A miRNA screen identifies a transcriptional program controlling adult stem cell maintenance in organoids

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

Organoids development relies on the self-organizing properties of adult stem cells to create structures which recapitulate the architecture, functionality, and genetic signature observed in original tissues. Little is known about of the exact nature of the intrinsic cell properties at the origin of organoid generation, and of the signaling pathways governing their plasticity and differentiation. Herein, we carried out a microRNA screen to functionally track adult stem cell from human mammary organoids epithelial cell culture. We uncover miR-106a-3p that enrich adult stem cell-like lineage, defined by CD44 and CD24 expression, and promotes organoids expansion. Transcriptomic analysis reveal that this miRNA acts through the regulation of a set of regulators of transcription *REST*, *CBFB* and *NF-YA*, and thus strengthening the adult stem cell maintenance of human epithelial cells. These data demonstrate that organoids can be directly generated from human epithelial cells by only one miRNA: miR-106a-3p. Hence, we reveal that a transcriptional program is clearly in place to elicit adult stem cells maintenance during organoids development.

Introduction

Three-dimensional (3D) human organoid culture models are appealing tools to study pathophysiological processes. These models have been described, by us and others, for lung[1, 2] as well as for numerous other organs.[3] The term "organoid" literally means organ-like, reflecting the ability of organoid culture conditions to prompt cells to self-organize into structures mimicking the architecture of the organ from which they derive. In contrast to organ explants, organoids can arise from a single primary cell,[1, 2, 4] thereby allowing the generation of human organoids from biopsies[1]. Non-tumor organoids are thought to arise from adult stem cells (aSCs), and therefore should in theory be capable of self-renewal and differentiation. However, there is currently a lack of understanding of the underlying epigenetic and genetic mechanisms that control organoid initiation, the maintenance of stem-cell and differentiation during organogenesis.

aSCs are highly specialized cells that can develop into more than one cell type, but can only produce the tissue in which they reside: as such, they are considered as 'multipotent', which corresponds to less polyvalence than pluripotent cells. Indeed, pluripotent cells are embryonic stem cells (ESC) and can give rise to all cell types. Consistent with this idea, primary cell populations enriched with known progenitor/stem cell markers are more efficient at forming organoids than the general cell populations.[5] The development of breast stem cells as well as their enrichment may rely on the orchestration of multiple critical transcription processes. Interestingly, increasing amounts of data show that the same molecular pathways manage self-renewal of normal stem cells and that of Cancer Stem Cells (CSCs) in tumors.[6] CSCs have been shown to express members of the Embryonic Stem Cells (ESC) core pluripotency complex, including octamer-binding transcription factor-4 (OCT4), SRY-Box Transcription Factor 2 (SOX2) and NANOG, which are positively correlated with the development of CSCs.[7, 8, 9, 10, 11, 12] The expression of OCT3/4 has been reported in differentiated normal and malignant human cells,[13] that of SOX2 in several malignant tissues[14] and that of NANOG in human neoplasms, including germ cell tumors, breast carcinoma, and osteosarcoma.[15] In addition, there is evidence that SOX2 is expressed in tumor-spheres, which can be generated from breast cancer tumors and cell lines.[16] These observations suggest that adult breast stem cells may contain cells with reminiscent properties of embryonic-like stem cells.[15, 17] So far, it is not known whether a comparable molecular signature of stemness — such as a minimal core transcriptional program — might be shared in adult stem cells.

In order to better understand mammalian development, as well as exploit the tremendous therapeutic potential of organoid models, it is necessary to identify and characterize the epigenetic and genetic mechanisms governing the fate of aSCs. MicroRNAs (miRNAs) have been shown to play an important role in regulating stem cell self-renewal and differentiation.[18] In general, one gene can be repressed by multiple miRNAs and one miRNA may repress multiple target genes, which results in the formation of complex regulatory networks. In a wide variety of developmental processes, miRNAs finely tune cellular identities by targeting important transcription factors of key pathways.[19] Hence, we sought to functionally investigate the contribution of miRNA-mediated gene regulation in the maintenance of adult stem cell-like lineage. This would allow to identify better the mechanisms responsible for controlling the cell-initiating subpopulation and improve tissue-specific organoid growth conditions. Since organoids derive from single stem cells, we performed a functional miRNA screen of human primary epithelial cells to identify the key factors in aSCs-derived organoids formation and expansion. In this study, we identified an uncharacterized miRNA, miR-106a-3p, together with its target genes that play a key role in such processes. Transcriptomic profiling of miR-106a-3p–transduced cells revealed genetic programs overlapping with those of other stem and progenitor cells, showing common features with ESC-like cells or some intermediate cellular states. Thus, using a gain-of-function approach, we discovered that the endogenous levels of three core transcription factors (OCT4, SOX2 and NANOG) were essential for generating organoids. Moreover, we identified the mechanism by which miR-106a-3p finely tunes the differentiation process through a set of transcription regulators (CBFB (Core-binding factor b), NF-YA (Nuclear Transcription Factor Y Subunit Alpha), REST (RE1 silencing transcription factor), and thus the maintenance of aSC-like lineage. Overall, our results highlight the importance of miR-106a-3p in the maintenance of aSC-derived organoids and provide transcriptional mechanisms underlying organogenesis.

Results

Organoid cultures of human mammary epithelial cells exhibit a CD44^{high}/CD24^{low} phenotype

This study was initiated to identify organoid-forming epithelial cell subpopulations that specify stem/progenitor cell functions in epithelial cells.[5] Stem cells are rare immortal cells within a mass of cultured cells. When dividing, they can both self-renew and give rise to other cell types through asymmetric mitoses. Previously, it had been shown that human primary mammary epithelial cells (HMECs) represent a multipotent stem cell population initially present in the basal layer of the gland.[20, 21, 22] These cells are estrogen-independent (tamoxifen-resistant) and display heterogeneous expression of luminal and myoepithelial lineage markers.[23] First, we characterized the properties of 3D-grown HMECs compared to conventional 2D culture. Cells were grown under organoid culture conditions as we previously described[1] and the cells[1] tested formed 3D-structured human organoids (Figure 1A). Approximately 3% of cells present in the culture featured the capacity to generate an organoid, suggesting the presence of stem cells within the mass culture. Next, the self-renewal capacity of organoid-initiating cells was assessed by serial organoid cultures to reflect adult stem cells maintenance from passage 5 to passage 11 (Figure 1B). Cells progressively lost self-renewal ability to form organoids upon serial propagation (Figure 1B), consistent with previously described loss of self-renewal potential of primary epithelial stem cells after a few passages.[24]

Previous studies have reported that HMECs with CD44^{high}/CD24^{low} phenotype have the highest progenitor ability compared to all other stem/progenitor cell subpopulations.[25] It has been shown also that this small cell subpopulation with the capacity of self-renewal and a high proliferation rate, originate from normal stem cells, and enhance sphere/organoid formation.[26, 27, 28, 29] Consequently, the CD44^{high}/CD24^{low} phenotype has been used as a reliable marker for the isolation of normal breast adult stem cells.[30, 31, 32, 33] In order to validate the generation of stem-like cells through 3D culture, flow cytometry was used to assess the expression of breast stem cell markers (CD44^{high}/CD24^{low}) in primary 3D-grown HMECs compared to conventional 2D culture (Figure 1C). In 2D cell culture, 85% of cells expressed both CD24 and CD44 at high levels (CD44^{high}/CD24^{high}) and 14% expressed the CD44^{high}/CD24^{low} phenotype (Figure 1C, top panel). In contrast, 3D cell cultures showed > 3-fold more CD44^{high}/CD24^{low} cells (~49%) (Figure 1C; lower panel, $p = 0.0268$, $n = 3$).

Together, these results indicate that cells grown as organoids acquired a $CD44^{high}/CD24^{low}$ expression pattern similar to that of stem/progenitor cells, which suggests that 3D organoids can be used to enrich breast stem cell markers for further screening.

A miRNA screening approach to selectively enhance the generation of organoids

To investigate whether miRNA-mediated gene regulation could enhance the organoid-forming ability, we designed a two-step functional screening strategy aiming at enrichment of the $CD44^{high}/CD24^{low}$ cell population and at increasing the proportion of organoids-generating cells. We monitored, as a tool, the expression of CD44 and CD24 following miRNA transfection into HMEC cells (Figure 2A-D). Following the quantitative image analysis of >100,000 cells at Passage 6 (P6), frequency distribution of CD44 intensity was compared in mass cultured cells (whole population), and in cells transfected by CD44 (Figure 2A) or by CD24 siRNAs (Figure 2B). CD44 and CD24 levels were lower in siRNA-transfected cells in comparison to the original population, which validates the specificity of our assay (Figure 2A-B). To identify miRNAs that play a role in the enrichment in $CD44^{high}/CD24^{low}$ cells, we performed an unbiased functional screen for miRNAs that modulate CD44/CD24 phenotypes in HMECs (Figure 2C). Using an approach similar to our genome-wide siRNA screen for p16 modulators,[34] we transfected actively proliferating cells (Passage 6, P6) with a collection of 837 miRNAs. siRNA-targeting siGLO ('cyclophilin B'; PPIB), CD44 or CD24 served as controls. We assigned cut-off values to define miRNA hits based on CD44 and CD24 cell integrated intensity. The raw screening data and quantitation of each phenotypic criterion are shown in Figure 2D. This strategy revealed that the miR-106a-3p shifts primary cells into a $CD44^{high}/CD24^{low}$ phenotype. This miRNA is a paralogue of the miR-17/92 cluster (Figure 2E). Next, to further confirm the results, we performed a secondary screen of the whole family cluster (Figure 2F). Twenty-eight miRNAs belonging to the cluster were retested, in triplicate, using the same method as in the primary screen (Figure 2C). A total of 4 miRNA hits were scored with Z-factor > 2 (Figure 2E) and prompted a shift in the $CD44^{high}/CD24^{low}$ population (Figure S1A). The top hit was miR-106a-3p (Figure 2F). We then confirmed that miR-106a-3p induced a $CD44^{high}/CD24^{low}$ phenotype using flow cytometry based on the expression of CD44 and CD24 (Figure 2G). In cells expressing the control mimic, the $CD44^{high}/CD24^{low}$ phenotype was ~10% of the total cell population (Figure 2G). Conversely, we observed a 5-fold increase of $CD44^{high}/CD24^{low}$ phenotype, reaching ~50% of the total cell population, in cells transfected with a miR106a-3p mimic (Figure 2G). Thus, mir-106a-3p

directly promotes the "transdifferentiation" of $CD44^{low}/CD24^{high}$ cell phenotype into $CD44^{high}/CD24^{low}$ cell phenotype.

In parallel, to correlate these data with the generation of stem-like cells in organoids, we assessed the frequency of organoid-initiation after transfection with each of the 28 miRNAs (miR-17/92 cluster) (Figure 2E). Out of 7 positive hits (Figure 2H), miR-106a-3p transfection enabled the highest organoid-initiating frequency (Figure 2H). Taken together, these results show that miR-106a-3p transfection (Figure 2I) is the only one that induces the two properties of 1) $CD44^{high}/CD24^{low}$ cells enrichment in the population and 2) organoid initiation enhancement.

miR-106a-3p drives the generation of human organoids

To further study miR-106a-3p function, we generated retroviral vectors of miR-106a as previously described[24] and evaluated its stable expression in HMECs (Figure 3 A-B). First, we examined the expression of miR-106a-5p and miR-106a-3p using RT-qPCR in control (miR-Vector) and miR-106a-infected cells (Figure 3A). We observed that miR-106a-5p was expressed in both infected cells, whereas miR-106a-3p was exclusively expressed in miR-106a infected cells, both by RT-qPCR and *in situ* hybridization (Figure 3A-B).

To further evaluate the impact of miR-106a on organoid architecture, organoids were analyzed using confocal microscopy. Apoptotic cells are present in organoids during lumen development of the mammary gland.[1] Immunofluorescence staining for the apoptosis marker, Caspase-3, showed that miR-106a did not impact luminal apoptosis during organogenesis (Figure 3C, Caspase-3). Moreover, organoids are characterized by a well-defined cell/Matrigel interface with a myoepithelial layer, which was not impacted by miR-106a overexpression (Figure 3C, CD44 and p63). In addition, organoids expressed β-catenin (a cadherin-based cell–cell junctions marker) and miR-106a overexpression did not show any effect on its localization and did not disrupt cell junctions (Figure 3C, β-catenin). These results show that miR-106a does not target the morphogenesis of organoids.

As expected, miR-106a stable overexpression in primary HMECs also greatly increased organoid-initiating frequency (Figure 3D). Next, to analyze miR-106a-3p and miR-106a-5p individual functions on organoid-initiating cells, miR-106a-3p or miR-106a-5p mimics were transfected into HMECs (Figure 3E). Overexpression of miR-106a-3p significantly increases organoids number by about 5-fold compared to control and miR-106a-

5p (Figure 3E). Our results indicate that miR-106a-5p does not impact the number of organoid-forming cells, demonstrating the specific requirement of miR-106a-3p for maintaining organoid-initiating cells (Figure 3E). Then, miR-106a stably transfected cells were transfected with LNA-anti-miR-106a-3p or LNA-control, allowing us to confirm that miR-106a-3p is required for organoid initiation (Figure 3F). Next, flow cytometry was used to evaluate the expression of CD44 and CD24 in miR–infected HMECs; 89.5% control HMECs (infected with the miR vector) presented a $CD44^{high}/CD24^{high}$ phenotype and 1.7% a $CD44^{high}/CD24^{low}$ phenotype (Figure 3G, left panel); in contrast, ~49.2% HMECs infected with miR-106a presented the in $CD44^{high}/CD24^{low}$ phenotype (Figure 3G, right panel).

To further elucidate the expression of miR-106a-3p in the different CD44/CD24 cell subpopulations, we generated CBX7-transfected HMECs because CBX7 is essential for the maintenance not only of ESCs[35, 36] but also of several aSC types.[37, 38, 39] CBX7-transfected HMECs presented enrichment in $CD44^{high}/CD24^{low}$ cells as compared to empty vector-transfected HMECs (Figure 4A-B). We then separated by flow cytometry $CD44^{high}/CD24^{low}$ (green) from $CD44^{high}/CD24^{high}$ cell populations (blue) to analyze the role of the endogenous expression of miR-106a-3p. The $CD44^{high}/CD24^{low}$ population was the only one to present an endogenous expression of miR-106a-3p (Figure 4C). Blocking the endogenous expression of miR-106a-3p with LNA-anti-miR-106a-3p or LNA-control (Figure 4D) impacted organoids generation (Figure 4E-F). From these experiments, we can conclude 1) that the increased $CD44^{high}/CD24^{low}$ phenotype in miR-106a-3p expressing cells is causative for the higher organoid number, 2) that miR-106a-3p promotes the $CD44^{high}/CD24^{low}$ phenotype and 3) that miR-106a-3p is required for organoids generation and thus participates in aSC-like lineage maintenance.

miR-106a-3p an actor of cell plasticity for differentiation

We then performed transcriptomic analyses to determine whether the miR-106a-3p impacts on cell plasticity to maintain adult stem-like cells or some intermediate cellular states. To study the effect of miR-106a-3p on global gene expression patterns, we isolated total RNAs from miR-106a-3p–transfected HMECs and performed a microarray analysis using Affymetrix chips (HG-U133 Plus 2.0). miR-106a-3p transfection induced significant changes in the expression of 1348 genes when compared to controls (Figure 5A). To gather more insights regarding the global pattern of transcriptional changes associated with miR-106a-3p, we compared our entire microarray dataset with already defined gene signatures of interest by

performing gene set enrichment analysis (GSEA). We observed an enrichment in a downregulated gene set involved in stem cell differentiation (Figure 5B). To explore the physiological role of miR-106a-3p in stem cell differentiation, we took advantage of the ability of human embryonic stem cells (hESCs) to differentiate more readily than aSCs. First, we evaluated the miR-106-3p endogenous expression profile in hESCs (Figure 5C). The expression of miR-106a-5p, miR-106a-3p as well as miR-302b (expressed most abundantly in slowly-growing hESCs)[40] was detectable in hESCs (Figure 5C). Since hESCs are pluripotent stem cells derived from blastocysts and have the property to proliferate indefinitely *in vitro* while maintaining the capacity to differentiate into derivatives of all three germ layers: ectoderm, mesoderm and endoderm,[41] we used hESC to derive early stages of endoderm, mesoderm and ectoderm (Figure 5D). hESCs were transfected with LNA-anti-miR-106a-3p (anti-miR106a-3p) or LNA-control (anti-miR-ctl) prior to endoderm (Figure 5E), mesoderm (Figure 5F) and ectoderm (Figure 5G) differentiation. We then applied classical protocols[42] to trigger directed differentiation of the three germ layers and determine whether blocking endogenous miR-106a-3p expression could change transcriptional expression level of *OCT4*, *SOX2* and *NANOG*. In anti-miR106a-3p transfected cells, *SOX2* expression decreased during endoderm, mesoderm and ectoderm differentiation (Figure 5H, I, and J), while *OCT4* expression decreased during endoderm and ectoderm differentiation (Figure 5H and J). Conversely, *NANOG* expression increased during mesoderm differentiation (Figure 5I). These data demonstrate that miR-106a-3p is involved in the early cell differentiation process into the three germ layers.

Given that *OCT4*, *SOX2* and *NANOG* expression are markers of hESC, we hypothesized that aSCs derived from organoids might also express these genes. In addition, to determine whether these three genes play a role in organoid initiation, we knocked down them individually and demonstrated that the three genes participate, individually, in the organoids generation (Figure 6A), suggesting that they are also related to organoid development. Next, in order to determine whether these genes were induced during organoids differentiation, their mRNA expressions were measured over time in organoid cultures (Figure 6B). Development of organoids can be separated into two stages: an early, highly proliferative stage in which single cells begin to form organoids (approximately the six first days of culture) followed by proliferation arrest, starting around day 8.[43] The expression of the three core regulator genes increased as a function of time, reaching a peak at day 6; then *OCT4* and *SOX2* decreased to low levels whereas *NANOG* remained at a high level (Figure 6B). Finally, to test whether

miR-106a-3p could contribute to modulate the transcriptional activity of these three core transcription factors, we assessed their mRNA (Figure 6C) and protein (Figure 6D-F) levels in miR-106a-3p expressing cells and control cells. All three genes were induced in miR-106a-3p–overexpressing cells as compared to control cells, both at the transcriptional (Figure 6C) and post-transcriptional levels (Figure 6D-F). We thus demonstrated the link between miR-106a-3p and *OCT4*, *SOX2* and *NANOG* expression levels and thus, confirming the specific role of miR-106a-3p via the expression of these transcription factors in organoid initiation. These results also suggest that organoids contain reminiscent properties of embryonic-like stem cells.

Identification of miR-106a-3p targets by in silico research and transcriptomic analysis

In the last part of this study, we sought to identify the genes and mechanisms by which miR-106a-3p enhances organoids generation. To establish whether the significant expression changes observed upon miR-106a-3p transfection (1,348 differential expressed (DE) genes, see materials and methods) was overlapping with the data obtained from prediction algorithms, we used the computational tool MicroT_CDS and identified 707 genes as predicted targets of the hsa-miR-106a-3p. Both datasets were intersected, resulting in a number of 144 common genes (Figure 6G). Subsequently, the 144 genes-functional enrichment analysis using REACTOME pathways and Gene Ontology for these 144 genes was used to identify 32 hub genes, and generate an expression heatmap to investigate the expression patterns in the microarray sample (see materials and methods). Based on these analyses, 32 hub genes related with miR-106a-3p were detected (Figure 6H), 30 genes being down-regulated and 2 up-regulated. Among the 30 upregulated genes, we identified 4 genes encoding transcription factors (*REST*, *CBFB*, *NFYA*, and *GATA3*), which were screened in a subsequent analysis (Figure 6I-M). Transcript levels of these 4 putative targets were quantified by RT-qPCR in the miR-106a-3p–transfected cells; each transcript showed a statistically significant decrease after miR-106a-3p transfection (Figure 6I). Next, we monitored the expression of the set of 4 transcription factors during organoid development (Figure 6J-M); the expression of *REST* (Figure 6J), *CBFB* (Figure 6K) and *NFYA* (Figure 6L) strongly decreased after 4 days of organoid development whereas *GATA3* expression showed moderate changes (Figure 6M).

We then next screened for the relevance of these 3 putative targets by measuring 1) the impact on OCT4, SOX2 and NANOG mRNA expression (Figure 7B,E and H), and 2) the

organoid-initiating frequency (Figure 7C, F and I). We hypothesized that the siRNA-mediated individual knock-down of the three genes differentially regulated during organoid development, *REST*, *CBFB*, and *NFYA*, would impact the expressions of *OCT4*, *SOX2* and *NANOG*, and thus restore the organoid-generating frequency observed after miR-106a-3p transfection. We observed that *REST* knock-down (Figure 7A) was accompanied by an increase in the expression of *OCT4* (Figure 7B), whereas *NFYA* knock-down (Figure 7D) induced an increase in *OCT4* and *SOX2* expressions (Figure 7E), and the knock-down of *CBFB* (Figure 7G) an increase in *NANOG* expression and a moderate decrease in *SOX2* expression (Figure 7H). We also showed that the knock-down of these three genes was able to restore organoid formation, thus mimicking the effect of miR-106a-3p overexpression (Figure 7C, F and I).

Discussion

Organoids are very powerful self-organizing cellular systems that can grow in 3D from human adult or pluripotent stem cells. Organoids possess the exciting potential of modeling key aspects of human development and disease processes, enabling advances towards precision medicine and human disease modeling. Pivotal for the success of organoid cultures is the understanding of signaling pathways that control lineage specification in tissues. Although it can be argued that identifying the stem cells is not critical for the culture of primary tissues, the understanding of the stem cell niche is essential for the improvement of tissue-specific organoid growth conditions as well as the maintenance and indefinite propagation of organoid cultures. We thus aimed at uncovering the key factors that are essential in promoting aSC–derived organoids. This involves first the identification and characterization of the organoid-initiating cell populations, and then the analysis of the regulation of the transcriptional processes involved in organoid generation. A more complete understanding of the development of organoids would enhance their relevance as models to study organ morphology, function and disease, and would open new avenues for drug development and regenerative medicine.

Herein, we focused on miRNAs as privileged factors able to regulate organoid generation and maintenance. We used an unbiased screening procedure which led us to the identification of the previously uncharacterized miR-106a-3p as a master regulator of the stem cell–like lineage which specifies the organoid-initiating cell population from human normal mammary epithelial cells. Collectively, in the experimental conditions investigated, aSC-derived organoids are controlled by the expression of a single miRNA, miR-106a-3p. In a number of deep sequencing studies, co-existence of 5p/3p pairs has been demonstrated in about half of the miRNA populations analyzed and the relative concentrations of the 5p/3p species may be comparable or extensively variable.[44, 45] Notably, the minor miRNA species, be they 5p or 3p, are evolutionarily conserved in the seed sequences, which plays in favor of biological high significance.[46, 47, 48] miR-106a-3p had been reported in literature to be expressed and to play a role in follicular lymphoma,[49] gastric cancer,[50] and renal carcinoma,[51] confirming thus the biological importance of this miRNA. Mir-106a-3p has also been reported in literature to be expressed and play a role in follicular lymphoma [49], gastric cancer[50] and in renal carcinoma[51]; thus confirming the biological significance of this miRNA.

For a cell to differentiate and adopt the identity of a specific cell lineage, transcriptional mechanisms, a series of specific genes must be switched on whereas others must be switched off. To date, it was not known whether a transcriptional program existed for establishing the adult stem cell maintenance. aSC identity, plasticity, and homeostasis are precisely orchestrated by lineage-restricted epigenetic and transcriptional regulatory networks. Therefore, by coupling gene array to miRNA/siRNA screening approaches, we further investigated the transcriptional mechanisms which specify the organoid-initiating cell population from human primary cells expressing miR-106a-3p. Gene expression profiling demonstrated that 3 related transcription factors that are target genes of miR-106a-3p (*CBFB*, *REST* and *NF-YA*) govern the generation of organoid-initiating cells and thus maintain stem cell self-renewal properties (Figure 7J). Interestingly, knocking down each of these 3 genes phenocopied the effects of the miR-106a-3p overexpression by modulating both the three core transcription factors (OCT4, SOX2 and NANOG) and the organoid initiating cell population.

REST is a transcription factor which has been previously reported to be expressed in epithelial cells,[52] and to be part of the OCT4/SOX2/NANOG transcriptional network.[53]. Indeed, REST shares a significant number of target genes with OCT4, SOX2 and NANOG, and several of these genes encode essential factors for cell maintenance in ESCs.[53]. It binds a 21-bp DNA recognition sequence and has two repressor domains that recruit corepressor 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.[54] 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. [55, 56] There is only one CBFB subunit while the other subunit is encoded by three mammalian genes: *RUNX1*, *RUNX2*, and *RUNX3* which all require CBFB for their function. The targeted inactivation of CBFB abrogates the activity of all RUNX complexes.[57] Targeted knock-out of RUNX genes has revealed distinct roles for these proteins in development, RUNX1 being required for hematopoiesis,[58] RUNX2 being required for osteogenesis,[59] and RUNX3 for neurogenesis and the control of gastric epithelial cell proliferation.[60, 61] The RUNX/CBFB complexes have been shown to play a role function in the maintenance of stem cells by activating FGF signaling loops between the epithelium

and mesenchyme. [57] Therefore, it is not surprising that controlling CBFB through miR-106a-3p expression participates in the aSC maintenance.

NF-Y, a ubiquitously expressed trimeric transcription factor, has a dual role as activator and repressor of transcription.[62 62]. 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 the complex assembly and sequence-specific DNA binding. NF-Y has previously been identified as the marker of CSCs in hepatocellular carcinoma and embryonic carcinoma cells.[63, 64, 65] In addition, it has been shown to regulate the expression of several human *SOX* genes, including *SOX2*.[66] NF-Y has been shown to regulate the expression of several human SOX genes, including SOX2[66], SOX9[67] , and SOX18[68]. 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 the maintenance of characteristics of CSCs. Interestingly, NF-Y has been shown to regulate ATF6 expression [69] which is involved in protein homeostasis. Recently it has been proposed that cell proteostasis restrains protein synthesis for maintenance of stem cells [70] thus, we could speculate that NF-YA might also participate in stem cell maintenance through the regulation of cell proteostasis. Herein, we found that NFYA regulates the expression of SOX2 for the maintenance of breast aSCs, consistent with the literature.

Mechanistically, miR-106a-3p targets a specific set of genes, namely *CBFB*, *REST* and *NFYA* to regulate, *in fine*, OCT4, SOX2 and NANOG expressions, therefore reducing heterogeneity within the organoid-initiating cell population and favoring organogenesis (Figure 7J). Thus, a complex mechanism is clearly in place in order to finely tune the expression of miR-106a-3p both in organoids and upon differentiation, which is conserved throughout development in adult and embryonic stem cells. Recent reports showed that differentiation of human aSCs[27] and mouse ESCs[22] are modulated through post-transcriptional attenuation of key factors such as OCT4, SOX2 and NANOG. It has been speculated that the same set of transcription factors plays an important role in the maintenance of multipotency

and self-renewal of aSCs. A recent study by Doffou, has also reported a similar phenomenon, showing that OCT4 expression is induced, starting at day 6, in an organoid hepatocyte model.[71] 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.[72].

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- 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 in the pluripotent state and rapidly decreases upon differentiation stimuli.[9] By contrast, pro-differentiation miRNAs, such as let-7, miR-296, miR-134 and miR-470, regulate the differentiation processes in pluripotent cells.[1] These miRNAs are upregulated during ESCs differentiation and inhibit the expression of pluripotency factors, including NANOG and SOX2.[1] Our data show that miR-106a-3p features as a self-renewal miRNA participating in the modulation of the core factor network (OCT4, SOX2 and NANOG) in aSCs, which in turn inhibits differentiation and favors maintenance (Figure 7J). Indeed, miR-106a-3p is sufficient by itself in targeting a specific set of genes *CBFB, REST* and *NF-YA*, to induce expression of OCT4, SOX2 and NANOG. Finally, the role of the 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 gives us some clues on how the stem cell status may be specified in mammary cells. Our work highlights the transcription factor network cooperativity for the establishment of stem cell identity and lineage commitment, and provides comprehensive regulatory principles for human epithelial homeostasis. To date, it was not known whether a transcriptional program existed for establishing the aSC maintenance, and this is what we demonstrate in this study (Figure 7J). Deciphering signaling cascades that control organoid development would enhance their relevance as models to study organ function, disease and therapy.

Acknowledgments

We gratefully acknowledge the members from ARTiSt group 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.[23, 34] HUES cells (HUES9) were cultured as previously described.[42]

Cell culture

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) organoids, cells were grown in laminin-rich basement membrane growth factor-reduced Matrigel (BDBiosciences) (Matrigel) as we previously described.[5]

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 46 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). High-content images were acquired with the Cytation3 (Bioteck) at 4× magnification, and analysis was performed using the Analysis software (Bioteck). The Z-factor 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 CD44high/CD24low 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; CD24 conjugated with phycoerythrin, PE) 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 analyzed 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.

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 analyzed 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 the 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.

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 [6]. 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 48 hr 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 Supp. Table 1.

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 analyzed using Leica software. Primary antibodies used were MoαCD44 (BD Biosciences); RbαCD44 (Abcam), MoαCD24 (BD Biosciences), Rbαcleaved Caspase-3 ((Asp175), Cell Signaling), Moαbeta-catenin (BD Biosciences), Moαp63 (clone 4A4; Santa CruzBiotechnology), cleaved-caspase 3 (Cell Signaling). 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, 36 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 (46 SSC with 0.1% Tween 20), II (26 SSC), and III (16 SSC) at the hybridization temperature. The 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. The 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[42]. 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 5 µM CHIR 99021 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 analyzed using an unpaired Student's t test. A difference at $p < 0.05$ was considered statistically significant.

Figure 1

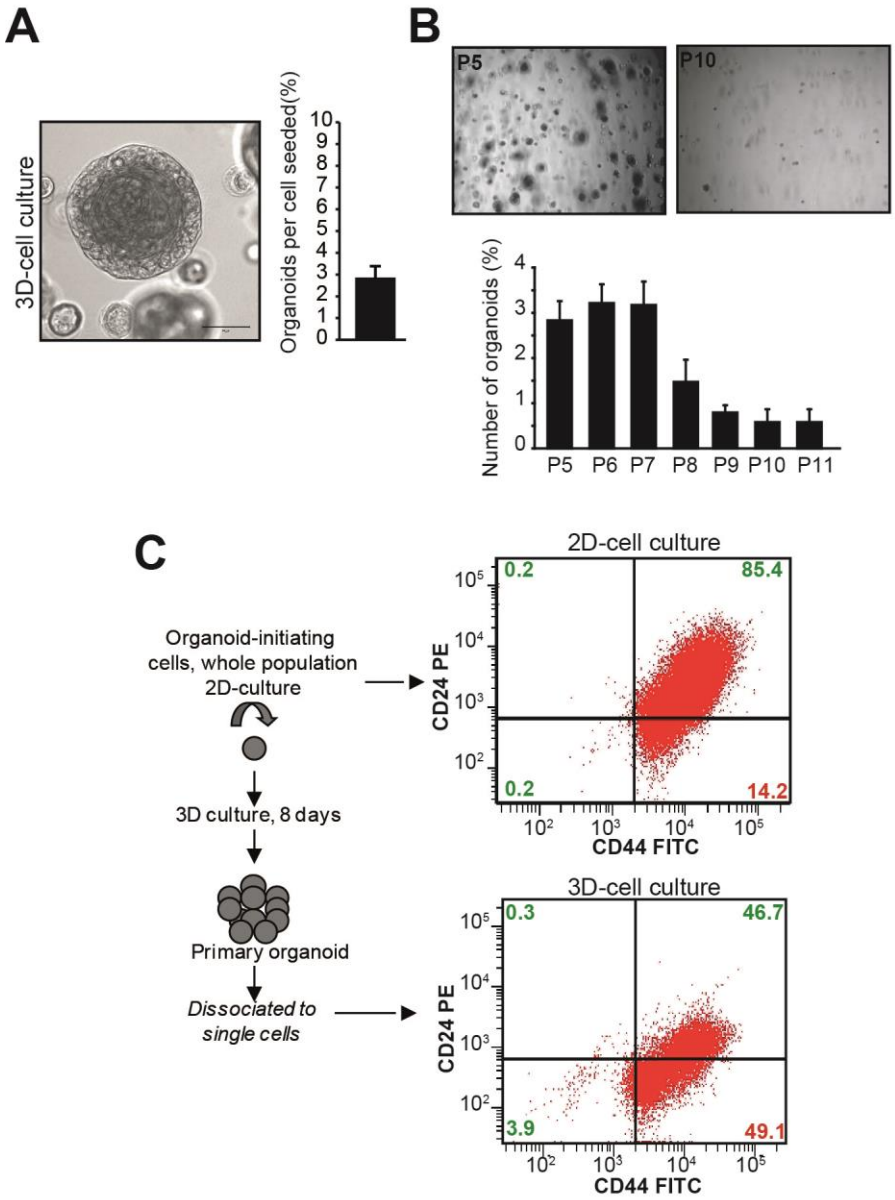


Figure 1. 3D-culture confers original cell properties as compared to 2D-culture

A, Morphology of organoids from HMEC cells cultivated in Matrigel. HMECs are primary cells obtained by dissociation of purified non-diseased human mammary gland. Brightfield microscopic image of one organoid, bar length 50 µm. The bar graph shows the mean of organoids per cell seeded (mean ± SEM) after 10 days of culture from three independent experiments. **B**, Representative brightfield pictures of organoids grown in 3D at passage 5 (P5) and passage 10 (P10). The bar graph shows the mean ± SEM of organoids per well, from passage P5 to passage P11. Data are from three independent experiments for each passage. **C**, Flow cytometric analyses of CD44/CD24 in HMEC cells derived from 2D-cell culture (top)

or from primary organoids culture in 3D (bottom). The expression of $CD44^{high}/CD24^{low}$ in dissociated organoids was higher than in 2D cultured cells. A minimum of 10,000 events were collected per sample.

Figure 2

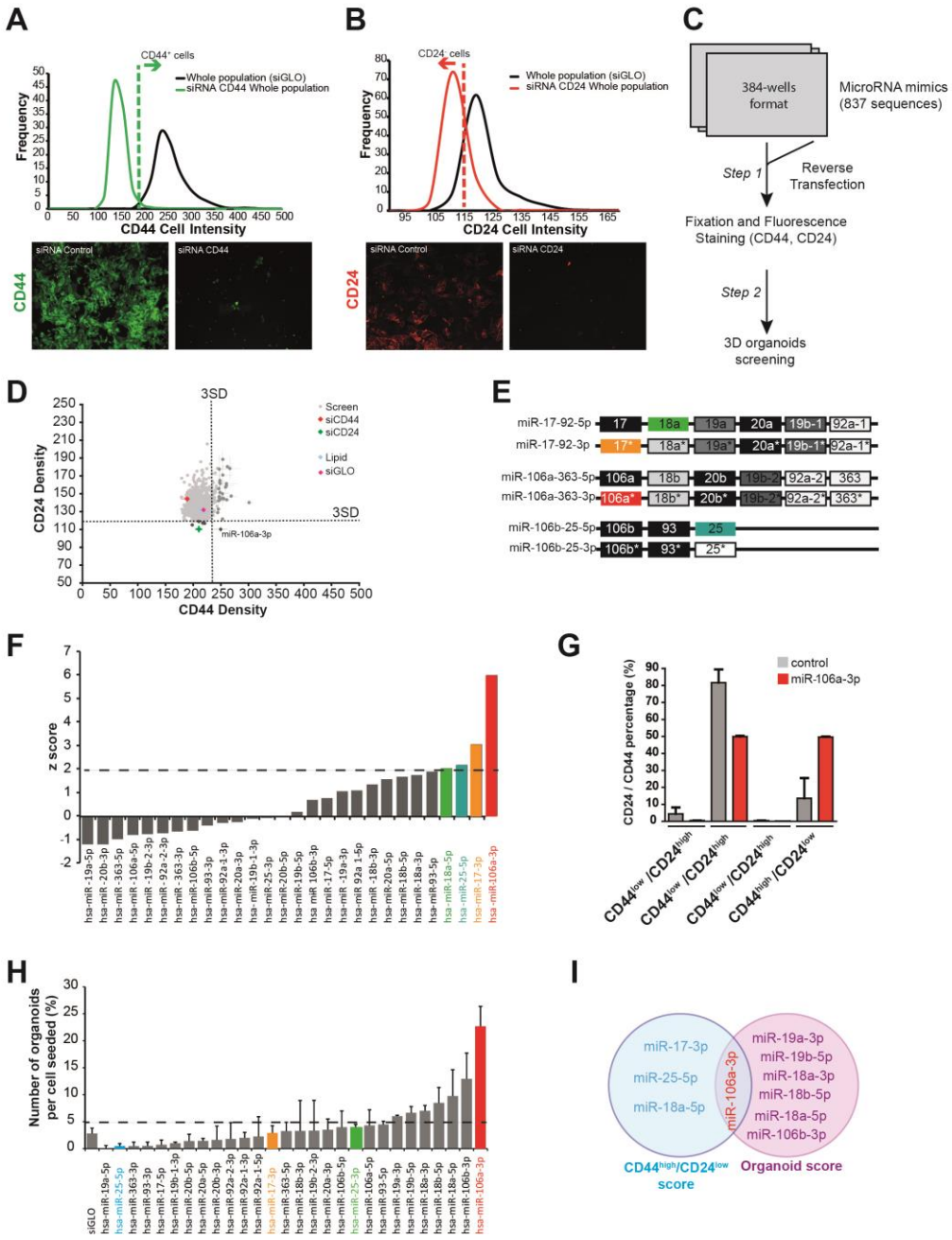
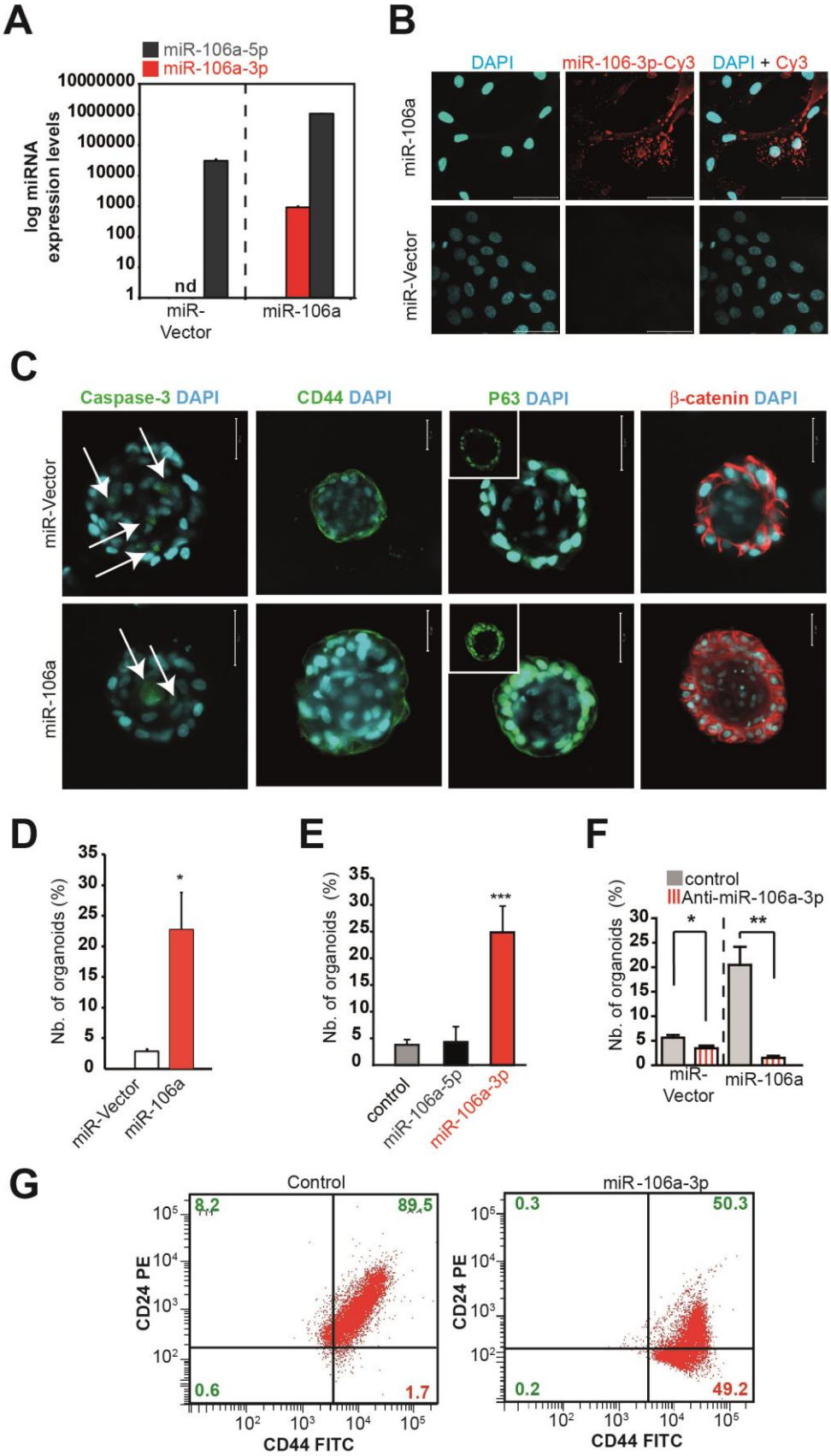


Figure 2. Identification of miR-106a-3p as the predominant miRNA in cells growing in 3D

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 HTS of the 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,** Mean numbers of organoids per cell seeded for each miRNA transfected in HMECs as compared to cells transfected with siRNA control (siGLO). **I,** Venn diagram depicting the overlap of miRNA scoring in common between the $CD44^{high}/CD24^{low}$ and organoid scores. Note that the overall number of miRNAs in common would be the overlap of the intersect of these two Venn diagrams.

Figure 3


A
miR-106a-5p
miR-106a-3p
log miRNA expression levels
10000000
1000000
100000
10000
1000
100
10
1
nd
miR-Vector
miR-106a
B
DAPI
miR-106-3p-Cy3
DAPI + Cy3
miR-106a
miR-Vector
C
Caspase-3 DAPI
CD44 DAPI
P63 DAPI
β-catenin DAPI
miR-Vector
miR-106a
D
Nb. of organoids (%)
miR-Vector
miR-106a
E
Nb. of organoids (%)
control
miR-106a-5p
miR-106a-3p
F
control
Anti-miR-106a-3p
Nb. of organoids (%)
miR-Vector
miR-106a
G
Control
miR-106a-3p
CD24 PE
CD44 FITC
8.2
89.5
0.6
1.7
0.3
50.3
0.2
49.2

663

Figure 3. 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. **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,** Confocal cross-sections of stable HMEC-miR106a organoids 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. **D,** The bar graphs show the mean number of organoids 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. **E,** Mean number of organoids per well in HMEC transfected with either control, miR106a-5p or miR106a-3p mimics. **F,** Percentage of organoids formed by cells seeded for either stable miR-vector transfcted with anti-miR control or stable miR-106a-HMEC transfected with anti-miR106a-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. **G**, Flow cytometric analyses of CD44/CD24 in HMEC transfected with either control or miR106a-3p mimics. A minimum of 10,000 events were collected per sample.

Figure 4

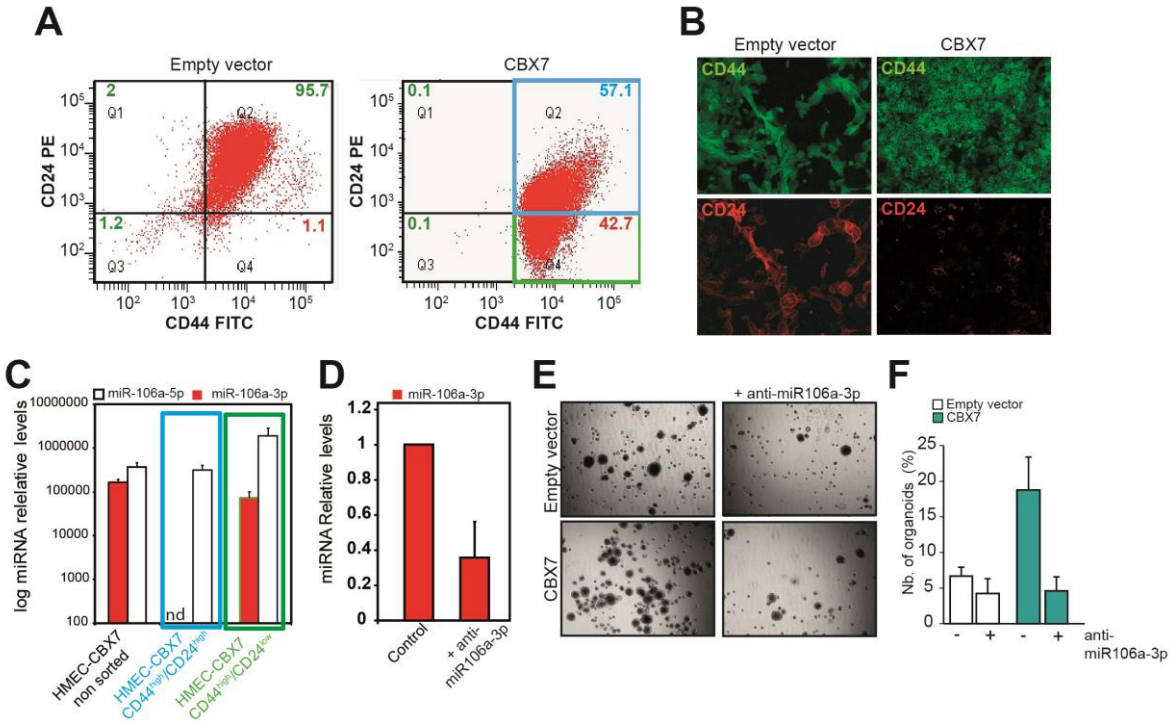


Figure 4. miR-106a-3p improves the maintenance of organoid-initiating cells

A, Flow cytometric analyses of CD44/CD24 in stable HMEC- vector as compared to HMEC-CBX7. A minimum of 10,000 events were collected per sample. **B**, Representative immunofluorescence pictures of HMEC- control (empty vector) and HMEC-CBX7 stained with CD44 antibody (green) and CD24 antibody (red). **C**, Subpopulations of $CD44^{high}/CD24^{low}$ (green) and $CD44^{high}/CD24^{high}$ (blue) were sorted using a flow cytometer; the relative miR-106a-3p and miR-106a-5p expression levels were determined by RT-qPCR for each subpopulation. **D**, Relative miR-106a-3p expression levels determined by RT-qPCR in stable cell lines obtained by retroviral infection of HMECs with empty-Vector or CBX7. **E**, Representative brightfield pictures of organoids grown in 3D in qPCR in stable cell lines obtained by retroviral infection of HMECs with empty-Vector or CBX7 transfected with an anti-miR-106a-3p mimic. **F**, Percentage of organoids formed HMEC-empty vector as compared to HMEC-CBX-7 in presence or not of anti-miR-106a-3p mimic.

Figure 5

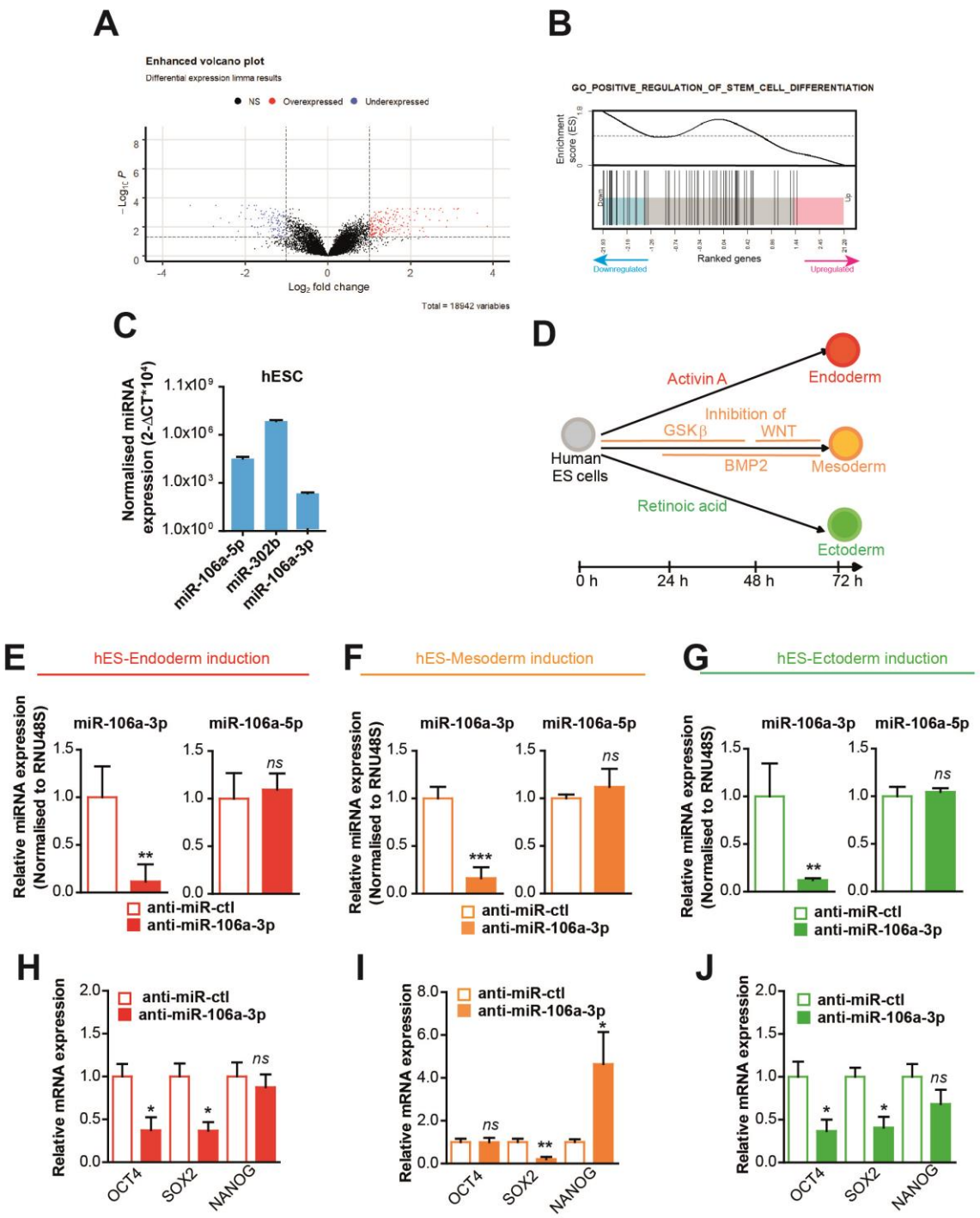


Figure 5. miR-106a-3p targets involved in downregulation of stem cell differentiation
A, Volcano plot comparing the gene expression fold changes and p-values between HMECs transfected with control mimic or miR106a-3p, using the R package EnhancedVolcano. The plot shows the distribution of the total number 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. **B**, 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. **C,** Relative miR-

106a-3p, miR-106a-5p and miR-302b expression levels determined by RT-qPCR in hESCs. **D**, Schematic representation of the human ES cell differentiation process, including timeline and key signaling pathways that are modulated. **E-G**, 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 (E), mesoderm (F) and ectoderm (G) differentiation. **H-J**, Relative mRNA expression levels of key regulators of pluripotency in hESCs transfected with control mimic or anti-miR106a-3p following induction of endoderm (H), mesoderm (I) and ectoderm (J) 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.

Figure 6

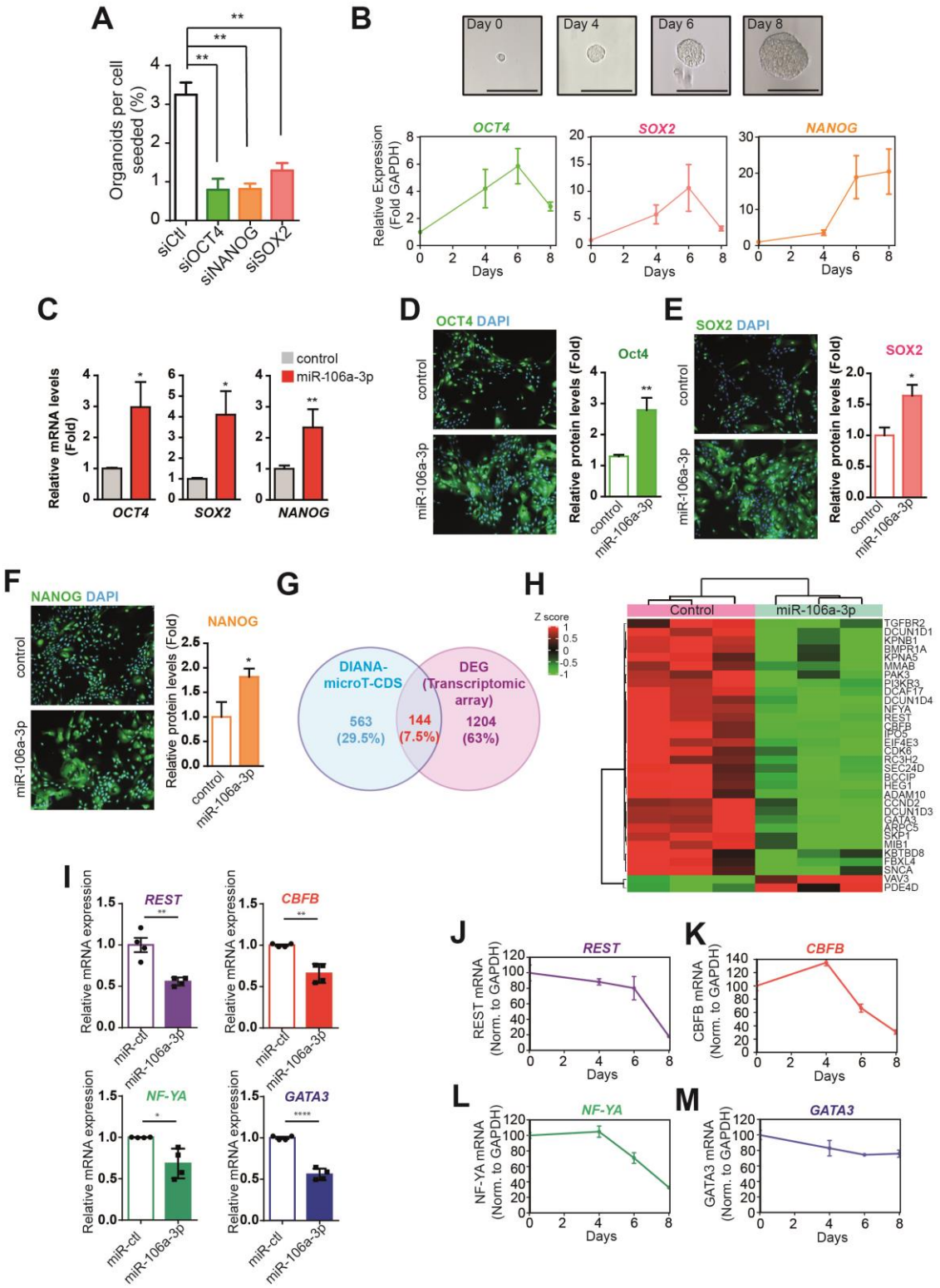


Figure 6. miR-106a-3p, a stem cell determining regulator for organoids and development

A, Percentage of organoids per seeded cell generated from HMECs transfected with siRNA control, siRNA-OCT4, siRNA-NANOG or siRNA-SOX2. **B**, Relative *OCT4*, *SOX2* and *NANOG* expression levels determined by RT-qPCR in 3D growth of HMECs at Days 0, 4, 6,

and 8. **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. **D-F**, Relative protein expression levels of OCT4 (D), SOX2 (E) and NANOG (F) determined by immunofluorescence analysis. **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. **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.

Figure 7

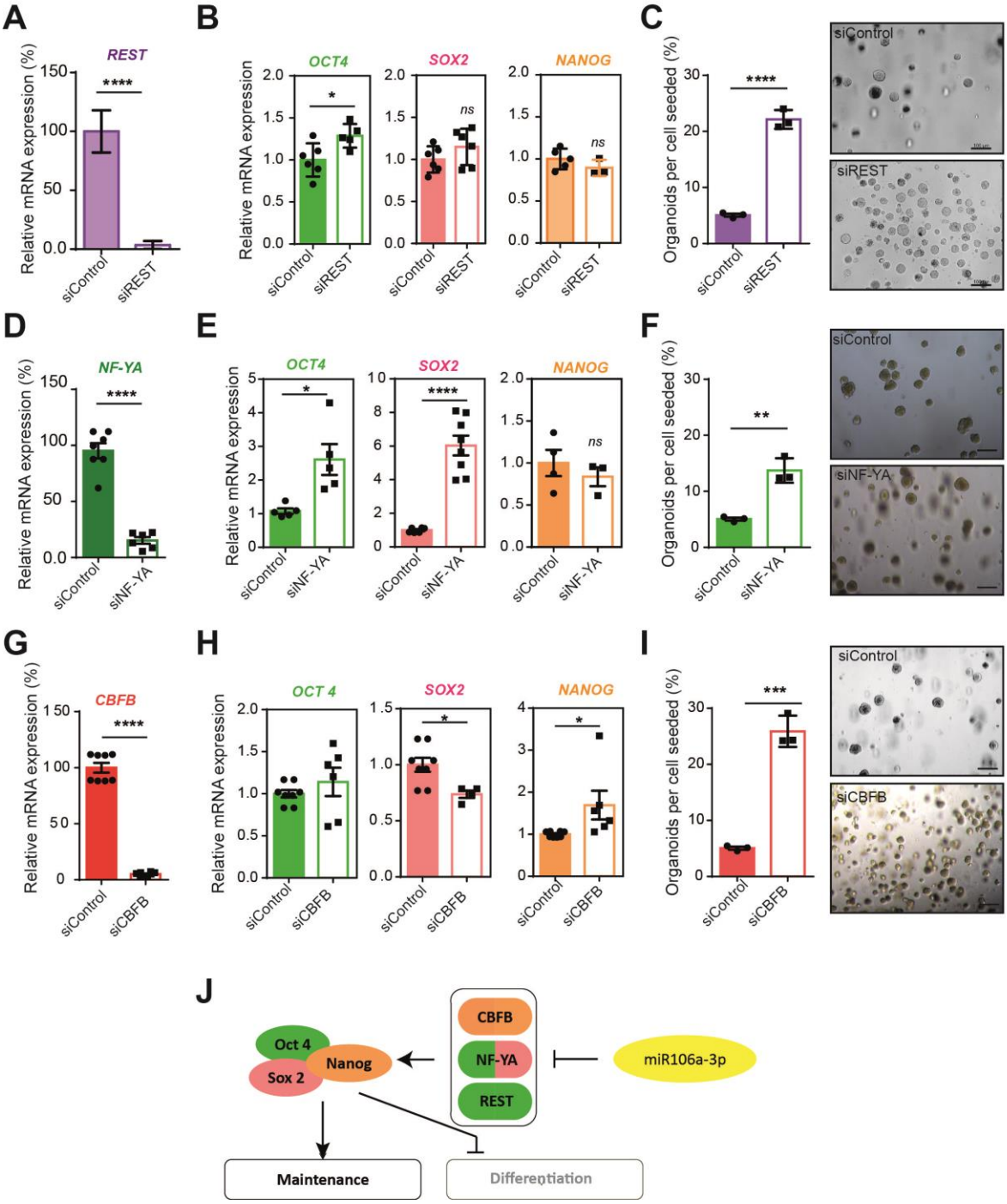


Figure 7. miR-106a-3p a barrier for cell differentiation program during mammary organoid initiation. A, Efficiency of *REST* gene silencing with specific human siRNA in transfected HMECs, as determined by RT-qPCR. Means and standard errors are shown. **B**, Impact of *REST* silencing (siREST) on *OCT4*, *SOX2* and *NANOG* gene expression as determined by RT-qPCR. **C**, Percentage of organoids formed by seeded for control-transfected HMECs (siRNA-ctl) as compared to siREST-transfected HMECs. Representative

brightfield pictures of organoids formed from control- and si-REST-transfected HMECs. **D**, Efficiency of *NFYA* gene silencing with specific human siRNA in transfected HMECs as determined by RT-qPCR. Means and standard errors are shown. **E**, Impact of *NFYA* silencing (siNF-YA) on *OCT4*, *SOX2* and *NANOG* gene expression as determined by RT-qPCR. **F**, Percentage of organoids formed by seeded for control-transfected HMECs (siRNA-ctl) as compared to siNFYA-transfected HMECs. Representative brightfield pictures of organoids formed from control- and siNFYA-transfected HMECs. **G**, Efficiency of *CBFB* gene silencing with specific human siRNA in transfected HMECs as determined by RT-qPCR. Means and standard errors are shown. **H**, Impact of *CBFB* silencing (siCBFB) on *OCT4*, *SOX2* and *NANOG* gene expression as determined by RT-qPCR. **I**, Percentage of organoids formed by seeded for control-transfected HMECs (siRNA-ctl) as compared to siCBFB-transfected HMECs. Representative brightfield pictures of organoids formed from HMECs control- and siCBFB-transfected HMECs . **J,** Schematic representation of the mechanism of organogenesis initiated by miR-106a-3p. miR-106a-3p acts on 3 main transcription factors CBFB, NF-YA and REST. These transcription factors regulate the expression of OCT4, SOX2 and NANOG, which in turns favors stem cell maintenance and block cell differentiation.

Supp. Figure 1

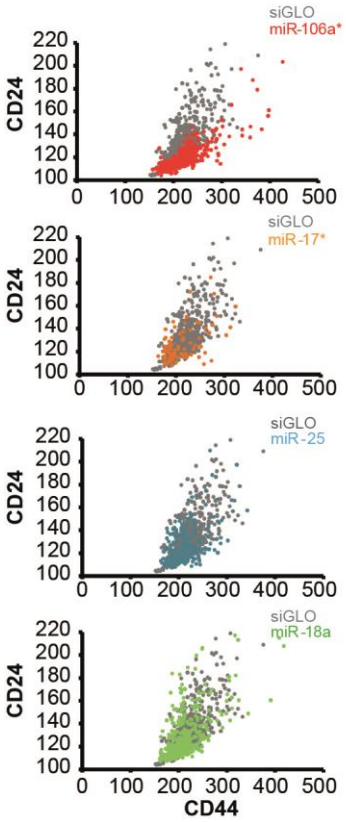


Figure S1

Scatter plots of CD44/CD24 intensity in normal HMEC transfected with miR-106a-3p (red); miR-17-3p (orange); miR-25-5p (blue); and miR-18-5p (green).

778

779 **Table S1** List of the primer sequences used in the study.

Gene Name	Sequences
CBFB-F	CCCGACCAGAGAAGCAAGTT
CBFB-R	GCCACAAAAGCGATTTCCGA
NFYA-F	CAGCAATAGTTCGACAGAGCAGA
NFYA-R	CTACCTGGAGGGTCTGGACTTG
REST-F	GGAGGAAACATTTAAGAAACCATTT
REST-R	CATGGCGGGTTACTTCATGT
GAPDH-F	CCTGGTATGACAACGAATTT
GAPDH-R	GTGAGGGTCTCTCTCTTCCT
OCT3/4-F	TTCAGCCAAACGACCATCTG
OCT3/4-R	CACGAGGGTTTCTGCTTTGC
SOX2-F	CCCACCTACAGCATGTCCTACTC
SOX2-R	TGGAGTGGGAGGAAGAGGTAAC
NANOG-F	TTCCCTCCTCCATGGATCTG
NANOG-R	TGTTTCTTGACTGGGACCTTGTC

780

781

782 **References**

783 1. Fessart D, Begueret H, Delom F. Three-dimensional culture model to distinguish normal from
784 malignant human bronchial epithelial cells. *The European respiratory journal* **42**, 1345-1356
785 (2013).

786
787 2. Fessart D*, et al.* Secretion of protein disulphide isomerase AGR2 confers tumorigenic
788 properties. *eLife* **5**, (2016).

789
790 3. Schwarz JS, de Jonge HR, Forrest JN, Jr. Value of Organoids from Comparative Epithelia
791 Models. *The Yale journal of biology and medicine* **88**, 367-374 (2015).

792
793 4. Zeineddine D*, et al.* Oct-3/4 dose dependently regulates specification of embryonic stem cells
794 toward a cardiac lineage and early heart development. *Developmental cell* **11**, 535-546
795 (2006).

796
797 5. Mani SA*, et al.* The epithelial-mesenchymal transition generates cells with properties of stem
798 cells. *Cell* **133**, 704-715 (2008).

799
800 6. Pardal R, Clarke MF, Morrison SJ. Applying the principles of stem-cell biology to cancer.
801 *Nature reviews Cancer* **3**, 895-902 (2003).

802
803 7. Kim SY*, et al.* Role of the IL-6-JAK1-STAT3-Oct-4 pathway in the conversion of non-stem
804 cancer cells into cancer stem-like cells. *Cellular signalling* **25**, 961-969 (2013).

805
806 8. Zhao D*, et al.* VEGF drives cancer-initiating stem cells through VEGFR-2/Stat3 signaling to
807 upregulate Myc and Sox2. *Oncogene* **34**, 3107-3119 (2015).

808
809 9. Lu CS*, et al.* Chemotherapeutics-induced Oct4 expression contributes to drug resistance and
810 tumor recurrence in bladder cancer. *Oncotarget* **8**, 30844-30858 (2017).

811
812 10. Wang XQ*, et al.* Octamer 4 (Oct4) mediates chemotherapeutic drug resistance in liver cancer
813 cells through a potential Oct4-AKT-ATP-binding cassette G2 pathway. *Hepatology* **52**, 528-
814 539 (2010).

815
816 11. Loh YH, Ng JH, Ng HH. Molecular framework underlying pluripotency. *Cell cycle* **7**, 885-891
817 (2008).

818
819 12. Hall J*, et al.* Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem
820 cell self-renewal. *Cell stem cell* **5**, 597-609 (2009).

821

822 13. Tai MH, Chang CC, Kiupel M, Webster JD, Olson LK, Trosko JE. Oct4 expression in adult
823 human stem cells: evidence in support of the stem cell theory of carcinogenesis.
824 *Carcinogenesis* **26**, 495-502 (2005).

825
826 14. Schoenhals M, Kassambara A, De Vos J, Hose D, Moreaux J, Klein B. Embryonic stem cell
827 markers expression in cancers. *Biochemical and biophysical research communications* **383**,
828 157-162 (2009).

829
830 15. Ezeh UI, Turek PJ, Reijo RA, Clark AT. Human embryonic stem cell genes OCT4, NANOG,
831 STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. *Cancer* **104**,
832 2255-2265 (2005).

833
834 16. Jin T, Branch DR, Zhang X, Qi S, Youngson B, Goss PE. Examination of POU homeobox gene
835 expression in human breast cancer cells. *International journal of cancer* **81**, 104-112 (1999).

836
837 17. Chambers I*, et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in
838 embryonic stem cells. *Cell* **113**, 643-655 (2003).

839
840 18. Mathieu J, Ruohola-Baker H. Regulation of stem cell populations by microRNAs. *Advances in*
841 *experimental medicine and biology* **786**, 329-351 (2013).

842
843 19. Stefani G, Slack FJ. Small non-coding RNAs in animal development. *Nature reviews Molecular*
844 *cell biology* **9**, 219-230 (2008).

845
846 20. Taylor-Papadimitriou J*, et al.* Keratin expression in human mammary epithelial cells cultured
847 from normal and malignant tissue: relation to in vivo phenotypes and influence of medium.
848 *Journal of cell science* **94 (Pt 3)**, 403-413 (1989).

849
850 21. Stampfer MR, Yaswen P. Culture models of human mammary epithelial cell transformation.
851 *Journal of mammary gland biology and neoplasia* **5**, 365-378 (2000).

852
853 22. Labarge MA, Garbe JC, Stampfer MR. Processing of human reduction mammoplasty and
854 mastectomy tissues for cell culture. *Journal of visualized experiments : JoVE*, (2013).

855
856 23. Garbe JC*, et al.* Molecular distinctions between stasis and telomere attrition senescence
857 barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer*
858 *research* **69**, 7557-7568 (2009).

859
860 24. Manuel Iglesias J*, et al.* Mammosphere formation in breast carcinoma cell lines depends
861 upon expression of E-cadherin. *PloS one* **8**, e77281 (2013).

862

863 25. Ghebeh H*, et al.* Profiling of normal and malignant breast tissue show CD44high/CD24low
864 phenotype as a predominant stem/progenitor marker when used in combination with Ep-
865 CAM/CD49f markers. *BMC cancer* **13**, 289 (2013).

866
867 26. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification
868 of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the*
869 *United States of America* **100**, 3983-3988 (2003).

870
871 27. Sleeman KE, Kendrick H, Ashworth A, Isacke CM, Smalley MJ. CD24 staining of mouse
872 mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells.
873 *Breast cancer research : BCR* **8**, R7 (2006).

874
875 28. Cordenonsi M*, et al.* The Hippo transducer TAZ confers cancer stem cell-related traits on
876 breast cancer cells. *Cell* **147**, 759-772 (2011).

877
878 29. Dontu G*, et al.* In vitro propagation and transcriptional profiling of human mammary
879 stem/progenitor cells. *Genes & development* **17**, 1253-1270 (2003).

880
881 30. Ponti D*, et al.* Isolation and in vitro propagation of tumorigenic breast cancer cells with
882 stem/progenitor cell properties. *Cancer research* **65**, 5506-5511 (2005).

883
884 31. Wright MH, Calcagno AM, Salcido CD, Carlson MD, Ambudkar SV, Varticovski L. Brca1 breast
885 tumors contain distinct CD44+/CD24- and CD133+ cells with cancer stem cell characteristics.
886 *Breast cancer research : BCR* **10**, R10 (2008).

887
888 32. Perrone G*, et al.* In situ identification of CD44+/CD24- cancer cells in primary human breast
889 carcinomas. *PloS one* **7**, e43110 (2012).

890
891 33. Wang LB, He YQ, Wu LG, Chen DM, Fan H, Jia W. [Isolation and characterization of human
892 breast tumor stem cells]. *Xi bao yu fen zi mian yi xue za zhi = Chinese journal of cellular and*
893 *molecular immunology* **28**, 1261-1264 (2012).

894
895 34. Bishop CL*, et al.* Primary cilium-dependent and -independent Hedgehog signaling inhibits
896 p16(INK4A). *Molecular cell* **40**, 533-547 (2010).

897
898 35. Morey L*, et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in
899 embryonic stem cells. *Cell stem cell* **10**, 47-62 (2012).

900
901 36. O'Loghlen A*, et al.* MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs
902 during ESC differentiation. *Cell stem cell* **10**, 33-46 (2012).

903

904 37. Gargiulo G*, et al.* In vivo RNAi screen for BMI1 targets identifies TGF-beta/BMP-ER stress
905 pathways as key regulators of neural- and malignant glioma-stem cell homeostasis. *Cancer*
906 *cell* **23**, 660-676 (2013).

907
908 38. Mohn F*, et al.* Lineage-specific polycomb targets and de novo DNA methylation define
909 restriction and potential of neuronal progenitors. *Molecular cell* **30**, 755-766 (2008).

910
911 39. Klauke K*, et al.* Polycomb Cbx family members mediate the balance between haematopoietic
912 stem cell self-renewal and differentiation. *Nature cell biology* **15**, 353-362 (2013).

913
914 40. Suh MR*, et al.* Human embryonic stem cells express a unique set of microRNAs.
915 *Developmental biology* **270**, 488-498 (2004).

916
917 41. Wobus AM, Boheler KR. Embryonic stem cells: prospects for developmental biology and cell
918 therapy. *Physiological reviews* **85**, 635-678 (2005).

919
920 42. Stefanovic S, Abboud N, Desilets S, Nury D, Cowan C, Puceat M. Interplay of Oct4 with Sox2
921 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. *The*
922 *Journal of cell biology* **186**, 665-673 (2009).

923
924 43. Debnath J, Brugge JS. Modelling glandular epithelial cancers in three-dimensional cultures.
925 *Nature reviews Cancer* **5**, 675-688 (2005).

926
927 44. Jagadeeswaran G*, et al.* Deep sequencing of small RNA libraries reveals dynamic regulation of
928 conserved and novel microRNAs and microRNA-stars during silkworm development. *BMC*
929 *genomics* **11**, 52 (2010).

930
931 45. Kuchenbauer F*, et al.* Comprehensive analysis of mammalian miRNA* species and their role
932 in myeloid cells. *Blood* **118**, 3350-3358 (2011).

933
934 46. Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching.
935 *EMBO reports* **12**, 172-177 (2011).

936
937 47. Yang JS*, et al.* Widespread regulatory activity of vertebrate microRNA* species. *Rna* **17**, 312-
938 326 (2011).

939
940 48. Guo L, Lu Z. The fate of miRNA* strand through evolutionary analysis: implication for
941 degradation as merely carrier strand or potential regulatory molecule? *PloS one* **5**, e11387
942 (2010).

943
944 49. Wang W*, et al.* MicroRNA profiling of follicular lymphoma identifies microRNAs related to cell
945 proliferation and tumor response. *Haematologica* **97**, 586-594 (2012).

946
947 50. Guo W, Li W, Yuan L, Mei X, Hu W. MicroRNA-106a-3p Induces Apatinib Resistance and
948 Activates Janus-Activated Kinase 2 (JAK2)/Signal Transducer and Activator of Transcription 3
949 (STAT3) by Targeting the SOCS System in Gastric Cancer. *Medical science monitor :*
950 *international medical journal of experimental and clinical research* **25**, 10122-10128 (2019).

951
952 51. Ma Y*, et al.* miR-106a* inhibits the proliferation of renal carcinoma cells by targeting IRS-2.
953 *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and*
954 *Medicine* **36**, 8389-8398 (2015).

955
956 52. Westbrook TF*, et al.* SCFbeta-TRCP controls oncogenic transformation and neural
957 differentiation through REST degradation. *Nature* **452**, 370-374 (2008).

958
959 53. Johnson R*, et al.* REST regulates distinct transcriptional networks in embryonic and neural
960 stem cells. *PLoS biology* **6**, e256 (2008).

961
962 54. van Arensbergen J*, et al.* Derepression of Polycomb targets during pancreatic organogenesis
963 allows insulin-producing beta-cells to adopt a neural gene activity program. *Genome research*
964 **20**, 722-732 (2010).

965
966 55. Speck NA, Terryl S. A new transcription factor family associated with human leukemias.
967 *Critical reviews in eukaryotic gene expression* **5**, 337-364 (1995).

968
969 56. Tang L*, et al.* Crystal structure of the nuclear matrix targeting signal of the transcription
970 factor acute myelogenous leukemia-1/polyoma enhancer-binding protein 2alphaB/core
971 binding factor alpha2. *The Journal of biological chemistry* **274**, 33580-33586 (1999).

972
973 57. Kurosaka H*, et al.* Core binding factor beta functions in the maintenance of stem cells and
974 orchestrates continuous proliferation and differentiation in mouse incisors. *Stem cells* **29**,
975 1792-1803 (2011).

976
977 58. Growney JD*, et al.* Loss of Runx1 perturbs adult hematopoiesis and is associated with a
978 myeloproliferative phenotype. *Blood* **106**, 494-504 (2005).

979
980 59. Ducy P, Zhang R, Geoffroy V, Ridall AL, Karsenty G. Osf2/Cbfa1: a transcriptional activator of
981 osteoblast differentiation. *Cell* **89**, 747-754 (1997).

982
983 60. Inoue K*, et al.* Runx3 controls the axonal projection of proprioceptive dorsal root ganglion
984 neurons. *Nature neuroscience* **5**, 946-954 (2002).

985
986 61. Li QL*, et al.* Causal relationship between the loss of RUNX3 expression and gastric cancer. *Cell*
987 **109**, 113-124 (2002).

988

989 62. Serra E, Zemzoumi K, di Silvio A, Mantovani R, Lardans V, Dissous C. Conservation and
990 divergence of NF-Y transcriptional activation function. *Nucleic acids research* **26**, 3800-3805
991 (1998).

992
993 63. Dai C*, et al.* Transcriptional activation of human CDCA8 gene regulated by transcription factor
994 NF-Y in embryonic stem cells and cancer cells. *The Journal of biological chemistry* **290**, 22423-
995 22434 (2015).

996
997 64. Moeinvaziri F, Shahhoseini M. Epigenetic role of CCAAT box-binding transcription factor NF-Y
998 on ID gene family in human embryonic carcinoma cells. *IUBMB life* **67**, 880-887 (2015).

999
1000 65. Bungartz G, Land H, Scadden DT, Emerson SG. NF-Y is necessary for hematopoietic stem cell
1001 proliferation and survival. *Blood* **119**, 1380-1389 (2012).

1002
1003 66. Mojsin M, Topalovic V, Marjanovic Vicentic J, Stevanovic M. Transcription factor NF-Y inhibits
1004 cell growth and decreases SOX2 expression in human embryonal carcinoma cell line NT2/D1.
1005 *Biochemistry Biokhimiia* **80**, 202-207 (2015).

1006
1007 67. Shi Z*, et al.* Context-specific role of SOX9 in NF-Y mediated gene regulation in colorectal
1008 cancer cells. *Nucleic acids research* **43**, 6257-6269 (2015).

1009
1010 68. Petrovic I, Kovacevic-Grujicic N, Stevanovic M. ZBP-89 and Sp3 down-regulate while NF-Y up-
1011 regulates SOX18 promoter activity in HeLa cells. *Molecular biology reports* **36**, 993-1000
1012 (2009).

1013
1014 69. Li M*, et al.* ATF6 as a transcription activator of the endoplasmic reticulum stress element:
1015 thapsigargin stress-induced changes and synergistic interactions with NF-Y and YY1.
1016 *Molecular and cellular biology* **20**, 5096-5106 (2000).

1017
1018 70. Li C*, et al.* Amino acid catabolism regulates hematopoietic stem cell proteostasis via a GCN2-
1019 eIF2alpha axis. *Cell stem cell* **29**, 1119-1134 e1117 (2022).

1020
1021 71. Doffou M*, et al.* Oct4 Is Crucial for Transdifferentiation of Hepatocytes to Biliary Epithelial
1022 Cells in an In Vitro Organoid Culture Model. *Gene expression* **18**, 51-62 (2018).

1023
1024 72. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation,
1025 dedifferentiation or self-renewal of ES cells. *Nature genetics* **24**, 372-376 (2000).

1026

1027