A miRNA screen identifies a transcriptional program controlling the fate of adult stem cell

Jacques Robert [1], Efstathios Vlachavas [2,3], Gilles Lemaître [4], Aristotelis Chatziioannou [3,5], Michel Puceat [6], Frederic Delom [1] and Delphine Fessart [1*]

[1] ARTiSt, Univ. Bordeaux, INSERM U1312, F-33000 Bordeaux, France;
[2] Division of Molecular Genome Analysis (B050), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
[3] e-NIOS PC, Kallithea-Athens, Greece
[4] Laboratoire de Génomique et Radiobiologie de la Kératinopoïèse, CEA/DRF/IBFJ/IRCM, Evry; Université Evry / Université Paris-Saclay, France.
[5] Centre of Systems Biology, Biomedical Research Foundation of the Academy of Athens, 4 Soranou Ephessiou Street, 11527, Athens, Greece
[6] Université Aix-Marseille, INSERM UMR_1251, F-13385 Marseille, France.

Corresponding author: DF - E-mail: delphine.fessart@inserm.fr

Abstract

The 3D cultures provide more insight into cell-to-cell and cell-to-matrix interactions, better mimicking the environment where stem cells reside compared to traditional 2D cultures. Although the precise molecular pathways involved in the regulation of stem and progenitor cell fate remain unknown, it is widely accepted that transcription factors play a crucial role as intrinsic regulators in these fate decisions.

In this study, we carried out a microRNA screen to track the behaviour of adult stem/progenitor cells derived from human mammary epithelial cells grown in 3D cultures. We identified miR-106a-3p, which enriches the adult stem cell-like lineage and promotes the expansion of 3D cultures. Transcriptomic analysis showed that this miRNA regulates transcription factors such as REST, CBFB, NF-YA, and GATA3, thereby enhancing the maintenance of adult stem/progenitor cells in human epithelial cells. These data reveal a clear transcriptional program that governs the maintenance of adult stem/progenitor cells and controls their fate.

Introduction

Three-dimensional (3D) human culture models are appealing tools for studying pathophysiological processes. Such models have been developed by our team and others for the lung [1, 2], as well as for various other organs [3]. These *in vitro* models offer an effective method to assess stem cell behaviour and investigate their molecular regulation. Adult stem cells (aSCs) exhibit a high degree of specialisation, allowing them to differentiate into multiple cell types. However, they can only generate tissues in which they reside, classifying them as either ‘unipotent’ or 'multipotent'. This classification highlights their limited versatility compared to pluripotent cells. Pluripotent cells, such as embryonic stem cells (ESCs), can give rise to all cell types. Consistent with this notion, primary cell populations enriched with known progenitor/stem cell markers are more efficient at forming 3D structures than general cell populations [4]. The development and enrichment of breast stem cells likely depend on the coordination of multiple critical transcriptional processes. Interestingly, increasing amounts of data show that the same molecular pathways regulate the self-renewal of both normal stem cells and Cancer Stem Cells (CSCs) in tumours [5]. CSCs express components of the core pluripotency complex found in ESCs, including octamer-binding transcription factor-4 (OCT4), SRY-Box Transcription Factor 2 (SOX2), and NANOG. These factors are closely associated with CSC development [6-11]. OCT4 expression has been reported in both differentiated normal and malignant human cells [12]. SOX2 expression is observed in various malignant tissues [13], and NANOG expression is detected in human neoplasms, including germ cell tumours, breast carcinoma, and osteosarcoma [14]. Additionally, SOX2 has been found in tumour-spheres derived from breast cancer tumours and cell lines [15]. These observations suggest that adult breast stem cells may contain cells with properties similar to embryonic-like stem cells [14, 16]. Currently, it is not known whether adult stem cells share a comparable molecular signature of stemness, such as a minimal core transcriptional program.

MicroRNAs (miRNAs) have been shown to play a crucial role in regulating stem cell self-renewal and differentiation [17]. Generally, a single gene can be repressed by multiple miRNAs, while one miRNA may target and repress multiple genes, which results in the formation of complex regulatory networks. In many developmental processes, miRNAs finely regulate cellular identities by targeting key transcription factors in critical pathways [18]. To investigate the contribution of miRNA-mediated gene regulation in maintaining the adult stem cell-like lineage, we aimed to identify the mechanisms controlling the cell-initiating subpopulation and optimise tissue-specific 3D growth conditions. Since 3D

structures originate from individual stem/progenitor cells, we conducted a functional miRNA screen on human primary epithelial cells to identify factors critical to the formation and expansion of aSC-derived 3D structures. In this study, we identified miR-106a-3p and its target genes as key players in these processes. Transcriptomic profiling of miR-106a-3p–transduced cells revealed genetic programs overlapping with those found in other stem and progenitor cells, displaying common features with ESC-like cells or intermediate cellular states. Through a gain-of-function approach, we demonstrated that endogenous levels of three core transcription factors—OCT4, SOX2, and NANOG—are essential for generating 3D structures from primary cells. Furthermore, we identified the mechanism by which miR-106a-3p finely tunes the differentiation process through a set of transcription regulators, including CBFB (Core-binding factor b), NFYA (Nuclear Transcription Factor Y Subunit Alpha), GATA3 (GATA Binding Protein 3), and REST (RE1 silencing transcription factor). These regulators are critical for maintaining the aSC-like lineage. In conclusion, our results highlight the significant role of miR-106a-3p in sustaining the aSC-derived 3D structure and elucidate the transcriptional mechanisms underlying their growth.

.

Results

3D cultures derived from normal human mammary epithelial cells preserve some mammary epithelial lineages.

The mammary epithelium is composed of distinct cell lineages among which luminal epithelial and myoepithelial cells. The choice of growth media and culture methods is critical in shaping the characteristics of primary human mammary epithelial cell strains and determining experimental outcomes. Finite-lifespan human mammary epithelial cells (HMECs) were generously provided by M. Stampfer via the Human Mammary Epithelial Cell (HMEC) Bank and were cultured in M87A-type medium. M87A medium has been noted for its ability to support pre-stasis HMEC growth for up to 60 population doublings (PD) and maintain luminal cells through approximately 8 passages, equivalent to about 30 PD [19]. To evaluate the heterogeneity of pre-stasis HMECs cultured in M87A, we performed immunostaining for epithelial lineage markers, as previously described [19]. We confirmed the presence of luminal epithelial progenitor (LEP) and myoepithelial progenitor (MEP) cells using specific antibodies against lineage-specific markers (CK18 and MUC1 for LEP cells, and CK14 and CDK5 for MEP cells) (Figure 1A). In addition, we evaluated aldehyde dehydrogenase (ALDH) activity as a marker for normal and malignant human mammary stem cells, as well as a predictor of clinical outcomes, in line with the findings of Ginestier *et al.*[20]. This analysis demonstrated that under these culture conditions, cells preserve stem/progenitor properties (Figure 1B). Using the ALDEFLUOR assay, we identified that approximately 5.2% of normal mammary epithelial cells exhibit ALDH enzymatic activity (Figure 1B). To further explore the regenerative potential of these cells, we performed single-cell seeding in a 3D matrix (Matrigel). This process led to the formation of small, initially spherical structures (up to Day 8) that later differentiated into organoids, with branching initiation observed by Day 10. By Day 20, the structures exhibited budding and lobule formation (Figure 1C), resembling the functional Terminal Duct Lobular Unit (TDLU) found in human breast tissue (Figure 1C). Under these 3D culture conditions, HMECs retained self-renewal capabilities and the ability to regenerate secondary structures over three generations (Figure 1D).

Our next objective was to verify whether HMECs cultured in 3D Matrigel maintained key mammary structural characteristics. Confocal microscopy revealed that cells self-organised, forming a lumen within the 3D structures (Figure 1E). Tracking individual structures showed progressive lumen formation, reaching maximum size by Day 10 of culture (Figure 1F). Confocal sectioning of immuno-stained 3D structures for the luminal progenitor/mature

luminal cell marker EpCAM and the basal/myoepithelial markers cytokeratin 14 (CK14) and p63 revealed that these structures were composed of luminal cells, basal cells, or a mixture of both (Figure 1G). This finding indicates that both basal and luminal cells are maintained in these cultures. Thus, primary cells cultured under these conditions preserve mammary epithelial lineages and retain the expression patterns of key mammary markers.

3D cultures derived from human mammary epithelial cells exhibit a CD44^{high}/CD24^{low} phenotype.

This study aimed to identify epithelial cell subpopulations capable of forming 3D structures and specifying stem/progenitor cell functions [4]. In this context, stem/progenitor cells are rare, immortal cells within a population of cultured cells. They possess the ability to both self-renew and generate other cell types through asymmetric cell division. Previous research has demonstrated that HMECs represent a multipotent stem cell population located in the basal layer of the mammary gland [21-23]. These cells are estrogen-independent (tamoxifen-resistant) and display heterogeneous expression of both luminal and myoepithelial lineage markers (Figure 1A) [24]. Initially, we compared the properties of HMECs cultured in 3D environments to those grown in conventional 2D culture systems. Approximately 3% of the cells in culture demonstrated the capacity to generate 3D structures (Figure 2A). The self-renewal capacity of 3D structure-initiating cells was then evaluated through serial passages, reflecting the maintenance of adult stem/progenitor cells from passage 5 to passage 11 (Figure 2A). As cells were serially propagated, their ability to form 3D structures progressively diminished, consistent with the previously observed loss of self-renewal potential in primary epithelial stem/progenitor cells after a few passages [25].

Previous studies have highlighted that HMECs with a CD44^{high}/CD24^{low} phenotype possess the highest progenitor activity compared to other stem/progenitor subpopulations [26]. This small subpopulation, characterized by self-renewal and a high proliferation rate, originates from normal stem cells and drives the formation of 3D structures [27-30]. Consequently, the CD44^{high}/CD24^{low} phenotype has been widely used in combination with other markers as reliable marker for isolating normal breast adult stem cells [31-34]. To confirm the generation of stem-like cells in 3D culture, we used flow cytometry to assess the expression of breast stem cell markers (CD44^{high}/CD24^{low}) in 3D-grown HMECs, comparing them to cells grown in 2D cultures (Figure 2B). In 2D cultures, 85% of cells expressed high levels of both CD24 and CD44 (CD44^{high}/CD24^{high}), while 14% displayed the CD44^{high}/CD24^{low} phenotype (Figure 2B, top panel). By contrast, 3D cultures showed a more

than threefold increase in $CD44^{high}/CD24^{low}$ cells (~49%) (Figure 1C, lower panel; p = 0.0268, n = 3). To further compare lineage specification between 2D and 3D culture conditions, we analysed pre-stasis HMECs by FACS for $CD24^{low}$ and CD49f expression. CD49f, in combination with $CD44^{high}/CD24^{low}$, serves as a marker to segregate different cell lineages. In 3D cultures, we observed an enrichment of luminal progenitors, while in 2D cultures, basal/stem cells were more prevalent (Figure 2C). Specifically, 2D cultures exhibited significantly more basal/stem cells and less luminal progenitors, whereas 3D cultures displayed the opposite trend, with more luminal progenitors and less basal/stem cells (Figure 2D). Together, these findings suggest that cells grown under 3D culture conditions acquire a $CD44^{high}/CD24^{low}$ expression pattern similar to that of stem/progenitor cells. This indicates that 3D culture can be a valuable tool for enriching luminal stem/progenitor cell markers, facilitating further screening and research.

A miRNA screening approach to selectively enhance the generation of 3D structure.

To investigate whether miRNA-mediated gene regulation could enhance the ability to form 3D structures, we designed a two-step functional screening strategy aimed at enriching the $CD44^{high}/CD24^{low}$ cell population and increasing the number of 3D structure-generating cells. We monitored the expression of CD44 and CD24 after transfecting HMEC cells with miRNAs (Figure 3A-D). After performing quantitative image analysis on over 100,000 cells at Passage 6 (P6), the frequency distribution of CD44 intensity was compared between mass-cultured cells (whole population) and cells transfected with either CD44 (Figure 3A) or CD24 siRNAs (Figure 3B). CD44 and CD24 levels were significantly reduced in siRNA-transfected cells compared to the original population, thereby validating the specificity of our assay (Figure 3A-B). To identify miRNAs that play a role in enriching the $CD44^{high}/CD24^{low}$ cell population, we performed an unbiased functional screen to detect miRNAs capable of modulating the CD44/CD24 phenotype in HMECs (Figure 3C). Using a similar approach to our previous genome-wide siRNA screen for p16 modulators [35], we transfected actively proliferating cells (Passage 6, P6) with a library of 837 miRNAs, along with siRNA controls targeting siGLO ('cyclophilin B'; PPIB), CD44, or CD24. We established cut-off values to define miRNA hits based on the integrated intensity of CD44 and CD24 expression. The raw screening data and quantification for each phenotypic criterion are shown in Figure 3D. This strategy revealed that miR-106a-3p shifts primary cells toward a $CD44^{high}/CD24^{low}$ phenotype (Figure 3D). To validate these findings, we performed a secondary screen targeting the entire miRNA cluster (Figure 3E-F), retesting 28 miRNAs from the cluster using the same method

as the primary screen (Figure 3C). Four miRNA hits scored a Z-score > 2 (Figure 3F), all of which induced a shift in the $CD44^{high}/CD24^{low}$ population (Appendix Figure S1A). The top hit was miR-106a-3p (Figure 3F). We then confirmed that miR-106a-3p induces the $CD44^{high}/CD24^{low}$ phenotype using flow cytometry based on the expression of CD44 and CD24 (Figure 3G). In cells transfected with the control mimic, only ~10% exhibited the $CD44^{high}/CD24^{low}$ phenotype (Figure 3G). In contrast, cells transfected with the miR-106a-3p mimic demonstrated a fivefold increase in the $CD44^{high}/CD24^{low}$ phenotype, comprising approximately 50% of the total cell population (Figure 3G).

In parallel, to link these findings with the generation of stem/progenitor-like cells in 3D culture, we evaluated the frequency of 3D structure initiation after transfection with each of the 28 miRNAs from the miR-17/92 cluster (Figure 3E). Out of the 7 positive hits (Figure 3H), miR-106a-3p transfection showed the highest capacity for 3D structure formation (Figure 3H). Collectively, these results demonstrate that miR-106a-3p transfection induces two critical properties: 1) enrichment of $CD44^{high}/CD24^{low}$ cells and 2) enhanced initiation of 3D structures (Figure 3I).

miR-106a-3p drives the generation of human 3D structures.

To further investigate the role of miR-106a-3p, we generated retroviral vectors for miR-106a, as previously described [24], and evaluated its stable expression in 2D-cultured HMECs (Figure 4A-4B). First, we measured miR-106a-5p and miR-106a-3p expression using RT-qPCR in control cells (miR-Vector) and miR-106a-infected cells grown in 2D (Figure 4A). We observed that while miR-106a-5p was expressed in both infected groups, miR-106a-3p expression was exclusive to miR-106a-infected cells, as confirmed by both RT-qPCR and in situ hybridization (Figure 4A-B). Additionally, when cells were cultured in 3D conditions, they expressed endogenous miR-106a-3p (Figure 4C). To confirm the physiological relevance of this miRNA, we used the DIANA-miTED database, which revealed that miR-106a-3p is expressed in various human tissues and cell lines (Appendix Figure S2).

Next, we assessed the impact of miR-106a on 3D architecture using confocal microscopy. Immunofluorescence staining for Caspase-3, an apoptosis marker, demonstrated that miR-106a had no effect on luminal apoptosis (Figure 4D, Caspase-3). Furthermore, the well-defined cell/Matrigel interface and the myoepithelial layer, characteristic of 3D structures, remained intact despite miR-106a overexpression (Figure 4D, CD44 and p63). β-catenin, a marker of cadherin-based cell junctions, was also unaffected in terms of localization or cell junction integrity by miR-106a overexpression (Figure 4D, β-catenin). Taken together,

these results indicate that miR-106a does not alter the morphogenesis of 3D structures and preserves their structural integrity and cellular interactions.

As expected, stable overexpression of miR-106a in primary HMECs markedly enhanced 3D structure formation capacity (Figure 4E). We then analysed the distinct functions of miR-106a-3p and miR-106a-5p in 3D structure-initiating cells by transfecting HMECs with either miR-106a-3p or miR-106a-5p mimics (Figure 4F). Overexpression of miR-106a-3p resulted in a fivefold increase in the number of 3D structures compared to both the control and miR-106a-5p (Figure 4F). In contrast, miR-106a-5p did not affect 3D structure formation, indicating the specific role of miR-106a-3p in maintaining this capacity (Figure 4F). Additionally, using LNA-anti-miR-106a-3p or LNA-control, we depleted endogenous miR-106a-3p levels and observed a 50% reduction in 3D structure-forming capacity, suggesting that miR-106a-3p is essential for 3D structure generation (Figure 4G). Flow cytometry analysis of miR-infected HMECs revealed that 89.5% of control cells (infected with miR-Vector) exhibited a $CD44^{high}/CD24^{high}$ phenotype, while only 1.7% displayed the $CD44^{high}/CD24^{low}$ phenotype (Figure 4H, left panel). In contrast, miR-106a-infected HMECs showed a significantly higher percentage (49.2%) of $CD44^{high}/CD24^{low}$ cells (Figure 4H, right panel). To evaluate whether this shift in population ratio affected cell survival, we examined the impact of miR-106a on population doubling in culture. Control cells (miR-Vector) stopped growing after approximately 10 population doublings, indicating limited proliferative potential. However, miR-106a extended the cells' lifespan up to 30 doublings, suggesting a survival advantage (Figure 4I).

We then investigated whether miR-106a influences specific lineage sub-populations. Following lentiviral infection with either miR-106a-GFP or miR-Vector in primary HMECs, we analysed the heterogeneity of the miR-106a population via immunofluorescence. Cells expressing miR-106a-GFP (green) were mildly positive for cytokeratin 14 (CK14), a basal/myoepithelial marker, but negative for vimentin (Figure 5A). To determine if 3D structures generated in the presence of miR-106a exhibit enhanced self-renewal capacity compared to vector control, we evaluated serial passages of dissociated cells from single 3D structures. As shown in Figure 1D, most cells could generate a new 3D structure for at least one generation, and a few for up to three generations, indicating limited self-renewal potential. However, miR-106a expression enabled the same population to generate 3D structures for at least five generations, significantly extending their self-renewal capacity (Figure 5B).

Lastly, to investigate miR-106a-3p's role in epithelial-mesenchymal transition (EMT), we performed RT-qPCR analysis on miR-106a-3p-expressing cells, as previously described,[2], targeting key EMT-related genes[36]. The epithelial marker set, including β-catenin, E-cadherin, N-cadherin, and KRT14, revealed significant upregulation of β-catenin, E-cadherin, and KRT14, while N-cadherin and vimentin were downregulated, confirming the observed vimentin-negative phenotype (Figure 5C, blue set). Among the mesenchymal markers tested—Twist1, Snail1, Zeb1, Zeb2, Tgfbr2, Tgfβ1, Smad2, MMP9, and MMP2—only Zeb1 and Twist1 were significantly upregulated, while Snail1 was downregulated (Figure 5C, red set). From the signalling group, Tgfβ1 showed a slight upregulation. Among non-genomic EMT processes, only MMP9 was upregulated.

In conclusion, these findings demonstrate that 1) the increased $CD44^{high}/CD24^{low}$ phenotype in miR-106a-3p-expressing cells drives the higher number of 3D structures; 2) miR-106a-3p induces a partial EMT phenotype; and 3) miR-106a-3p is essential for generating 3D structures, contributing to the maintenance of aSC/progenitor-like lineages.

miR-106a-3p a key player of cell plasticity for differentiation.

We performed transcriptomic analyses to determine whether miR-106a-3p influences cell plasticity, allowing the maintenance of adult stem-like cells or intermediate cellular states. To investigate the effect of miR-106a-3p on global gene expression, we extracted total RNA from miR-106a-3p-transfected HMECs and performed a microarray analysis using Affymetrix chips (HG-U133 Plus 2.0). Transfection with miR-106a-3p resulted in significant changes in the expression of 1,348 genes compared to the control mimic (Figure 6A). We analysed the transcriptomic data to investigate the differential activation of major signalling pathways and transcription factors. The major pathways are depicted in Figures 6B and 6C, revealing a significant downregulation of the MAPK, WNT, and PI3K pathways, while other oncogenic pathways, such as hypoxia response, JAK/STAT, and p53 pathways, were upregulated (Figures 6B and 6C). Additionally, we observed a reduction in the Wnt and PI3K signalling pathways, both known for their roles in maintaining stem/progenitor cells [37]. Moreover, SMAD4 and SMAD3 transcription factors were downregulated, correlating with the suppression of the TGFβ pathway. We also observed a differential regulation of genes involved in mammary breast development (Appendix Figure S4). To further understand the global transcriptional changes associated with miR-106a-3p, we compared our entire microarray dataset with established gene signatures through gene set enrichment analysis (GSEA). We identified an enrichment of downregulated gene sets involved in stem cell

differentiation (Figure 6D). To explore the physiological role of miR-106a-3p in stem cell differentiation, we utilized human embryonic stem cells (hESCs), which express endogenous levels of miR-106a-3p (as observed in the DIANA-miTED database, Appendix Figure S2). Human ESCs differentiate more readily than aSCs. Throughout development, ESCs undergo epigenetic reprogramming, developmental patterning, and differentiation into various organs, each containing tissue-resident aSCs. Based on this, we hypothesized that miR-106a-3p in aSCs could be a remnant from ESCs. First, we evaluated the endogenous expression profile of miR-106-3p in hESCs (Figure 6E). Both miR-106a-5p, miR-106a-3p, and miR-302b (a marker of slowly-growing hESCs [38]) were detectable in hESCs (Figure 6E). Since hESCs are pluripotent and can indefinitely proliferate *in vitro* while maintaining the capacity to differentiate into all three germ layers (ectoderm, mesoderm, and endoderm) [39], we used hESCs to derive early stages of these germ layers (Figure 6F). We transfected hESCs with LNA-anti-miR-106a-3p (anti-miR106a-3p) or LNA-control (anti-miR-ctl) before initiating differentiation into endoderm (Figure 6G), mesoderm (Figure 6H), and ectoderm (Figure 6I). We then applied established protocols [40] to induce differentiation and determine whether blocking endogenous miR-106a-3p would alter transcriptional levels of OCT4, SOX2, and NANOG. In cells transfected with anti-miR-106a-3p, SOX2 expression decreased during differentiation into all three germ layers (Figure 6J, 6K, and 6L), while OCT4 expression decreased during endoderm and ectoderm differentiation (Figure 6H and 6J). NANOG expression, however, increased during mesoderm differentiation (Figure 6I). To further investigate the impact of miR-106a-3p depletion on hESC differentiation, we monitored the expression of specific genes upon inducing the three germ layers (Figure 6M, 6N, and 6O). Endodermal gene expression was minimally affected by miR-106a-3p levels (Figure 6M), while mesodermal and ectodermal gene expression increased upon miR-106a-3p downregulation (Figure 6N and 6O). These data demonstrate that miR-106a-3p plays a role in the early differentiation processes into the three germ layers, particularly the mesoderm and ectoderm layers.

There is ongoing debate about the presence of unipotent, bipotent, or multipotent stem cells in mammary gland tissue [41-43]. The transcription factors OCT4, SOX2, and NANOG are known to maintain pluripotency in embryonic stem cells [44]. Over recent years, evidence has accumulated supporting the existence of stem cells in both mouse and human mammary tissue [45]. Various strategies, such as FACS sorting based on cell surface antigen expression, have been used to identify and isolate human breast stem/progenitor cells. Additionally, *in vitro* cell culture systems have been developed, allowing human mammary epithelial cells to

proliferate in suspension as non-adherent mammospheres [30]. Simoe *et al*. demonstrated that stem cells isolated from both normal human breast and breast tumour cells express higher levels of embryonic stem cell genes, including NANOG, OCT4, and SOX2 [46]. Ectopic expression of these factors, particularly NANOG and SOX2, expands the stem cell population and enhances their ability to form mammospheres. Higher expression of NANOG, OCT4, and SOX2 was observed in CD44+CD24−/low and EMA+CALLA+ stem cell populations compared to other cell types. Overexpression of these genes led to an increase in stem cell populations, indicating their role in maintaining human mammary stem cells. Given the expression of *OCT4*, *SOX2*, and *NANOG* in hESCs and their presence in human breast cells, we hypothesized that aSCs derived from 3D structures might also express these genes. To test whether these transcription factors are involved in 3D structure initiation, we knocked down each gene and demonstrated their roles in 3D structure formation (Figure 7A). This suggests that *OCT4*, *SOX2*, and *NANOG* play significant roles in 3D growth. Furthermore, to determine whether these genes are induced during 3D differentiation, we monitored their mRNA levels over time in 3D cultures (Figure 7B). 3D structure development occurs in two phases: an initial phase where single cells proliferate to initiate 3D formation (around the first six days) and a proliferation arrest beginning around day 8 [47]. The expression of *OCT4*, *SOX2*, and *NANOG* peaked around day 6; subsequently, *OCT4* and *SOX2* levels decreased, while *NANOG* remained consistently high (Figure 7B). Finally, to investigate the potential modulation of these transcription factors by miR-106a-3p, we measured their mRNA (Figure 7C) and protein levels (Figures 7D-7F) in miR-106a-3p-expressing cells compared to controls. All three genes were upregulated in miR-106a-3p-overexpressing cells at both the transcriptional and post-transcriptional levels (Figure 7C-7F). This confirms that miR-106a-3p regulates *OCT4*, *SOX2*, and *NANOG* expression during 3D structure initiation.

Identification of miR-106a-3p targets through in silico analysis and transcriptomic profiling

In the last part of this study, we sought to identify the genes and mechanisms underlying the enhancement of 3D structure formation by miR-106a-3p. To determine if the significant expression changes observed after miR-106a-3p transfection (1,348 differentially expressed (DE) genes, as described in materials and methods) overlapped with data from predictive algorithms, we used the computational tool MicroT_CDS. This analysis identified 707 predicted targets of hsa-miR-106a-3p. The intersection of these datasets yielded 144

common genes (Figure 7G). Next, we conducted a functional enrichment analysis of the 144 common genes using REACTOME pathways and Gene Ontology, which led to the identification of 32 hub genes. An expression heatmap was generated to investigate the expression patterns of these genes within the microarray sample (see materials and methods). From this analysis, we identified 32 hub genes associated with miR-106a-3p, of which 30 were downregulated and 2 upregulated (Figure 7H). Among the downregulated genes, we identified 4 encoding transcription factors (*REST CBFB*, *NFYA*, and *GATA3*), which we subsequently examined in more detail (Figures 7I-7M). Transcript levels of these 4 putative targets were quantified by RT-qPCR in miR-106a-3p–transfected cells, revealing a statistically significant reduction in expression for each (Figure 7I). We then monitored the expression of these transcription factors during 3D structure development. *REST* (Figure 7J), *CBFB* (Figure 7K), and *NFYA* (Figure 7L) showed a significant decrease after 4 days of 3D structure development, while *GATA3* displayed more moderate changes (Figure 7M).

To validate the relevance of these 4 putative targets, we measured: 1) the impact on *OCT4*, *SOX2*, and *NANOG* mRNA expression (Figure 8B, 8E, 8H, and 8K), and 2) the frequency of 3D structure initiation (Figure 8C, 8F, 8I, and 8L). We hypothesised that siRNA-mediated knockdown of the four genes (*REST*, *CBFB*, *NFYA*, and *GATA3*) would impact *OCT4*, *SOX2*, and *NANOG* expression, thereby restoring the 3D structure-generating capacity observed after miR-106a-3p transfection. Each siRNA was first validated by its effect on the CD44/CD24 cell population (Appendix Figure S3), after which we assessed their impact on *OCT4*, *SOX2*, and *NANOG* expression. Knockdown of *REST* (Figure 8A) increased *OCT4* expression (Figure 8B), while knockdown of *NFYA* (Figure 8D) led to increased expression of both *OCT4* and *SOX2* (Figure 8E). *CBFB* knockdown (Figure 8G) resulted in increased *NANOG* expression and a moderate decrease in *SOX2* expression (Figure 8H), while *GATA3* knockdown led to higher *SOX2* and *NANOG* levels (Figure 8K). Finally, we demonstrated that knockdown of *REST*, *NFYA*, and *CBFB* could restore 3D structure formation, mimicking the effect of miR-106a-3p overexpression (Figure 8C, 8F, and 8I).

Discussion

While some argue that identifying stem/progenitor cells is unnecessary for primary tissue culture, understanding the stem cell niche is crucial for improving tissue-specific 3D structure growth conditions and sustaining 3D cultures indefinitely. In this study, we aimed to uncover the key factors promoting 3D structures derived from aSCs. This process involves identifying and characterising the cell populations initiating 3D structure formation, followed by an analysis of the transcriptional processes regulating 3D structure generation.

Here, we focused on miRNAs as key regulators capable of controlling both the generation and maintenance of 3D structures. Using an unbiased screening approach, we identified miR-106a-3p, a previously uncharacterised miRNA, as a master regulator of the stem/progenitor cell-like lineage that specifies the 3D structure-initiating cell population from human normal primary mammary epithelial cells. Our findings indicate that, in the experimental conditions explored, aSC-derived 3D structures are, at least in part, controlled by miR-106a-3p expression. Previous deep sequencing studies have shown that miRNA pairs (5p/3p) co-exist in roughly half of miRNA populations, often with varying concentrations of the 5p/3p species[48, 49]. Notably, the minor miRNA species—either 5p or 3p—exhibit evolutionary conservation in their seed sequences, underscoring their biological significance [50-52]. Several miRNAs have been found to either promote or inhibit stemness. For example, the loss of miR-205 expression induced a stemness phenotype in mammary epithelial cells, promoting EMT, altering cell polarity, and disrupting symmetric division by upregulating Zeb1/2 and Notch expression [53]. Additionally, miR-106a-3p has been linked to follicular lymphoma [54], gastric cancer [55], and renal carcinoma [56] , highlighting its biological importance.

In breast tissue, EMT has been associated with stem/progenitor cell properties, including the expression of the $CD44^{high}/CD24^{low}$ profile. We hypothesised that miR-106a-3p may confer EMT features. However, instead of inducing a complete epithelial-to-mesenchymal transition, evidence increasingly suggests the presence of intermediate states where cells express both epithelial and mesenchymal traits [57]. Indeed, we observed that miR-106a-3p induced a hybrid EMT state, characterised by the expression of both epithelial markers (e.g., β-catenin and E-cadherin) and mesenchymal markers (e.g., TWIST1, ZEB1, TGFβ1, and MMP9). Moreover, cells expressing miR-106a-3p exhibited features associated with stemness, such as elevated TWIST1 levels. Interestingly, previous studies have shown that TWIST1 alone is not sufficient to induce EMT-related morphological changes [58, 59],

suggesting that stem cell-like properties do not arise from the EMT programme itself but rather from a hybrid EMT state. ZEB1, another transcription factor upregulated by miR-106a-3p, also plays a key role in the hybrid EMT state and the acquisition of stem cell-like characteristics [60]. Our findings suggest that stem/progenitor-like properties are linked to a hybrid EMT state rather than a complete epithelial or mesenchymal phenotype. Future studies should further investigate the time-frame window of miR-106a-3p induction during the EMT process.

Since major pathways such as Wnt, TGFβ/BMP, and Notch have been reported to be critical for maintaining mammary aSCs, thus we evaluated the potential involvement of these pathways. Stem cells possess self-renewal activities and multipotency, characteristics that tend to be maintained under hypoxic microenvironments;[61] therefore, it is not surprising to observe an up-regulation of hypoxia pathways and activation of HIF1α transcription factor. Remarkably, it has been shown by cell sorting that cells bearing a stem cell phenotype ($CD44^{high}/CD24^{low}$) express a constitutive activation of JAK-STAT pathway,[62] which we observed to be up-regulated along with STAT1 and STAT2 transcription factors. In parallel, we observe a down-regulation of Wnt signaling as well as PI3K pathways. Wnt is known to play an important role in stem cells maintenance; its inhibition has been shown to lead to the inactivation of PI3 kinase signaling pathways to ensure a balance control of stem cell renewal.[37] Moreover, in our model, we observed a down-regulation of SMAD4 and SMAD3 transcription factors, correlated with a down-regulation of the TGFβ pathway. Signals mediated by TGF-β family members have been implicated in maintaining and differentiating various types of somatic stem cells. [63] Moreover, we identified p53 pathway, in mammary stem cells, p53 has been shown to be critical to control the maintenance of a constant number of stem cells pool.[64, 65]

As cells differentiate and assume a specific lineage identity, some transcriptional mechanisms must switch on or off particular genes. To date, it remains unclear whether a defined transcriptional programme governs the aSCs maintenance. aSC identity, plasticity, and homeostasis are regulated by epigenetic and transcriptional networks specific to each lineage. By integrating gene arrays with miRNA/siRNA screening approaches, we identified 4 transcription factors (*CBFB*, *REST*, *NFYA*, and *GATA3*) as targets of miR-106a-3p, which are crucial for the generation of 3D structure-initiating cells and the maintenance of stem cell self-renewal (Figure 8M). Interestingly, knocking-down each of these genes phenocopied the

effects of miR-106a-3p overexpression by modulating both the three core transcription factors (*OCT4*, *SOX2* and *NANOG*) and the 3D structure initiating cell population.

REST is a transcription factor expressed in epithelial cells [66], and plays a role OCT4/SOX2/NANOG transcriptional network [67]. It shares numerous target genes with OCT4, SOX2, and NANOG, several of which encode essential factors for cell maintenance in ESCs [67]. It binds a 21-bp DNA recognition sequence and has two repressor domains that recruit co-repressor complexes. REST binding sites in ESCs overlap with genomic regions that carry Polycomb-repressed chromatin in FACS-purified multipotent progenitors of the early embryonic pancreas [68]. Herein, REST also restrains differentiation into breast progenitors to favor aSCs maintenance in breast epithelial cells.

Core binding factor (CBF) is a heterodimeric transcription factor complex composed of a DNA-binding subunit, one of three runt-related transcription factor (RUNX) factors, and a non-DNA binding subunit, CBFB.[69, 70] There is only one CBFB subunit while the other subunit is encoded by three mammalian genes: *RUNX1*, *RUNX2*, and *RUNX3*, all of which require CBFB for their function. The targeted inactivation of CBFB abrogates the activity of all RUNX complexes[71]. Targeted knock-out of RUNX genes has revealed distinct roles for these proteins in development, with RUNX1 being required for hematopoiesis[72], RUNX2 for osteogenesis,[73] and RUNX3 for neurogenesis and the control of gastric epithelial cell proliferation.[74, 75]. The RUNX/CBFB complexes have been shown to play a role in the stem cells maintenance by activating FGF signaling loops between the epithelium and the mesenchyme [71]. Additionally, CBFB has been found to be upregulated in the lactating mammary gland and is necessary for lobulo-alveolar development [76]. Furthermore, CBFB knockout has been shown to lead to a down-regulation of SNAIL and VIM expression during the hybrid-EMT, indicating a more epithelial state in breast cell lines [77]. These data confirm our observations, indicating that miR-106a-3p, through the down-regulation of CBFB, induces a down-regulation of both SNAIL1 and VIM in our model. Therefore, it is not surprising that controlling CBFB through miR-106a-3p expression participates in the maintenance of aSCs.

NF-Y, a ubiquitously expressed trimeric transcription factor, has a dual role as an activator and repressor of transcription [78]. The heterodimer protein complex comprises three subunits (NF-YA, NF-YB, and NF-YC). NF-YA is considered as the limiting regulatory subunit of the trimer, since it is required for complex assembly and sequence-specific DNA binding. NF-Y has previously been identified as a marker of CSCs in hepatocellular carcinoma and embryonic carcinoma cells [79-81]. In addition, it has been shown to regulate

the expression of several human *SOX* genes, including *SOX2* [82], NF-Y also regulates the expression of several human SOX genes, including *SOX2*[82], *SOX9* [83], and *SOX18*[84]. This transcriptional activation function of NF-Y is mediated, at least in part, by direct binding to CCAAT boxes within promoters of target genes and by making complex interplay with other factors involved in transcriptional regulation of human *SOX* genes. It has also been shown that the NF-Y binding site CCAAT within the proximal region of the human *SOX2* gene promoter plays a key role in regulating SOX2 expression in cervical CSCs, establishing that NF-YA is essential for maintaining CSCs characteristics. Interestingly, NF-Y has been shown to regulate ATF6 expression [85] which is involved in protein homeostasis. Recently, it has been proposed that cell proteostasis restrains protein synthesis for the maintenance of stem cells [86]. As a consequence, we can speculate that NF-YA might also participate in stem cell maintenance through the regulation of cell proteostasis. Herein, we found that NF-YA regulates the expression of SOX2 for the maintenance of breast aSCs, consistent with the literature.

The GATA transcription factors play critical roles in the gene regulatory networks governing cell fates specification and maintenance. Among them, the zinc finger transcription factor GATA-3 is known to serve as a key regulator of commitment and maturation of the luminal breast epithelial lineage [87] and displays essential roles in the morphogenesis of the mammary gland during embryonic and adult stages. Notably, GATA-3 has been established to be a critical regulator of luminal differentiation, and its deficiency leads to an expansion of luminal progenitors and a concurrent block in differentiation [87]. These data confirmed our observations that miR-106a-3p could act on differentiation through the down-regulation of the GATA3 transcription factor.

It is now well recognized that the lineage hierarchy of the mammary epithelium is composed of a basal/myoepithelial lineage that can function as multipotent mammary stem cells, capable of generating multilineage functional mammary epithelia *in vivo* [43, 88]. These cells have thus been viewed as the primary hierarchy of the mammary epithelium, giving rise to more restricted lineage-specific progenitors, and being responsible for the continuous generation of all mammary epithelial lineages in adults. However, it remains unclear whether cells possessing mammary stem cell potential in the basal layer are distinct from differentiated, non-stem cell basal cells or if stem cell potential is a general feature of cells present in this layer.

The luminal compartment has been shown to contain mature differentiated luminal cells, unable to generate 3D structures, as well as a subpopulation of luminal progenitor cells

capable of giving rise to structures containing luminal cells [43, 88]. We thus performed lineage-tracing experiments using tissue dissociation and surface marker staining and sorting to identify three main populations: mature luminal cells, luminal progenitors, and basal/stem cells. We found that cells growing in 3D are enriched in luminal progenitors and miR-106a enriches this same population. Interestingly, it has been shown that these luminal cells are also able to give rise to ductal structures containing both lineages [89], indicating that a more limited mammary stem cell potential may be present in this lineage as well. It has been suggested that mammary stem cells might possess lineage characteristics that are at an intermediate point on the basal–luminal axis. Indeed, fetal mammary stem cells (fMaSCs), which give rise to the entire mammary gland, do indeed appear in fact to express both luminal and basal markers, representing an intermediate position on this axis, and their gene expression signature is closer to that of adult luminal progenitors than adult basal cells [41, 90].

Mechanistically, miR-106a-3p targets a specific set of genes, namely *CBFB*, *REST*, *NFYA* and *GATA3*, to finely regulate the expression of OCT4, SOX2, and NANOG, thereby reducing heterogeneity within the 3D structure-initiating cell population (Figure 8M). Consequently, a complex mechanism is clearly established to finely modulate the expression of key transcription factors in 3D culture, a process conserved throughout development in adult and embryonic stem cells. Recent reports have demonstrated that the differentiation of human aSCs [91] and mouse ESCs [92] is modulated through post-transcriptional attenuation of key factors such as OCT4, SOX2, and NANOG . A recent study by Doffou *et al.*, has also reported a similar phenomenon in an organoid hepatocyte model, showing that OCT4 expression is induced, starting at day 6 [93]. This induction is seen in hepatocytes in anticipation of trans-differentiation. This increase of OCT4 could be attributed to differentiation in our model, based on previous studies showing that both under- and over-expression of OCT4 could lead to cell differentiation [94].

miRNA-directed regulations provide a way to finely tune aSCs' self-renewal and differentiation. Indeed, miRNAs play an important role in gene regulation for ESCs' pluripotency, self-renewal, and differentiation. These miRNAs can be divided into two subgroups: pluripotency-inducing miRNAs and pro-differentiation miRNAs. The first subgroup, including miR-137, miR-184, miR-200, miR-290, miR-302, and miR-9, is exclusively expressed during the pluripotent state and rapidly decreases upon differentiation stimuli [17, 18]. By contrast, pro-differentiation miRNAs, such as let-7, miR-296, miR-134,

and miR-470, regulate the differentiation processes in pluripotent cells [38].These miRNAs are upregulated during ESCs differentiation and inhibit the expression of pluripotency factors, including NANOG and SOX2 [38]. Even though aSCs are epigenetically and temporally very distant from ESCs, we found that the miR-106a-3p impaired differentiation in ESCs and 3D-structure formation in aSCs suggesting that this miRNA might be a remnant of ESCs. Our data show that miR-106a-3p features as a miRNA participating in the modulation of the core factor network (OCT4, SOX2, and NANOG), which in turn inhibits differentiation and favors maintenance of stem/progenitor cells (Figure 8M). Indeed, miR-106a-3p is sufficient by itself in targeting a specific set of genes *CBFB, REST, NF-YA,* and *GATA3*, and to induce expression of *OCT4*, *SOX2*, and *NANOG*. Finally, the role of miR-106a-3p is of particular interest in explaining how mammary epithelial cells acquire stem cell-like properties in normal conditions. Indeed, the capacity of miR-106a-3p to promote stem cell-like behavior in 3D gives us some clues on how the stem cell status may be specified in mammary cells. Our study underscores the cooperative action of transcription factor networks in establishing stem/progenitor cell identity, providing comprehensive regulatory principles for human epithelial homeostasis. To date, it was not known whether a transcriptional program existed to maintain aSCs, a question addressed in this study (Figure 8M). Understanding the transcriptional networks and signalling pathways involved in 3D structure formation could enhance the use of organoids as models for studying organ function, disease, and therapy.

Acknowledgments

We gratefully acknowledge the members of the ARTiSt Lab for their critical remarks. FD was supported by grants from the “Ligue Contre le Cancer Gironde” and from the “Site de recherche intégrée sur le cancer de Bordeaux” (SIRIC Brio). This work has been supported by grants from the “Région Nouvelle-Aquitaine” (DF).

Author Contributions

Conceptualization, D.F., F.D., G.L., and A.C; Methodology, D.F., F.D., and M.P.; Investigation, D.F., F.D., E.V, and M.P.; Writing – Original Draft, D.F.; Writing – Review & Editing, D.F., F.D., J.R., E.V and M.P; Resources, D.F. and M.P.; Supervision, D.F., and M.P.

Declaration of Interests

None

Materials and Methods

Contact for reagent and resource sharing

Further information and requests for resources and reagents may be directed to and will be fulfilled by the Lead Contact, Delphine Fessart (delphine.fessart@inserm.fr).

Experimental model details

Cell lines

Normal HMECs with finite life-span have been previously fully described [24, 35]. Finite-lifespan HMECs were provided by the Human Mammary Epithelial Cell (HMEC) Bank from M. Stampfer. M87A-type media was used [19]. M87A media has been reported to support pre-stasis HMEC growth for as much as 60 population doublings (PD), and to maintain luminal cells for as many as 8 passages (~30 PD). It has been shown that M87A supports pre-stasis HMEC growth, including myoepithelial, luminal, and progenitor cells for 30–60 PD [95]. HUES cells (HUES9) were cultured as previously described [40].

Cell culture

For standard 2D culture, primary human mammary epithelial cells at passage 4 were established and maintained in M87A medium as previously described [24]. HMECs cells at Passage 6 (P6) were used for the miRNA screening and follow-up miRNA studies, unless otherwise stated. For three-dimensional culture (3D), cells were grown in laminin-rich basement membrane growth factor-reduced Matrigel (BDBiosciences) (Matrigel) as we previously described.[5] Mycoplasma testing was performed prior to all experiments in this study.

Methods

High-content miRNA screening

The miRNA screen was performed in triplicate, using the Human pre-miR miRNA library (Ambion), consisting of 837 miRNAs, together with control small interfering RNAs (siRNAs) targeting Cyclophilin B (Dharmacon), CD44, and CD24 (Qiagen). HMECs at P6 were reverse-transfected with 30 nM miRNA in 384-well format using HiperFect (QIAGEN), in triplicate using Janus apparatus (Perkin Elmer) of the POETIC plateform. Plates were incubated for 48 h, medium was changed and fixed/stained 72 h later with CD44-FITC conjugated antibody (Abcam), CD24 antibody (BD Biosciences) and GtαMo AlexaFluor546 (Invitrogen), 4′,6-diamidino-2-phenylindole (DAPI, Sigma). For the screening in 3D culture, cells were fixed following 8 days after transfection. High-content images were acquired with

the Cytation3 (Bioteck) at 4× magnification, and analysis was performed using the Analysis software (Bioteck). The Z-score provides a metric of the median absolute deviation by which an individual miRNA transfected condition (averaged over three replicates) differs from the population median (median percentage $CD44^{high}/CD24^{low}$ population).

Flow cytometry

Following trypsinization, cells were strained through a 40 μm nylon mesh to ensure single cells are obtained and suspended in ice-cold solution to obtain a density of 1×10^6 cells/ml. Antibodies (CD44 conjugated with FITC (Abcam, Ab19622); CD24 conjugated with phycoerythrin, PE (BD Biosciences, 555428); CD-24 (BD Biosciences, 55046); CD49f conjugated to Cy5 (BD Biosciences, 551129) were added to the cell suspension at concentrations suggested by the manufacturer and cells were incubated at 4°C in the dark for 45 min. These labeled cells were washed twice, suspended in PBS and analysed using a flow cytometer (Becton Dickinson). The cells were stained with either isotype-matched control antibodies or with no primary antibody as negative controls. No difference was observed between these two controls. To separate the ALDH positive population by FACS, the ALDEFLUOR kit (StemCell Technologies) was used. Cells were suspended in ALDEFLUOR assay buffer containing ALDH substrate (BAAA, 1 μM per 1×10^6 cells) and incubated during 40 minutes at 37°C. As negative control, for each sample of cells an aliquot was treated with 50mM diethylaminobenzaldehyde (DEAB), a specific ALDH inhibitor.

RNA isolation and miRNA microarray

Total RNAs were isolated from three independent samples of HMEC-transfected cells using the miRNeasy Kit (Qiagen) according to the manufacturer's instructions. The quantity and size of RNAs were analysed for concentration, purity and integrity by using spectrophotometric methods in combination with the Agilent Bioanalyzer (Agilent Technologies).

Microarray analyses were performed on 3 independent replicates of mimic control transfected cell samples (control), 3 independent replicates of miR-106a-3p transfected cell samples.

Complete gene expression analysis was performed with R (R version 3.6.1)/Bioconductor software (https://doi.org/10.1038/nmeth.3252). Initially, the raw data were imported with oligo R package (1.48.0) (https://doi.org/doi:10.18129/B9.bioc.oligo) and processed (background subtraction, quantile normalization and summarization with median polish method) using the RMA algorithm (https://doi.org/10.1093/biostatistics/4.2.249). In addition, for the annotation process, the R packages affycoretools (1.56.0) and hgu133plus2.db (3.2.3)

were used to map probe sets to gene symbols (https://doi.org/doi:10.18129/B9.bioc.affycoretools). Next, after the removal of control features, a non-specific intensity filtering procedure was applied to remove probesets that were not expressed at least in one of the two conditions (control or transfected samples). Finally, aiming to identify differentially expressed genes between transfected and non-transfected samples, linear models were fitted and statistical inference was estimated using the limma R package (3.40.6) (https://doi.org/10.1093/nar/gkv007). Differentially expressed genes were identified using an FDR value < 0.01 & an absolute value of log2- fold change > log2(1.5). Regarding the visualization part of the differential expression analysis, volcano plots were created with the EnhancedVolcano R package (1.2.0) (https://doi.org/doi:10.18129/B9.bioc.EnhancedVolcano), whereas the R package ComplexHeatmap (2.0.0) was utilized for the creation of the gene expression heatmaps based on selected gene signatures (https://doi.org/10.1093/bioinformatics/btw313). The gene expression data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-6594. The R code implemented for the analysis is available upon request. We have analysed our transcriptomic data, using the PROGENy R package (version 1.16.0) (https://doi.org/10.1038/s41467-017-02391-6) and DoRothEA (version 1.6.0)-decoupleR (version 2.1.6) computational pipeline (https://doi.org/10.1101/gr.240663.118, https://doi.org/10.1093/bioadv/vbac016), to investigate the differential activation of major signaling pathways and transcriptional factors.

Functional enrichment analysis

To exploit the biological mechanisms involved in the miR-106a-3p transfection effect, the BioInfoMiner interpretation web platform was used (10.4018/IJMSTR.2016040103; https://doi.org/10.15252/emmm.201707929). BioInfoMiner implements an automated and robust network analysis of functional terms, by the integration of semantic information through different biomedical vocabularies, aiming to elucidate the significantly perturbed biological processes, and critical genes with centrality role affected in the studied phenotype (https://doi.org/10.1038/s41467-022-30159-0). Furthermore, in order to unravel if specific mechanisms related to stemness and differentiation are significantly altered in the transfected samples, a customized enrichment analysis approach was applied through rotation gene set tests (https://doi.org/10.1093/bioinformatics/btq401), using the limma R package (mroast function). For the gene set signatures, we initially selected from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/),

(https://doi.org/10.1073/pnas.0506580102) the ontology and curated gene sets (version 7.1). Then, as a final step we kept only the terms/pathways that included the phrases "STEM" or "NOTCH" and had at least 10 genes in the respective signature. Enrichment analysis was performed using REACTOME pathways (https://doi.org/10.1093/nar/gkab1028) and Gene Ontology (https://doi.org/10.1093/nar/gkaa1113) to identify 32 hub genes, and generate an expression heatmap to investigate the expression patterns in the microarray sample.

miRNA target identification

The miRNA targets predictions based on miRanda, DianaMT, miRDB and miRWalk were downloaded from www.microrna.org (August 2010 release), http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/ and from http://mirdb.org/miRDB/. We used the computational tool MicroT_CDS (https://doi.org/10.1093/nar/gkt393)

miRNA target stem cells signature analysis

Gene set stem cells enrichment analysis for predicted miRNA targets was carried out using the web interface of Stem checker (http://stemchecker.sysbiolab.eu/) using default settings.

miRNA and antigomiR transfections

HMECs were transfected with 30 nM miRNA or 30 nM antigomiR (anti-miRNA) in 384-well plates using HiperFect (Qiagen), and the protocol described above for 'High-content miRNA Screening' was followed. For siRNA transfections, pools of three siRNA per target were purchased (Qiagen) and the cell were transfected with Hyperfect (Qiagen) according to manufacturer instructions.

Quantitative reverse transcriptase-polymerase chain reaction

Methodology for quantitative reverse transcriptase-polymerase chain reaction (RT-qPCR) has been described previously. Quantitative RT-PCR reactions were performed with SYBR Green Master Mix (ABI). For siRNA knockdown experiments, three siRNA per targets were used and, RNA was extracted from 1×10^5 cells 48hr post-transfection. *GAPDH* levels were quantified for each cDNA sample in separate qPCR reactions and were used as an endogenous control. Target gene-expression levels were quantified using target specific probes. Values were normalized to the internal *GAPDH* control and expressed relative to *siGLO* transfected control levels (100%). All qPCR reactions were run in triplicate from three independent samples. Primers used for qPCR are in Appendix Table S1.

Retroviral stable cell lines

106a-5p/-3p miRNA hit was cloned into MirVec as previously described.[3] After sequence verification, 5 mg of plasmid DNA was transfected into HMEC P5 was transduced into Phoenix packaging cells using Fugene (Roche, Basel, Switzerland). Viral supernatant was

harvested 48 h after transfection. Target HMECs were seeded in a six-well plate at a density of 5000 cells/cm^2 and spinfected the following day at 32 °C, 350 r.p.m. for 1 h with viral supernatant in the presence of 8 mg/ml polybrene. Cells were selected with blasticidin (3 mg/ml). Cells were harvested for RT-qPCR analysis.

Immunofluorescence

Fixed cells were permeabilized with 0.1% Triton X-100 (Sigma) for 30 min at room temperature (RT) cells were stained for 2 h at RT with a primary antibody followed by a secondary antibody staining for 1 h at RT (AlexaFlour-488-conjugated goat anti-mouse antibody (Invitrogen). Cells were imaged on Leica Dmi8 microscope. Images were analysed using Leica software. Primary antibodies used were CD44 (BD Biosciences, 550392); CD44 (Abcam, Ab19622), CD24 (BD Biosciences, 550426), cleaved Caspase-3 ((Asp175), Cell Signaling, 9664S), beta-catenin (BD Biosciences, 610153), p63 (clone 4A4; Santa CruzBiotechnology, sc8431), MoαCK14 (Ozyme), CK18 (Santa CruzBiotechnology, sc32722), EpCAM (EBA-1; BD Biosciences, 347197); MUC1 (Ozyme, BTMBNC80960); CK5 (Abcam, Ab17130), Vimentin (Sigma, V6389). Secondary antibodies were the appropriate AlexaFluor-488 or AlexaFluor-546 antibody (Invitrogen). DAPI and CellMask Deep Red (Invitrogen) were also included. Images were collected with the Dmi8 microscope (Leica) or the Zeiss 510 Meta Confocal microscope (Zeiss) and Developer Software (Leica) used for image analysis.

In situ hybridization (ISH) and microscopy

ISH was performed by using specific DIG-labeled miRNA LNAprobes from Exiqon. Briefly, cells were fixed in 4% paraformaldehyde for 30 min, followed by 70% ethanol for at least 16 h at 4°C. Cells were then permeabilized with 0.1% Triton X-100 for 10 min. The washed cells were then pre-hybridized with a prehybridization buffer (46 SSC, 25% formamide, Denhardt's solution, 2% blocking reagents, 0.25 mg/ml yeast tRNA, 0.25 mg/ml salmon sperm DNA) for 30 min at room temperature, followed by hybridization at 23 °C below the Tm of the LNA probe for 2 h. The cells were subsequently washed with Washing Buffer I (2X SSC with 0.1% Tween 20), II (2X SSC), and III (0.5X SSC) at the hybridization temperature. Cells were blocked with a signal enhancer (Lifetechnologies) for 1 h at room temperature, and then incubated with a mouse anti-DIG antibody at a dilution of 1:1000 at 4°C overnight. Cells were washed with PBS three times to remove unbounded mouse anti-DIG antibody. Then, cells were incubated with a fluorescently labeled secondary antibody. To confirm that the ISH signals were indeed from the specific hybridization of the probes with the target RNA, the cells stained with a specific miR-scramble DIG-labeled miRNA

LNAprobes from Exiqon. The DNA was stained with DAPI. The samples were mounted on a fluorescent mounting medium (Dako). The images were taken with a LSM-510 Meta (Zeiss) confocal microscope.

ESCs differentiation

HUES cells (HUES9) were cultured as previously described [40]. Endoderm was induced by treating the cells for 3 days with 100 ng activin A (Peprotech, France) in DMEM supplemented with 10% FCS. Mesoderm was induced by culturing the cells in RPMI supplemented with 20% B27 (Thermofisher, France) and added with 5 µM CHIR 99021 (Stem cell, France) for 24 hr, then with BMP2 (10 ng/ml, Thermofisher, France) and CHIR 99021 (5 µM) the second day and finally IWR1 (2 µM) and BMP2 (10 ng/ml) the third day. Ectoderm was induced in RPMI supplemented with N2 medium (Thermofisher) and 0.5 µM retinoic acid for three days.

Quantification and statistical analyses

Quantification data are presented as means ± SEM. Statistical significance was analysed using an unpaired Student's t test. A difference at $p < 0.05$ was considered statistically significant.

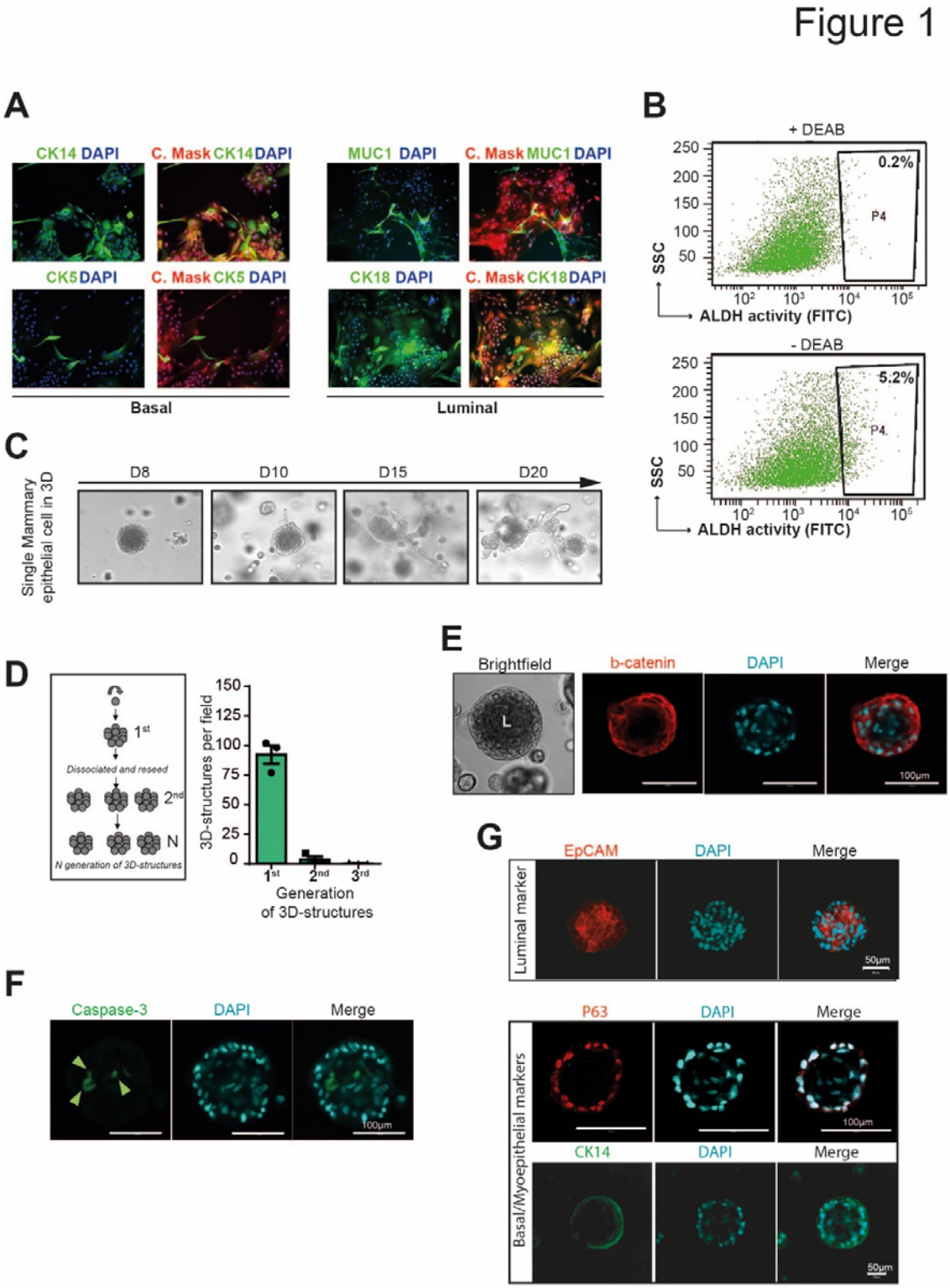

Figure 1 : 3D cultures of mammary primary epithelial cells.

A, Representative immunofluorescence (IF) analysis depicting basal/myoepithelial markers stained for CK14 (green), CK5 (in green), entire cell with Cell Mask (red), and nuclei with DAPI (blue). Additionally, luminal markers stained for CK18 (green), MUC1 (green), cell mask (red), and DAPI. **B,** Representative Fluorescence-Activated Cell Sorting (FACS)

analysis of normal breast epithelial cells using the ALDEFLUOR assay. Cells incubated with ALDEFLUOR substrate (BAAA) and the specific ALDH inhibitor, DEAB, established the baseline fluorescence and defined the ALDEFLUOR-positive region (P4) (n=3, mean ± SEM). **C,** Development of primary breast cells (single MECs) in 3D Matrigel following 20 days. Results are representative of at least three independent experiments. **D,** Quantification of primary, secondary, and 3rd generation 3D structure development. Data are presented as mean values ± SD, n=3. **E,** Immuno-fluorescent staining of β-catenin (red) and DAPI nuclear staining (blue) in 3D Matrigel organoids. Results are representative of three independent repeats of this experiment. **F,** Confocal cross-sections of 3D structure stained with active Caspase-3 and DAPI (blue) for the nucleus. Scale bars, 50 μm. **G,** Confocal cross-sections of 3D structures stained with Epcam (red) as the luminal marker and p63 (red), CK14 (green) as basal/myoepithelial markers, and DAPI (blue) for the nucleus. Scale bars, 50 μm.

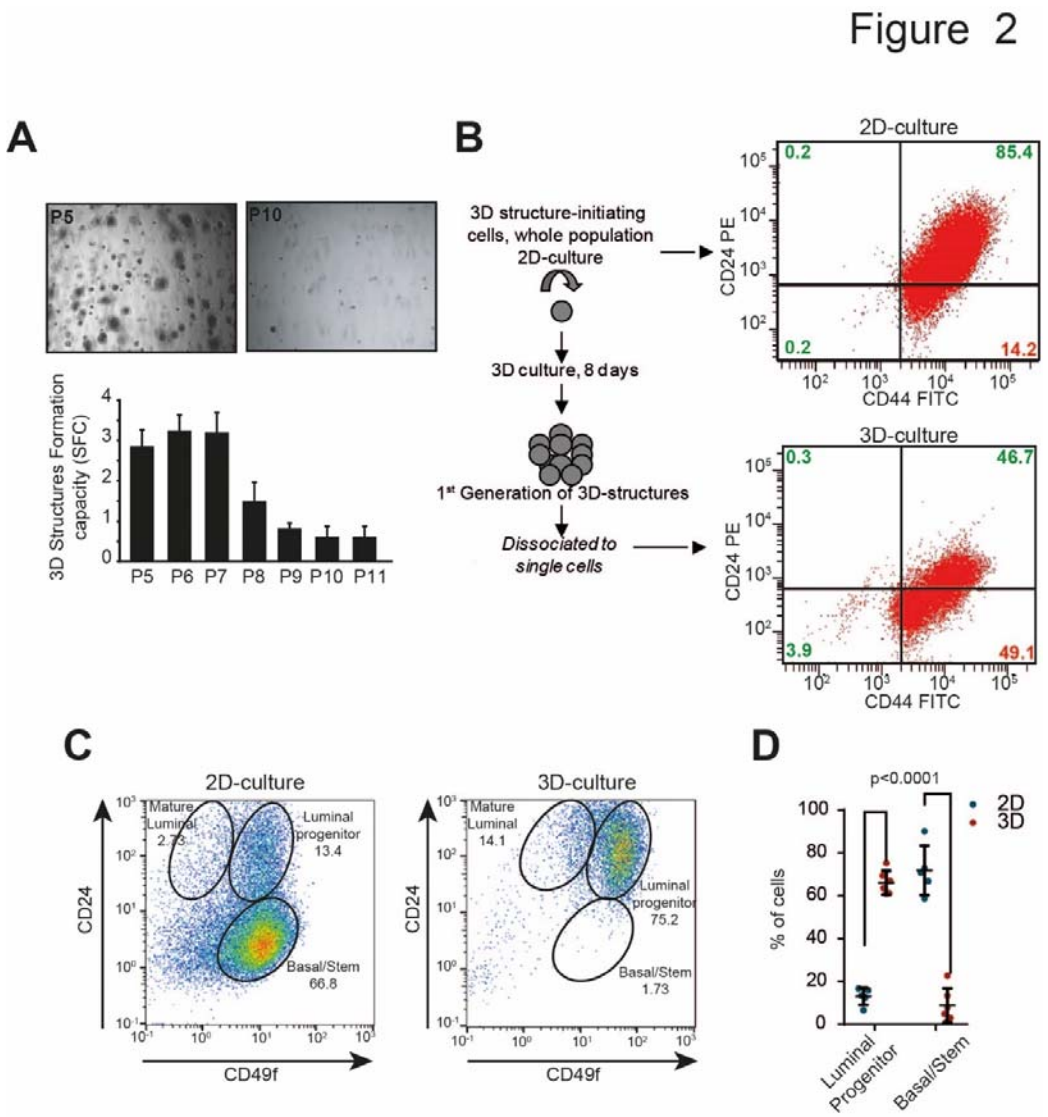

Figure 2. 3D-culture confers original cell properties as compared to 2D-culture

A, Representative brightfield pictures of primary cells grown in 3D at passage 5 (P5) and passage 10 (P10). The bar graph shows the SFC, the mean ± SEM of 3D structure per well, from passage P5 to passage P11. Data are from three independent experiments for each passage. **B**, Flow cytometry analyses of CD44/CD24 in HMEC cells derived from 2D-cell culture (top) or from primary culture in 3D (bottom). The expression of $CD44^{high}/CD24^{low}$ in dissociated 3D structures was higher than in 2D cultured cells. A minimum of 10,000 events were collected per sample. (n=3, mean ± SEM) **C**, Representative flow cytometry analysis of CD24/CD49f in HMEC culture in 2D (left) compared to 3D (right). Mature luminal, luminal progenitor, and basal/stem populations are indicated. (n=3, mean ± SEM) **D**, Mammary epithelial cell population changes between cell culture conditions. Percent of Luminal progenitor ($C24^{high}/CD49f^{high}$) and basal/stem cells ($CD24^{low}/CD49f^{high}$) are quantified between culture conditions. Data are from four to five independent experiments. (mean±SEM).

Figure 3

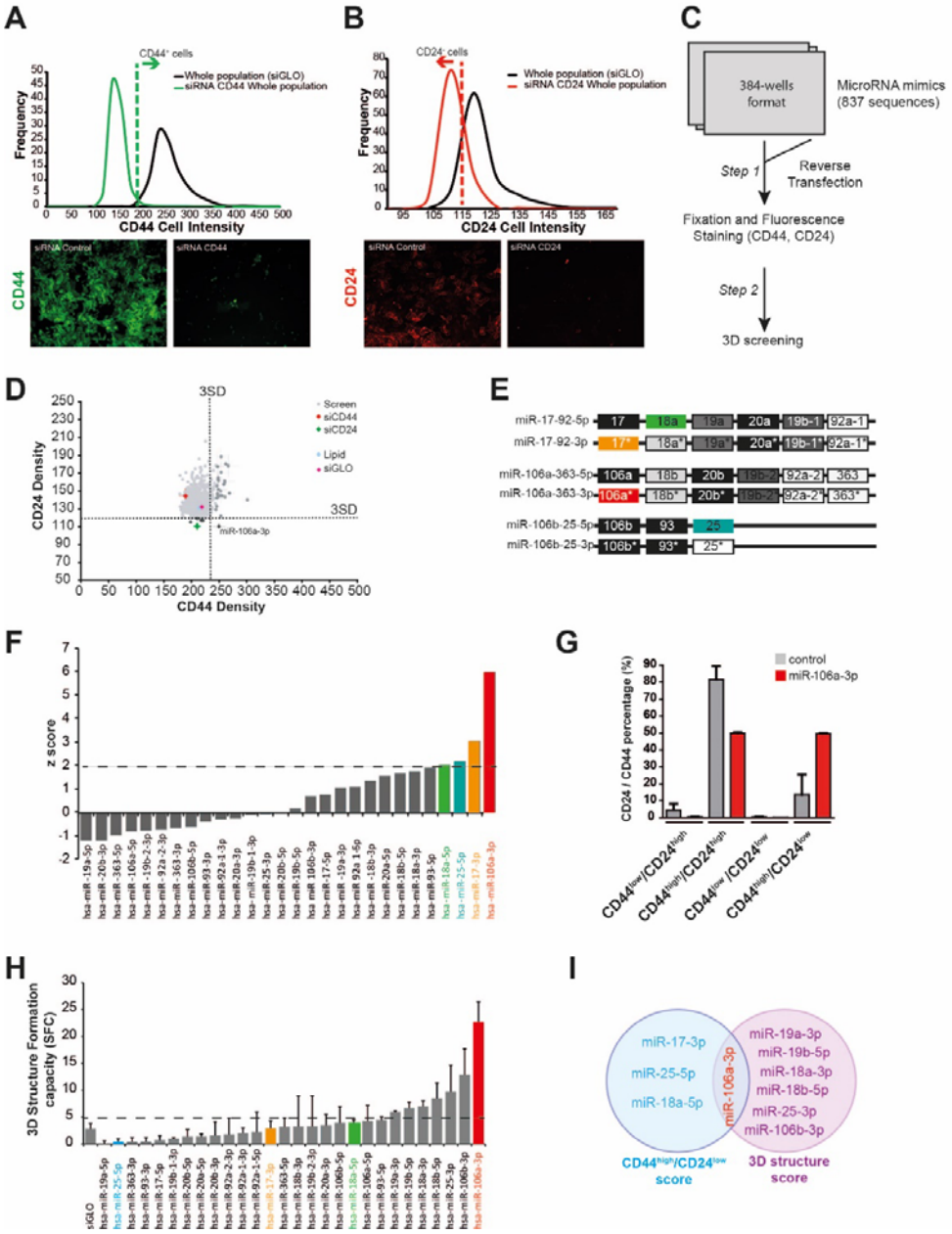

Figure 3. Identification of miR-106a-3p as the predominant miRNA in cells growing in 3D

A, Frequency distributions of CD44 intensity in HMECs at P6 (whole population, siGLO) as compared to HMEC-CD44-siRNA-depleted cells. Bottom panels are representative immunofluorescence pictures of HMECs stained with CD44 antibody (green): siGLO (left) and CD44 siRNA-knocked-down cells (right). **B,** Frequency distributions of CD24 intensity in HMECs at P6 (whole population, siGLO) as compared to HMEC CD24 siRNA knock down. Bottom panels are representative immunofluorescence pictures of HMECs stained with CD24 antibody (red): siGLO (left) and CD24 siRNA-knocked-down cells. **C,**Workflow for image-based miRNA screening for $CD44^{high}/CD24^{low}$ enhancers in primary human HMECs.

HMECs were plated in 384-well plates and subjected to screening of miRNA libraries using optimized immunofluorescence staining for CD44 and CD24. **D,** Screening dot-plot showing the relationship between CD44 and CD24 intensities. Based on the frequency distributions generated for each phenotypic criterion (CD44 and CD24 intensity levels), we assigned highly stringent cutoffs for scoring positive hits in the genome-wide screen (dashed lines, 3 standard deviations (3SD) from the siGLO negative control. **E**, Members of the miR-17/92 cluster and its two paralogues miR-106a/363 and miR-106b/25. Red: miR-106a-3p; blue: miR-25-5p; green: miR-18-5p; orange: miR-17-3p. **F,** HMECs were transfected with miRNA mimics of the miR-17/92 cluster and its two paralogues and screened using conditions identical to the full screen. Z-Scores were calculated for individual miRNA mimics and plotted according to rank order. Dashed lines indicate 2 standard deviations (2SD) above the mean of the distribution. In colors are the miRNA above the 2SD. **G,** Mean percentages ± SEM of CD44/CD24 subpopulations from at least three independent sorting experiments in HMEC transfected with miR-106a-3p mimic as compared to control. **H,** 3D-SFC represents the mean numbers of 3D structures per cell seeded for each miRNA transfected in HMECs as compared to cells transfected with siRNA control (siGLO). Data are from three independent experiments. (mean±SEM). **I,** Venn diagram depicting the overlap of miRNA scoring in common between the $CD44^{high}/CD24^{low}$ and 3D structures scores. Note that the overall number of miRNAs in common would be the overlap of the intersect of these two Venn diagrams.

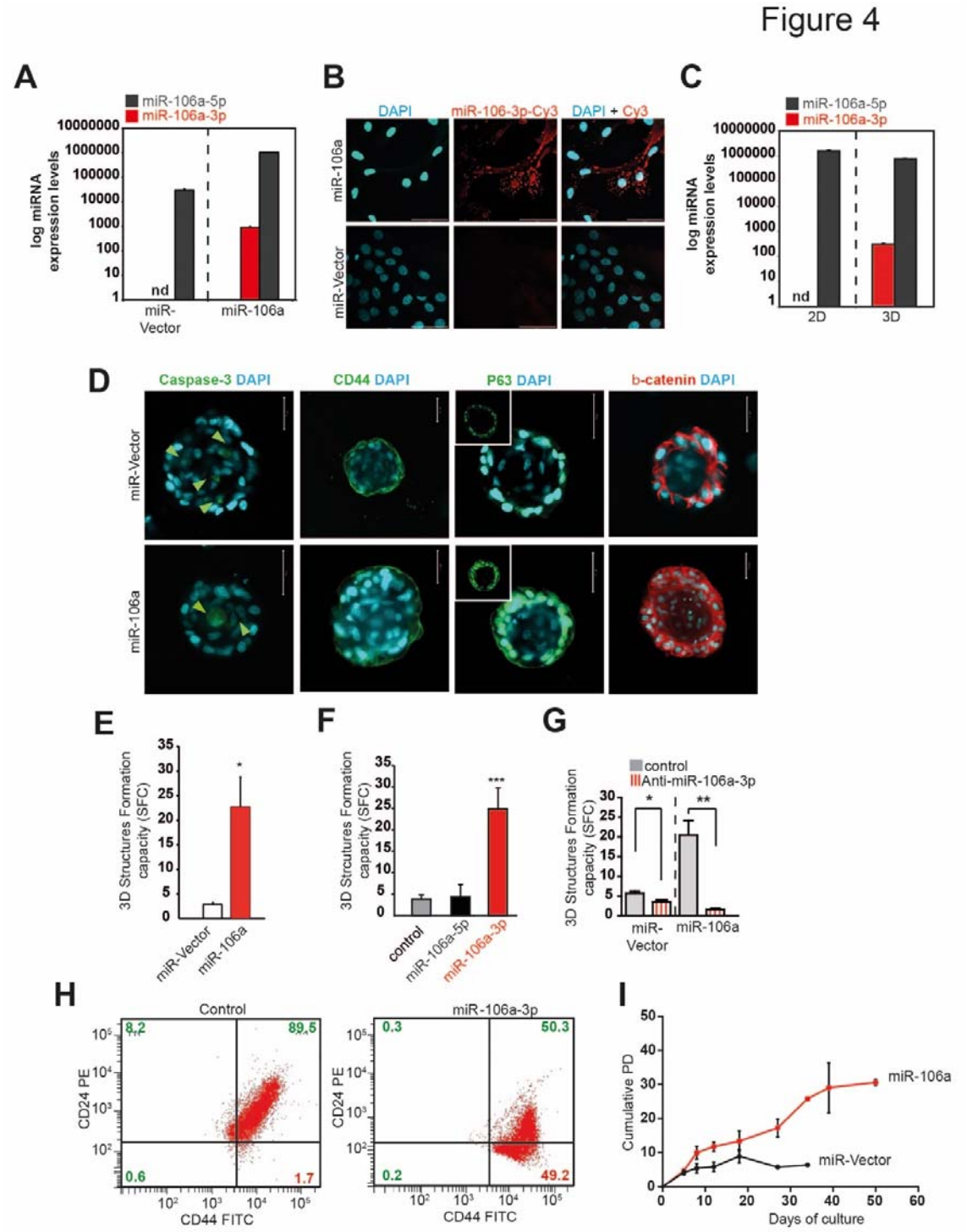

Figure 4. Properties of miR-106a-3p

A, Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in HMECs after retroviral infection with miR-Vector or miR-106a. n.d., not detectable. Data are from at least three independent experiments. (mean±SEM). **B,** FISH detection of miR-106a-3p in HMEC-miR106a stable cell lines. miR-106a-3p positive signals are visualized in red. Scale bar: 50 µm. **C,** Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in HMECs grown in 2D as compared to 3D; n.d., not detectable. Data are from at least three independent experiments. (mean±SEM). **D,** Confocal cross-sections of stable

HMEC-miR106a 3D structures as compared to miR-Vector organoids stained with respectively active Caspase-3, CD44, p63 or β-catenin, and DAPI (blue) for nucleus. The arrows indicate apoptotic cells. Scale bars, 50 µm. **E** The bar graphs show the SFC, the mean number of 3D structures per well for miR-106a stable HMECs as compared to cells with miR-Vector. Statistical significance by Student's t test is indicated by one ($p < 0.05$), two ($p < 0.01$), or three ($p < 0.001$) asterisks. Data are from at least three independent experiments. (mean±SEM). **F,** SFC represents the mean number of 3D structures per well in HMEC transfected with either control, miR-106a-5p or miR106a-3p mimics. Data are from at least three independent experiments. (mean±SEM). **G,** SFC as the percentage of 3D strcutures formed by cells seeded for either stable miR-Vector transfected with anti-miR control or stable miR-106a-HMEC transfected with anti-miR-106a-3p. Statistical significance by Student's t test is indicated by one ($p < 0.05$), two ($p < 0.01$), or three ($p < 0.001$) asterisks. Data are from at least three independent experiments. (mean±SEM). **H**, Flow cytometric analyses of CD44/CD24 in HMEC transfected with either control or miR-106a-3p mimics. A minimum of 10,000 events were collected per sample. (n=3, mean ± SEM) **I,** Growth kinetics of human primary cells expressing either –miR-Vector or –miR-106a in long-term culture. The curve relationship between cumulative population doubling (PD) and duration of culture demonstrates a relatively linear PD rate with the progression of time. Data are from at least three independent experiments. (mean±SEM).

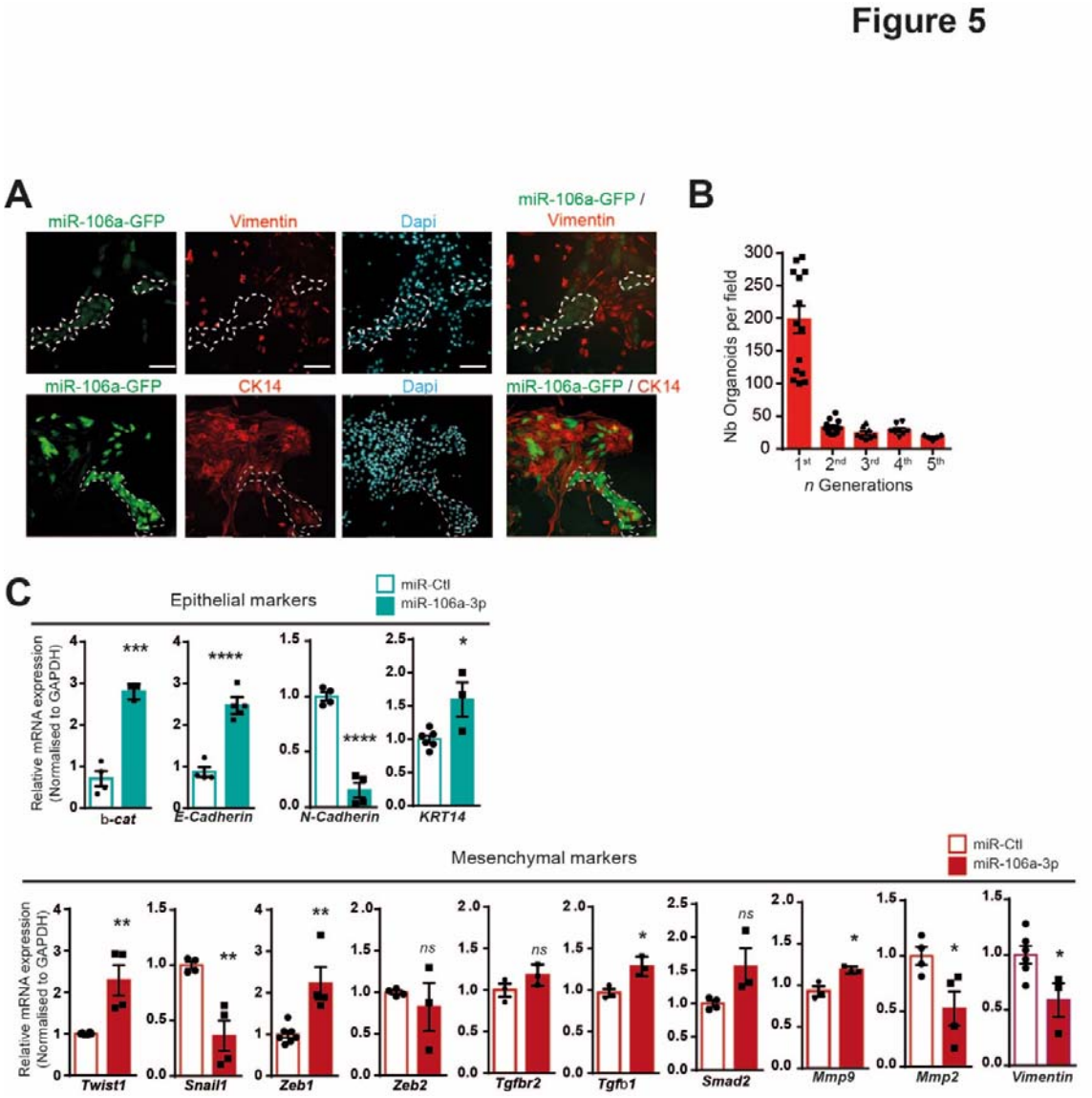

Figure 5. miR-106a-3p converts cell to a partial EMT-hybrid stade.
A, Representative immunofluorescence (IF) analysis depicting loss of Vimentin expression stained for Vimentin (red), in miR-106a-GFP infected cells (green) as compared to non-infected cells, and nuclei with DAPI (blue). Additionally, basal/myoepithelial markers stained for CK14 (green), and DAPI. **B,** Quantification n generation of 3D structure development following miR-106a cell infection of cells. Data are presented as mean values□±□SD. Data are from at least four independent experiments. (mean±SEM). **C**, Relative mRNA expression levels of key regulators of EMT determined by RT-qPCR in cells transfected with control miRNA (miR-Vector) or miR-106a-3p. In all graphs, means and standard errors are shown, and statistical significance by Student's t test is indicated by one ($p < 0.05$), two ($p < 0.01$), or four ($p < 0.0001$) asterisks. Data are from at least three independent experiments. (mean±SEM).

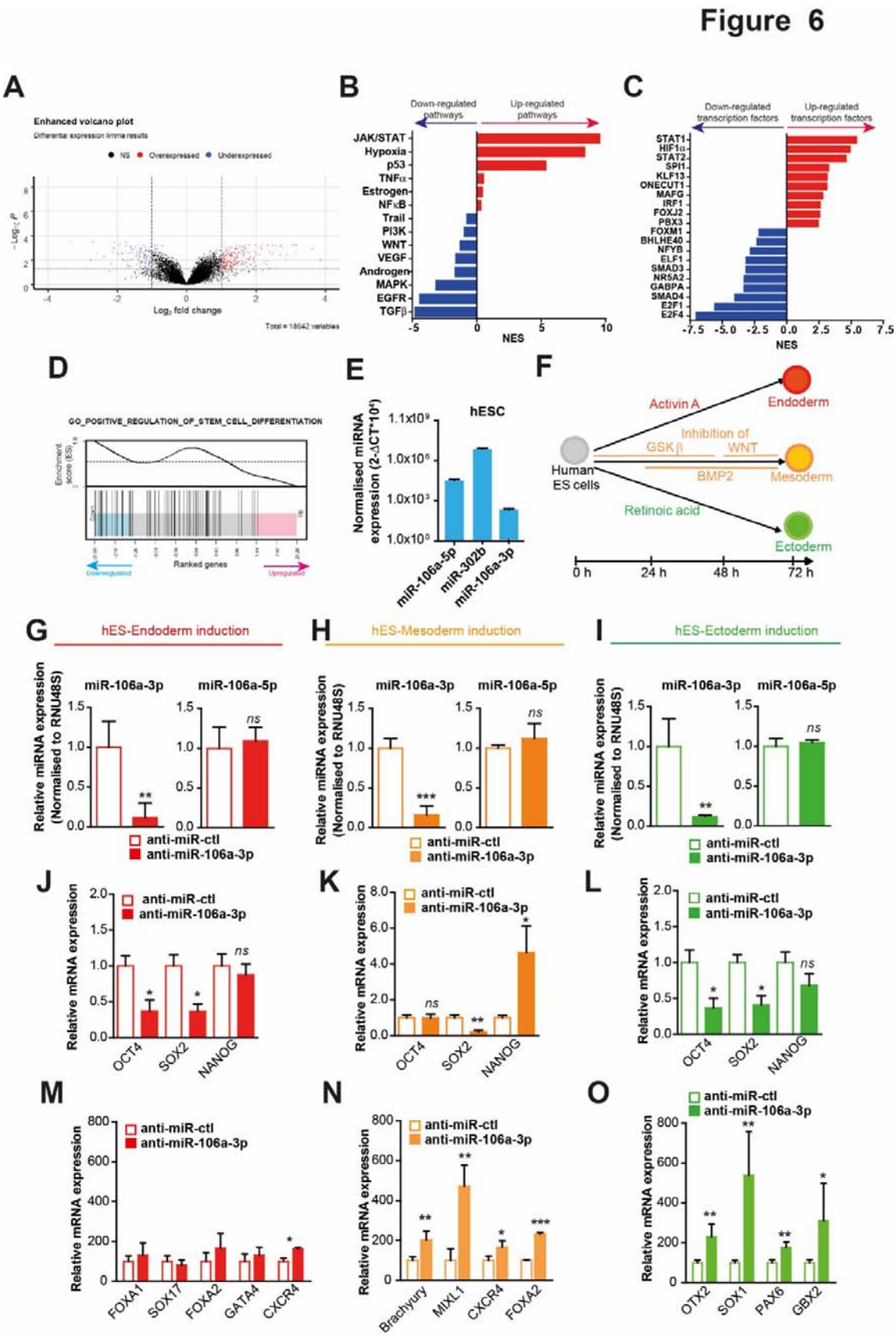

Figure 6. miR-106a-3p targets involved in downregulation of stem cell differentiation

A, Volcano plot comparing the gene expression fold changes and p-values between HMECs transfected with control mimic or miR-106a-3p, using the R package Enhanced Volcano. The plot shows the distribution of the total number of 18,942 genes in the core gene category were tested. Genes with fold change > 1 and FDR value < 0.05 are indicated in red, and genes with fold change < −1 and FDR value < 0.05 are indicated in blue. Data resulted in significant changes in the expression of 1,348 genes. **B-C**, Barplots illustrating 14 main pathways and the top-10 regulated TFs (up and down-regulated) following miR-106a-3p expression. The visualized numeric values are normalized enrichment scores (NES) of the top100 most responsive genes for each indicated pathway, whereas for the relative TF activities, the signed transcription factor (TF) - target gene interactions when utilized based on the DoRothEA database (weighted mean statistic). The relative molecular footprint alterations in control as compared to miR-106a-3p are represented by each bar, where NES < 0 represents down-regulation of specified pathway (B) or transcription factors (C). NES > 0 represents upregulation of specified pathway (B) or transcription factors (C). An absolute value of 2/-2 denotes the relative significance of the differential activity level. **D**, Barcode enrichment plot from the gene set enrichment analysis results, indicating down-regulation of stem cell differentiation pathway-related genes in the miR-106-3p-transfected group. The statistical significance of the enrichment is an FDR value of 0.024 and it is based on the multiple rotation gene-set testing (mroast) from the Limma R package (https://doi.org/10.1093/bioinformatics/btq401). **E,** Relative miR-106a-3p, miR-106a-5p and miR-302b expression levels determined by RT-qPCR in hESCs. Data are from at least three independent experiments. (mean±SEM). **F**, Schematic representation of the human ESCs differentiation process, including timeline and key signaling pathways that are modulated. **G-I**, Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in hESCs cells transfected with control mimic or anti-miR-106a-3p following induction of the endoderm (G), mesoderm (H) and ectoderm (I) differentiation. (n=3, mean ± SEM) **J-L**, Relative mRNA expression levels of key regulators of pluripotency in hESCs transfected with control mimic or anti-miR-106a-3p following induction of endoderm (J), mesoderm (K) and ectoderm (L) differentiation. In all graphs, means and standard errors are shown, and statistical significance by Student's t test is indicated by one ($p < 0.05$), two ($p < 0.01$), or three ($p < 0.001$) asterisks. (n=3, mean ± SEM) **M-O,** Relative mRNA expression levels of differentiation target genes in hESCs transfected with control mimic or anti-miR-106a-3p following Endoderm induction (M), .Mesoderm induction (N) and Ectoderm induction (O). (n=3, mean ± SEM).

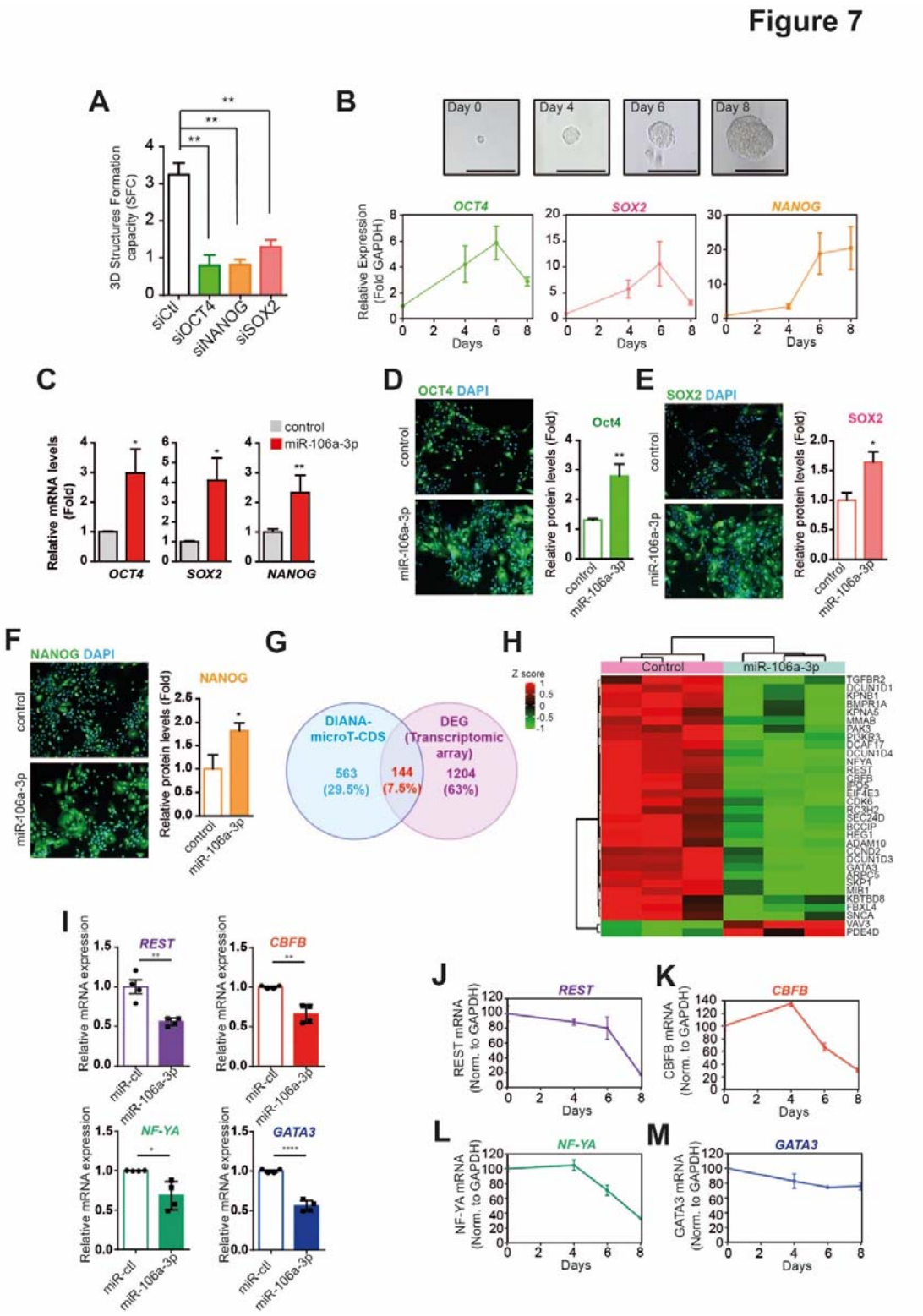

Figure 7. miR-106a-3p, a stem cell determining regulator for 3D structure development
A, SFC represents the percentage of 3D structure per seeded cell generated from HMECs transfected with siRNA control, siRNA-OCT4, siRNA-NANOG or siRNA-SOX2. (n=3, mean ± SEM) **B**, Relative *OCT4*, *SOX2* and *NANOG* expression levels determined by RT-qPCR in 3D growth of HMECs at Days 0, 4, 6, and 8. Data are from at least three independent

experiments. (mean±SEM). **C**, Relative *OCT4*, *SOX2* and *NANOG* expression levels (measured by RT-qPCR and normalized to RNA48) following miR-106a-infected HMECs transfection with LNA-control (grey) or LNA- miR-106a-3p (red). In all graphs, means and standard errors are shown, and statistical significance by Student's t test is indicated by one ($p < 0.05$), two ($p < 0.01$), or three ($p < 0.001$) asterisks. Data are from at least three independent experiments. (mean±SEM). **D-F**, Relative protein expression levels of OCT4 (D), SOX2 (E) and NANOG (F) determined by immunofluorescence analysis. Data are from at least three independent experiments. (mean±SEM). **G**, Venn diagram depicting the overlap of differentially expressed genes in the microarray experiment (1348 genes), with the predicted targets of the hsa-miR-106a-3p using the computational tool MicroT_CDS (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=MicroT_CDS/index) (707 genes), resulting in a number of 144 common genes. **H**, Heatmap of the 32 hub genes, resulted from the union of the REACTOME and GO biological processes, using the R package ComplexHeatmap **I**, *REST*, *CBFB*, *NFYA* and *GATA3* mRNA expression levels (measured by RT-qPCR and normalized to *RNA48*) following miR-106a-infected HMECs transfection with LNA-control (miR-ctl) or LNA-miR-106a-3p. Data are from at least three independent experiments. (mean±SEM). **J-M**, Relative *REST* (J), *CBFB* (K), *NFYA* (L) and *GATA3* (M) expression levels determined by RT-qPCR in 3D growth of HMECs at days 0, 4, 6, and 8. Data are from at least three independent experiments. (mean±SEM).

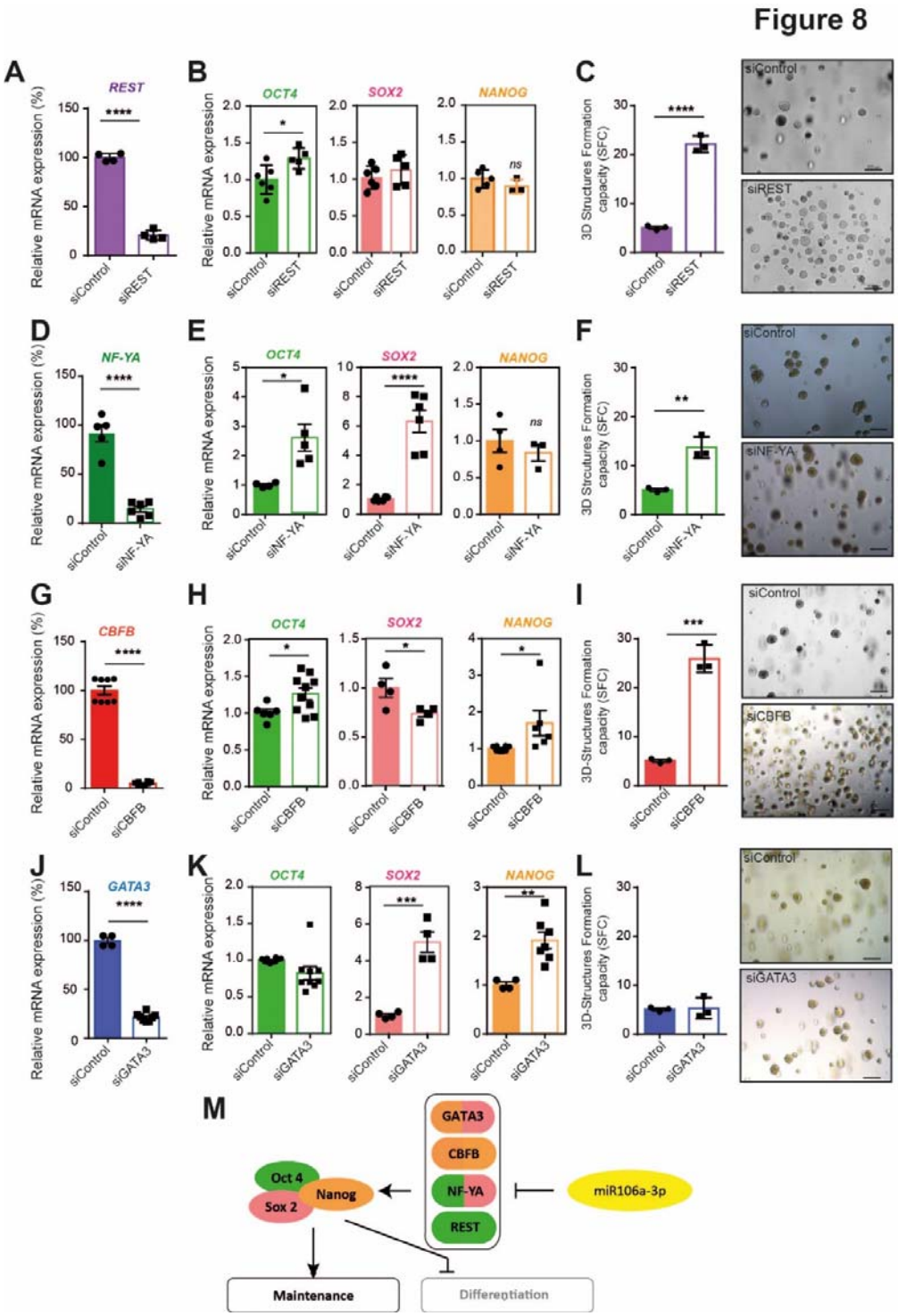

Figure 8. miR-106a-3p a barrier for cell differentiation program during mammary 3D structure initiation. A, Efficiency of *REST* gene silencing with specific human siRNA in transfected HMECs, as determined by RT-qPCR. Means and standard errors are shown. Data are from at least three independent experiments. (mean±SEM). **B**, Impact of *REST* silencing (siREST) on *OCT4*, *SOX2* and *NANOG* gene expression as determined by RT-qPCR. Data are from at least three independent experiments. (mean±SEM). **C**, SFC represents the percentage of 3D structures formed by seeded for control-transfected HMECs (siControl) as compared to siREST-transfected HMECs. Representative brightfield pictures of organoids from si-Control-

and si-REST-transfected HMECs. Data are from at least three independent experiments. (mean±SEM). **D**, Efficiency of *NFYA* gene silencing with specific human siRNA in transfected HMECs as determined by RT-qPCR. Means and standard errors are shown. Data are from at least three independent experiments. (mean±SEM). **E**, Impact of *NFYA* silencing (siNFYA) on *OCT4*, *SOX2* and *NANOG* gene expression as determined by RT-qPCR. **F**, SFC represents the percentage of 3D structures formed by seeded for control-transfected HMECs (siControl) as compared to siNFYA-transfected HMECs. Representative brightfield pictures of 3D structures from control- and siNFYA-transfected HMECs. Data are from at least three independent experiments. (mean±SEM). **G**, Efficiency of *CBFB* gene silencing with specific human siRNA in transfected HMECs as determined by RT-qPCR. Data are from at least three independent experiments. (mean±SEM).**H**, Impact of *CBFB* silencing (siCBFB) on *OCT4*, *SOX2* and *NANOG* gene expression as determined by RT-qPCR. Data are from at least three independent experiments. (mean±SEM). **I**, SFC represents the percentage of 3D structures formed by seeded for control-transfected HMECs (si-control) as compared to siCBFB-transfected HMECs. Representative brightfield pictures of 3D structures formed from HMECs control- and siCBFB-transfected HMECs. Data are from at least three independent experiments. (mean±SEM). **J**, Efficiency of *GATA3* gene silencing with specific human siRNA in transfected HMECs as determined by RT-qPCR. Data are from at least three independent experiments. (mean±SEM). **K**, Impact of *GATA3* silencing (siGATA3) on *OCT4*, *SOX2* and *NANOG* gene expression as determined by RT-qPCR. Data are from at least three independent experiments. (mean±SEM). **L**, SFC represents the percentage of 3D structures formed by seeded for control-transfected HMECs (si-control) as compared to siGATA3-transfected HMECs. Representative brightfield pictures of 3D structures formed from HMECs control- and siGATA3-transfected HMECs. Data are from at least three independent experiments. (mean±SEM). **M,** Schematic representation of the mechanism initiated by miR-106a-3p. miR-106a-3p acts on 4 main transcription factors CBFB, NF-YA, GATA3 and REST. These transcription factors regulate the expression of OCT4, SOX2 and NANOG, which in turns favors stem cell maintenance and block cell differentiation.

1007 **References**

1008 1. Fessart, D., Begueret, H. & Delom, F. (2013) Three-dimensional culture model to distinguish
1009 normal from malignant human bronchial epithelial cells, *The European respiratory journal.* **42**, 1345-
1010 56.
1011 2. Fessart, D., Domblides, C., Avril, T., Eriksson, L. A., Begueret, H., Pineau, R., Malrieux, C., Dugot-
1012 Senant, N., Lucchesi, C., Chevet, E. & Delom, F. (2016) Secretion of protein disulphide isomerase
1013 AGR2 confers tumorigenic properties, *eLife.* **5**.
1014 3. Schwarz, J. S., de Jonge, H. R. & Forrest, J. N., Jr. (2015) Value of Organoids from Comparative
1015 Epithelia Models, *The Yale journal of biology and medicine.* **88**, 367-74.
1016 4. Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F.,
1017 Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J. & Weinberg, R. A. (2008) The
1018 epithelial-mesenchymal transition generates cells with properties of stem cells, *Cell.* **133**, 704-15.
1019 5. Pardal, R., Clarke, M. F. & Morrison, S. J. (2003) Applying the principles of stem-cell biology to
1020 cancer, *Nature reviews Cancer.* **3**, 895-902.
1021 6. Kim, S. Y., Kang, J. W., Song, X., Kim, B. K., Yoo, Y. D., Kwon, Y. T. & Lee, Y. J. (2013) Role of the IL-6-
1022 JAK1-STAT3-Oct-4 pathway in the conversion of non-stem cancer cells into cancer stem-like cells,
1023 *Cellular signalling.* **25**, 961-9.
1024 7. Zhao, D., Pan, C., Sun, J., Gilbert, C., Drews-Elger, K., Azzam, D. J., Picon-Ruiz, M., Kim, M., Ullmer,
1025 W., El-Ashry, D., Creighton, C. J. & Slingerland, J. M. (2015) VEGF drives cancer-initiating stem cells
1026 through VEGFR-2/Stat3 signaling to upregulate Myc and Sox2, *Oncogene.* **34**, 3107-19.
1027 8. Lu, C. S., Shieh, G. S., Wang, C. T., Su, B. H., Su, Y. C., Chen, Y. C., Su, W. C., Wu, P., Yang, W. H.,
1028 Shiau, A. L. & Wu, C. L. (2017) Chemotherapeutics-induced Oct4 expression contributes to drug
1029 resistance and tumor recurrence in bladder cancer, *Oncotarget.* **8**, 30844-30858.
1030 9. Wang, X. Q., Ongkeko, W. M., Chen, L., Yang, Z. F., Lu, P., Chen, K. K., Lopez, J. P., Poon, R. T. &
1031 Fan, S. T. (2010) Octamer 4 (Oct4) mediates chemotherapeutic drug resistance in liver cancer cells
1032 through a potential Oct4-AKT-ATP-binding cassette G2 pathway, *Hepatology.* **52**, 528-39.
1033 10. Loh, Y. H., Ng, J. H. & Ng, H. H. (2008) Molecular framework underlying pluripotency, *Cell cycle.* **7**,
1034 885-91.
1035 11. Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P.,
1036 Mansfield, W., Walker, R., Tomlinson, S. & Smith, A. (2009) Oct4 and LIF/Stat3 additively induce
1037 Kruppel factors to sustain embryonic stem cell self-renewal, *Cell stem cell.* **5**, 597-609.
1038 12. Tai, M. H., Chang, C. C., Kiupel, M., Webster, J. D., Olson, L. K. & Trosko, J. E. (2005) Oct4
1039 expression in adult human stem cells: evidence in support of the stem cell theory of carcinogenesis,
1040 *Carcinogenesis.* **26**, 495-502.
1041 13. Schoenhals, M., Kassambara, A., De Vos, J., Hose, D., Moreaux, J. & Klein, B. (2009) Embryonic
1042 stem cell markers expression in cancers, *Biochemical and biophysical research communications.* **383**,
1043 157-62.
1044 14. Ezeh, U. I., Turek, P. J., Reijo, R. A. & Clark, A. T. (2005) Human embryonic stem cell genes OCT4,
1045 NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma, *Cancer.* **104**,
1046 2255-65.
1047 15. Jin, T., Branch, D. R., Zhang, X., Qi, S., Youngson, B. & Goss, P. E. (1999) Examination of POU
1048 homeobox gene expression in human breast cancer cells, *International journal of cancer.* **81**, 104-12.
1049 16. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. & Smith, A. (2003)
1050 Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells, *Cell.*
1051 **113**, 643-55.
1052 17. Mathieu, J. & Ruohola-Baker, H. (2013) Regulation of stem cell populations by microRNAs,
1053 *Advances in experimental medicine and biology.* **786**, 329-51.
1054 18. Stefani, G. & Slack, F. J. (2008) Small non-coding RNAs in animal development, *Nature reviews*
1055 *Molecular cell biology.* **9**, 219-30.
1056 19. Garbe, J. C., Pepin, F., Pelissier, F. A., Sputova, K., Fridriksdottir, A. J., Guo, D. E., Villadsen, R.,
1057 Park, M., Petersen, O. W., Borowsky, A. D., Stampfer, M. R. & Labarge, M. A. (2012) Accumulation of

1058 multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia,
1059 *Cancer research.* **72**, 3687-701.
1060 20. Ginestier, C., Hur, M. H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J.,
1061 Viens, P., Kleer, C. G., Liu, S., Schott, A., Hayes, D., Birnbaum, D., Wicha, M. S. & Dontu, G. (2007)
1062 ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor
1063 clinical outcome, *Cell stem cell.* **1**, 555-67.
1064 21. Taylor-Papadimitriou, J., Stampfer, M., Bartek, J., Lewis, A., Boshell, M., Lane, E. B. & Leigh, I. M.
1065 (1989) Keratin expression in human mammary epithelial cells cultured from normal and malignant
1066 tissue: relation to in vivo phenotypes and influence of medium, *Journal of cell science.* **94 (Pt 3)**, 403-
1067 13.
1068 22. Stampfer, M. R. & Yaswen, P. (2000) Culture models of human mammary epithelial cell
1069 transformation, *Journal of mammary gland biology and neoplasia.* **5**, 365-78.
1070 23. Labarge, M. A., Garbe, J. C. & Stampfer, M. R. (2013) Processing of human reduction
1071 mammoplasty and mastectomy tissues for cell culture, *Journal of visualized experiments : JoVE*.
1072 24. Garbe, J. C., Bhattacharya, S., Merchant, B., Bassett, E., Swisshelm, K., Feiler, H. S., Wyrobek, A. J.
1073 & Stampfer, M. R. (2009) Molecular distinctions between stasis and telomere attrition senescence
1074 barriers shown by long-term culture of normal human mammary epithelial cells, *Cancer research.* **69**,
1075 7557-68.
1076 25. Manuel Iglesias, J., Beloqui, I., Garcia-Garcia, F., Leis, O., Vazquez-Martin, A., Eguiara, A., Cufi, S.,
1077 Pavon, A., Menendez, J. A., Dopazo, J. & Martin, A. G. (2013) Mammosphere formation in breast
1078 carcinoma cell lines depends upon expression of E-cadherin, *PloS one.* **8**, e77281.
1079 26. Ghebeh, H., Sleiman, G. M., Manogaran, P. S., Al-Mazrou, A., Barhoush, E., Al-Mohanna, F. H.,
1080 Tulbah, A., Al-Faqeeh, K. & Adra, C. N. (2013) Profiling of normal and malignant breast tissue show
1081 CD44high/CD24low phenotype as a predominant stem/progenitor marker when used in combination
1082 with Ep-CAM/CD49f markers, *BMC cancer.* **13**, 289.
1083 27. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. (2003) Prospective
1084 identification of tumorigenic breast cancer cells, *Proceedings of the National Academy of Sciences of*
1085 *the United States of America.* **100**, 3983-8.
1086 28. Sleeman, K. E., Kendrick, H., Ashworth, A., Isacke, C. M. & Smalley, M. J. (2006) CD24 staining of
1087 mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells,
1088 *Breast cancer research : BCR.* **8**, R7.
1089 29. Cordenonsi, M., Zanconato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., Inui, M.,
1090 Montagner, M., Parenti, A. R., Poletti, A., Daidone, M. G., Dupont, S., Basso, G., Bicciato, S. & Piccolo,
1091 S. (2011) The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells, *Cell.*
1092 **147**, 759-72.
1093 30. Dontu, G., Abdallah, W. M., Foley, J. M., Jackson, K. W., Clarke, M. F., Kawamura, M. J. & Wicha,
1094 M. S. (2003) In vitro propagation and transcriptional profiling of human mammary stem/progenitor
1095 cells, *Genes & development.* **17**, 1253-70.
1096 31. Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., Pilotti, S., Pierotti, M.
1097 A. & Daidone, M. G. (2005) Isolation and in vitro propagation of tumorigenic breast cancer cells with
1098 stem/progenitor cell properties, *Cancer research.* **65**, 5506-11.
1099 32. Wright, M. H., Calcagno, A. M., Salcido, C. D., Carlson, M. D., Ambudkar, S. V. & Varticovski, L.
1100 (2008) Brca1 breast tumors contain distinct CD44+/CD24- and CD133+ cells with cancer stem cell
1101 characteristics, *Breast cancer research : BCR.* **10**, R10.
1102 33. Perrone, G., Gaeta, L. M., Zagami, M., Nasorri, F., Coppola, R., Borzomati, D., Bartolozzi, F.,
1103 Altomare, V., Trodella, L., Tonini, G., Santini, D., Cavani, A. & Muda, A. O. (2012) In situ identification
1104 of CD44+/CD24- cancer cells in primary human breast carcinomas, *PloS one.* **7**, e43110.
1105 34. Wang, L. B., He, Y. Q., Wu, L. G., Chen, D. M., Fan, H. & Jia, W. (2012) [Isolation and
1106 characterization of human breast tumor stem cells], *Xi bao yu fen zi mian yi xue za zhi = Chinese*
1107 *journal of cellular and molecular immunology.* **28**, 1261-4.

1108 35. Bishop, C. L., Bergin, A. M., Fessart, D., Borgdorff, V., Hatzimasoura, E., Garbe, J. C., Stampfer, M.
1109 R., Koh, J. & Beach, D. H. (2010) Primary cilium-dependent and -independent Hedgehog signaling
1110 inhibits p16(INK4A), *Molecular cell.* **40**, 533-47.
1111 36. Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A.,
1112 Casanova, J., Christofori, G., Dedhar, S., Derynck, R., Ford, H. L., Fuxe, J., Garcia de Herreros, A.,
1113 Goodall, G. J., Hadjantonakis, A. K., Huang, R. Y. J., Kalcheim, C., Kalluri, R., Kang, Y., Khew-Goodall, Y.,
1114 Levine, H., Liu, J., Longmore, G. D., Mani, S. A., Massague, J., Mayor, R., McClay, D., Mostov, K. E.,
1115 Newgreen, D. F., Nieto, M. A., Puisieux, A., Runyan, R., Savagner, P., Stanger, B., Stemmler, M. P.,
1116 Takahashi, Y., Takeichi, M., Theveneau, E., Thiery, J. P., Thompson, E. W., Weinberg, R. A., Williams, E.
1117 D., Xing, J., Zhou, B. P., Sheng, G. & Association, E. M. T. I. (2020) Guidelines and definitions for
1118 research on epithelial-mesenchymal transition, *Nature reviews Molecular cell biology.* **21**, 341-352.
1119 37. He, X. C., Zhang, J., Tong, W. G., Tawfik, O., Ross, J., Scoville, D. H., Tian, Q., Zeng, X., He, X.,
1120 Wiedemann, L. M., Mishina, Y. & Li, L. (2004) BMP signaling inhibits intestinal stem cell self-renewal
1121 through suppression of Wnt-beta-catenin signaling, *Nature genetics.* **36**, 1117-21.
1122 38. Suh, M. R., Lee, Y., Kim, J. Y., Kim, S. K., Moon, S. H., Lee, J. Y., Cha, K. Y., Chung, H. M., Yoon, H.
1123 S., Moon, S. Y., Kim, V. N. & Kim, K. S. (2004) Human embryonic stem cells express a unique set of
1124 microRNAs, *Developmental biology.* **270**, 488-98.
1125 39. Wobus, A. M. & Boheler, K. R. (2005) Embryonic stem cells: prospects for developmental biology
1126 and cell therapy, *Physiological reviews.* **85**, 635-78.
1127 40. Stefanovic, S., Abboud, N., Desilets, S., Nury, D., Cowan, C. & Puceat, M. (2009) Interplay of Oct4
1128 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate, *The*
1129 *Journal of cell biology.* **186**, 665-73.
1130 41. Van Keymeulen, A., Rocha, A. S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N.,
1131 Dekoninck, S. & Blanpain, C. (2011) Distinct stem cells contribute to mammary gland development
1132 and maintenance, *Nature.* **479**, 189-93.
1133 42. Deome, K. B., Faulkin, L. J., Jr., Bern, H. A. & Blair, P. B. (1959) Development of mammary tumors
1134 from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H
1135 mice, *Cancer research.* **19**, 515-20.
1136 43. Shackleton, M., Vaillant, F., Simpson, K. J., Stingl, J., Smyth, G. K., Asselin-Labat, M. L., Wu, L.,
1137 Lindeman, G. J. & Visvader, J. E. (2006) Generation of a functional mammary gland from a single stem
1138 cell, *Nature.* **439**, 84-8.
1139 44. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G.,
1140 Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. & Young, R. A.
1141 (2005) Core transcriptional regulatory circuitry in human embryonic stem cells, *Cell.* **122**, 947-56.
1142 45. LaMarca, H. L. & Rosen, J. M. (2008) Minireview: hormones and mammary cell fate--what will I
1143 become when I grow up?, *Endocrinology.* **149**, 4317-21.
1144 46. Simoes, B. M., Piva, M., Iriondo, O., Comaills, V., Lopez-Ruiz, J. A., Zabalza, I., Mieza, J. A., Acinas,
1145 O. & Vivanco, M. D. (2011) Effects of estrogen on the proportion of stem cells in the breast, *Breast*
1146 *cancer research and treatment.* **129**, 23-35.
1147 47. Debnath, J. & Brugge, J. S. (2005) Modelling glandular epithelial cancers in three-dimensional
1148 cultures, *Nature reviews Cancer.* **5**, 675-88.
1149 48. Jagadeeswaran, G., Zheng, Y., Sumathipala, N., Jiang, H., Arrese, E. L., Soulages, J. L., Zhang, W. &
1150 Sunkar, R. (2010) Deep sequencing of small RNA libraries reveals dynamic regulation of conserved
1151 and novel microRNAs and microRNA-stars during silkworm development, *BMC genomics.* **11**, 52.
1152 49. Kuchenbauer, F., Mah, S. M., Heuser, M., McPherson, A., Ruschmann, J., Rouhi, A., Berg, T.,
1153 Bullinger, L., Argiropoulos, B., Morin, R. D., Lai, D., Starczynowski, D. T., Karsan, A., Eaves, C. J.,
1154 Watahiki, A., Wang, Y., Aparicio, S. A., Ganser, A., Krauter, J., Dohner, H., Dohner, K., Marra, M. A.,
1155 Camargo, F. D., Palmqvist, L., Buske, C. & Humphries, R. K. (2011) Comprehensive analysis of
1156 mammalian miRNA* species and their role in myeloid cells, *Blood.* **118**, 3350-8.
1157 50. Griffiths-Jones, S., Hui, J. H., Marco, A. & Ronshaugen, M. (2011) MicroRNA evolution by arm
1158 switching, *EMBO reports.* **12**, 172-7.

1159 51. Yang, J. S., Phillips, M. D., Betel, D., Mu, P., Ventura, A., Siepel, A. C., Chen, K. C. & Lai, E. C. (2011)
1160 Widespread regulatory activity of vertebrate microRNA* species, *Rna.* **17**, 312-26.
1161 52. Guo, L. & Lu, Z. (2010) The fate of miRNA* strand through evolutionary analysis: implication for
1162 degradation as merely carrier strand or potential regulatory molecule?, *PloS one.* **5**, e11387.
1163 53. Chao, C. H., Chang, C. C., Wu, M. J., Ko, H. W., Wang, D., Hung, M. C., Yang, J. Y. & Chang, C. J.
1164 (2014) MicroRNA-205 signaling regulates mammary stem cell fate and tumorigenesis, *The Journal of*
1165 *clinical investigation.* **124**, 3093-106.
1166 54. Wang, W., Corrigan-Cummins, M., Hudson, J., Maric, I., Simakova, O., Neelapu, S. S., Kwak, L. W.,
1167 Janik, J. E., Gause, B., Jaffe, E. S. & Calvo, K. R. (2012) MicroRNA profiling of follicular lymphoma
1168 identifies microRNAs related to cell proliferation and tumor response, *Haematologica.* **97**, 586-94.
1169 55. Guo, W., Li, W., Yuan, L., Mei, X. & Hu, W. (2019) MicroRNA-106a-3p Induces Apatinib Resistance
1170 and Activates Janus-Activated Kinase 2 (JAK2)/Signal Transducer and Activator of Transcription 3
1171 (STAT3) by Targeting the SOCS System in Gastric Cancer, *Medical science monitor : international*
1172 *medical journal of experimental and clinical research.* **25**, 10122-10128.
1173 56. Ma, Y., Zhang, H., He, X., Song, H., Qiang, Y., Li, Y., Gao, J. & Wang, Z. (2015) miR-106a* inhibits
1174 the proliferation of renal carcinoma cells by targeting IRS-2, *Tumour biology : the journal of the*
1175 *International Society for Oncodevelopmental Biology and Medicine.* **36**, 8389-98.
1176 57. Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van
1177 Keymeulen, A., Brown, D., Moers, V., Lemaire, S., De Clercq, S., Minguijon, E., Balsat, C., Sokolow, Y.,
1178 Dubois, C., De Cock, F., Scozzaro, S., Sopena, F., Lanas, A., D'Haene, N., Salmon, I., Marine, J. C., Voet,
1179 T., Sotiropoulou, P. A. & Blanpain, C. (2018) Identification of the tumour transition states occurring
1180 during EMT, *Nature.* **556**, 463-468.
1181 58. Beck, B., Lapouge, G., Rorive, S., Drogat, B., Desaedelaere, K., Delafaille, S., Dubois, C., Salmon, I.,
1182 Willekens, K., Marine, J. C. & Blanpain, C. (2015) Different levels of Twist1 regulate skin tumor
1183 initiation, stemness, and progression, *Cell stem cell.* **16**, 67-79.
1184 59. Schmidt, J. M., Panzilius, E., Bartsch, H. S., Irmler, M., Beckers, J., Kari, V., Linnemann, J. R.,
1185 Dragoi, D., Hirschi, B., Kloos, U. J., Sass, S., Theis, F., Kahlert, S., Johnsen, S. A., Sotlar, K. & Scheel, C.
1186 H. (2015) Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1
1187 activation, *Cell reports.* **10**, 131-9.
1188 60. Sacchetti, A., Teeuwssen, M., Verhagen, M., Joosten, R., Xu, T., Stabile, R., van der Steen, B.,
1189 Watson, M. M., Gusinac, A., Kim, W. K., Ubink, I., Van de Werken, H. J., Fumagalli, A., Paauwe, M.,
1190 Van Rheenen, J., Sansom, O. J., Kranenburg, O. & Fodde, R. (2021) Phenotypic plasticity underlies
1191 local invasion and distant metastasis in colon cancer, *eLife.* **10**.
1192 61. Semenza, G. L. (2016) Dynamic regulation of stem cell specification and maintenance by hypoxia-
1193 inducible factors, *Molecular aspects of medicine.* **47-48**, 15-23.
1194 62. Hernandez-Vargas, H., Ouzounova, M., Le Calvez-Kelm, F., Lambert, M. P., McKay-Chopin, S.,
1195 Tavtigian, S. V., Puisieux, A., Matar, C. & Herceg, Z. (2011) Methylome analysis reveals Jak-STAT
1196 pathway deregulation in putative breast cancer stem cells, *Epigenetics.* **6**, 428-39.
1197 63. Watabe, T. & Miyazono, K. (2009) Roles of TGF-beta family signaling in stem cell renewal and
1198 differentiation, *Cell research.* **19**, 103-15.
1199 64. Solozobova, V. & Blattner, C. (2011) p53 in stem cells, *World journal of biological chemistry.* **2**,
1200 202-14.
1201 65. Cicalese, A., Bonizzi, G., Pasi, C. E., Faretta, M., Ronzoni, S., Giulini, B., Brisken, C., Minucci, S., Di
1202 Fiore, P. P. & Pelicci, P. G. (2009) The tumor suppressor p53 regulates polarity of self-renewing
1203 divisions in mammary stem cells, *Cell.* **138**, 1083-95.
1204 66. Westbrook, T. F., Hu, G., Ang, X. L., Mulligan, P., Pavlova, N. N., Liang, A., Leng, Y., Maehr, R., Shi,
1205 Y., Harper, J. W. & Elledge, S. J. (2008) SCFbeta-TRCP controls oncogenic transformation and neural
1206 differentiation through REST degradation, *Nature.* **452**, 370-4.
1207 67. Johnson, R., Teh, C. H., Kunarso, G., Wong, K. Y., Srinivasan, G., Cooper, M. L., Volta, M., Chan, S.
1208 S., Lipovich, L., Pollard, S. M., Karuturi, R. K., Wei, C. L., Buckley, N. J. & Stanton, L. W. (2008) REST
1209 regulates distinct transcriptional networks in embryonic and neural stem cells, *PLoS biology.* **6**, e256.

1210 68. van Arensbergen, J., Garcia-Hurtado, J., Moran, I., Maestro, M. A., Xu, X., Van de Casteele, M.,
1211 Skoudy, A. L., Palassini, M., Heimberg, H. & Ferrer, J. (2010) Derepression of Polycomb targets during
1212 pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program,
1213 *Genome research.* **20**, 722-32.
1214 69. Speck, N. A. & Terryl, S. (1995) A new transcription factor family associated with human
1215 leukemias, *Critical reviews in eukaryotic gene expression.* **5**, 337-64.
1216 70. Tang, L., Guo, B., Javed, A., Choi, J. Y., Hiebert, S., Lian, J. B., van Wijnen, A. J., Stein, J. L., Stein, G.
1217 S. & Zhou, G. W. (1999) Crystal structure of the nuclear matrix targeting signal of the transcription
1218 factor acute myelogenous leukemia-1/polyoma enhancer-binding protein 2alphaB/core binding
1219 factor alpha2, *The Journal of biological chemistry.* **274**, 33580-6.
1220 71. Kurosaka, H., Islam, M. N., Kuremoto, K., Hayano, S., Nakamura, M., Kawanabe, N., Yanagita, T.,
1221 Rice, D. P., Harada, H., Taniuchi, I. & Yamashiro, T. (2011) Core binding factor beta functions in the
1222 maintenance of stem cells and orchestrates continuous proliferation and differentiation in mouse
1223 incisors, *Stem cells.* **29**, 1792-803.
1224 72. Growney, J. D., Shigematsu, H., Li, Z., Lee, B. H., Adelsperger, J., Rowan, R., Curley, D. P., Kutok, J.
1225 L., Akashi, K., Williams, I. R., Speck, N. A. & Gilliland, D. G. (2005) Loss of Runx1 perturbs adult
1226 hematopoiesis and is associated with a myeloproliferative phenotype, *Blood.* **106**, 494-504.
1227 73. Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L. & Karsenty, G. (1997) Osf2/Cbfa1: a transcriptional
1228 activator of osteoblast differentiation, *Cell.* **89**, 747-54.
1229 74. Inoue, K., Ozaki, S., Shiga, T., Ito, K., Masuda, T., Okado, N., Iseda, T., Kawaguchi, S., Ogawa, M.,
1230 Bae, S. C., Yamashita, N., Itohara, S., Kudo, N. & Ito, Y. (2002) Runx3 controls the axonal projection of
1231 proprioceptive dorsal root ganglion neurons, *Nature neuroscience.* **5**, 946-54.
1232 75. Li, Q. L., Ito, K., Sakakura, C., Fukamachi, H., Inoue, K., Chi, X. Z., Lee, K. Y., Nomura, S., Lee, C. W.,
1233 Han, S. B., Kim, H. M., Kim, W. J., Yamamoto, H., Yamashita, N., Yano, T., Ikeda, T., Itohara, S.,
1234 Inazawa, J., Abe, T., Hagiwara, A., Yamagishi, H., Ooe, A., Kaneda, A., Sugimura, T., Ushijima, T., Bae,
1235 S. C. & Ito, Y. (2002) Causal relationship between the loss of RUNX3 expression and gastric cancer,
1236 *Cell.* **109**, 113-24.
1237 76. Visvader, J. E. & Lindeman, G. J. (2003) Transcriptional regulators in mammary gland
1238 development and cancer, *The international journal of biochemistry & cell biology.* **35**, 1034-51.
1239 77. Brown, M. S., Abdollahi, B., Wilkins, O. M., Lu, H., Chakraborty, P., Ognjenovic, N. B., Muller, K. E.,
1240 Jolly, M. K., Christensen, B. C., Hassanpour, S. & Pattabiraman, D. R. (2022) Phenotypic heterogeneity
1241 driven by plasticity of the intermediate EMT state governs disease progression and metastasis in
1242 breast cancer, *Science advances.* **8**, eabj8002.
1243 78. Serra, E., Zemzoumi, K., di Silvio, A., Mantovani, R., Lardans, V. & Dissous, C. (1998) Conservation
1244 and divergence of NF-Y transcriptional activation function, *Nucleic acids research.* **26**, 3800-5.
1245 79. Dai, C., Miao, C. X., Xu, X. M., Liu, L. J., Gu, Y. F., Zhou, D., Chen, L. S., Lin, G. & Lu, G. X. (2015)
1246 Transcriptional activation of human CDCA8 gene regulated by transcription factor NF-Y in embryonic
1247 stem cells and cancer cells, *The Journal of biological chemistry.* **290**, 22423-34.
1248 80. Moeinvaziri, F. & Shahhoseini, M. (2015) Epigenetic role of CCAAT box-binding transcription
1249 factor NF-Y on ID gene family in human embryonic carcinoma cells, *IUBMB life.* **67**, 880-7.
1250 81. Bungartz, G., Land, H., Scadden, D. T. & Emerson, S. G. (2012) NF-Y is necessary for
1251 hematopoietic stem cell proliferation and survival, *Blood.* **119**, 1380-9.
1252 82. Mojsin, M., Topalovic, V., Marjanovic Vicentic, J. & Stevanovic, M. (2015) Transcription factor NF-
1253 Y inhibits cell growth and decreases SOX2 expression in human embryonal carcinoma cell line
1254 NT2/D1, *Biochemistry Biokhimiia.* **80**, 202-7.
1255 83. Shi, Z., Chiang, C. I., Labhart, P., Zhao, Y., Yang, J., Mistretta, T. A., Henning, S. J., Maity, S. N. &
1256 Mori-Akiyama, Y. (2015) Context-specific role of SOX9 in NF-Y mediated gene regulation in colorectal
1257 cancer cells, *Nucleic acids research.* **43**, 6257-69.
1258 84. Petrovic, I., Kovacevic-Grujicic, N. & Stevanovic, M. (2009) ZBP-89 and Sp3 down-regulate while
1259 NF-Y up-regulates SOX18 promoter activity in HeLa cells, *Molecular biology reports.* **36**, 993-1000.

1260 85. Li, M., Baumeister, P., Roy, B., Phan, T., Foti, D., Luo, S. & Lee, A. S. (2000) ATF6 as a transcription
1261 activator of the endoplasmic reticulum stress element: thapsigargin stress-induced changes and
1262 synergistic interactions with NF-Y and YY1, *Molecular and cellular biology.* **20**, 5096-106.
1263 86. Li, C., Wu, B., Li, Y., Chen, J., Ye, Z., Tian, X., Wang, J., Xu, X., Pan, S., Zheng, Y., Cai, X., Jiang, L. &
1264 Zhao, M. (2022) Amino acid catabolism regulates hematopoietic stem cell proteostasis via a GCN2-
1265 eIF2alpha axis, *Cell stem cell.* **29**, 1119-1134 e7.
1266 87. Asselin-Labat, M. L., Sutherland, K. D., Barker, H., Thomas, R., Shackleton, M., Forrest, N. C.,
1267 Hartley, L., Robb, L., Grosveld, F. G., van der Wees, J., Lindeman, G. J. & Visvader, J. E. (2007) Gata-3
1268 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation, *Nature*
1269 *cell biology.* **9**, 201-9.
1270 88. Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H. I. & Eaves, C. J. (2006)
1271 Purification and unique properties of mammary epithelial stem cells, *Nature.* **439**, 993-7.
1272 89. Vaillant, F., Lindeman, G. J. & Visvader, J. E. (2011) Jekyll or Hyde: does Matrigel provide a more
1273 or less physiological environment in mammary repopulating assays?, *Breast cancer research : BCR.*
1274 **13**, 108.
1275 90. Spike, B. T., Engle, D. D., Lin, J. C., Cheung, S. K., La, J. & Wahl, G. M. (2012) A mammary stem cell
1276 population identified and characterized in late embryogenesis reveals similarities to human breast
1277 cancer, *Cell stem cell.* **10**, 183-97.
1278 91. Yi, R., Poy, M. N., Stoffel, M. & Fuchs, E. (2008) A skin microRNA promotes differentiation by
1279 repressing 'stemness', *Nature.* **452**, 225-9.
1280 92. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. &
1281 Jones, J. M. (1998) Embryonic stem cell lines derived from human blastocysts, *Science.* **282**, 1145-7.
1282 93. Doffou, M., Adams, G., Bowen, W. C., Paranjpe, S., Parihar, H. S., Nguyen, H., Michalopoulos, G.
1283 K. & Bhave, V. S. (2018) Oct4 Is Crucial for Transdifferentiation of Hepatocytes to Biliary Epithelial
1284 Cells in an In Vitro Organoid Culture Model, *Gene expression.* **18**, 51-62.
1285 94. Niwa, H., Miyazaki, J. & Smith, A. G. (2000) Quantitative expression of Oct-3/4 defines
1286 differentiation, dedifferentiation or self-renewal of ES cells, *Nature genetics.* **24**, 372-6.
1287 95. Lee, J. K., Bloom, J., Zubeldia-Plazaola, A., Garbe, J. C., Stampfer, M. R. & LaBarge, M. A. (2018)
1288 Different culture media modulate growth, heterogeneity, and senescence in human mammary
1289 epithelial cell cultures, *PloS one.* **13**, e0204645.
1290