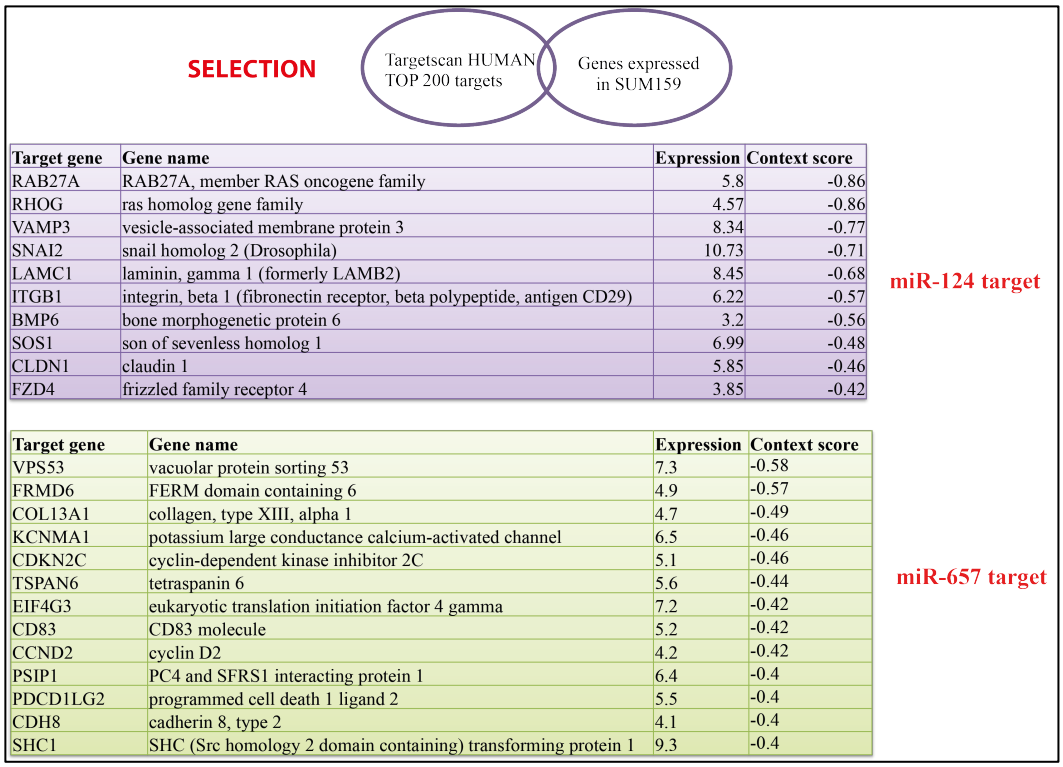


Figure 34 - Efficacy of miRNA oligo transfection. Left panels, SUM159 cells were transiently transfected with oligos (50 nM) mimicking either miR-124 (upper part) or miR-657 (lower part). A control oligo (scrambled, SCR) was used as a negative control. Shown, the qRT-PCR analysis performed to assess the efficacy of transfection in three independent biological replicas. miRNA expression was calculated as Log2fold change over SCR. Results are expressed as the mean ± sdev. The right panels report the quantification of miRNA expression from the lentiviral vectors used in the screening and in the validation experiments.

Before performing the "unbiased" analysis of the transcriptional changes by RNA-seq, we performed a "biased" approach, selecting the top predicted target genes of miR-124 and miR-657, and looking by RT-qPCR at their transcriptional changes upon miRNA overexpression. This analysis also served to establish whether there was

a significant transcriptional response under the experimental conditions used. For each miRNA, we selected the top 200 predicted target genes using the TargetScan6.2 algorithm (www.targetscan.org). This algorithm identifies potential miRNA targets based on the presence of conserved miRNA Responsive Elements (MREs), complementarity to the "seed" region of miRNAs, in the 3'UTR of messenger RNAs (mRNASs). Putative target genes are ranked according to a context-score (CS), which takes into account the type of complementarity (7mer), the location of the MRE within the 3'UTR, and potential secondary structures: the lower the CS value the stronger the predicted interaction between miRNA and its target. We, thus, restricted the "biased" analysis to potential targets with a CS < -0.4, selecting those with the lowest value. As a further filter, we considered only those targets that were expressed in our model system (namely SUM159 cells), using an in-house expression dataset. We obtained a final list of 10 predicted target genes of miR-124 and 13 predicted targets of miR-657 (listed in Figure 35).



Target gene	Gene name	Expression	Context score
RAB27A	RAB27A, member RAS oncogene family	5.8	-0.86
RHOG	ras homolog gene family	4.57	-0.86
VAMP3	vesicle-associated membrane protein 3	8.34	-0.77
SNAI2	snail homolog 2 (Drosophila)	10.73	-0.71
LAMC1	laminin, gamma 1 (formerly LAMB2)	8.45	-0.68
ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29)	6.22	-0.57
BMP6	bone morphogenetic protein 6	3.2	-0.56
SOS1	son of sevenless homolog 1	6.99	-0.48
CLDN1	claudin 1	5.85	-0.46
FZD4	frizzled family receptor 4	3.85	-0.42

miR-124 target

Target gene	Gene name	Expression	Context score
VPS53	vacuolar protein sorting 53	7.3	-0.58
FRMD6	FERM domain containing 6	4.9	-0.57
COL13A1	collagen, type XIII, alpha 1	4.7	-0.49
KCNMA1	potassium large conductance calcium-activated channel	6.5	-0.46
CDKN2C	cyclin-dependent kinase inhibitor 2C	5.1	-0.46
TSPAN6	tetraspanin 6	5.6	-0.44
EIF4G3	eukaryotic translation initiation factor 4 gamma	7.2	-0.42
CD83	CD83 molecule	5.2	-0.42
CCND2	cyclin D2	4.2	-0.42
PSIP1	PC4 and SFRS1 interacting protein 1	6.4	-0.4
PDCD1LG2	programmed cell death 1 ligand 2	5.5	-0.4
CDH8	cadherin 8, type 2	4.1	-0.4
SHC1	SHC (Src homology 2 domain containing) transforming protein 1	9.3	-0.4

miR-657 target

Figure 35 – A "biased" approach for identification of miR-124 or miR-657 target genes. The upper diagram summarizes the strategy used to select putative targets of miR-124 and miR-657. Starting from predicted targets (according to the Targetscan 6.2 algorithm), the first 200 ranked targets (with a context-score < -0.4) were chosen and further selected by checking their expression levels in the SUM159 cell line (only targets expressed in the model system were considered). The top predicted genes for each miRNA (10 for miR-124 and 13 for miR-657) are listed with the Gene Symbol (Target gene), full gene name, expression value (according to an in-house microarray dataset of SUM159 cells), and context-score (by TargetScan 6.2).

These target genes were then analyzed by qRT-PCR in SUM159 cells transfected with the miRNA oligo mimics. We observed that the majority of miR-124 predicted target genes (7 out of 10) were effectively downregulated upon miR-124 overexpression (Figure 36). Of note, these validated target genes include known regulators of the epithelial-to-mesenchymal transition (EMT) program, SNAI2 and the integrin ITGB1, which have been previously associated with stem cell traits in breast cancer cells. Regarding the miR-657 predicted target genes, 3 out of 13 were

highly regulated and 4 other genes showed a mild transcriptional effect (Figure 37). Thus, the "biased" approach indicated that a significant transcriptional response occurred upon miR-124 and miR-657 mimics transfection. In particular, miR-124 was able to induce the repression of many genes.

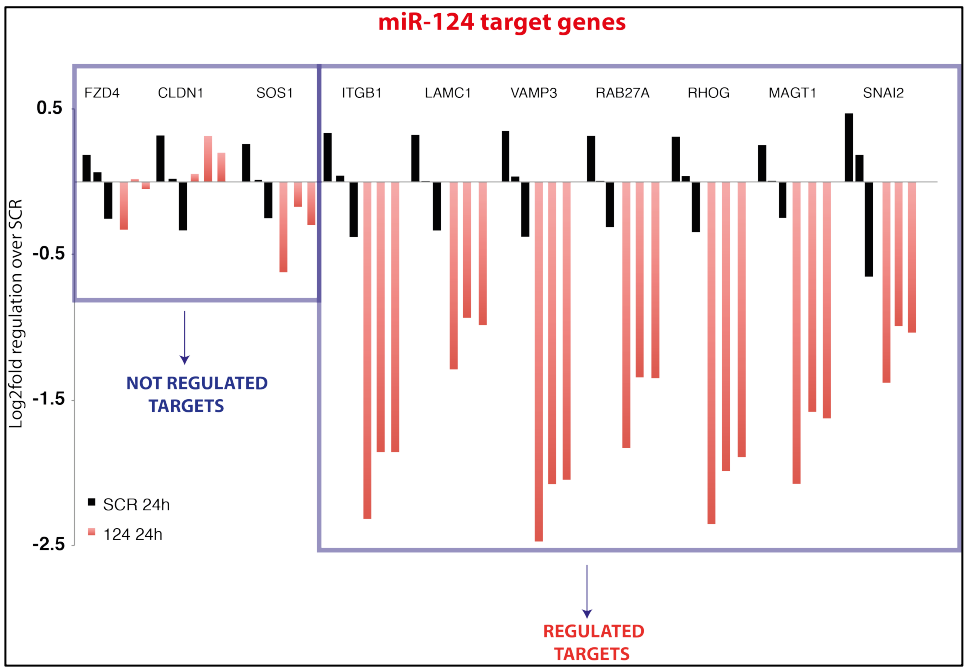


Figure 36 – Expression of putative miR-124 target genes upon miR-124 overexpression in SUM159 cells. A qRT-PCR analysis was performed to evaluate the change in expression of 10 putative miR-124 target genes 24 h post-transfection with miR-124 oligos in SUM159 cells. Expression levels are shown as log2fold change regulation over the control oligo (SCR) in three independent experiments. Black and red bars represent SCR and miR-124 overexpressing samples, respectively. Seven out of the 10 analyzed target genes were consistently downregulated by miR-124 (Regulated Targets), while the remaining 3 were not regulated (Not Regulated Targets).

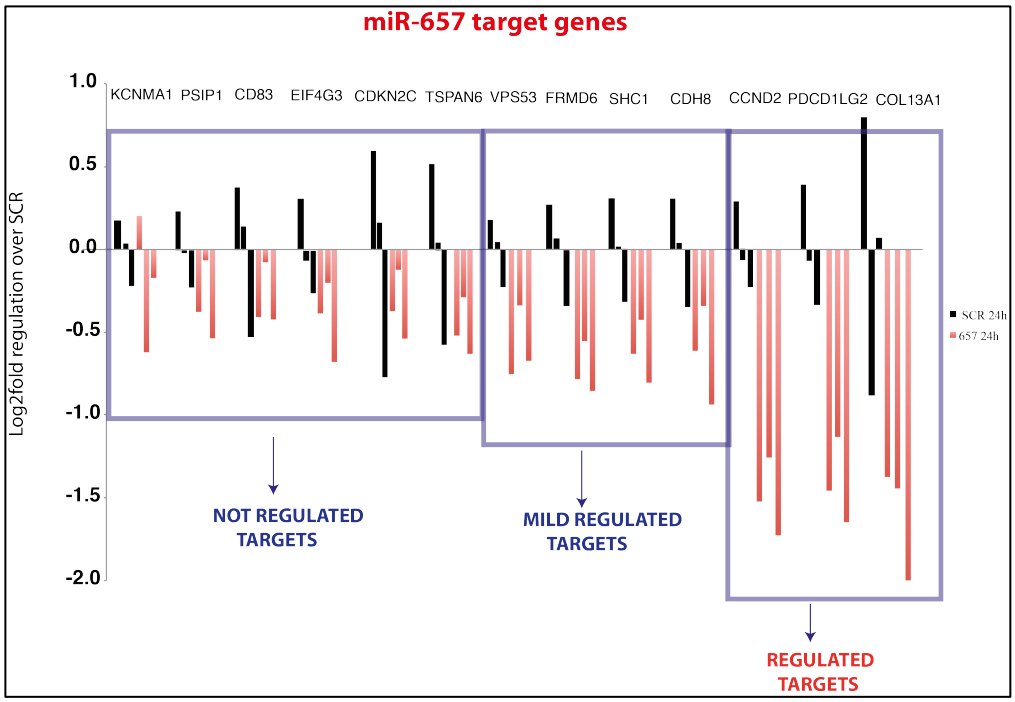


Figure 37 - Expression of putative miR-657 target genes upon miR-657 overexpression in SUM159 cells. A qRT-PCR analysis was performed to evaluate the change in expression of 13 putative miR-657 targets upon miR-657 overexpression in SUM159 cells. Expression levels are shown as log2fold change in regulation over the control oligo (SCR) in three independent experiments. Black and red bars represent SCR and miR-657 overexpressing samples, respectively. Three out of the 13 analyzed target genes were consistently downregulated by miR-657 (Regulated Targets), while 4 targets displayed a mild downregulation (Mild Regulated Targets). The remaining 6 targets were not regulated (Not Regulated Targets).

3.8.3 RNA- sequencing: data analysis and quality control of samples

We, next, implemented an "unbiased" analysis of transcriptional changes induced by miR-124 and miR-657 by means of RNA-sequencing, performed at 50 bp length and single read mode with a depth of 80 million reads (for details on library preparation see "Materials and Methods"). We obtained about 90 million reads from each sample, which were analyzed by a bioinformatician of our group (Benedetta Cerruti) using an in-house pipeline. This pipeline consists of standard processing and filtering steps

including: i) sequencing quality controls, ii) removal of adapters and duplicated reads, and iii) alignment on the reference (human) genome. At the end of the procedure, the final data consisted of ~ 30 million aligned reads per sample, making a total of > 270 million reads. Expression data were also normalized to library size and transcript length, and were expressed as fragment per kb of exon per million reads mapped [FPKM = n^o of reads / (gene dimension {kb}*n^o of total reads)*10^9]. The reference genome includes 23,368 transcripts, most of which are poorly expressed or not expressed at all and barely distinguishable from technical noise. We, thus, eliminated transcripts that were not significantly represented, using a standard cut-off for this type of experiment (FPKM >2; meaning that there are at least 2 reads per kb of gene, every million of mapped reads in all nine samples). Using this approach, we obtained a total of 10,950 expressed transcripts, which is in line with similar analyses performed by our and other groups.

Initially, we performed a principal component analysis (PCA) to examine the global transcriptional differences between all samples, without any *a priori* definition of groups. According to the PCA, the samples are visualized in multidimensional data sets (typically 2D or 3D graphs) according to the main sources of variability (components; Figure 38). In particular, in our case, this analysis helped us in assessing whether the experimental conditions (i.e. miRNA overexpression) influence the transcriptome of each sample more than the technical and biological variability. We observed that all the biological triplicates were clustered together, meaning that their transcriptional profiles are similar (Figure 38). Moreover, the different treatments (transfections) were far away each other, indicating that we are dealing with different expression profiles as a consequence of a profound change in the gene expression upon miRNA overexpression (Figure 38).

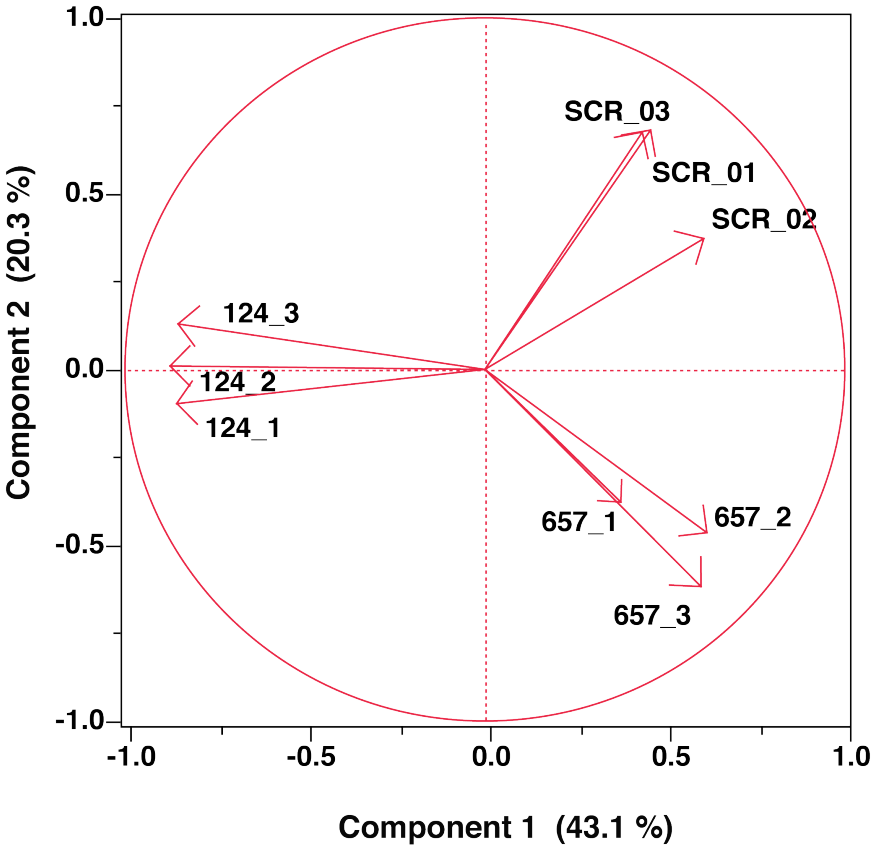


Figure 38 - RNA-sequencing analysis of transcriptional effects induced by either miR-124 or miR-657 overexpression: Principal component analysis (PCA). SUM159 cells were transiently transfected with miR-124 or miR-657 in three independent experiments. A control oligo (scrambled, SCR) was also used as a negative control. The transcriptional effects induced by miRNA overexpression were evaluated by RNA-sequencing. A PCA was performed on approx. 10,000 genes that were expressed in SUM159 (FPKM>2) revealing similarities and differences between individual samples. As shown in the graph, samples treated with the same way grouped together, suggesting that the biological differences between samples were far greater than technical noise observed between replicates.

Based on the evidence that mammalian miRNAs predominantly act to decrease target mRNA levels (Guo et al., 2010), we reasoned that the transcriptional changes observed could be primarily due to mRNA silencing mechanisms. Indeed, preliminary analysis of selected putative target genes for miR-124 and miR-657 by RT-qPCR revealed that both miRNAs were able to induce downregulation of their target genes

(see Figure 36 and Figure 37). We exploited a computational tool, available on the web, known as SylArray (Bartonicek and Enright, 2010), which searches for overrepresented MREs in the 3′UTR sequences of transcripts coming from a given dataset. The algorithm returns the distribution of the MREs for any miRNA within the selected gene expression dataset. If the distribution of the MREs is asymmetric, then the corresponding miRNA is affecting gene expression.

Figure 39 shows the SylArray analysis of the SUM159-miR-124 dataset. As expected, there was a significant enrichment (red line in plot) of the sequence matching the miR-124 seed, with a remarkable enrichment score (MAX P-value 10-95) asymmetrically distributed towards the downregulated genes (Figure 39). This means that the miR-124 target genes are frequently downregulated, and almost completely depleted from the "not regulated" or "upregulated" part of the dataset. Similarly, SylArray analysis of the SUM159-miR-657 dataset indicated that the miR-657 seed was enriched among the downregulated genes, although the enrichment was less pronounced than that of the miR-124 seed in the SUM-miR-124 dataset (MAX P-value 10-13; Figure 40). Curiously, SylArray analysis of the mi-124 and miR-657 datasets revealed a mild (MAX P-value 10-10; Figure 39) or strong enrichment (MAX P-value 10-37; Figure 40) of the miR-2054 MRE among upregulated genes, respectively. These results could indicate that these genes are upregulated as the result of downregulation of miR-2054 upon miR-124 or miR-657 overexpression; a hypothesis that we are currently investigating.

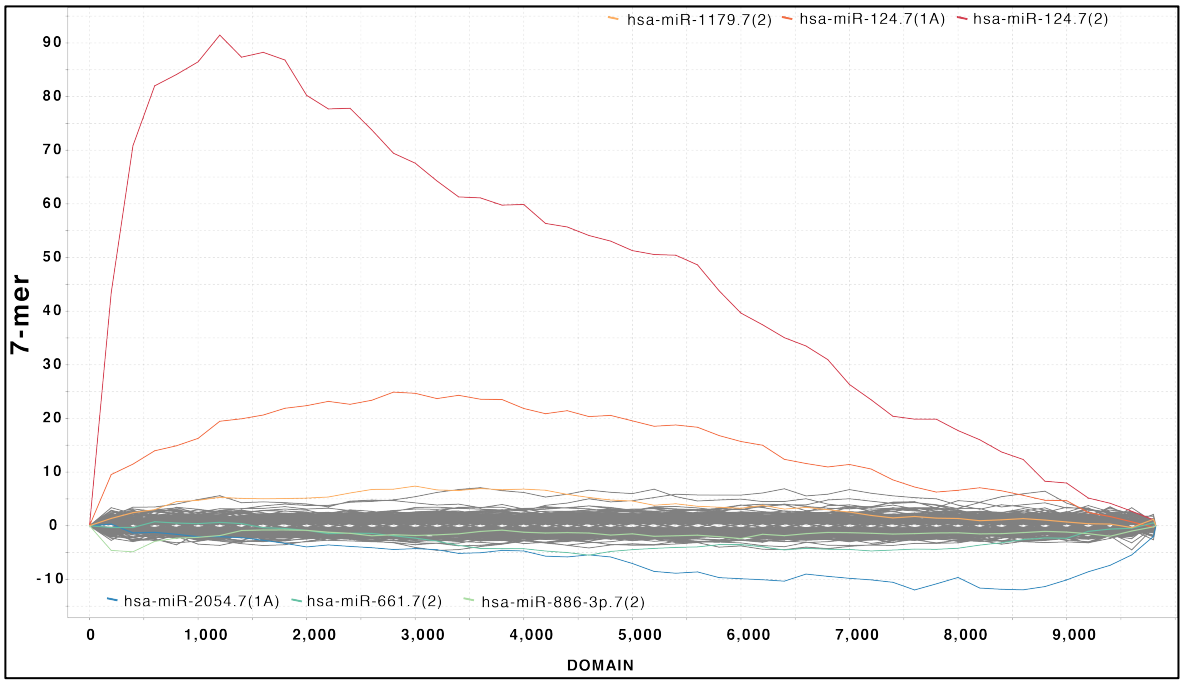


Figure 39 - RNA-sequencing analysis of transcriptional effects induced by miR-124: Sylarray analysis of the SUM159 miR-124 dataset. Seed enrichment analysis was performed using the Sylarray algorithm, which performs an unbiased search for miRNA binding sites (MREs) in the 3'-UTR of the sequenced genes. The graph reports the distribution of MREs in the RNA-seq dataset generated from SUM159 cells transfected with the miR-124 oligo, 24 h post-transfection. On the x-axis, the sequenced genes are ordered from the most downregulated to the most upregulated genes; on the y-axis, corresponding enrichment score of the 7mer MREs for each sequenced gene, expressed as a P-value. The three miRNAs with the most significantly enriched MREs among: 1) the upregulated genes, indicative of downregulated miRNAs, are shown in shades of blue (below x-axis); 2) the downregulated genes, indicative of upregulated miRNAs, are highlighted in shades of red.

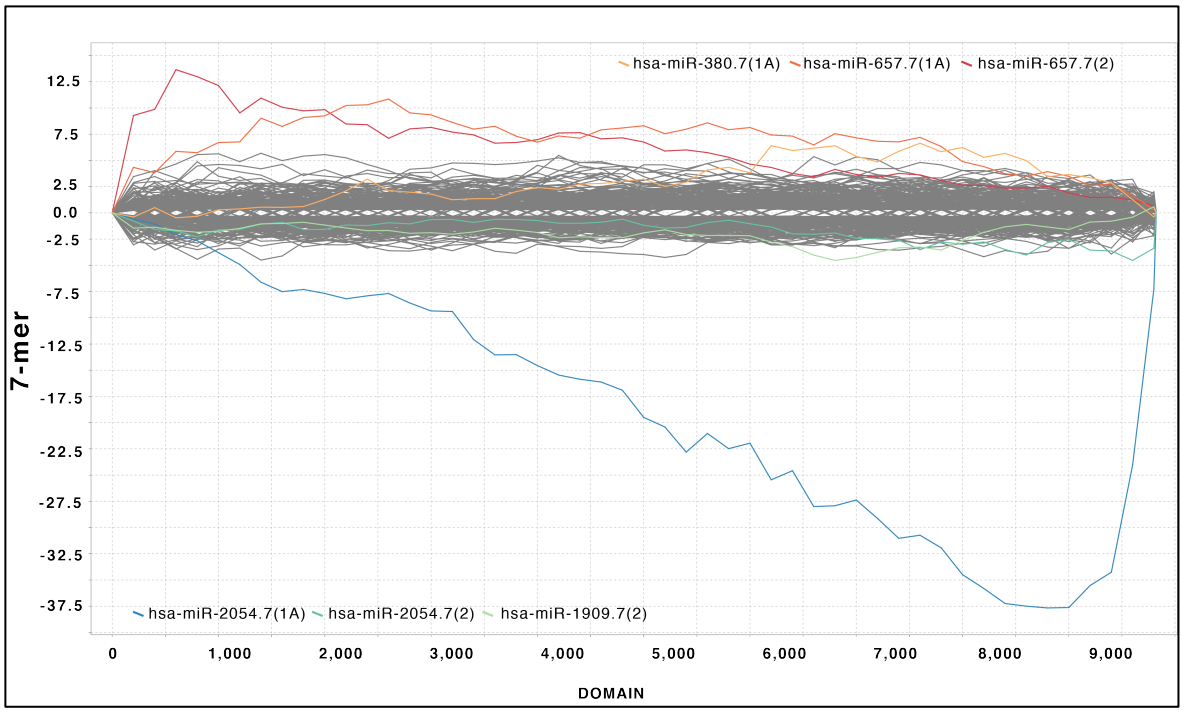


Figure 40 - RNA-sequencing analysis of transcriptional effects induced by miR-657. Sylarray analysis of the SUM159 miR-657 dataset. Seed enrichment analysis was performed using the Sylarray algorithm, which performs an unbiased search for miRNA binding sites (MREs) in the 3'-UTR of the sequenced genes. The graph reports the distribution of MREs in the RNA-seq dataset generated from SUM159 cells transfected with the miR-657 oligo, 24 h post-transfection. On the x-axis, the sequenced genes are ordered from the most downregulated to the most upregulated genes; on the y-axis, corresponding enrichment score of the 7mer MREs for each sequenced gene, expressed as a P-value. The three miRNAs with the most significantly enriched MREs among: 1) the upregulated genes, indicative of downregulated miRNAs, are shown in shades of blue (below x-axis); 2) the downregulated genes, indicative of upregulated miRNAs, are highlighted in shades of red.

3.8.4 Identification of miR-124 regulated genes and pathways involved in the control of mammary stem cell self-renewal

Next, we identified the miRNA-regulated transcripts by comparing the gene expression regulation between miRNA-transfected samples and controls (i.e., SCR *vs.* miR-124 and SCR *vs.* miR-657) using a statistical tool (DeSEQ2, (Anders and Huber,

2010)). We selected only genes that showed an upregulation or downregulation greater than |0.2| $\log_2$fold change with a false discovery rate (FDR) < 0.01. We identified a list of 769 genes regulated upon miR-124 overexpression (455 downregulated and 314 upregulated), and a set of 182 genes regulated upon miR-657 overexpression (130 downregulated, 52 upregulated), as shown in the heatmap reported in Figure 41A. To reveal the function of the miR-124 down-regulated genes, we also performed a functional analysis on those genes (Figure 41B) and found out that a great number of them are involved in eight different cellular functions: self-renewal, cellular homeostasis, proliferation of cells, expression of DNA, cell movement, migration of tumor cells, cell movement of tumor cells and migration of cells. Notably, these different functions can be classified in three main categories: i) self-renewal; ii) proliferation/apoptosis; iii) migration, suggesting that miR-124 could directly act on these mechanisms to exert its anti-tumoral function.

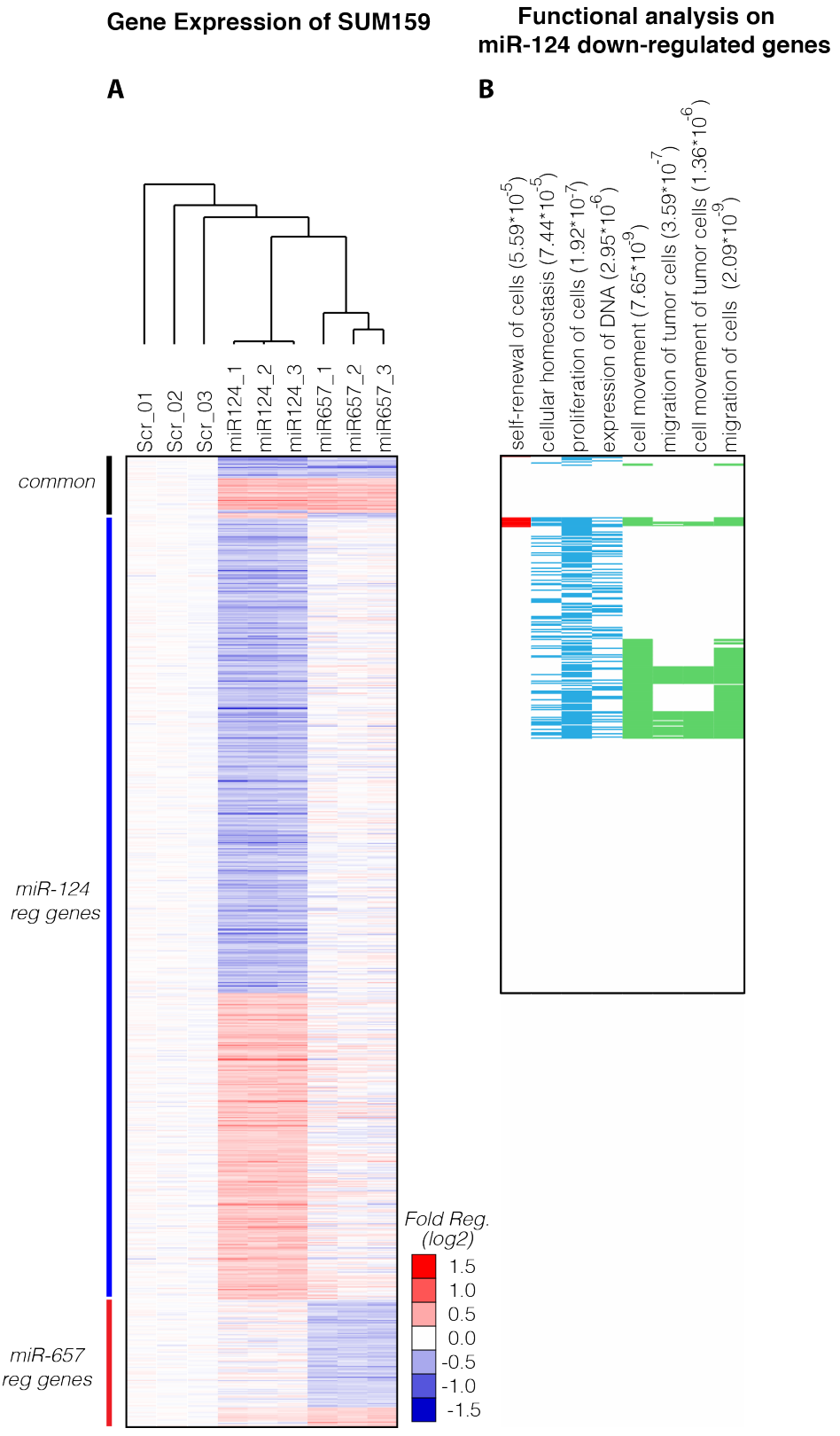


Figure 41 – Identification of miR-124 and miR-657 target genes regulated upon overexpression of these miRNAs in SUM159 cells. A) Gene expression data were analysed using the DeSEQ2 statistical

tool. Genes DISPLAYING an upregulation or downregulation greater than |0.2| log_2fold change, with a false discovery rate (FDR) < 0.01 were considered. The heatmap shows the target genes regulated upon miR-124 overexpression (middle), miR-657 overexpression (lower part) and common to both miR-124 and miR-657 overexpression (upper part). The colour intensity reflects the log_2fold regulation *vs.* SCR oligos: red indicates maximum upregulation, blue maximum downregulation. B) Functional analysis of miR-124 downregulated genes. For each function, a p-value annotation enrichment is reported in brackets.

To reveal critical hubs affected by the two miRNAs and the signaling pathways to which they belong, we performed an upstream Ingenuity Pathways Analysis (IPA) in collaboration with Dr. Matteo Marzi in our lab. We focused initially on miR-124, since it showed the most consistent transcriptional effects in the RNA-seq analysis. We performed a "Core Analysis" by IPA using the 769 regulated genes above-mentioned as the query gene list (for details see "Material and Methods"). We searched for all the possible gene sets that significantly overlap with the query dataset, including gene sets associated with gene ontology (molecular functions or biological processes), molecular pathways, upstream regulators, and disease-associated signatures (such as cancer). The analysis also implements a series of statistical calculations to measure the significance of the overlap identified. For significantly associated pathways, IPA analysis assigns a Z-score, which reflects the type and the strength of regulation. If the pathway is activated the Z-score is positive, if it is inhibited the Z-score is negative. The higher the value of the Z-score (absolute value), the higher the level of regulation of the pathway (activated or inhibited). Results from the IPA analysis revealed that several pathways are significantly associated with miR-124 transcriptional changes (Figure 42). These pathways can be

grouped into three main effects: inhibition of stem cell self-renewal, inhibition of cell migration, and cell cycle/apoptosis regulation (Figure 42).

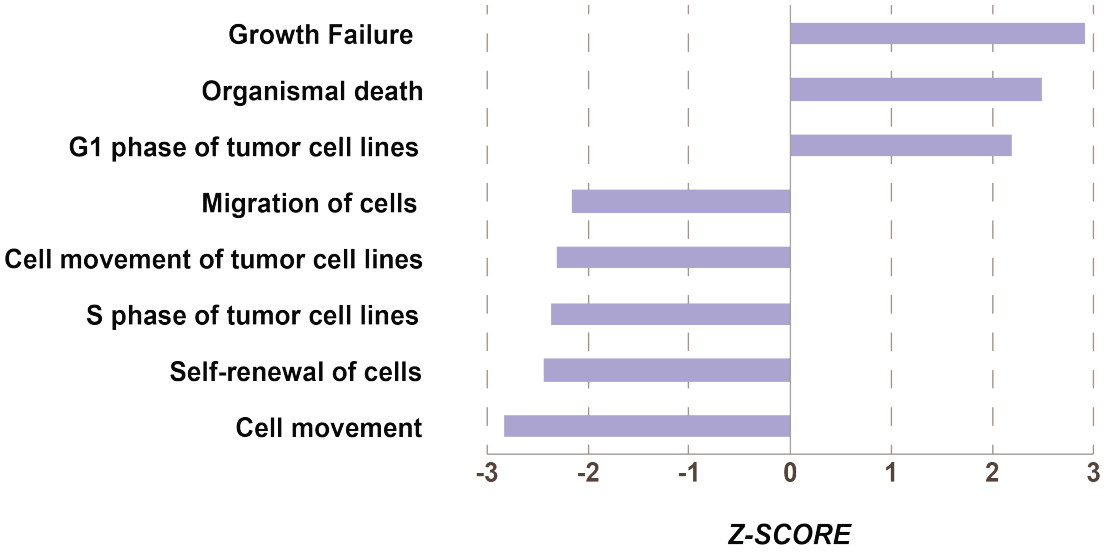


Figure 42 – Analysis of the transcriptional effects induced by miR-124 overexpression by IPA. We used the list of 769 regulated genes upon miR-124 overexpression as the input for the IPA-based core analysis. The most significantly affected pathways (z-score> |2| and with a p-value> 10^{-3}) upon miR-124 overexpression are reported in the graph: a negative Z-score indicates downregulation, while a positive Z-score indicates upregulation. Three main functional categories were distinguished, related to self-renewal (self-renewal of cells), migration (migration of cells, cell movement of tumor cells, cell movement) and cell cycle/apoptosis regulation (growth failure, organismal death, G1 phase of tumor cell lines, S phase of tumor cell lines).

We next focused on the "self-renewal of cells" pathway, which fits with our original aim, and generated a network comprising 13 stem cell genes found in the miR-124 overexpression signature (Figure 43). This analysis revealed that the majority of the genes that induce self-renewal, such as JAG1, STAT3, SMAD5, GLI3, are downregulated by miR-124. Indeed these genes are associated to relevant stem cell self-renewal pathway such us Notch (through Jag1), Hedgehog (through GLI3), non-

canonical TGF-β (through SMAD5) or inflammation pathways that converge to STAT3.

Overall, the effect of miR-124 is to largely inhibit self-renewal through the repression of several hubs of stem cell self-renewal.

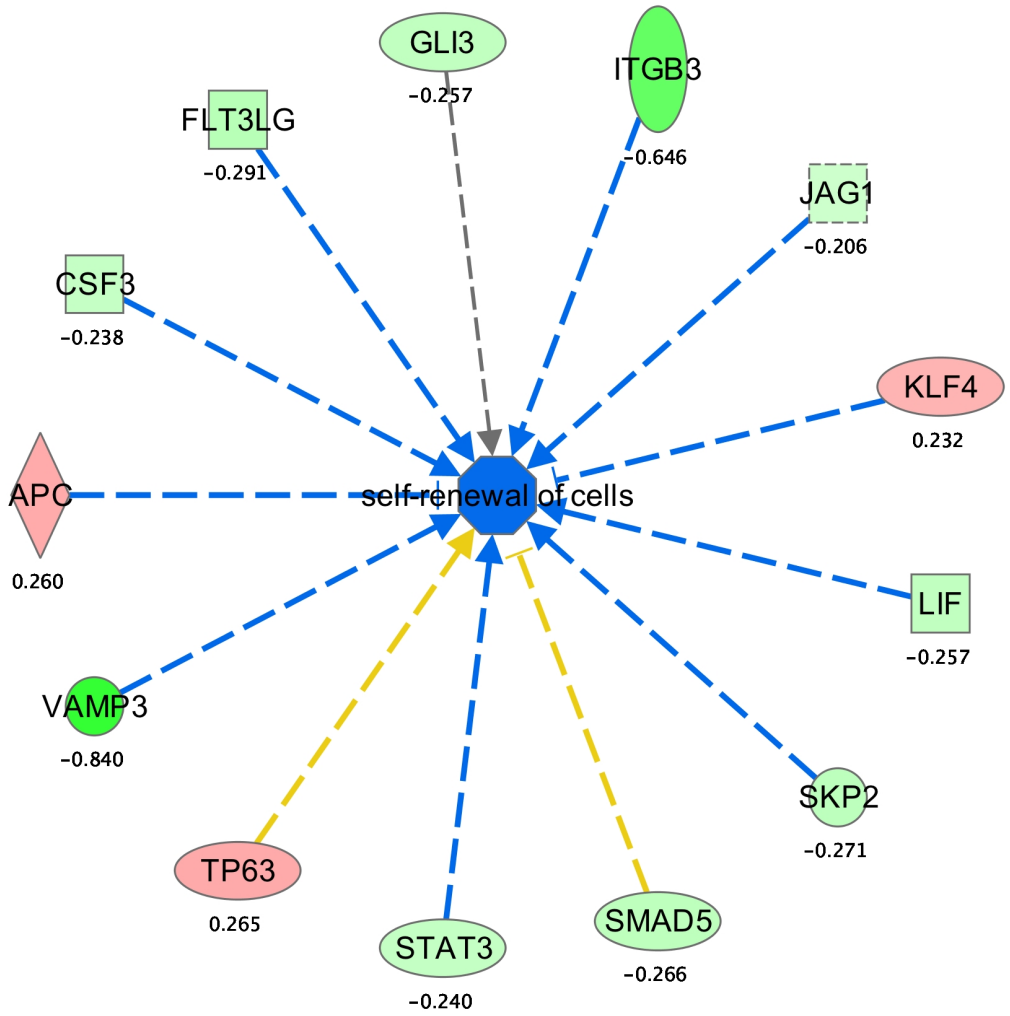


Figure 43 – Analysis of miR-124 regulated genes belonging to the "self-renewal of cells" pathway. The image shows a network of 13 miR-124 regulated genes belonging to the IPA "self-renewal of cells" pathway (see Figure 32). The transcriptional effect induced by miR-124 is indicated by a color code (red, induction of expression; green, repression. Higher is the intensity of color, higher is its regulation) and a value which corresponds to the log_2fold regulation. The arrows in the network are also colored to indicate whether the effect on gene expression was "consistent" (blue) or "inconsistent" (yellow) with the overall effect of self-renewal inhibition.

We reasoned that the transcriptional changes induced by miR-124 treatment might be related to stem cell-associated gene expression programs. Thus, we retrieved from the literature, gene expression signatures related to breast stem cell enriched populations (Stem cell-signatures) isolated with different approaches. These included: 1) a signature derived from human normal and breast cancer stem cells isolated as lin-$CD44^{+}/CD24^{-}$ cells, hereafter referred to as the "Polyak" signature (Shipitsin et al., 2007); 2) a signature derived from mammary repopulating units (MRU) isolated from murine normal mammary tissue as $CD49f^{high}/CD24^{med}$ cells, hereafter referred to as the "Stingl" signature (Stingl et al., 2006); 3) a signature derived from mammary stem cells (MaSCs) isolated from murine normal mammary tissues as $Sca1^{-}/CD49f^{hi}/CD29^{hi}/CD24^{+}$ cells, hereafter called the "Visvader" signature (Lim et al., 2010); 4) a signature derived from quiescent PKH26 label-retaining cells isolated from human normal primary mammospheres, hereafter referred to as the "Pece" signature (Pece et al., 2010). Since two signatures were derived from murine tissues (Stingl and Visvader), we considered only those genes conserved between human and mouse for our analysis.

We used a computational tool known as Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) to analyze the correlation of each stem cell-signature (gene set) with the gene expression changes that occurred upon miR-124 overexpression. Briefly, each gene set is ordered according to the regulation of a query dataset (in our case, the miR-124 or miR-657 dataset) from the most upregulated to the most downregulated genes. The negative hypothesis (non correlation) implies that genes are distributed randomly in the query dataset. However, if two datasets are correlated (positively or negatively), the distribution will be asymmetric. The overlap between the gene sets is expressed by the normalized enrichment score (NES) with a FDR (q-

value) as a measure of the statistical significance of the correlation. Of note, a positive NES indicates gene set enrichment at the top of the ranked list; a negative NES indicates gene set enrichment at the bottom of the ranked list. The GSEA analysis protocol suggests that an enrichment is significant when the FDR is at least less than 0.25 (cut-off value), meaning that the result is likely to be valid 3 out of 4 times.

We performed the GSEA analysis on the miR-124 dataset, splitting each stem cell-signature in two, according to their regulation in the original dataset: i) UP, gene induced in stem cells; and ii) DOWN, gene repressed in stem cells. We observed that two of the stem cell-signatures were significantly correlated with the miR-124 dataset, in particular, the "Polyak UP" (NES = -1.51, FDR = 0.03) and the "Visvader UP" (NES = -1.26, FDR = 0.18) signatures (Figure 44). This result suggests that the expression of miR-124 could inhibit the expression of genes specifically involved in sustaining and maintaining stem cell functions ("UP" signatures). No significant enrichment (FDR > 0.25) was observed for other two SC signatures, Stingl and Pece (Figure 45).

Taken together, these analyses confirmed that miR-124 negatively regulates a transcriptional program associated with the acquisition of stem cell traits and self-renewal.

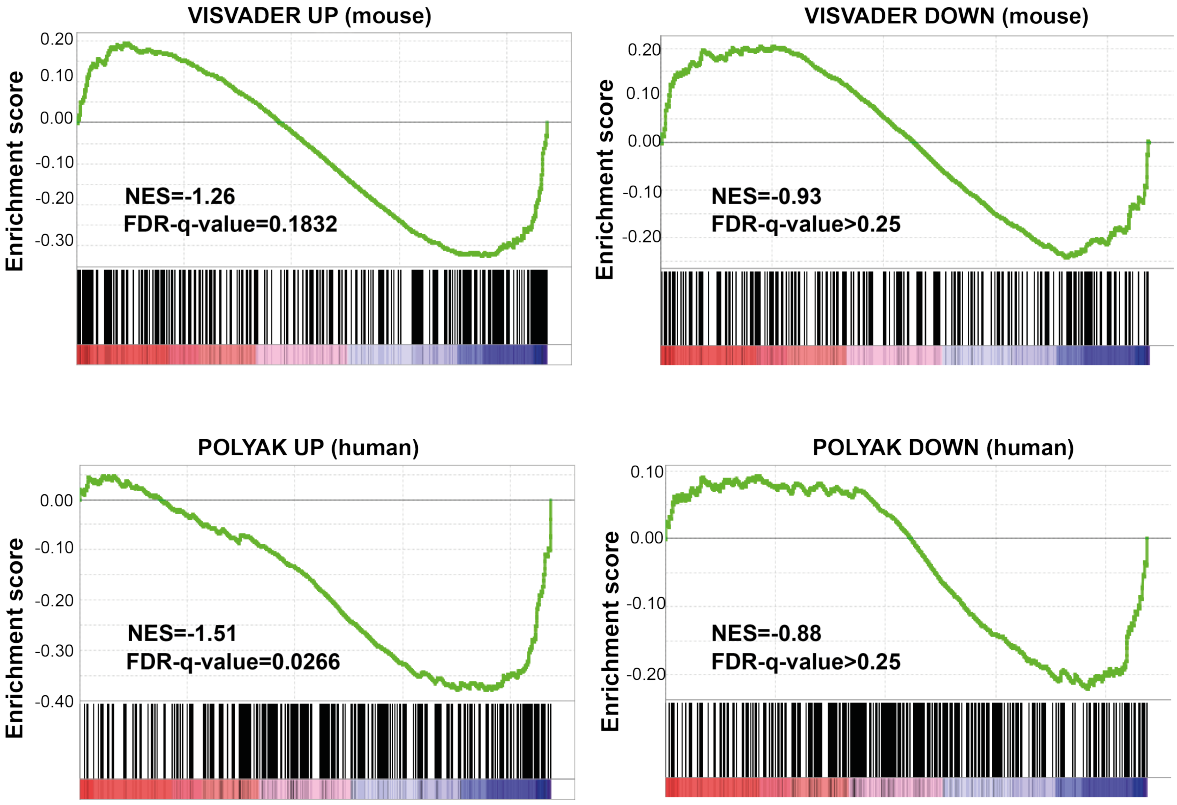


Figure 44 - GSEA analysis of the miR-124 dataset against the Visader and Polyak mammary stem cell signatures. GSEA was used to correlate stem cell signatures with the gene expression changes observed upon miR-124 overexpression in SUM159 cells. The enrichment plot together with the normalized enrichment score (NES) and significance (FDR q-value) are shown for the Visvader (UP and DOWN) and the Polyak (UP and DOWN) signatures used as gene sets. The enrichment was considered significant when the FDR was < 0.25.

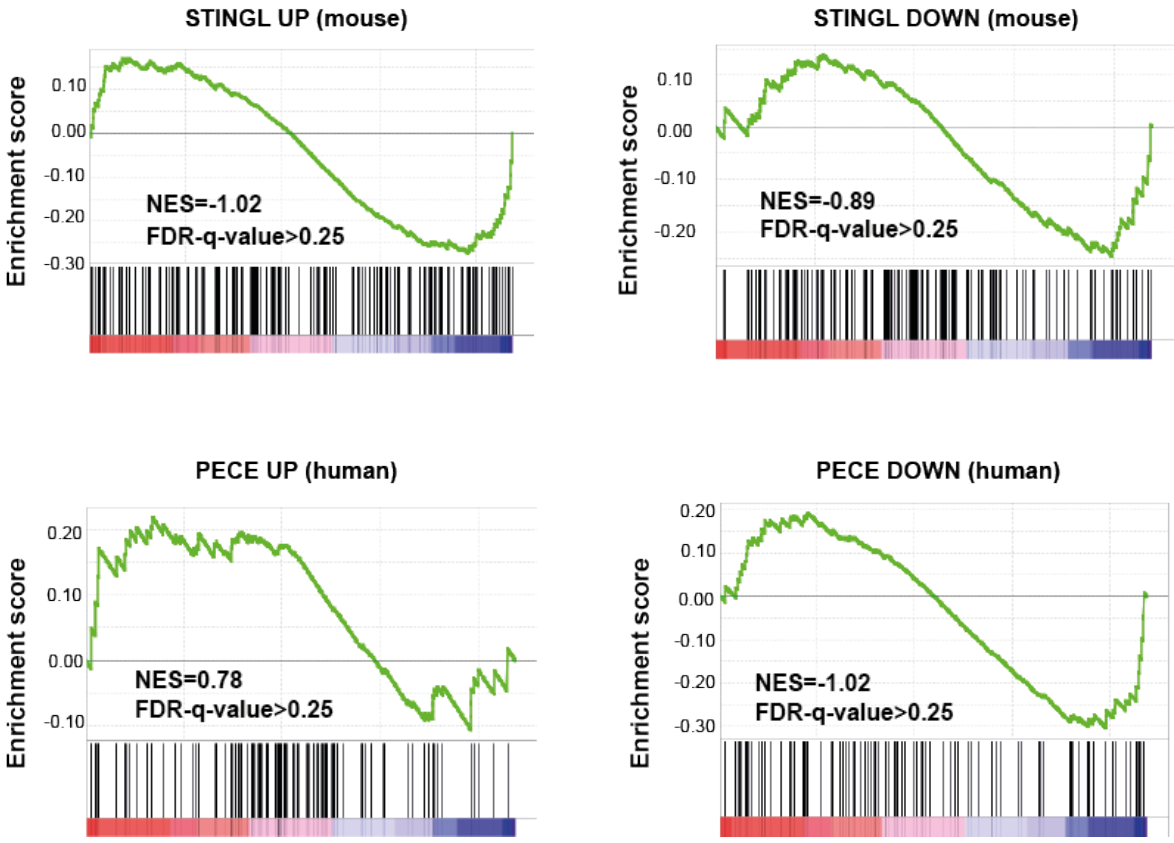


Figure 45 - GSEA analysis of the miR-124 dataset against the Stingl and Pece mammary stem cell signatures. GSEA was used to correlate stem cell signatures with the gene expression changes observed upon miR-124 overexpression in SUM159 cells. The enrichment plot together with the normalized enrichment score (NES) and significance (FDR q-value) are shown for the Stingl (UP and DOWN) and the Pece (UP and DOWN) signatures used as gene sets. The enrichment was considered significant when the FDR was < 0.25.

We, next, set out to identify those genes belonging to stem cell pathways that are directly targeted by miR-124. Hence, we retrieved the complete list of predicted miR-124 targets using the two most widely used prediction algorithms, namely *Targetscan6.2* and *miRanda*. These two algorithms use quite different metrics to calculate the candidate targets, such as the presence of conserved seed sequences that bind with perfect complementarity to the 3'UTR of mRNAs (TargetScan) or free energy binding contribution to non-perfect complementary binding (miRanda). We

obtained 3482 miR-124 targets by Targetscan and 3156 miR-124 targets by miRanda; these targets were largely overlapping. We filtered these target predictions using the genes regulated upon miR-124 overexpression, and obtained 279 genes predicted by both Targetscan and miRanda that were also transcriptionally downregulated by miR-124 overexpression in SUM159 cells (Figure 46).

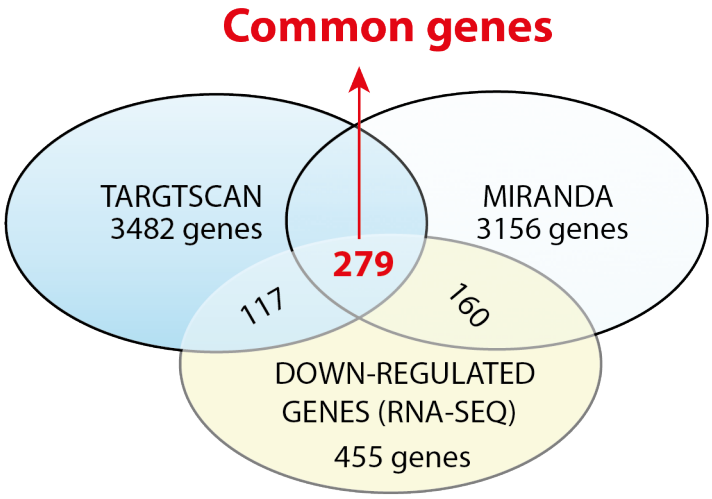


Figure 46 - Identification of miR-124 putative target genes. The Venn diagram shows the strategy used to identify miR-124 putative target genes by combining predictive algorithms (TargetScan6.2 and miRanda) with the transcriptional effects observed upon miR-124 overexpression in SUM159 cells by RNA-seq. A group of putative common target genes (279) was identified, which comprised more than 50% of all downregulated genes following miR-124 overexpression.

Among these 279 "common" genes, we searched for those that could be related to stem cells functions, such as: i) genes belonging to any of the four stem cell signatures; ii) genes belonging to a home-made signature obtained by comparing SUM159 grown in non-adherent conditions (3D vs. adherent conditions (2D). iii) genes belonging to the "self-renewal of cells" function by IPA. We obtained a list of 35 genes associated with mammary stem cells and self-renewal function that could explain the phenotype induced by miR-124 overexpression (Figure 47). Among them, a

top candidate is JAG1, a gene concomitantly regulated in all the above-mentioned analysis, which has already been identified as a validated miR-124 target in neurogenesis (Baud et al., 2011; Cheng et al., 2009).

Of course, further experiments are needed to establish whether the activity of miR-124 is effectively mediated by JAG1 regulation or also involves other stem cell-related targets. Moreover, an extensive validation of the tumor-suppressive effects of miR-124 on primary cancer stem cell (CSC)-rich human tumors is also required.

GENE SYMBOL	POLYAK	VISVADER	STINGL	PECE UP	3DVs2D	IPA
JAG1	YES	YES	YES	YES	YES	YES
AMOTL1	YES	YES	YES			
COL4A1	YES	YES	YES		YES	
CTGF	YES	YES	YES		YES	
GSN	YES	YES			YES	
ITGB1	YES	YES				
LAMC1		YES	YES			
SNAI2		YES	YES			
ADAMTS9	YES				YES	
B4GALT1	YES					
C3orf58		YES				
CCND2		YES				
CLDND1	YES					
CTDSP1	YES					
EGR1	YES				YES	
ELK3		YES			YES	
GLIS2	YES					
HDAC4		YES				
MYADM	YES				YES	
NRP2		YES			YES	
OAF	YES					
P4HA1			YES		YES	
QKI		YES				
RBMS1			YES		YES	
REEP3			YES			
SPHK1		YES			YES	
SSH1		YES			YES	
TSKU		YES			YES	
WIPF1		YES			YES	
ZFP36L2	YES				YES	
GLI3						YES
ITGB3					YES	YES
SMAD5						YES
STAT3						YES
VAMP3						YES

Figure 47 - Identification of miR-124 stem cell target genes: the "35-gene list". The list shows the 35 putative direct miR-124 target genes related to mammary stem cells or self-renewal. The list was obtained by combining the 279 common regulated genes (Miranda + Targetscan+ RNA seq) with the stem cell signatures (Polyak, Visader, Stingl, Pece UP) and IPA "self-renewal of cell" pathway. The table shows the list of 35 genes (gene symbol) and the signatures which they belong to (indicated by 'YES'). A fifth in-house list was also included (3D *vs.* 2D), which was obtained by analyzing the

transcriptional changes measured in 3D wild type (WT) SUM159 mammospheres *vs.* 2D WT SUM159 cells. The master regulators identified by IPA were included.

4 DISCUSSION

4.1 Phenotype screening to unravel miRNA functions in complex phenotypes

Emerging evidence indicates that, in the breast cancer, a rare subpopulation of cells within the tumor bearing stem cell (SC)-like properties, named CSCs, is responsible for the degree of aggressiveness of the tumor (Pece et al., 2010), as well as the emergence of therapeutic resistance and disease relapse (Liu and Wicha, 2010; Visvader and Lindeman, 2008). Therefore, we reasoned that the identification of key molecules able to regulate the properties of CSCs is of primary importance to our understanding of the underlying mechanisms of "stemness" and to the development of novel therapeutic approaches to fight even the most aggressive and resistant forms of cancer. miRNAs are extremely attractive candidates for molecules that could have a key role in the regulation of "stemness" properties. They are critical hubs in gene expression networks that control cell fate specification and stem cell properties. Indeed, a growing body of evidence supports a role for miRNAs as critical regulators of self-renewal for many stem cell types, including adult tissue stem cells and CSCs (Greene et al., 2010b) (Xu et al., 2009). Nevertheless, research on miRNAs and their relationship to MaSCs is still in its infancy. In particular, the identity of miRNAs that regulate "stemness" and differentiation in the breast, and their related cellular pathways are still largely unknown. Equally importantly, we do not know whether it is possible to harness miRNA biological action to control CSC expansion and, therefore, neutralize the most aggressive forms of breast cancer.

To improve our knowledge on miRNAs that are able to specifically act on BCSCs, we decided to quantitatively and systematically characterize individual miRNAs of the human genome by performing an unbiased genetic screening. Our approach was

based on the assumption that individual miRNAs could alter the "stemness" properties of breast cancer cells and, hence, be selected (either positively or negatively) in a stem cell phenotype screening assay.

Compared to other approaches, phenotype screenings have several advantages. Firstly, they are unbiased and no prior knowledge on the function of specific miRNAs is required. The selection is based on the "pure" ability of miRNAs to alter the transcriptome in such a way that cells acquire a new property, or improve an existing property, and are therefore positively selected. Conversely, miRNAs could inhibit an existing property and confer a disadvantage to cells, thereby resulting in their negative selection. Of course, the stronger the effect of the miRNA, the more rapid and powerful the selection will be.

Another advantage of the screening approach is that it is possible to screen almost all of the human miRNAs in one experiment due to the relatively low complexity of the miRNA genome. Moreover, since miRNAs are enriched for critical regulators of cell fate and stem cell properties, the probability of scoring key regulators of these properties is relatively high. Furthermore, miRNAs are themselves appealing therapeutic molecules. Indeed, the potential usefulness of a microRNA-based therapy in cancer is now being exploited, with the attempt to modulate their expression, reintroducing microRNAs lost in cancer or inhibiting oncogenic microRNAs by using anti-microRNA oligonucleotides (Li and Rana, 2014) (Chen et al., 2010) (Esposito et al., 2014). Hence, in addition to providing information on the mechanisms that control "stemness" properties in tumors, our screening could also highlight microRNA-candidate therapeutics.

4.2 A stem cell self-renewal phenotype screening strategy to target BCSCs

A major restrain when studying cancer stem cells is their rareness and their hard manipulation, so that a critical issue in setting up our phenotype screening strategy was to employ a suitable biological assay.

To date, the only approach to enrich breast cancer stem cells *in-vitro* is culturing cells as mammospheres. Indeed, it was shown that rare, single-founder stem cells can form multicellular sphere structures under serum-free suspension conditions that are enriched for stem and early progenitors cells (Dontu et al., 2003).

As demonstrated by our lab and others, this assay reproduces the behavior of normal and cancer stem cells *in-vitro*: i) it selectively measures the clonogenic ability of cells with stem cell-like properties, i.e., those able to withstand anoikis and to generate actively proliferating and differentiating progenitors that constitute the bulk of the sphere; ii) it measures the self-renewal potential of such cells upon the serial propagation of sphere cultures (Dontu et al., 2003) (Pece et al., 2010). We, therefore, reasoned that miRNA selected with this assay might influence either the ability of stem cells to survive or the ability of stem cells to self-renew.

For samples derived from human breast cancers, the mammosphere forming ability correlates with the CSC content of the tumor, established by *in-vivo* transplantation assay (Pece et al., 2010). Moreover, the ability of breast cancer cells to be serially propagated as mammospheres *in-vitro* correlates with their self-renewal potential determined *in-vivo*. This result suggests that the serial propagation of mammospheres can be considered as an *in-vitro* proxy for the content and self-renewal ability of CSCs ((Pece et al., 2010) and Pece personal communication). For these reasons, we strongly believe that the identification of key molecules able to regulate self-renewal mechanisms of CSCs will provide fundamental insights into tumorigenesis, in turn

leading to efficient methods to control stem cell self-renewal for cancer therapy.

Here, we exploited the enrichement of CSCs in mammospheres to establish a high-throughput unbiased pooled screening system suitable to identify microRNAs that are specifically involved in the regulation of those rare cells CSCs.

Indeed, by combining the extremely powerful genome-wide screening with the challenging and sensitive next-generation sequencing methods, we could uncover microRNAs capable of affecting the self-renewal ability of breast cancer stem cells. As far as we know, this is the first whole-genome screening approach specifically performed on a stem cell population, enriched by exploiting the arduous mammosphere assay, which lead to the identification of possible microRNA therapeutic molecules that could specifically target breast cancer stem cells.

Taken together, our findings confirm the utility of the presented screening strategy to identify processes with specific relevance to CSCs and also, our results gave an insight into the molecular mechanisms that govern CSC properties through the control of microRNA molecules.

4.3 Set-up of the screening system

The main drawback of phenotype screenings is that they are technically challenging, in the sense that a number of steps have to be improved and optimized to obtain a high enough reproducibility, robustness and reliability of results. As described in detail in the results section of this thesis, we performed an extensive optimization of almost all the steps of the screening, trying to keep to a minimum, confounding factors (such as those due to multiple integration events or transgene silencing). Whenever possible, we performed pilot experiments to establish the feasibility of the procedure or to assess the effect of modifications of the protocol (e.g. selection of

GFP^+ cells in the assay). Indeed, one of the main aims in this optimization was to avoid the selection of stochastic events. This means that the driver of the positive or negative selection of the miRNA is not the miRNA function, but a random event already present in the initial cell population (cancer cells are heterogeneous) or generated by the procedure (e.g. by transgene integration events).

To circumvent this problem, we devised two strategies: i) the screening was duplicated and only results obtained in the two independent experiments were considered meaningful; ii) we included in the analysis only those miRNAs that were significantly represented in the initial population, so that their positive or negative selection is the result of miRNA-driven effects, rather than random effects. Regarding this latter point, for each miRNA we calculated its *"Coverage"* (C), i.e., number of cells in the initial population that, as a result of independent integration events, express the particular miRNA. In our analysis, we included only those miRNAs with a coverage above a certain threshold.

The choice of the threshold was absolutely arbitrary, although it was immediately clear that the higher the coverage the more robust the effect of the miRNA in the screening assay. However, a high coverage has the disadvantage that fewer miRNAs will be considered in the screening. We therefore decided for a coverage threshold of $C>100$, which allowed the analysis of 150 different miRNAs from the pool of 650 constructs present in the library. As we mapped about 500 miRNA precursors in our initial population, this means that overall we discarded results for almost 350 miRNAs that were poorly represented in the initial transfected population. However, the distribution of results in the two replicas was quite reproducible for the 150 clones with $C>100$, while it was much more variable for the pool of 350 clones with $C<100$ (data not shown), suggesting that stochastic events were taking place when coverage

was below our threshold. Nonetheless, it could be interesting to select a few of the 350 discarded miRNAs to see if any of them are positively and negatively selected, and eventually compare their validation efficiency with that of the hit candidates selected from the C>100 pool.

Another limitation of the screening assay is that its results are very much dependent on the cell model used. Cancer cell lines bear specific mutations/genetic alterations that reproduce only some of all the possible alterations occurring in the breast cancer; hence, screening results likely reflect the modifiers related to those specific genetic alterations. To circumvent this limitation, we also considered further alternative breast cancer cell lines (i.e., luminal MCF7 and metastatic MDA-MB-231), thus encompassing multiple genetic backgrounds. However, the screening assay was not feasible with these alternative models, mainly due to a low efficiency of mammosphere formation (which negatively impinges on the coverage) and to the requirement of cell type-specific optimizations. For these reasons, the screening was performed on SUM159 cells, which have the desired qualities needed to ensure the feasibility of the screening assay. Firstly, SUM159 cells display a remarkably high efficiency of mammosphere formation (SFE > 10%), meaning that the frequency of cells with stem cell properties is high, in line with previous findings from other labs (Fillmore and Kuperwasser, 2008). Secondly, SUM159 cell reproduce one of the most adverse subtypes of breast cancer, the claudin-low subtype harboring mutant p53 and lacking the ER, PR and HER2 receptors; this breast cancer subtype typically has a high number of CSCs with immortal behavior (Visvader and Stingl, 2014). Thirdly, SUM159 cells are one of the few cell lines in which conversion from CSCs to non-CSCs (and *vice versa*) has been observed and shown to rely on a specific transcriptional program (Gupta et al., 2011) (Scheel et al., 2011) (Chaffer et al.,

2013). Thus, in our screening assay it is possible that a miRNA could influence this transcriptional program directly.

Another technical issue of the screening assay relates to the analysis of gDNA and clone deconvolution. Two approaches could be used to count miRNA clones in the initial transfected cell population and over the different passages, in order to reveal those that were positively or negatively selected during the assay: i.e., microarray analysis or NGS. Microarrays are widely used for genome-wide experiments, in particular, for transcriptome analysis or the identification of genetic alterations/variants; but these kind of arrays cannot be used for clone deconvolution, which requires custom-made arrays as described by other groups (Izumiya et al., 2010) (Tsuchiya et al., 2011). Moreover, microarrays have a limited specificity and sensitivity, and a poor dynamic range in quantification experiments. In contrast, NGS could be easily applied to clone deconvolution, as it generates qualitative and quantitative information with an incredibly high throughput. Furthermore, due to the possibility of combining multiple samples in just one sequencing experiment (multiplexing), costs would be massively reduced and technical variability would be minimized with NGS.

4.4 Positive and negative miRNA modifiers of CSCs

The screening provided reliable results for 152 miRNAs, which were significantly represented (C>100) in the initial population (P0) in the two biological replicas. This pool of miRNAs was further analyzed in the parallel 2D and 3D assays to establish which miRNAs may influence the behavior of cancer cells in respect to proliferation of cells in adherent condition (2D) and self-renewal of cells in non-adherent condition (3D). Overall we noticed that most miRNAs had a strong negative effect in the 3D assay, while being neutral or have a slightly negative effect in the 2D assay. The

simplest (and more likely) interpretation is that cells grown in condition selective for CSCs (3D) are more sensitive to even the slightest transcriptional interference (produced by the expression of any miRNA) than cells simply grown in adherent condition. Nevertheless, we were able to identify some clones that were significantly enriched or depleted in both 3D replicas, potentially inhibiting or increasing CSC self-renewal, respectively. We identified up to 20 candidate miRNAs, most of which (18 miRNA) depleted plus a couple of miRNAs that were enriched. This uneven distribution of miRNA candidates (depleted miRNA clones > increased miRNA clones) likely reflects specific biological properties of miRNAs. Indeed, multiple experimental evidences suggest that most miRNAs tend to have a tumour suppressor function (Kumar et al., 2007). Accordingly, miRNAs that could inhibit the growth and/or self-renewal of CSCscould be considered with tumour-suppressive function. An alternative possibility is that, given the high CSC content of SUM159 cells, it is easier to find modifiers that decrease rather than increase the efficiency of mammosphere formation.

A number of so-called “stem-inhibitor” miRNAs have been already identified, and some even in the breast. For instance, let-7 and miR-200c, whose levels are typically low in CSCs with respect to non-CSC cells, could both inhibit CSC growth and self-renewal by targeting critical cell fate determinants, such as BMI (Shimono et al., 2009) or HMGA2 (Yu et al., 2007). Notably, our screening was able to confirm these data: both let-7a-1 and miR-200a were found to be consistently depleted in the 3D mammosphere assay. In particular, let-7 is among the 20 candidates, as it possesses a coverage value>100 (see bubble plot 3D vs 2D with C>100), miR-200a is not present within the 20 candidate list, but is included in the negative-selected clones when we consider a lower coverage cut-off (see bubble plot 3D vs 2D with C>50).

Combining the results of the 3D with the 2D experiments, we could stratify the 20 candidate miRNA modifiers into three groups: i) a group of 3 common miRNAs, miR-1-1, miR-124-1 and miR-889, which were negatively selected in both assays; ii) a group of 15 miRNAs which had a stronger effect in the non-adherent (3D) conditions compared to the 2D ones; iii) 2 enriched miRNAs specific for the 3D assay. These results suggested that proliferation and CSC growth/self-renewal likely rely on distinct mechanisms, targeted by different miRNAs. The few miRNAs that were in common (miR-1; miR-124 and miR-889) could either have independent functions on both processes or could impact on self-renewal through proliferation. Of course, it is not possible to discriminate between these two possibilities without analyzing the effect induced by such miRNAs in individual assays. Indeed, we performed a proof-of-principle validation by analyzing candidate miRNAs in SUM159 directly. We selected 10 out of 20 miRNAs in the validation steps, of which half (5 miRNAs) displaying the strongest effects in the 3D screening (the " top hit candidates": miR-1, miR-124, miR-889, miR-657, miR-664). Overall 6 clones confirmed to have an effect on the formation and propagation of SUM159 mammospheres, suggesting that the screening had a 60% validation rate. Of note, most of the clones (3/5) of the "top hit" list confirmed a very strong effects even in this individual validation step. In particular, miR-1 and miR-124 profoundly and significantly decreased the SFE capability of SUM159 throughout the generations, with miR-124 showing an almost complete block in sphere formation capability. Micro-RNA-1 is a muscle-enriched miRNA that inhibits proliferation of progenitor cells and promotes myogenesis. In lung and prostate cancers, miR-1 has been reported as a tumor suppressor miRNA, being downregulated in tumor cells and able to inhibit cancer cell growth when overexpressed (Hudson et al., 2012; Nasser et al., 2008).

Indeed, miR-1 could target many cell cycle genes plus (Zhao et al., 2007) the oncogene MET, a receptor tyrosine kinase, which is frequently upregulated in tumors and associated to migration, metastasis and epithelial-to-mesenchymal transition (EMT) (Nasser et al., 2008).

miR-124 is an important regulator of neuronal differentiation during neurogenesis, where it controls the expression of the self-renewal marker Sox9 (Cheng et al., 2009). In cancer, miR-124 appears to be a potential tumor suppressor miRNA, because it is downregulated in multiple cancer types, such as hepatocellular, endometrial, colorectal and prostate carcinomas, and glioma (Lu et al., 2013) (Lang, 2012)(Li et al., 2014) (Wang et al., 2013) (Shi et al., 2013) (Xia et al., 2012) (Liang et al., 2013). In some cancer cells, the expression of miR-124 inhibited tumor growth both *in vitro* and *in vivo* (Lu et al., 2013). In hepatocellular carcinoma, miR-124 was found to target important oncogenic pathways, such as the signal transducer and activator of transcription 3 (STAT3) pathway (Lu et al., 2013) and the PI3K/Akt pathway through the phosphoinositide 3-kinase catalytic subunit alpha (PIK3CA) (Lang, 2012 #1173). miR-124 also targets the androgen receptor (AR) in prostate cancer, which in turn induces an upregulation of the tumor suppressor p53 (Shi et al., 2013), whereas in glioma, miR-124 targets SNAI2, thereby impairing tumorigenicity and neurosphere formation (Xia et al., 2012). In breast cancer, miR-124 expression is inversely correlated with histological grade, and has been described to inhibit EMT by targeting SLUG in the aggressive breast cancer cell lines, MDA-MB-231 and BT-549 (Liang, 2013). These data suggest an important role for miR-124 as a strong tumor suppressor molecule with anti-metastatic potential and as a possible biomarker for diagnosis.

In contrast to miR-1 and miR-124, miR-657 enhanced the sphere forming capability of SUM159 cells over several passages suggesting that it could act as an oncogenic

molecule. Intriguingly, miR-657 is a poorly characterized miRNA, reported to have altered expression in cancers of the cervix (Ding, 2014 #1180) and a potential oncogenic function in liver cancer (Zhang et al., 2013a).

Three other miRNAs (miR-144, miR-153 and miR-204) were also validated, showing a significant albeit limited effect. The expression of miR-144 has been found to be downregulated in thyroid cancer and its restoration is able to suppress invasion and migration *in vitro* by targeting the E-cadherin suppressors, ZEB1 and ZEB2 (Guan et al., 2014). Moreover, in nasopharyngeal carcinoma, miR-144 was found to promote cell proliferation, migration and invasion, and to repress the tumor suppressor PTEN (Zhang et al., 2013b). miR-153 has been shown to be significantly downregulated in lung cancer, and its overexpression in these cancer cells inhibited their proliferation, migration and growth as xenograft tumors *in vivo* (Yuan et al., 2014). Finally, miR-204 was found to be downregulated in gastrointestinal tumor tissues and its overexpression in these tumor cells inhibited cell proliferation *in vitro* (Zhang et al., 2014).

It is worth mentioning that even the clones that were not confirmed in the validation, displayed effects on SUM159 cells that could explain why they were positively or negatively selected in our screening assay. For instance, miR-664, a miRNA positively selected in the screening, did not increase the SFE of SUM159 cells, but did increase the proliferation of cells in the 3D mammosphere assay over multiple generations, as evidenced by an increase in mean sphere size (data not shown). This observation can be explained by an effect of miR-664 on progenitor proliferation in 3D condition. Hence, during the 3D competition screen, the number of cells that bears miR-664 increases over time. Conversely, other miRNAs affected SFE only in the first generation of spheres (F1), but not in the following generations. Hence, these

miRNAs likely inhibit the ability of SUM159 cells to adapt to non-adherent growth conditions rather than CSC growth and self-renewal.

Taken together, our findings strongly support the use of miRNA-based phenotype competition screenings to identify potential regulators of complex biological functions. To our knowledge, this is one of the few miRNA-based screening approaches able to select both positive and negative regulators of a specific phenotype. We speculate that this kind of approach could be extended to other phenotypes, or even, to other classes of molecules, including long non-coding RNAs (lncRNAs), whose functions have still to be elucidated.

4.5 miR-124 targets multiple pathways involved in CSC maintenance

Our screening highlighted some miRNAs as promising regulators of stemness properties in breast cancer and, therefore, as potential therapeutic targets. In particular, miR-124 and miR-657, the miRNAs that most markedly inhibited or enhanced the sphere-forming ability of SUM159 cells, respectively, were the most interesting candidates.

We were interested in characterizing the mechanism through which these miRNAs influence CSC self-renewal, as observed in the SUM159 model system. Therefore, we analyzed the transcriptional changes induced upon miRNA overexpression by RNA-seq, based on the assumption that miRNA effects occur predominantly at the transcriptional level (Guo et al., 2010). A comparison of the effects of the two miRNAs on the transcriptomes of SUM159 cells revealed that they had largely independent effects, with different sets of genes being regulated by them miRNAs. As a whole, the miR-124 transcriptional effects were far larger than those of miR-657 (769 *vs.* 182 regulated genes, respectively). The observed transcriptional changes are likely the consequence of direct effects of the miRNAs, as shown by the remarkable

enrichment of genes containing the corresponding MREs, identified by Sylamer analysis. Intriguingly, the set of downregulated genes following miR-124 overexpression appeared to be enriched in miR-124 target genes. With the use of Bioinformatics tools to predict miRNA targets (miRanda and Targetscan), we identified 279 predicted miR-124 target genes among the set of 455 downregulated genes upon miR-124 overexpression. In contrast, miR-657 induced fewer transcriptional changes, involving fewer putative direct target genes than miR-124. This difference in the magnitude of the transcriptional effects of miR-124 and miR-657 might reflect the different evolutionary histories of these two miRNAs. miR-657 is a recently evolved miRNA that exists only in humans, whereas, miR-124 is highly conserved across many species and could have thus evolved many miRNA:target interactions to affect multiple genes and/or multiple pathways at the same time, as underlined by our experiments.

Functional analysis by IPA revealed at least three main biological effects associated with miR-124 transcriptional changes: inhibition of self-renewal, inhibition of migration and cell-cycle/apoptosis regulation. Each effect was associated with largely different gene sets indicating that miR-124 can inhibit multiple pathways simultaneously. This peculiar property of miR-124 might explain why it was selected both in the 3D screen (where it regulates self-renewal) and in the 2D screen (where it regulates the cell proliferation).

Regarding the self-renewal inhibition exerted by miR-124 overexpression, 13 stem-related genes were found to be under the control of this miRNA, belonging to different pathways associated with CSCs: EMT signaling, NOTCH pathway, Hedgehog, STAT3 and TGFbeta/Wnt signaling. Of note, 8 out of 13 stem cell genes were downregulated by miR-124 (such as JAG1, STAT3, SMAD5, and GLI3), in line

with the observed effect of this miRNA on self-renewal. We, thus, searched for the possible miR-124 direct targets related to stem cells and self-renewal functions and generated a list of 35 candidate genes (the "35-gene list") that could explain the self-renewal phenotype regulated by miR-124. Among these, JAG1 appears to be one of the top candidates. This gene has been shown to be a direct target of miR-124 (Baud et al., 2011) and is involved in the Notch pathway (JAG1 is a NOTCH receptor ligand), which is one of the pathways activated by the stem cell to tightly control self-renewal and the balance between asymmetric/symmetric cell division in human MaSCs (Dontu et al., 2004).

Given the fact that the MaSC is a complex system in which individual pathways interact extensively to maintain the delicate balance between self-renewal, proliferation and differentiation (Izrailit and Reedijk, 2012; Liu et al., 2005), and that miR-124 targets a great number of genes, it is reasonable to suggest that the miR-124 might coordinate multiple signaling pathways. Therefore, miR-124 might act as key controller by operating at multiple levels on gene regulation, impinging on genes that are tightly interconnected to modulate CSC self-renewal. Indeed, the IPA analysis also revealed a possible role for miR-124 as negative regulator of cell migration. This function was not directly assayed in our screening, but is indeed relevant to CSCs. Together, our results suggest that miR-124 could be a multifunctional tumor suppressor, with the ability to act on the tumor bulk, reducing cell proliferation and inducing apoptosis, and also on CSCs, limiting their self-renewal potential, and inhibiting migration ability and invasion potential, thus reducing metastasis.

4.6 Future plans

In order to develop target CSC therapies, it is essential to understand the molecular mechanisms of CSC properties. To study those mechanisms, we developed a high-

throughput unbiased screening system and provided evidences that it is suitable to identify microRNAs involved in the regulation of CSCs.

Our results mainly identified miR-124 to act as putative tumor suppressor gene, specifically targeting breast cancer stem cells.

Further evidences are now needed to corroborate the biological role of this miRNA as regulator of BCSC properties.

Firstly, *in-vivo* experiments must be carried out to reveal whether this miRNA is able to counteract tumor formation. To this aim, we are exploiting patient-derived xenografts (PDXs) models, where surgically resected tumor samples are engrafted directly into immune-compromised mice. Numerous tumor-specific PDX models have been established in our laboratory and, importantly, they are all biologically stable as they maintain the molecular, genetic and histological heterogeneity typical of tumors of origin through serial passaging in mice.

Moreover, PDXs provide an excellent *in-vivo* platform to study cancer stem-cell biology and to assess novel cancer therapeutics.

For our goals, *in-vivo* experiments support a more relevant system to clinically test our hypotheses that miR-124 should impair human tumor development by employing PDX models rather than cell lines. *In-vivo* experiments using one of the triple negative breast cancers PDX available in our lab are now ongoing. We expect that PDXs overexpressing miR-124, upon lentiviral infection, should not be longer able to sustain tumor growth and propagation through serial passaging in mice compared to control treated PDX. We claim that miR-124 treatment *in-vivo* should 1) block stem-cell self-renewal, thus directly targeting BCSCs 2) impede tumor cell proliferation by targeting the tumor bulk population 3) prevent metastasis.

Secondly, the mechanisms through which this miRNA exerts its function must be

clarified. The "35 gene-list" could represent an interesting starting point to uncover new pathways related to the self-renewal of breast cancer stem cells. We are currently developing a genetic screening strategy, based on RNA interference, using a pooled library constituted of shRNA targeting the possible 35 direct target genes of miR-124, which are related to its self-renewal mechanism.

In parallel, the recognition and the study of the genes associated to the proliferation mechanisms as well as those having an anti-migratory function must be elucidated.

Thirdly, the physiological role of miR-124 needs to be illustrated. To this aim, miR-124 sensor that specifically measures miR-124 levels could be used to reveal the endogenous expression of this microRNA *in-vivo* using "friend virus B type (FVB)" mice, which possess a normal genetic background. Moreover, miR-124 inhibitors, based on anti-sense oligonucleotide sequences, could be used to knock down this microRNA *in-vivo* in FVB mice and the mammosphere assay could then be exploited to reveal whether miR-124 absence promotes mammosphere formation by enhancing self-renewal mechanisms in normal conditions.

Lastly, we have planned to measure miR-124 levels on different primary tumor samples from patients through the use of specific probes and take advantage of the in-situ hybridization technique. This would indicate us whether miR-124 expression could be exploited to stratify breast cancers for their grade of aggressiveness and thus be used not only as a therapeutic molecule, but also as a diagnostic marker.

For all these reasons, we strongly believe that miR-124 could represent an "ideal drug" acting on different pathways and involving several mechanisms and thus, its role in the context of breast cancers obviously requires further investigations.

5 REFERENCES

1. Al-Hajj, M. (2003). From the Cover: Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences *100*, 3983-3988.
2. Ali, N., Karlsson, C., Aspling, M., Hu, G., Hacohen, N., Scadden, D.T., and Larsson, J. (2009). Forward RNAi screens in primary human hematopoietic stem/progenitor cells. Blood *113*, 3690-3695.
3. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology *11*, R106.
4. Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.
5. Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215-233.
6. Bartonicek, N., and Enright, A.J. (2010). SylArray: a web server for automated detection of miRNA effects from expression data. Bioinformatics *26*, 2900-2901.
7. Basyuk, E., Suavet, F., Doglio, A., Bordonne, R., and Bertrand, E. (2003). Human let-7 stem-loop precursors harbor features of RNase III cleavage products. Nucleic Acids Res *31*, 6593-6597.
8. Baud, O., Liu, X.S., Chopp, M., Zhang, R.L., Tao, T., Wang, X.L., Kassis, H., Hozeska-Solgot, A., Zhang, L., Chen, C.*, et al.* (2011). MicroRNA Profiling in Subventricular Zone after Stroke: MiR-124a Regulates Proliferation of Neural Progenitor Cells through Notch Signaling Pathway. PLoS One *6*, e23461.
9. 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. Nature genetics *35*, 215-217.
10. Bonci, D., Coppola, V., Musumeci, M., Addario, A., Giuffrida, R., Memeo, L., D'Urso, L., Pagliuca, A., Biffoni, M., Labbaye, C.*, et al.* (2008). The miR-15a–miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. Nature Medicine *14*, 1271-1277.
11. Borchert, G.M., Lanier, W., and Davidson, B.L. (2006). RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol *13*, 1097-1101.
12. Bracken, C.P., Gregory, P.A., Kolesnikoff, N., Bert, A.G., Wang, J., Shannon, M.F., and Goodall, G.J. (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. Cancer Res *68*, 7846-7854.
13. Bushati, N., and Cohen, S.M. (2007). microRNA functions. Annual review of cell and developmental biology *23*, 175-205.
14. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K.*, et al.* (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A *99*, 15524-15529.
15. 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.

16. Chaffer, Christine L., Marjanovic, Nemanja D., Lee, T., Bell, G., Kleer, Celina G., Reinhardt, F., D'Alessio, Ana C., Young, Richard A., and Weinberg, Robert A. (2013). Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity. Cell *154*, 61-74.
17. Chang, T.-C., Yu, D., Lee, Y.-S., Wentzel, E.A., Arking, D.E., West, K.M., Dang, C.V., Thomas-Tikhonenko, A., and Mendell, J.T. (2007). Widespread microRNA repression by Myc contributes to tumorigenesis. Nature Genetics *40*, 43-50.
18. 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. Journal of Clinical Investigation *124*, 3093-3106.
19. Chekulaeva, M., and Filipowicz, W. (2009). Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. Current Opinion in Cell Biology *21*, 452-460.
20. Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83-86.
21. Chen, T., Heller, E., Beronja, S., Oshimori, N., Stokes, N., and Fuchs, E. (2012). An RNA interference screen uncovers a new molecule in stem cell self-renewal and long-term regeneration. Nature *485*, 104-108.
22. Chen, Y., Zhu, X., Zhang, X., Liu, B., and Huang, L. (2010). Nanoparticles modified with tumor-targeting scFv deliver siRNA and miRNA for cancer therapy. Molecular therapy : the journal of the American Society of Gene Therapy *18*, 1650-1656.
23. 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.
24. 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.
25. Costinean, S., Zanesi, N., Pekarsky, Y., Tili, E., Volinia, S., Heerema, N., and Croce, C.M. (2006). Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E -miR155 transgenic mice. Proceedings of the National Academy of Sciences *103*, 7024-7029.
26. Creighton, C.J., Nagaraja, A.K., Hanash, S.M., Matzuk, M.M., and Gunaratne, P.H. (2008). A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. Rna *14*, 2290-2296.
27. Ding, H., Wu, Y., *et al.* (2014). Characterization of the MicroRNA expression profile of cervical squamous cell carcinoma metastases. Asian Pac J Cancer Prev, 15 (4), 1675-1679.
28. 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 Dev *17*, 1253-1270.
29. Dontu, G., Al Hajj,M Abdallah,W, Clarke,MF and Wicha,M (2003). Stem cells in normal breast development and breast cancer.
30. 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 Res *6*, R605-615.

31. 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.
32. Esposito, C.L., Cerchia, L., Catuogno, S., De Vita, G., Dassie, J.P., Santamaria, G., Swiderski, P., Condorelli, G., Giangrande, P.H., and de Franciscis, V. (2014). Multifunctional aptamer-miRNA conjugates for targeted cancer therapy. Mol Ther *22*, 1151-1163.
33. Esquela-Kerscher, A., and Slack, F.J. (2006). Oncomirs — microRNAs with a role in cancer. Nature Reviews Cancer *6*, 259-269.
34. Esquela-Kerscher, A., Trang, P., Wiggins, J.F., Patrawala, L., Cheng, A., Ford, L., Weidhaas, J.B., Brown, D., Bader, A.G., and Slack, F.J. (2008). The let-7 microRNA reduces tumor growth in mouse models of lung cancer. Cell Cycle *7*, 759-764.
35. Eulalio, A., Mano, M., Dal Ferro, M., Zentilin, L., Sinagra, G., Zacchigna, S., and Giacca, M. (2012). Functional screening identifies miRNAs inducing cardiac regeneration. Nature *492*, 376-381.
36. Fang, Z., and Rajewsky, N. (2011). The impact of miRNA target sites in coding sequences and in 3'UTRs. PLoS One *6*, e18067.
37. Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Reviews Genetics *2008*, 102-114.
38. 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 Res *10*, R25.
39. 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.
40. Garzon, R., Calin, G.A., and Croce, C.M. (2009). MicroRNAs in Cancer. Annu Rev Med *60*, 167-179.
41. Gjorevski, N., and Nelson, C.M. (2011a). Integrated morphodynamic signalling of the mammary gland. Nature reviews Molecular cell biology *12*, 581-593.
42. Greene, S.B., Gunaratne, P.H., Hammond, S.M., and Rosen, J.M. (2010). A putative role for microRNA-205 in mammary epithelial cell progenitors. Journal of Cell Science *123*, 606-618.
43. Guan, H., Liang, W., Xie, Z., Li, H., Liu, J., Liu, L., Xiu, L., and Li, Y. (2014). Down-regulation of miR-144 promotes thyroid cancer cell invasion by targeting ZEB1 and ZEB2. Endocrine.
44. Guiu, S., Michiels, S., Andre, F., Cortes, J., Denkert, C., Di Leo, A., Hennessy, B.T., Sorlie, T., Sotiriou, C., Turner, N.*, et al.* (2012). Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement. Ann Oncol *23*, 2997-3006.
45. 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.
46. Guo, L., Zhao, R.C., and Wu, Y. (2011). The role of microRNAs in self-renewal and differentiation of mesenchymal stem cells. Exp Hematol *39*, 608-616.

47. 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.
48. Gupta, P.B., Onder, T.T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R.A., and Lander, E.S. (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell *138*, 645-659.
49. 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.
50. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J.*, et al.* (2005). A microRNA polycistron as a potential human oncogene. Nature *435*, 828-833.
51. Hinske, L.C., Galante, P.A., Kuo, W.P., and Ohno-Machado, L. (2010). A potential role for intragenic miRNAs on their hosts' interactome. BMC Genomics *11*, 533.
52. Houbaviy, H., Michael F. Murray,and Phillip A. Sharp (2003). Embryonic Stem Cell-Specific MicroRNAs.
53. 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.
54. Hudson, R.S., Yi, M., Esposito, D., Watkins, S.K., Hurwitz, A.A., Yfantis, H.G., Lee, D.H., Borin, J.F., Naslund, M.J., Alexander, R.B.*, et al.* (2012). MicroRNA-1 is a candidate tumor suppressor and prognostic marker in human prostate cancer. Nucleic Acids Res *40*, 3689-3703.
55. Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M.*, et al.* (2005). MicroRNA gene expression deregulation in human breast cancer. Cancer Res *65*, 7065-7070.
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. Izumiya, M., Okamoto, K., Tsuchiya, N., and Nakagama, H. (2010). Functional screening using a microRNA virus library and microarrays: a new high-throughput assay to identify tumor-suppressive microRNAs. Carcinogenesis *31*, 1354-1359.
58. Jansson, M.D., and Lund, A.H. (2012). MicroRNA and cancer. Mol Oncol *6*, 590-610.
59. Jolicoeur, F. (2005). Intrauterine breast development and the mammary myoepithelial lineage. Journal of mammary gland biology and neoplasia *10*, 199-210.
60. Judson, R.L., Babiarz, J.E., Venere, M., and Blelloch, R. (2009a). Embryonic stem cell-specific microRNAs promote induced pluripotency. Nat Biotechnol *27*, 459-461.
61. Kalluri, R., and Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. J Clin Invest *119*, 1420-1428.

62. 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. Proc Natl Acad Sci U S A *104*, 19291-19296.
63. Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.
64. Kluiver, J., Haralambieva, E., de Jong, D., Blokzijl, T., Jacobs, S., Kroesen, B.J., Poppema, S., and van den Berg, A. (2006). Lack of BIC and microRNA miR-155 expression in primary cases of Burkitt lymphoma. Genes Chromosomes Cancer *45*, 147-153.
65. Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res *42*, D68-73.
66. Kumar, M.S., Lu, J., Mercer, K.L., Golub, T.R., and Jacks, T. (2007). Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nature Genetics *39*, 673-677.
67. Kwak, P.B., Iwasaki, S., and Tomari, Y. (2010). The microRNA pathway and cancer. Cancer Sci *101*, 2309-2315.
68. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.
69. Lang, Q., and Ling, C. (2012). MiR-124 suppresses cell proliferation in hepatocellular carcinoma by targeting PIK3CA.
70. Lau, N.C. (2001). An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans. Science *294*, 858-862.
71. 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.
72. 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.
73. Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862-864.
74. 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.
75. Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J *21*, 4663-4670.
76. Li, C., Xiong, Q., Zhang, J., Ge, F., and Bi, L.J. (2012). Quantitative proteomic strategies for the identification of microRNA targets. Expert Rev Proteomics *9*, 549-559.
77. Li, Y., Zhang, Z., Liu, X., Huang, T., He, W., Shen, Y., Hong, K., and Cao, Q. (2014). miR-124 functions as a tumor suppressor in the endometrial carcinoma cell line HEC-1B partly by suppressing STAT3. Molecular and cellular biochemistry *388*, 219-231.
78. Li, Z., and Rana, T.M. (2014). Therapeutic targeting of microRNAs: current status and future challenges. Nature reviews Drug discovery *13*, 622-638.
79. Liang, Y.J., Wang, Q.Y., Zhou, C.X., Yin, Q.Q., He, M., Yu, X.T., Cao, D.X., Chen, G.Q., He, J.R., and Zhao, Q. (2013). MiR-124 targets Slug to regulate

epithelial-mesenchymal transition and metastasis of breast cancer. Carcinogenesis *34*, 713-722.
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. Lin, R.J., Lin, Y.C., Chen, J., Kuo, H.H., Chen, Y.Y., Diccianni, M.B., London, W.B., Chang, C.H., and Yu, A.L. (2010). microRNA signature and expression of Dicer and Drosha can predict prognosis and delineate risk groups in neuroblastoma. Cancer Res *70*, 7841-7850.
82. Lindenbergh-van der Plas, M., Martens-de Kemp, S.R., de Maaker, M., van Wieringen, W.N., Ylstra, B., Agami, R., Cerisoli, F., Leemans, C.R., Braakhuis, B.J., and Brakenhoff, R.H. (2013). Identification of lethal microRNAs specific for head and neck cancer. Clin Cancer Res *19*, 5647-5657.
83. Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., Martin-Trevino, R., Shang, L., McDermott, S.P., Landis, M.D.*, et al.* (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports *2*, 78-91.
84. Liu, S., Dontu, G., and Wicha, M.S. (2005). Mammary stem cells, self-renewal pathways, and carcinogenesis. Breast Cancer Research *7*, 86.
85. Liu, S., and Wicha, M.S. (2010). Targeting breast cancer stem cells. J Clin Oncol *28*, 4006-4012.
86. Lu, Y., Yue, X., Cui, Y., Zhang, J., and Wang, K. (2013). MicroRNA-124 suppresses growth of human hepatocellular carcinoma by targeting STAT3. Biochemical and biophysical research communications *441*, 873-879.
87. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.
88. Luo, J., Emanuele, M.J., Li, D., Creighton, C.J., Schlabach, M.R., Westbrook, T.F., Wong, K.K., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell *137*, 835-848.
89. Ma, Y., Yao, N., Liu, G., Dong, L., Liu, Y., Zhang, M., Wang, F., Wang, B., Wei, X., Dong, H.*, et al.* (2014). Functional screen reveals essential roles of miR-27a/24 in differentiation of embryonic stem cells. The EMBO Journal.
90. Manavathi, B., Samanthapudi, V.S.K., and Gajulapalli, V.N.R. (2014). Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development. Frontiers in Cell and Developmental Biology *2*.
91. Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. Cell *134*, 521-533.
92. Martin, E.C., Krebs, A.E., Burks, H.E., Elliott, S., Baddoo, M., Collins-Burow, B.M., Flemington, E.K., and Burow, M.E. (2014). miR-155 induced transcriptome changes in the MCF-7 breast cancer cell line leads to enhanced mitogen activated protein kinase signaling. Genes Cancer *5*, 353-364.
93. Maziere, P., and Enright, A.J. (2007). Prediction of microRNA targets. Drug discovery today *12*, 452-458.

94. Merritt, W.M., Lin, Y.G., Han, L.Y., Kamat, A.A., Spannuth, W.A., Schmandt, R., Urbauer, D., Pennacchio, L.A., Cheng, J.F., Nick, A.M., *et al.* (2008). Dicer, Drosha, and outcomes in patients with ovarian cancer. N Engl J Med *359*, 2641-2650.
95. Nasser, M.W., Datta, J., Nuovo, G., Kutay, H., Motiwala, T., Majumder, S., Wang, B., Suster, S., Jacob, S.T., and Ghoshal, K. (2008). Down-regulation of Micro-RNA-1 (miR-1) in Lung Cancer: SUPPRESSION OF TUMORIGENIC PROPERTY OF LUNG CANCER CELLS AND THEIR SENSITIZATION TO DOXORUBICIN-INDUCED APOPTOSIS BY miR-1. Journal of Biological Chemistry *283*, 33394-33405.
96. Nazarov, P.V., Reinsbach, S.E., Muller, A., Nicot, N., Philippidou, D., Vallar, L., and Kreis, S. (2013). Interplay of microRNAs, transcription factors and target genes: linking dynamic expression changes to function. Nucleic Acids Res *41*, 2817-2831.
97. Niu, Z., Goodyear, S.M., Rao, S., Wu, X., Tobias, J.W., Avarbock, M.R., and Brinster, R.L. (2011). MicroRNA-21 regulates the self-renewal of mouse spermatogonial stem cells. Proc Natl Acad Sci U S A *108*, 12740-12745.
98. Olive, V., Jiang, I., and He, L. (2010). mir-17-92, a cluster of miRNAs in the midst of the cancer network. Int J Biochem Cell Biol *42*, 1348-1354.
99. Ota, A., Tagawa, H., Karnan, S., Tsuzuki, S., Karpas, A., Kira, S., Yoshida, Y., and Seto, M. (2004). Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res *64*, 3087-3095.
100. Ozen, M., Creighton, C.J., Ozdemir, M., and Ittmann, M. (2007). Widespread deregulation of microRNA expression in human prostate cancer. Oncogene *27*, 1788-1793.
101. Park, S.M., Gaur, A.B., Lengyel, E., and Peter, M.E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. Genes Dev *22*, 894-907.
102. 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.
103. Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res *12*, R68.
104. 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.
105. Rodriguez, A., Vigorito, E., Clare, S., Warren, M.V., Couttet, P., Soond, D.R., van Dongen, S., Grocock, R.J., Das, P.P., Miska, E.A., *et al.* (2007). Requirement of bic/microRNA-155 for Normal Immune Function. Science *316*, 608-611.
106. Sabatier, R., Finetti, P., Guille, A., Adelaide, J., Chaffanet, M., Viens, P., Birnbaum, D., and Bertucci, F. (2014). Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. Mol Cancer *13*, 228.
107. Scheel, C., Eaton, E.N., Li, S.H., Chaffer, C.L., Reinhardt, F., Kah, K.J., Bell, G., Guo, W., Rubin, J., Richardson, A.L., *et al.* (2011). Paracrine and

autocrine signals induce and maintain mesenchymal and stem cell states in the breast. Cell *145*, 926-940.

108. Shi, X.B., Xue, L., Ma, A.H., Tepper, C.G., Gandour-Edwards, R., Kung, H.J., and deVere White, R.W. (2013). Tumor suppressive miR-124 targets androgen receptor and inhibits proliferation of prostate cancer cells. Oncogene *32*, 4130-4138.
109. 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.
110. 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.
111. Song, Su J., Ito, K., Ala, U., Kats, L., Webster, K., Sun, Su M., Jongen-Lavrencic, M., Manova-Todorova, K., Teruya-Feldstein, J., Avigan, David E.*, et al.* (2013). The Oncogenic MicroRNA miR-22 Targets the TET2 Tumor Suppressor to Promote Hematopoietic Stem Cell Self-Renewal and Transformation. Cell Stem Cell *13*, 87-101.
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. Stingl, J.a.C., C (2007). Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis.
114. 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. Proc Natl Acad Sci U S A *102*, 15545-15550.
115. TGCA (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61-70
116. Tiscornia, G., and Izpisua Belmonte, J.C. (2010). MicroRNAs in embryonic stem cell function and fate. Genes Dev *24*, 2732-2741.
117. Tsuchiya, N., Izumiya, M., Ogata-Kawata, H., Okamoto, K., Fujiwara, Y., Nakai, M., Okabe, A., Schetter, A.J., Bowman, E.D., Midorikawa, Y.*, et al.* (2011). Tumor Suppressor miR-22 Determines p53-Dependent Cellular Fate through Post-transcriptional Regulation of p21. Cancer Res *71*, 4628-4639.
118. 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.
119. Ventura, A., and Jacks, T. (2009). MicroRNAs and cancer: short RNAs go a long way. Cell *136*, 586-591.
120. 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.
121. Visvader, J.E., and Lindeman, G.J. (2012). Cancer stem cells: current status and evolving complexities. Cell Stem Cell *10*, 717-728.
122. Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev *28*, 1143-1158.

123. 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 *103*, 2257-2261.
124. Voorhoeve, P.M., le Sage, C., Schrier, M., Gillis, A.J.M., Stoop, H., Nagel, R., Liu, Y.-P., van Duijse, J., Drost, J., Griekspoor, A.*, et al.* (2006). A Genetic Screen Implicates miRNA-372 and miRNA-373 As Oncogenes in Testicular Germ Cell Tumors. Cell *124*, 1169-1181.
125. Wang, M.J., Li, Y., Wang, R., Wang, C., Yu, Y.Y., Yang, L., Zhang, Y., Zhou, B., Zhou, Z.G., and Sun, X.F. (2013). Downregulation of microRNA-124 is an independent prognostic factor in patients with colorectal cancer. International journal of colorectal disease *28*, 183-189.
126. Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J.E., Baehner, L., and Blelloch, R. (2008). Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nature genetics *40*, 1478-1483.
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. Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.
129. Watson, C.J., and Khaled, W.T. (2008). Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development *135*, 995-1003.
130. 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.
131. 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.
132. Wu, S., Aksoy, M., Shi, J., and Houbaviy, H.B. (2014). Evolution of the miR-290-295/miR-371-373 cluster family seed repertoire. PLoS One *9*, e108519.
133. Xia, H., Cheung, W.K.C., Ng, S.S., Jiang, X., Jiang, S., Sze, J., Leung, G.K.K., Lu, G., Chan, D.T.M., Bian, X.W.*, et al.* (2012). Loss of Brain-enriched miR-124 MicroRNA Enhances Stem-like Traits and Invasiveness of Glioma Cells. Journal of Biological Chemistry *287*, 9962-9971.
134. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A., and Kosik, K.S. (2009). MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells. Cell *137*, 647-658.
135. Yi, R., Poy, M.N., Stoffel, M., and Fuchs, E. (2008). A skin microRNA promotes differentiation by repressing 'stemness'. Nature *452*, 225-229.
136. 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.
137. Yuan, Y., Du, W., Wang, Y., Xu, C., Wang, J., Zhang, Y., Wang, H., Ju, J., Zhao, L., Wang, Z.*, et al.* (2014). Suppression of AKT expression by miR-153 produced anti-tumor activity in lung cancer. Int J Cancer.

138. Zhang, B., Yin, Y., Hu, Y., Zhang, J., Bian, Z., Song, M., Hua, D., and Huang, Z. (2014). MicroRNA-204-5p inhibits gastric cancer cell proliferation by downregulating USP47 and RAB22A. Medical Oncology *32*.
139. Zhang, L., Yang, L., Liu, X., Chen, W., Chang, L., Chen, L., Loera, S., Chu, P., Huang, W.C., Liu, Y.R.*, et al.* (2013a). MicroRNA-657 promotes tumorigenesis in hepatocellular carcinoma by targeting transducin-like enhancer protein 1 through nuclear factor kappa B pathways. Hepatology *57*, 1919-1930.
140. Zhang, L.Y., Ho-Fun Lee, V., Wong, A.M., Kwong, D.L., Zhu, Y.H., Dong, S.S., Kong, K.L., Chen, J., Tsao, S.W., Guan, X.Y.*, et al.* (2013b). MicroRNA-144 promotes cell proliferation, migration and invasion in nasopharyngeal carcinoma through repression of PTEN. Carcinogenesis *34*, 454-463.
141. Zhao, Y., Ransom, J.F., Li, A., Vedantham, V., von Drehle, M., Muth, A.N., Tsuchihashi, T., McManus, M.T., Schwartz, R.J., and Srivastava, D. (2007). Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. Cell *129*, 303-317.
142. Zuber, J., McJunkin, K., Fellmann, C., Dow, L.E., Taylor, M.J., Hannon, G.J., and Lowe, S.W. (2010). Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. Nat Biotechnol *29*, 79-83.