Direct programming of human mammary self-organised organoids by miR-106a-3p

Delom F.[1*], Puceat M.[2], and Fessart D.[1*]

1 Univ. Bordeaux, F-33000 Bordeaux, France.

2 Université Aix-Marseille, INSERM UMR_1251, F-13385 Marseille, France.

* **Corresponding author:**

Delphine Fessart - Bordeaux University - 146, Léo-Saignat – 33076 Bordeaux – France. Phone: (00 33) +5 56 33 04 31. Fax: (00 33) +5 56 33 33 30. E-mail: delphine.fessart@yahoo.fr

Frederic Delom - Bordeaux University - 146, Léo-Saignat – 33076 Bordeaux – France. Phone: (00 33) +5 56330431. Fax: (00 33) +5 56333330. E-mail: frederic.delom@yahoo.fr

Lead contact: Delphine Fessart, delphine.fessart@yahoo.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 differentiation. Herein, we carried out a functional microRNA screen to identify miRNAs at the origin of organoid generation from human epithelial cell culture. We uncover miR-106a-3p that initiates and promotes organoids. This miRNA acts as a master inducer of the expression of the three core pluripotency transcription factors (NANOG, OCT4 and SOX2) through the regulation of a set of 10 genes, and thus strengthening the reprogramming and cell differentiation of human epithelial cells into organoids. These data demonstrate that organoids can be directly generated from human epithelial cells by only one miRNA: miR-106a-3p. Hence, we appear to have identified a new determinant of organoid identity, which plays a role in reprogramming, cell differentiation and tissue engineering.

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 the lung [5,6] as well as for numerous other tissues [19]. 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 were derived. In contrast to organ explants, organoids can arise from a single primary cell [5,6,28], thereby allowing the generation of human organoids from biopsies [5]. Non-tumor organoids are thought to arise from adult stem cells (aSCs), and therefore should in theory be capable of self-renewal and differentiation. Consistent with this idea, primary cells enriched with known progenitor/stem cell markers are more efficient at forming organoids than the general cell population [13]. However, currently there is a lack of understanding of the underlying epigenetic and genetic mechanisms that control organoid-initiating frequency, self-renewal and differentiation during organogenesis.

To better understand mammalian development, as well as to exploit the tremendous therapeutic potential of organoid models, it is necessary to identify and characterize the genetic mechanisms governing the fate of aSCs. MicroRNAs (miRNA) have been recently shown to play an important role in regulating stem cell self-renewal and differentiation [15]. 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 or restrict cellular identities by targeting important transcription factors or key pathways [20]. Hence, we sought to investigate the contribution of miRNA-mediated gene regulation in the enrichment of progenitor/stem cell markers. This would allow us to better characterize the mechanisms responsible for controlling the initiating cell sub-population and thus improving tissue-specific organoid growth conditions. Consequently, we performed a miRNA screen in human primary epithelial cells to identify the mediators influencing the initiation of stem cell derived organoids. We identified a previously uncharacterized miRNA, miR-106a-3p, and its target genes that play a key role in such process. Using a gain of function approach, we discovered that the endogenous levels of three core transcription factors (OCT4, SOX2 and NANOG) were post-transcriptionally controlled by miR-106a-3p in human aSCs. Moreover, we discovered that miR-106a-3p is necessary and sufficient to fine tune the differentiation process, and thus the pluripotent state through a specific transcriptional regulatory network. Overall, our results

highlight the importance of miR-106a-3p in the initiation of stem-cell derived organoids and provide some clues about the mechanism underlying organogenesis.

Results

Organoid culture of Human Epithelial Cells exhibits a CD44high/CD24low phenotype

This study was initiated to identify organoid-initiating epithelial cell subpopulations that specify stem/progenitor cell functions in epithelial cells [13]. One of the main characteristics of stem cell is to be rare immortal cells within a mass culture that can both self-renew by dividing and give rise to many cell types. First, we characterized the properties of human primary mammary epithelial cells (HMEC) grown in 3D compared to conventional 2D culture. As a control, cells were grown under organoid culture conditions as we previously described [5] and the cell lines tested formed 3D-structured human organoids (Figure 1A). Approximately 3% of cells present in the culture featured the capacity to reconstitute an organoid (Figure 1A), suggesting the presence of stem cells within the mass culture. Next, the self-renewal capacity of organoid-initiating cells was assessed by serial organoid formation 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 few passages [14].

Previous studies have reported that human mammary epithelial cells with CD44high/CD24low phenotype have the highest progenitor ability compared to all other stem/progenitor subpopulations [7]. Therefore, we analyzed the expression of CD44 and CD24 in 2D cell culture compared to 3D using flow cytometry (Figure 1C). In 2D cell culture, 85% expressed both CD24 and CD44 at high levels and 14% expressed CD24 at low levels together with high levels of CD44 (Figure 1C, Top panel). In contrast, 3D cell culture showed more than 3-fold increase in CD44high/CD24low phenotype cells (~49%) compared to 2D (~14%) (Figure 1C; lower panel, $p=0.0268$, $n=3$). We then analyzed the cells from 2D culture using standard immunofluorescence (Figure 1D) to determine the expression of CD24 and CD44 cell-surface markers. The overlaid images showed a mix of cell populations: CD24 cells (green), CD44 cells (red) and CD44/CD24 co-expressing cells (yellow) (Figure 1D). Together, these results indicate that cells grown as organoids acquired a CD44high/CD24low expression pattern similar to stem/progenitor cell that can be used for further screening.

A miRNA screening approach to selectively favor organoids formation

To investigate whether miRNA-mediated gene regulation could promote organoid formation, we monitored, as a tool, the expression of CD44 and CD24 following miRNA transfection into HMEC cells (Figure 2A-D). Following quantitative image analysis of >100,000 cells at Passage 6 (P6), frequency distributions of CD44 intensity were compared in mass culture (whole population), and mass culture exposed to CD44 siRNA (Figure 2A) or exposed to CD24 siRNA (Figure 2B). CD44 and CD24 levels were lower in siRNA-depleted cells in comparison to the whole population, which validates the specificity of our assay (Figure 2A-B). To identify miRNAs that play a role in the enrichment of CD44high/CD24low cell phenotype, we performed an unbiased functional screen for miRNAs that modulate CD44/CD24 phenotypes in HMEC (Figure 2C). Using an approach similar to our genome-wide small interfering RNA (siRNA) screen for p16 modulators [2], we transfected actively proliferating cells (Passage 6, P6) with 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 density. 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 CD44high/CD24low 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 2E). Twenty-eight miRNAs belonging to the cluster were retested, in triplicate, using the same method as in the primary screen (Figure 1C). A total of 4 hits were scored as those miRNAs with Z-factor >2 (Figure 2F) and prompted a shift in the CD44high/CD24low population (Figure S1B). The top hit was miR-106a-3p (Figure 2F and Figure S1A). We then confirmed miR-106a-3p induced a CD44high/CD24low phenotype using flow cytometry based on the expression of CD44 and CD24 (Figure S1B and Figure 2G). In cells expressing the control mimic, the CD44high/CD24low population ($CD24^-/CD44^+$) was ~10% of the total cell population (Figure 2G). Conversely, we observed a 5-fold increase of CD44high/CD24low population, ~50% of the total cell population, in cells transfected with a mimic miR106a-3p (Figure 2G).

In parallel, to correlate these data with organoid development, we assessed the organoid-initiating frequency of each of the 28 miRNAs (miR-17/92 cluster) (Figure 2E). Out of the total of 7 positive hits (Figure 2H), miR-106a-3p displayed the highest organoid-initiating frequency (Figure 2H). Taken together, these results show that miR-106a-3p (Figure 2I), is the only miRNA that exhibits the two properties of 1) enriching CD44high/CD24low cells and 2) favoring organoid development.

The development of human organoids is driven by miR-106a-3p.

Next, we questioned whether miR-106a-3p is endogenously expressed in organoids compared to 2D culture. We found that only the 3D culture of organoids expressed miR-106a-3p, thus reinforcing its potential role in organoid formation (Figure 3A). To further study miR-106a-3p function, we generated retroviral vectors of miR-106a as previously described [24] evaluated its stable expression in HMECs (Figure 3B-E). 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 3B) and observed that miR-106a-5p was expressed in both conditions. On the contrary, both RT-qPCR and *in situ* hybridization showed that miR-106a-3p was exclusively expressed in miR-106a cells (Figure 3B-C) at levels similar to those observed in 3D cultures (Figure 3A).

As expected, miR-106a stable overexpression greatly increased organoid-initiating frequency (Figure 3D). 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 [4]. Immunofluorescence staining for the apoptosis marker, Caspase-3, demonstrated that miR-106a did not impact on luminal apoptosis during organogenesis (Figure 3E, 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 3E, CD44 and p63). In addition, organoids expressed β-catenin (cell junction marker) adjacent to the plasma membranes and miR-106a overexpression did not show any effect on its localization and did not disrupt cell junctions (Figure 3E, β-catenin). These results demonstrate that miR-106a does not disrupt the structure of organoids.

To test miR-106a-3p and miR-106a-5p individual functions on capacity of organoid-initiating cells, miR-106a-3p or miR-106a-5p mimics were transfected in HMEC cells (Figure 3F-G). We examined miR-106a-5p and miR-106a-3p expression by RT-qPCR and observed that miR-106a-5p is expressed both in control, miR-106a-5p and miR-106a-3p cells (Figure 3F). As expected, miR-106a-3p is only expressed in miR-106a-3p transfected cells (Figure 3F). Next, we studied the individual role of miR-106a-5p and miR-106a-3p overexpression on organoid-initiating frequency (Figure 3G). The overexpression of miR-106a-3p significantly increases organoids number by about 5-fold compared to control and miR-106a-5p (Figure 3G). Our results indicate that miR-106a-5p does not impact on organoids frequency, demonstrating the specific requirement of miR-106a-3p to mediate the self-renewal capacity

of organoid-initiating cells (Figure 3G). Taken together, these results demonstrate that i) miR-106a-3p expression's is only restricted to organoids, and ii) miR-106a-3p, but not miR-106a-5p, is required for the development of human organoids.

Identification of miR-106a-3p targets

Since a single miRNA can potentially target hundreds of genes [10], we next cross-referenced the predicted targets of miR-106a-3p using four different algorithms (miRanda, miRDB, DianaMT and miRWalk). We found 67 genes common to the four algorithms (Figure 4A). In parallel, to study the effect of the over-expression of miR-106a-3p on global gene expression patterns, we isolated total RNA from HMEC cells transfected with miR-106a-3p mimic and performed microarray analysis (HG-U133 Plus 2.0). The results indicated that transfection of miR-106a-3p induced significant changes in the expression of 6465 genes (p value <0.05; Table S1) when compared to controls. To establish whether the global expression changes observed upon miR-106a-3p overexpression correlated with the data from prediction algorithms (Figure 4A), both datasets were intersected (Table S1). The results show that on an average < 7% of the genes differentially expressed following miR-106a-3p transfections are direct or indirect regulatory targets of this miRNA (Figure S2 and Figure 4B). Almost half of the differentially expressed genes were down-regulated (3153, ~48%) following miR-106a-3p transfection (Table S1). Of the 3153, only 35 (1.1%) genes were found to be direct targets of miR-106a-3p by the aforementioned four different algorithms (Figure 4B).

We next screened for the relevance of these putative targets in two different assays including 1) the increase in organoid-initiating frequency (Figure 4C and D), and 2) the enrichment in the CD44high/CD24low cell population (Figure 4E). We hypothesized that the depletion of the target with siRNA would increase organoid initiating frequency and induce a CD44high/CD24low phenotype. Results showed that, out of the thirty-five targets tested, ten candidates had the capacity to increase organoid initiating frequency (Figure 4C and D). Therefore, depletion of *ADD3*, *B4GALT6*, *C12Orf14*, *CCAR1*, *MCM10*, *PANK3*, *PP6R3*, *PXMP3*, *TLE4*, and *TSPYL2* genes increased the number of organoids and exhibited a similar effect as miR-106a-3p (Figure 4C and D). To determine if the individual depletion of each of these ten genes shifted towards a CD44high/CD24low phenotype, we measured CD44 and CD24 levels in cells depleted for each individual transcript. We observed that the depletion of these genes increased, similar to what observed with miR-106a-3p overexpression, a cell

population CD44high/CD24low (Figure 4E). Taken together, these results demonstrate that miR-106a-3p repress multiple target genes, with downregulation of individual targets recapitulating the total miRNA effects required for both CD44high/CD24low phenotype and organoid development.

miR-106a-3p and its targets regulate the expression of OCT4, SOX2 and NANOG

Our results demonstrate that miR-106a-3p promotes the enrichment of CD44high/CD24low cells and thereby enhancing stem/progenitor cell properties. To identify common features among different human pluripotent cells and to search for clues into the genetics of human germ cell, we next compared the miR-106-3p endogenous expression profile in a series of cancer cell lines, normal cell lines as well as germ cells, including pure hESCs, that express high levels of endogenous pluripotency markers (Figure S4 and Figure 5A). The miR-106a expression pattern from 20 different cell lines was examined by RT-qPCR (Figure S4 and Figure 5A). The results showed that miR-106a-3p was expressed exclusively in 2 human pluripotent embryonic carcinoma lines (NCC-IT, TERA1) which are derived from poorly differentiated germ cell tumors and in human embryonic stem cells (hESCs) (Figure 5A).

To further investigate whether the ten target genes of miR-106a-3p were associated with the stemness properties, the expression of the ten genes *ADD3*, *B4GALT6*, *C12Orf14*, *CCAR1*, *MCM10*, *PANK3*, *PP6R3*, *PXMP3*, *TLE4*, and *TSPYL2* was analyzed using StemChecker, a web-based tool to explore stemness in a gene set. The resulting chart showed that this gene set is significantly enriched in targets of NANOG, SOX2 and OCT4 (the human pluripotency master regulators) as well as E2F4 (Figure 5B). To test the effect of miR-106a-3p on pluripotent transcriptional activity, we assessed the mRNA and protein levels of three core pluripotent transcription factors in miR-106a-3p expressing cells compared to control cells. The three core pluripotent transcription factors were induced in miR-106a-3p overexpressing cells as compared to those expressing mimic control miRNA, both at the genomic (Figure 5C) and proteomic levels (Figure 5D).

Finally, we analyzed the expression of these three core pluripotent transcription factor genes upon individual knockdown of each of the ten target genes (Figure 5E-G), previously validated by qRT-PCR (Figure S3). We observed that the depletion of *TSPYL2*, *PXMP3*,

PP6R3, *PANK3* increased the expression of the three core pluripotent transcription factors (Figure 5H). *B4GALT6*, *ADD3*, *C12ORF14*, *CCAR1* when depleted, increased *OCT4* and *SOX2*; while depletion of *TLE4* increased gene expression of *SOX2* and *NANOG* (Figure 5H). Finally, the depletion of *MCM10* gene induced *OCT4* expression only (Figure 5H). These data demonstrate that miR-106a-3p plays a role in the process of pluripotency by regulating the core of master regulators OCT4, SOX2, and NANOG through the control of a set of specific genes.

miR-106a-3p controls human organoid development

To further determine whether miR-106a-3p is required for self-renewal capacity of organoids-initiating cells, miR-106a cells were transfected with LNA-anti-miR-106a-3p or LNA-control (Figure 6A-B). We confirmed that miR-106a-3p and not miR-106a-5p expression was decreased in miR-106a cells transfected with LNA-antimiR-106a-3p using RT-qPCR (Figure 6A). As expected, the levels of miR-106a-3p in cells transfected with anti-miR-106a-3p were considerably low compared to control cells (Figure 6A), whereas miR-106a-5p levels remained unchanged (Figure 6A). To further confirm the specific requirement of miR-106a-3p for self-renewal of organoid-initiating cells, miR-106a-infected and control cells were transfected with LNA-anti-miR-106a-3p or LNA-control and were grown in organoids. In control cells, anti-miR-106a-3p did not repress the formation of organoids (Figure 6B), since miR-106a-3p is not expressed in primary HMEC (Figure S4). As expected, we observed that suppression of miR-106a-3p abrogated the self-renewal capacity of organoids-initiating cells (Figure 6B).

Since we demonstrated that miR-106a-3p contributed to pluripotency (Figure 5) and self-renewal (Figure 6B) in human adult stem cells, we next compared the miR-106-3p endogenous expression profile in miR-106a-3p transfected cells, in control cells and in hESCs, that express high levels of endogenous pluripotency markers. The expression of miR-106a-5p as well as miR-302b (a gold standard marker for pluripotency) was detectable in HMEC control (HMEC+ctl), HMEC transfected with miR-106a-3p mimic (HMEC+miR-106a-3p) and hESCs cells (Figure 6C). Interestingly, the expression of miR-106a-3p was only detectable in HMEC miR-106a-3p and hESCs (Figure 6C). 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 [18,23]. We used human ES cells to derive early stages of endoderm,

mesoderm and ectoderm (Figure S5A). Interestingly, the expression of the miR-106a-3p was significantly upregulated upon mesoderm and ectoderm differentiation compared to embryonic stem cells (hESCs) (Figure 6D). To explore how mir-106a-3p could control differentiation, we next took advantage of the ability of hESCs to differentiate more readily than aSCs. hESCs were transfected with LNA-anti-miR-106a-3p (anti-miR106a-3p) or LNA-control (anti-miR-ctl) (Figure S5B, D and F) prior to endoderm, mesoderm and ectoderm differentiation. The decreased level of miR-106a-3p in anti-miR106a-3p transfected cells as compared to control cells was validated by RT-qPCR (Figure S5B, D and F). Next, we applied directed differentiation protocols to trigger the three germ layers to find out whether blocking miR-106a-3p expression could change expression of transcription factor *OCT4*, *SOX2* and *NANOG*. Expression of *SOX2* decreased during endoderm, mesoderm and ectoderm differentiation, while *OCT4* decreased during endoderm and ectoderm differentiation in anti-miR106a-3p transfected cells compared to control cells (Figure S 5C, E and G). Conversely, expression of *NANOG* increased during mesoderm differentiation in anti-miR-106a-3p cells (Figure S5E). To further understand the impact of miR-106a-3p depletion on hESCs differentiation, we monitored expression of specific genes upon induction of the three embryonic germ layers (Figure 6E, F and G). Expression of endodermal genes was not or weakly affected by the level of expression of miR-106a-3p (Figure 6B). In contrast, expression of mesoderm- and ectoderm-specific genes increased upon miR-106a-3p down-regulation (Figure 6D and F). Collectively, these data demonstrate that miR-106a-3p is involved in the differentiation process and is essential for human organoid development.

Discussion

Organoids are very powerful self-organizing cellular systems that have been grown in 3D from human adult or pluripotent stem cells. Organoids show the exciting potential of modeling key aspects of human development and disease processes, as well as advance efforts towards precision medicine and human disease modeling. Central to the success of organoid cultures is the understanding of the endogenous stem cell niche and signaling pathways that control lineage specification in tissues. Although it can be argued that identifying the stem cells is not critical for culturing primary tissue, the understanding of the stem cell niche is essential for the sustenance and indefinite propagation of cultures. Therefore, our aim was to uncover key factors, such as miRNAs, essential in promoting stem cell derived organoids. This involves reducing heterogeneity within the organoid-initiating cell population through a better characterization of initiating cell types and improvement of tissue-specific organoid growth conditions. 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 combined organoid analyses and miRNA screening to identify the previously uncharacterized miR-106a-3p as a master regulator of the stem/progenitor cell pools which specify the organoid-initiating cell population from human primary cells (Figure 6H). By coupling gene array to siRNA screening approaches, we further identified ten target genes of miR-106a-3p (*ADD3*, *B4GALT6*, *C12Orf14*, *CCAR1*, *MCM10*, *PANK3*, *PP6R3*, *PXMP3*, *TLE4*, and *TSPYL2*) which govern self-renewal capacity of organoid-initiating cells and thus maintain stem/progenitor cell properties. Interestingly, knocking down each of these ten genes phenocopied the effects of the miR-106a-3p overexpression. A recurring observation in organoid models is that the signaling pathways governing organoid formation are identical to those utilized during *in vivo* organ development and homeostasis. Specifically, from the targets identified in this study, *TLE4* is a transcription factor which has been previously shown to play a role in early embryogenesis [11]. Indeed, Tle4-knockout mice die at around four weeks with defects in bone development and bone marrow aplasia [25]. *TLE4* expression has also been shown to increase upon LIF withdrawal and loss of *TLE4* leads to increased pluripotency marker expression and inhibits ESC differentiation towards both the epiblast and endoderm lineages [11]. These data are consistent with our observation that miR-106a-3p inhibits *TLE4*. *C12Orf14* (*FAM60A*) gene is a regulator of SIN3-HDAC function and gene expression [16]. Indeed, *C12Orf14* is a subunit of the Sin3 deacetylase complex and

resides in active Histone deacetylase 1 and 2 (HDAC1/2). HDAC1-null embryos die before E10.5, showing that the *HDAC1* gene is essential for embryonic development. Hence, these observations are consistent with the effects of miR-106a-3p on *C12Orf14*. *MCM10* is exclusive to eukaryotes and is essential for both initiation and elongation phases of nuclear DNA replication [12]. The physiological function of *MCM10* protein has been shown in the Mcm10-knockout mouse model and reveals that *MCM10* expression is required for early embryogenesis. Thus, the effect of miR-106a-3p on this gene is in agreement with its previously reported function.

Collectively, in each of the experimental conditions investigated, aSC-derived organoids are controlled by the expression of a single miRNA, miR-106a-3p. This miRNA targets a specific set of genes to regulate, *in fine*, OCT4, SOX2 and NANOG, therefore, reduces heterogeneity within the organoid-initiating cell population to favor organogenesis (Figure 6H). Hence, a complex mechanism (Figure 6H) is clearly in place in order to fine-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 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 aSCs. Although hESC pluripotency requires OCT4 and SOX2, the consequence of elevated Oct4 and Sox2 levels on hESCs renewal and pluripotency have been overlooked. A less than 2-fold increase of Oct4 protein turns murine and human ESCs into primitive endoderm and mesoderm [17,29] and more specifically mesendoderm [21]. Regulation by miRNA provides a way to finely tune hESC self-renewal and differentiation. Indeed, miRNAs play an important role in gene regulation during pluripotency, self-renewal and differentiation of ESCs. miRNAs can be divided into two subgroups: pluripotent miRNAs and pro-differentiation miRNAs. Pluripotent miRNAs have been found to be involved in maintaining self-renewal and pluripotency of ESCs. This class of miRNAs, 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, have been found to regulate the differentiation processes in pluripotent cells [1]. These miRNAs are found to be upregulated during differentiation in ESCs and inhibited the expression of pluripotency factors, including Nanog and Sox2 [1]. A miRNA has two arms: miR-5p and miR-3p (miR-5p/-3p). Depending on the

tissue or cell types, both arms can become functional. Indeed, selection of either or both of the 5p or 3p miRNA species has been reported to be dependent on temporal, spatial, physiological and pathological conditions [8,26]. Our data demonstrate that miR-106a-3p features an unexpected biological function in modulating the human pluripotency factor network (OCT4, SOX2 and NANOG) and, in turn, in regulating differentiation. Indeed, a low level of endogenous miR-106a-3p is sufficient to induce expression of OCT4, SOX2 and NANOG. Upon differentiation, miR-106a-3p is elevated and therefore reinforces ESC pluripotency at the expense of differentiation, and more specifically towards mesoderm and ectoderm or mesectoderm. Finally, the role of the miR-106a-3p is of particular interest in yield predictions on how mammary cells acquire stem cell-like properties in normal state. Indeed, the capacity of miR-106a-3p to promote stem cell-like behavior gives us some clues on how stem/progenitor cell states may be specified in mammary cells. Future studies are necessary to define by which precise mechanism miR-106a-3p controls the human pluripotent stem cells, and how cells can decide or control which variant of the miRNA to express under what circumstances.

Acknowledgments

This work was supported in part by grant from "La Region Nouvelle-Aquitaine" is warmly thanked for financial support. We thank Dr. E. Chevet and Prs J. Robert and T. Hupp for critical reading of the manuscript, Pr. P. Soubeyran for leadership at the SIRIC BRIO, Dr. V.alérie LeMorvan for help in setting up the RT-qPCR system, and Dr. R. Nookala of Institut Bergonié for the medical writing service.

Author Contributions

Conceptualization, D.F. and F.D.; Methodology, D.F., F.D., and M.P.; Investigation, D.F., F.D., and M.P.; Writing – Original Draft, D.F.; Writing – Review & Editing, D.F., F.D., and M.P; Resources, D.F. and M.P.; Supervision, D.F., and M.P.

Declaration of Interests

None

386 Reference List

387

388 [1] F. Anokye-Danso, *et al.*, "Highly efficient miRNA-mediated reprogramming of mouse and
389 human somatic cells to pluripotency," **8**(4), 376 (2011).

390 [2] C. L. Bishop, *et al.*, "Primary cilium-dependent and -independent Hedgehog signaling inhibits
391 p16(INK4A)," **40**(4), 533 (2010).

392 [3] V. Borgdorff, *et al.*, "Multiple microRNAs rescue from Ras-induced senescence by inhibiting
393 p21(Waf1/Cip1)," **29**(15), 2262 (2010).

394 [4] J. Debnath and J. S. Brugge, "Modelling glandular epithelial cancers in three-dimensional
395 cultures," **5**(9), 675 (2005).

396 [5] D. Fessart, H. Begueret, and F. Delom, "Three-dimensional culture model to distinguish normal
397 from malignant human bronchial epithelial cells," **42**(5), 1345 (2013).

398 [6] D. Fessart, *et al.*, "Secretion of protein disulphide isomerase AGR2 confers tumorigenic
399 properties," **5** (2016).

400 [7] H. Ghebeh, *et al.*, "Profiling of normal and malignant breast tissue show CD44high/CD24low
401 phenotype as a predominant stem/progenitor marker when used in combination with
402 Ep-CAM/CD49f markers," **13**, 289 (2013).

403 [8] L. Guo and Z. Lu, "The fate of miRNA* strand through evolutionary analysis: implication for
404 degradation as merely carrier strand or potential regulatory molecule?," **5**(6), e11387
405 (2010).

406 [9] H. B. Houbaviy, M. F. Murray, and P. A. Sharp, "Embryonic stem cell-specific MicroRNAs," **5**(2),
407 351 (2003).

408 [10] A. Jacobsen, *et al.*, "Analysis of microRNA-target interactions across diverse cancer types,"
409 **20**(11), 1325 (2013).

410 [11] A. F. Laing, S. Lowell, and J. M. Brickman, "Gro/TLE enables embryonic stem cell differentiation
411 by repressing pluripotent gene expression," **397**(1), 56 (2015).

412 [12] H. J. Lim, *et al.*, "Targeted disruption of Mcm10 causes defective embryonic cell proliferation
413 and early embryo lethality," **1813**(10), 1777 (2011).

414 [13] S. A. Mani, *et al.*, "The epithelial-mesenchymal transition generates cells with properties of
415 stem cells," **133**(4), 704 (2008).

416 [14] Iglesias J. Manuel, *et al.*, "Mammosphere formation in breast carcinoma cell lines depends
417 upon expression of E-cadherin," **8**(10), e77281 (2013).

418 [15] J. Mathieu and H. Ruohola-Baker, "Regulation of stem cell populations by microRNAs," **786**,
419 329 (2013).

[16] I. M. Munoz, *et al.*, "Family with sequence similarity 60A (FAM60A) protein is a cell cycle-fluctuating regulator of the SIN3-HDAC1 histone deacetylase complex," **287**(39), 32346 (2012).

[17] H. Niwa, J. Miyazaki, and A. G. Smith, "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells," **24**(4), 372 (2000).

[18] B. E. Reubinoff, *et al.*, "Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro," **18**(4), 399 (2000).

[19] J. S. Schwarz, H. R. de Jonge, and J. N. Forrest, Jr., "Value of Organoids from Comparative Epithelia Models," Yale J. Biol. Med. **88**(4), 367 (2015).

[20] G. Stefani and F. J. Slack, "Small non-coding RNAs in animal development," **9**(3), 219 (2008).

[21] S. Stefanovic, *et al.*, "Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate," **186**(5), 665 (2009).

[22] Y. Tay, *et al.*, "MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation," **455**(7216), 1124 (2008).

[23] J. A. Thomson, *et al.*, "Embryonic stem cell lines derived from human blastocysts," **282**(5391), 1145 (1998).

[24] P. M. Voorhoeve, *et al.*, "A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors," **124**(6), 1169 (2006).

[25] J. C. Wheat, *et al.*, "The corepressor Tle4 is a novel regulator of murine hematopoiesis and bone development," **9**(8), e105557 (2014).

[26] J. S. Yang, *et al.*, "Widespread regulatory activity of vertebrate microRNA* species," **17**(2), 312 (2011).

[27] R. Yi, *et al.*, "A skin microRNA promotes differentiation by repressing 'stemness'," **452**(7184), 225 (2008).

[28] X. Yin, *et al.*, "Engineering Stem Cell Organoids," **18**(1), 25 (2016).

[29] D. Zeineddine, *et al.*, "Oct-3/4 dose dependently regulates specification of embryonic stem cells toward a cardiac lineage and early heart development," **11**(4), 535 (2006).

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@yahoo.fr).

EXPERIMENTAL MODEL DETAILS

Cell lines

Normal finite lifespan HMECs were obtained from Lonza and were grown in MEBM phenol red-free medium supplemented with MEGM Single Quots (Lonza, Basel, Switzerland). HUES cells (HUES9) were cultured as previously described [21]

Cell culture

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].

METHOD DETAILS

High-content miRNA screening

The miRNA screen was performed in triplicate, using the Human pre-miR miRNA library (Ambion), consisting of 328 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. 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 DMI8 microscope (Leica) at 10× magnification, and analysis was performed using the Analysis software (Leica). 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 analysis of gene expression

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. Data were analyzed and normalized using the Rosetta Resolver Error Model. In order to remove systemic noise from the data, genes with low intensity values that are close to background were filtered out. We use standard deviation of the background (σ) value to estimate background. We then filter-extract the genes identified as being significantly differentially expressed between the conditions by fold change applied to the genes passing the background filtering criteria. Differentially expressed probe sets were identified using a p-value <0.05. 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.

miRNA target download

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/.

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.

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, 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.

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 [21]. 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.

DATA AVAILABILITY

Data that support the findings of this study have been deposited have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-6594. All other relevant data are available from the corresponding authors.

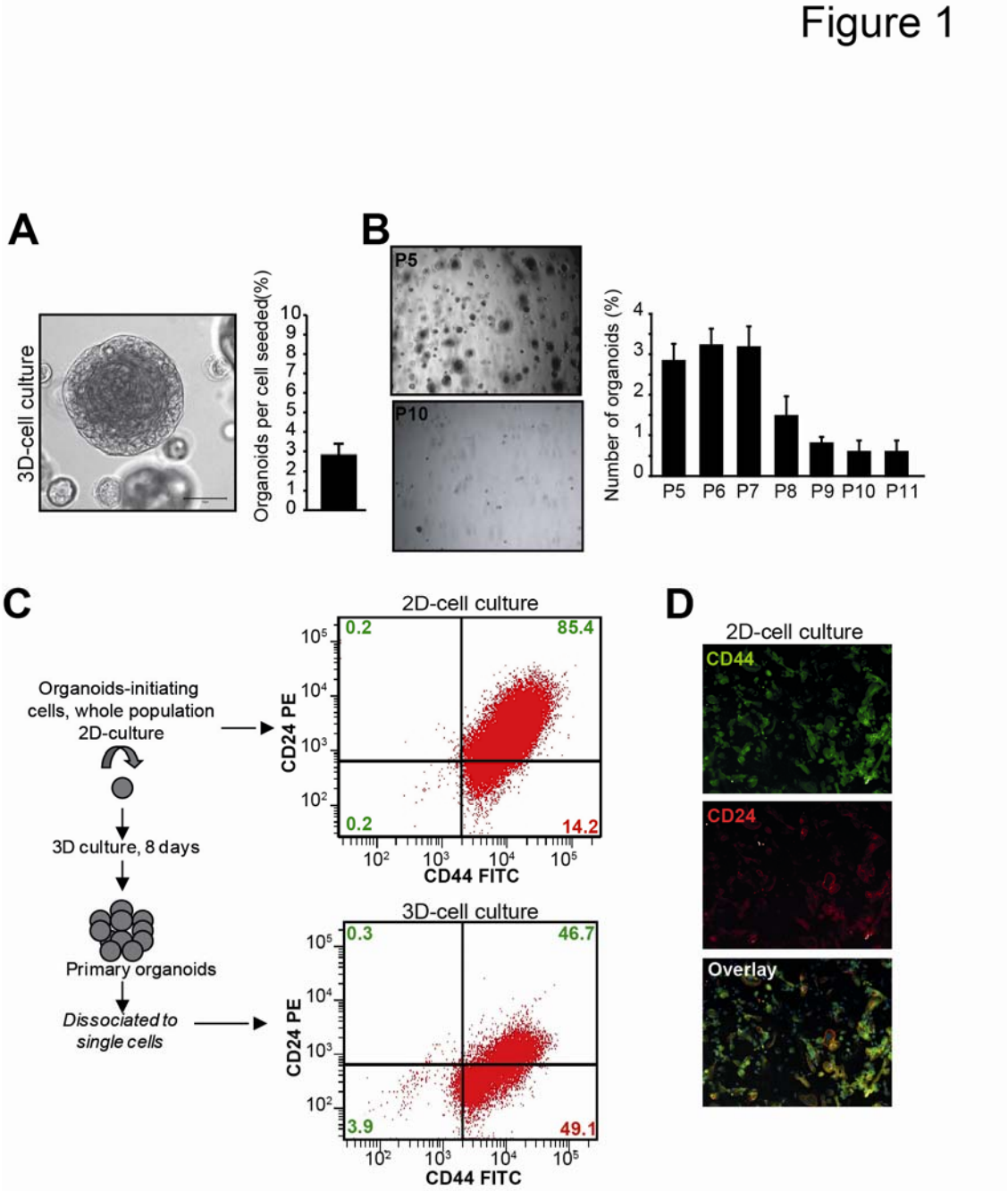


Figure 1. 3D-culture confers cells properties different from 2D-culture

A, Morphology of organoids from HMEC cells cultivated in Matrigel. HMECs are primary cells obtained by dissociation of purified non-diseased human mammary. Brightfield microscopic image of one organoid, bar length 50 µm. The bar graph shows the mean of organoids per well (mean ± SEM.) after 10 days of culture from three independent experiments. **B**, Representative brightfield pictures of organoids per well 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 following primary organoids culture (3D-cell culture - bottom). The expression of CD44high/CD24low in organoids cells was compared with the 2D culture cells. A minimum of 10,000 events were collected per sample. **D**, Analysis by immunofluorescence of CD44 (green) and CD24 (red) expression levels in HMEC from 2D culture.

Figure 2

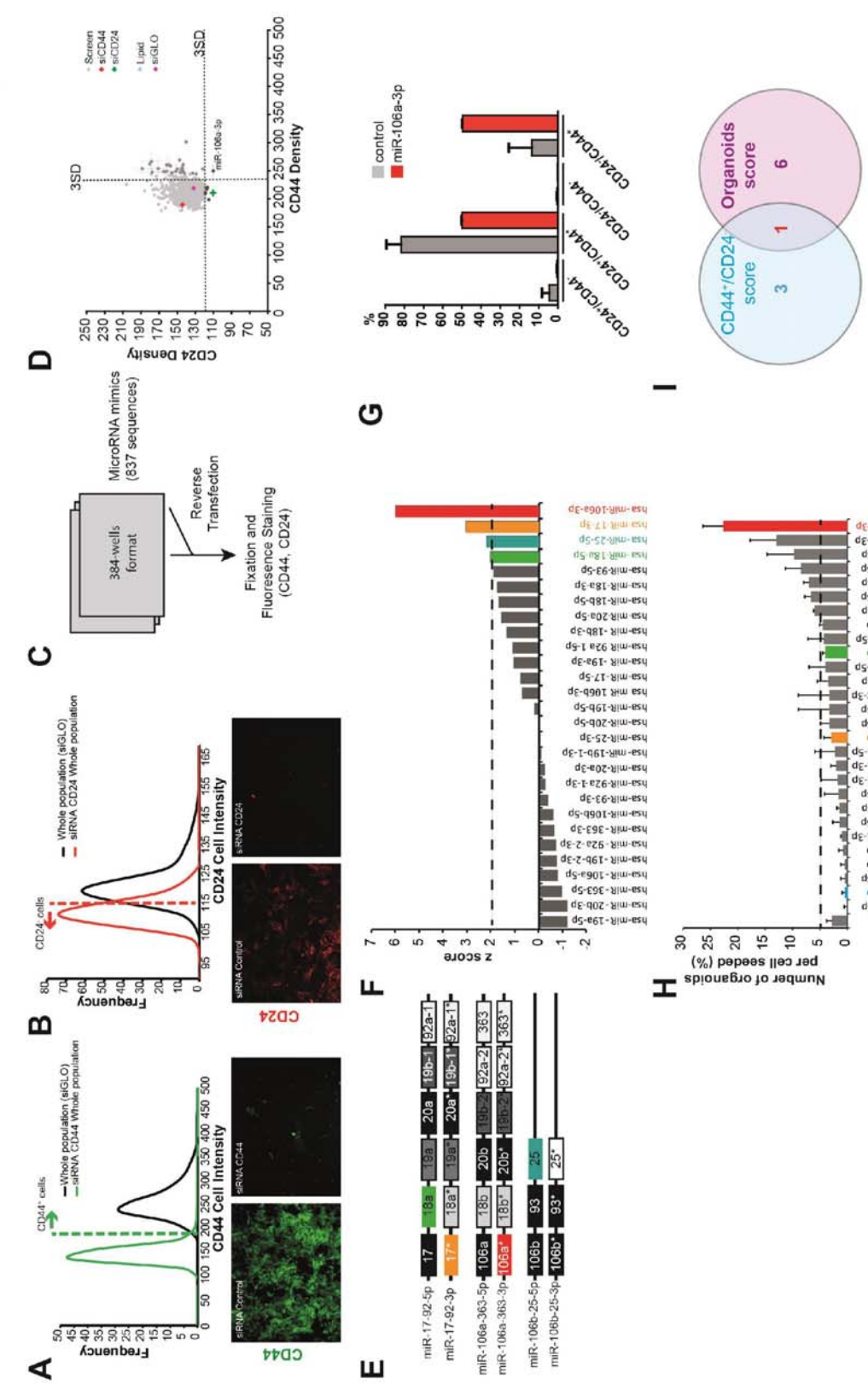

A
B
C
D
E
F
G
H
I
CD44+ cells
Whole population (siGLO)
siRNA CD44 Whole population
Frequency
CD44 Cell Intensity
siRNA Control
siRNA CD44
CD44
CD24 cells
Whole population (siGLO)
siRNA CD24 Whole population
CD24 Cell Intensity
siRNA CD24
MicroRNA mimics
(837 sequences)
384-wells
format
Reverse
Transfection
Fixation and
Fluoresence Staining
(CD44, CD24)
Screen
siCD44
siCD24
Lipid
siGLO
3SD
miR-106a-3p
CD24 Density
CD44 Density
control
miR-106a-3p
CD24+/CD44-
CD24-/CD44-
CD24-/CD44+
CD24+/CD44+
z score
Number of organoids
per cell seeded (%)
Organoids
score
6
1
CD44+/CD24-
score
3

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) as compared to HMEC-CD44-siRNA- depleted cells. **B,** Frequency distributions of CD24 intensity in HMECs at P6 (whole population, siGLO) as compared to HMEC-CD24-siRNA-depleted cells. Bottom panels are representative immunofluorescence pictures of HMECs stained with CD24 antibody (red) as compared to HMEC-CD24-siRNA-depleted cells. **C,**Workflow for image-based screening of miRNA screening for CD44high/CD24low 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 intensity. Based on the frequency distributions generated for each of phenotypic criteria (CD44 and CD24 intensity levels), we assigned highly stringent cutoffs for scoring positive hits in the genome-wide screen (dashed lines, 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 above the mean of the distribution. In colors are the miRNA above the 2SD. **G,** The graph shows mean of the percent of CD44/CD24 subpopulations with ± SEM of at least three independent sorting experiments in HMEC transfected with miR-106a-3p mimic as compared to control. **H,** The bar graphs show the mean number of organoids per well for each miRNA transfected in HMECs as compared to cells transfected with siRNA control. **I,** Venn diagram depicting the overlap of miRNA scoring in common between the CD44high/CD24low and organoids scores. Note that the overall number of miRNAs in common would be the overlap of the intersect of these two Venn diagrams.

Figure 3

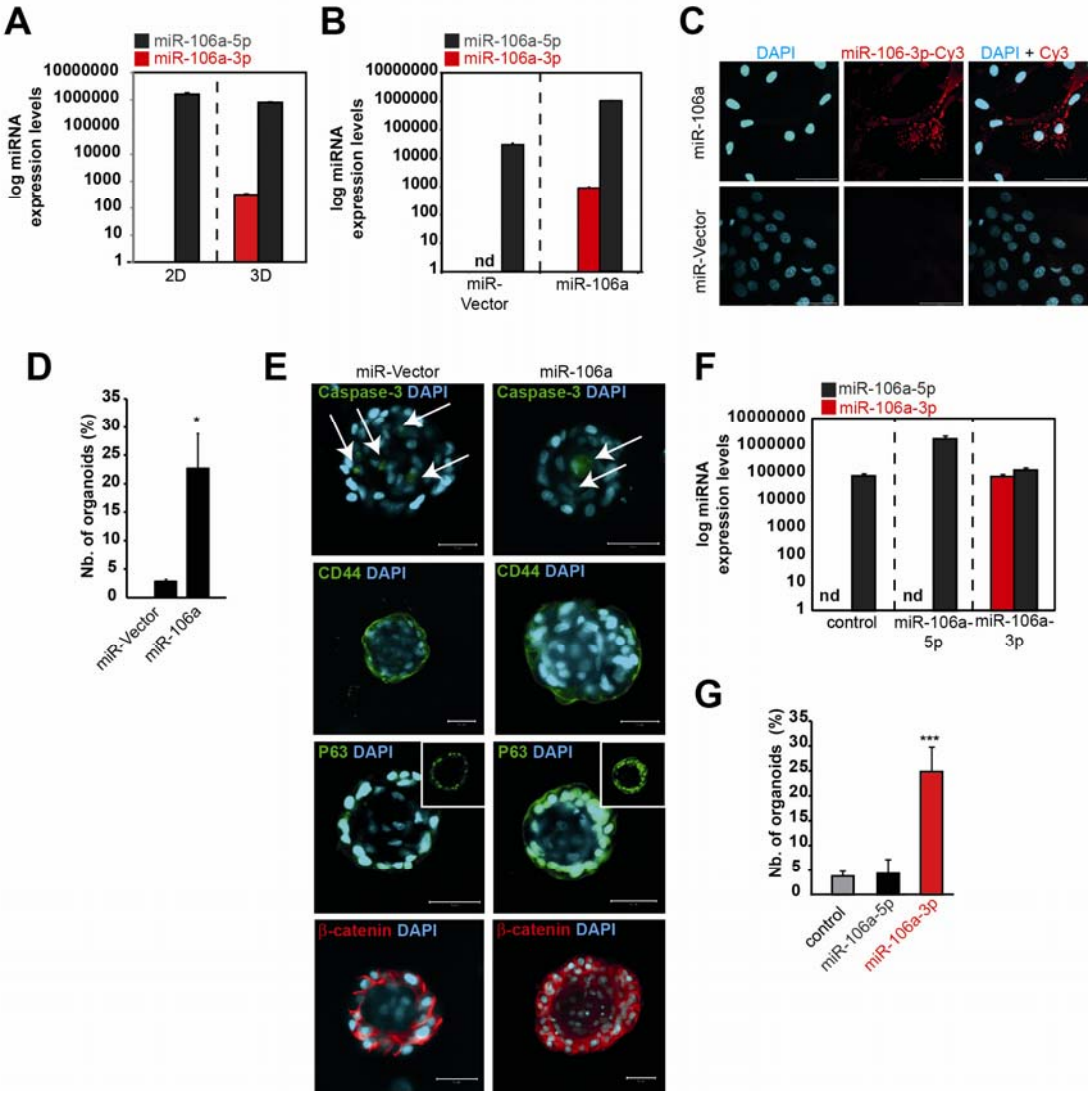
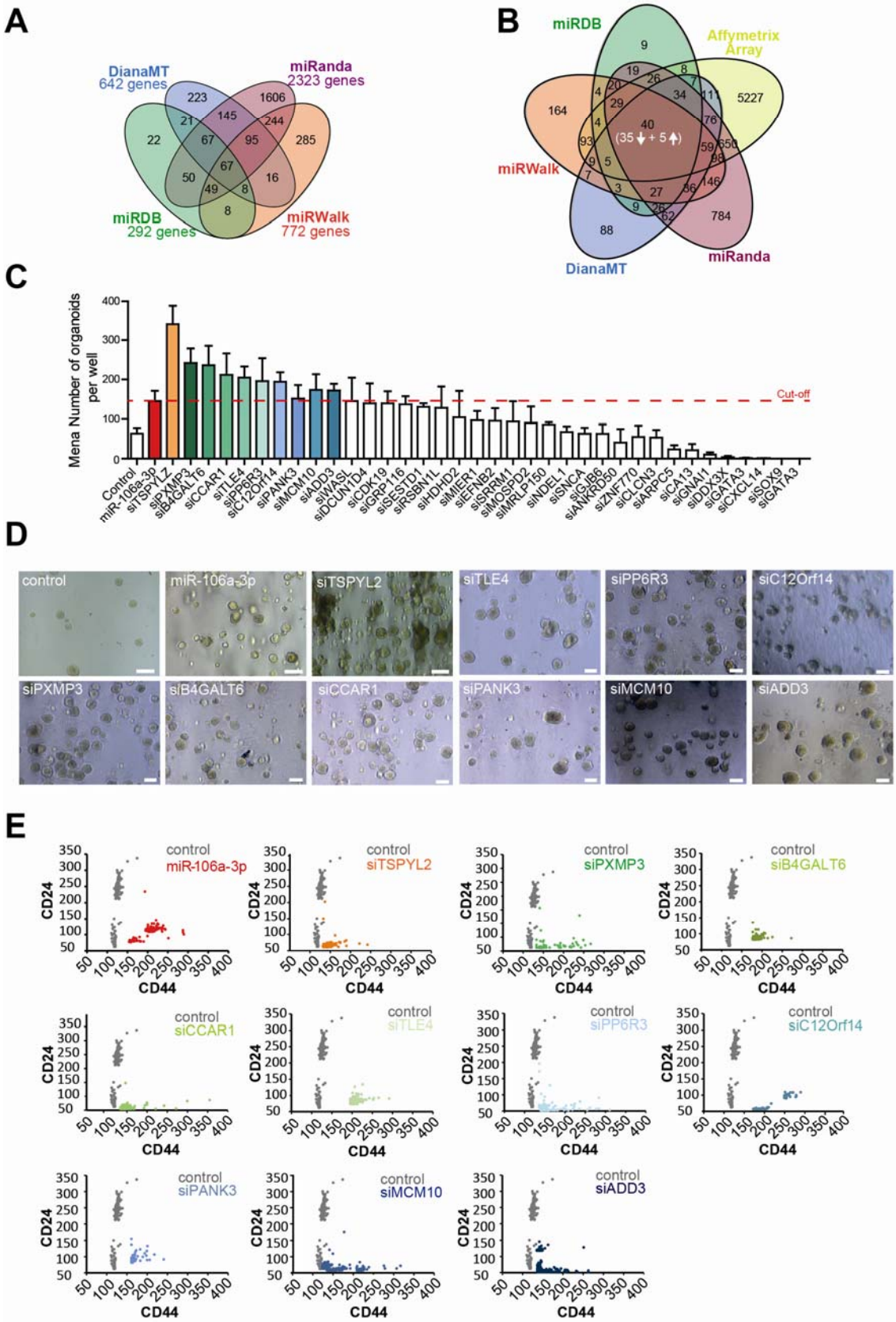


Figure 3. Properties of miR-106a-3p

A, Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in HMECs grown in 2D as compared to 3D-culture. **B,** Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in stable cell lines obtained by retroviral infection of HMECs with miR-Vector or miR-106a. **C,** 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. **D,** The bar graphs show the mean number of organoids per well for miR-106a transfected HMECs as compared to cells transfected 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,** 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. **F,** Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in HMEC transfected with either control mimic, miR106a-5p or miR106a-3p. n.d. means not detectable. **G,** Percentage of organoids formed by cells seeded for HMEC transfected with either control mimic, or miR106a-5p or 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.

Figure 4


A
DianaMT
642 genes
miRanda
2323 genes
miRDB
292 genes
miRWalk
772 genes
B
miRDB
Affymetrix Array
miRWalk
DianaMT
miRanda
(35 ↓ + 5 ↑)
C
Mena Number of organoids per well
Cut-off
Control
miR-106a-3p
siTSPYL2
siPXMP3
siB4GALT6
siCCAR1
siTLE4
siPP6R3
siC12Orf14
siPANK3
siMCM10
siADD3
siWASL
siDCUN1D4
siCDK19
siGRP116
siSESTD1
siRSBN1L
siHDHD2
siMIER1
siEFNB2
siSRRM1
siMOSPD2
siMRLP150
siNDEL1
siSNCA
siGJB6
siANKRD50
siZNF770
siCLCN3
siARPC5
siCA13
siGNAI1
siDDX3X
siGATA3
siCXCL14
siSOX9
siGATA3
D
control
miR-106a-3p
siTSPYL2
siTLE4
siPP6R3
siC12Orf14
siPXMP3
siB4GALT6
siCCAR1
siPANK3
siMCM10
siADD3
E
control
miR-106a-3p
siTSPYL2
siPXMP3
siB4GALT6
siCCAR1
siTLE4
siPP6R3
siC12Orf14
siPANK3
siMCM10
siADD3
CD24
CD44

651

Figure 4. Identification and validation of miR-106a-3p targets

A, Venn diagram showing the distribution of the computed miR-106a-3p targets identified through 4 algorithms DianaMT, miRanda, miRDB and miRWalk. **B,** Venn diagram showing the overlap between the computed miR-106a-3p targets from (A) and Affymetrix expression profiling array. Among all genes expressed in HMEC-miR106a-3p-transfected cells, 35 were both predicted miR-106a-3p targets and downregulated as revealed by Affymetrix array. **C,** The bar graph shows the mean of organoids per well (mean ± SEM.) in siRNA-transfected cells, three independent experiments for siRNA **D,** Representative brightfield pictures of organoids per well grown in 3D in siRNA-transfected cells. **E,** Scatter plots of CD44/CD24 intensity in normal HMEC transfected with miR-106a-3p (red); and each siRNA - potential targets of miR-106a-3p.

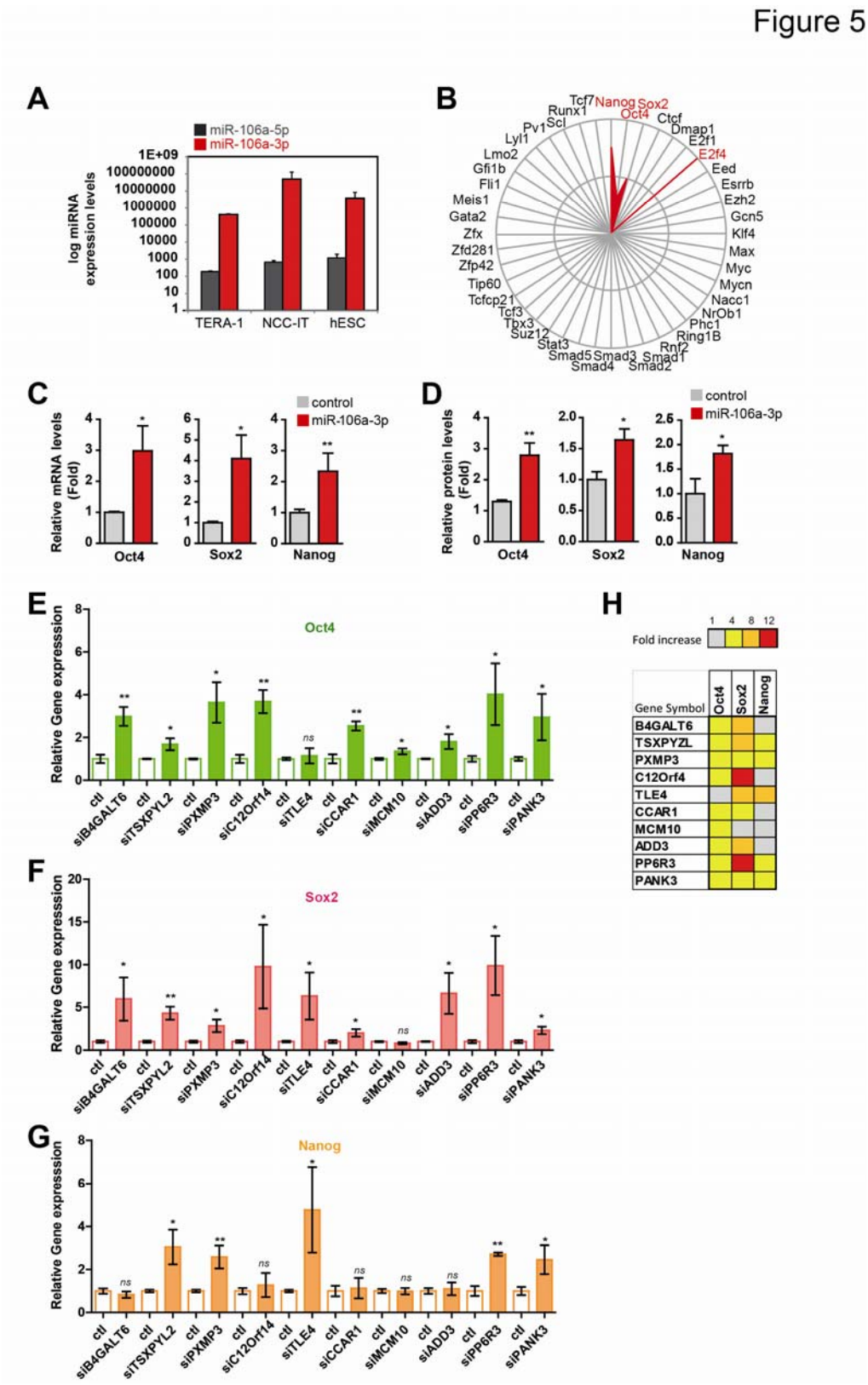
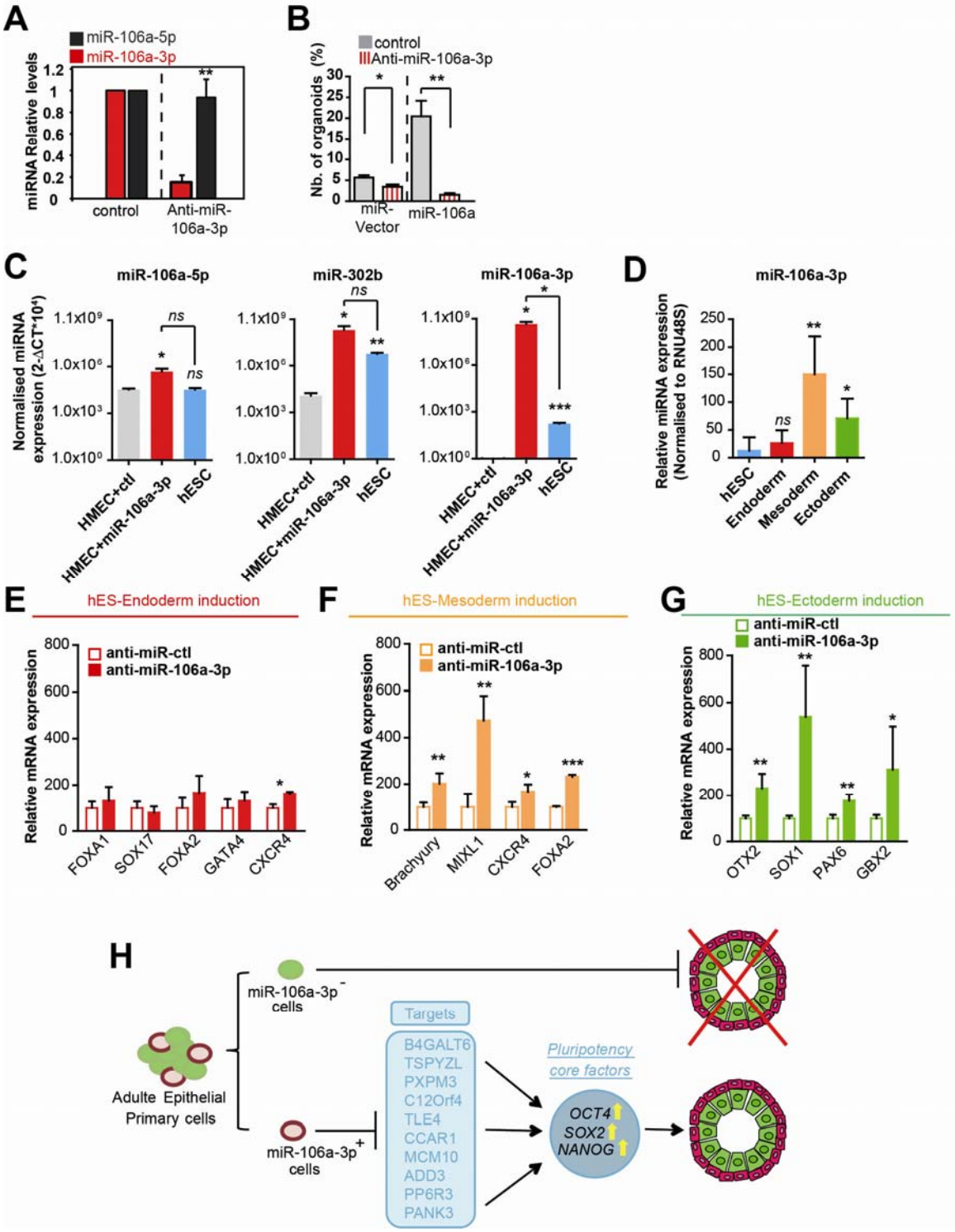


Figure 5. miR-106a-3p targets regulate pluripotency master regulators

A, Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in TERA-1, NCCIT and hESCs. **B,** The graph displayed the overlap significance and the overlap

percentage of the input gene lists with the sets of Transcription Factors (TF) target genes determined by ChIP-chip and ChIP-Seq studies. Each spoke represents the merged set of target genes for a specific TF. The radar chart shows that a high percentage of the input genes are known targets of the pluripotency master regulators (Nanog, Sox2, Oct4 and E2f4). **C,** RT-qPCR measurements of the mRNA levels of key regulators of pluripotency in HMEC transfected with control mimic or miR106a-3p. **D,** Relative protein expression levels of key regulators of pluripotency in HMEC transfected with control mimic or miR106a-3p. **E,** RT-qPCR analysis of OCT4 levels in siRNA-transfected cells (n = 3 independent transfection experiments). The data are shown relative to the control samples transfected with nontargeting "scrambled" RNAi sequence. **F,** RT-qPCR analysis of SOX2 levels in siRNA-transfected cells (n = 3 independent transfection experiments). The data are shown relative to the control samples transfected with nontargeting "scrambled" RNAi sequence. **G,** RT-qPCR analysis of NANOG levels in siRNA-transfected cells (n = 3 independent transfection experiments). The data are shown relative to the control samples transfected with nontargeting "scrambled" RNAi sequence. 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. **H,** Top, Color code used to illustrate the number of fold increase of the experimental siRNA value from the siRNA control in the RT-qPCR experiments. Bottom, Heat maps of the fold increase scores for each gene (Oct4, Sox2, Nanog) following transfection with the respective siRNAs.

Figure 6



692

693 **Figure 6. miR-106a-3p, a stem cell determinant for oganoids and development**
694 **A,** miR106a-3p expression levels (measured by RT-qPCR and normalized to RNA48)
695 following miR-106a-infected HMEC cell transfection with LNA-control (left) or LNA-anti-
696 miR-106a-3p (right). **B,** Percentage of organoids per cell seeded formed by miR-106a-

infected HMEC cells transfected with LNA-control (left) or LNA-anti-miR-106a-3p (right). **C,** Relative miR-106a-5p, miR-302b and miR-106a-3p expression levels determined by RT-qPCR in HMEC transfected with either control mimic or miR106a-3p, as compared to hESCs cells. 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,** Relative endogenous miR-106a-3p expression levels determined by RT-qPCR in hESCs cells following the 3 germ layer induction. 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. **E,** Relative mRNA expression levels of differentiation target genes in hESCs transfected with control mimic or anti-miR106a-3p following Endoderm induction. 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. **F,** Relative mRNA expression levels of differentiation target genes in hESCs transfected with control mimic or anti-miR106a-3p following Mesoderm induction. 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. **G,** Relative mRNA expression levels of differentiation target genes in hESCs transfected with control mimic or anti-miR106a-3p following Ectoderm induction. 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. **H,** Schematic representation of the mechanism of organogenesis initiated by miR-106a-3p.

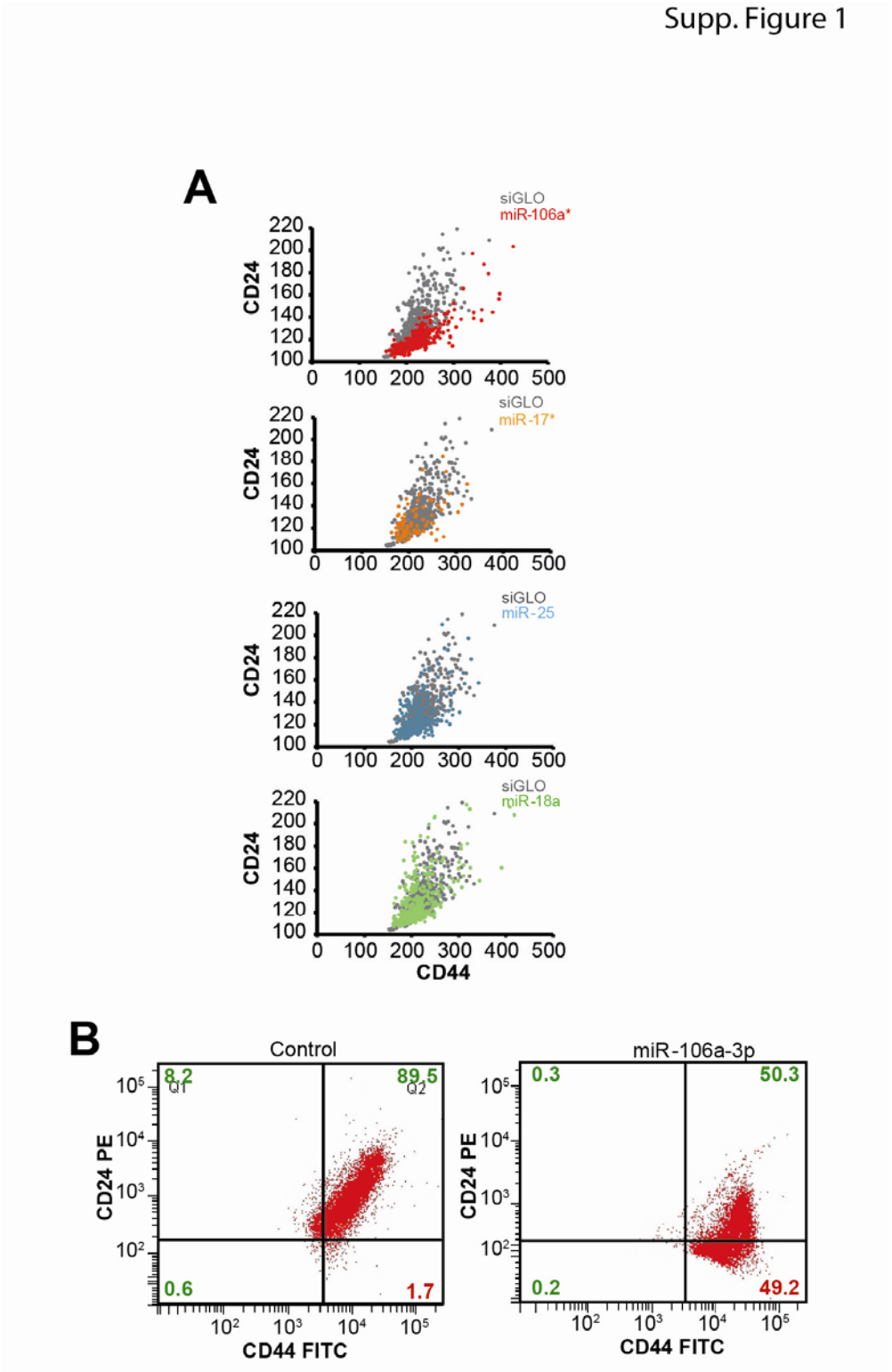
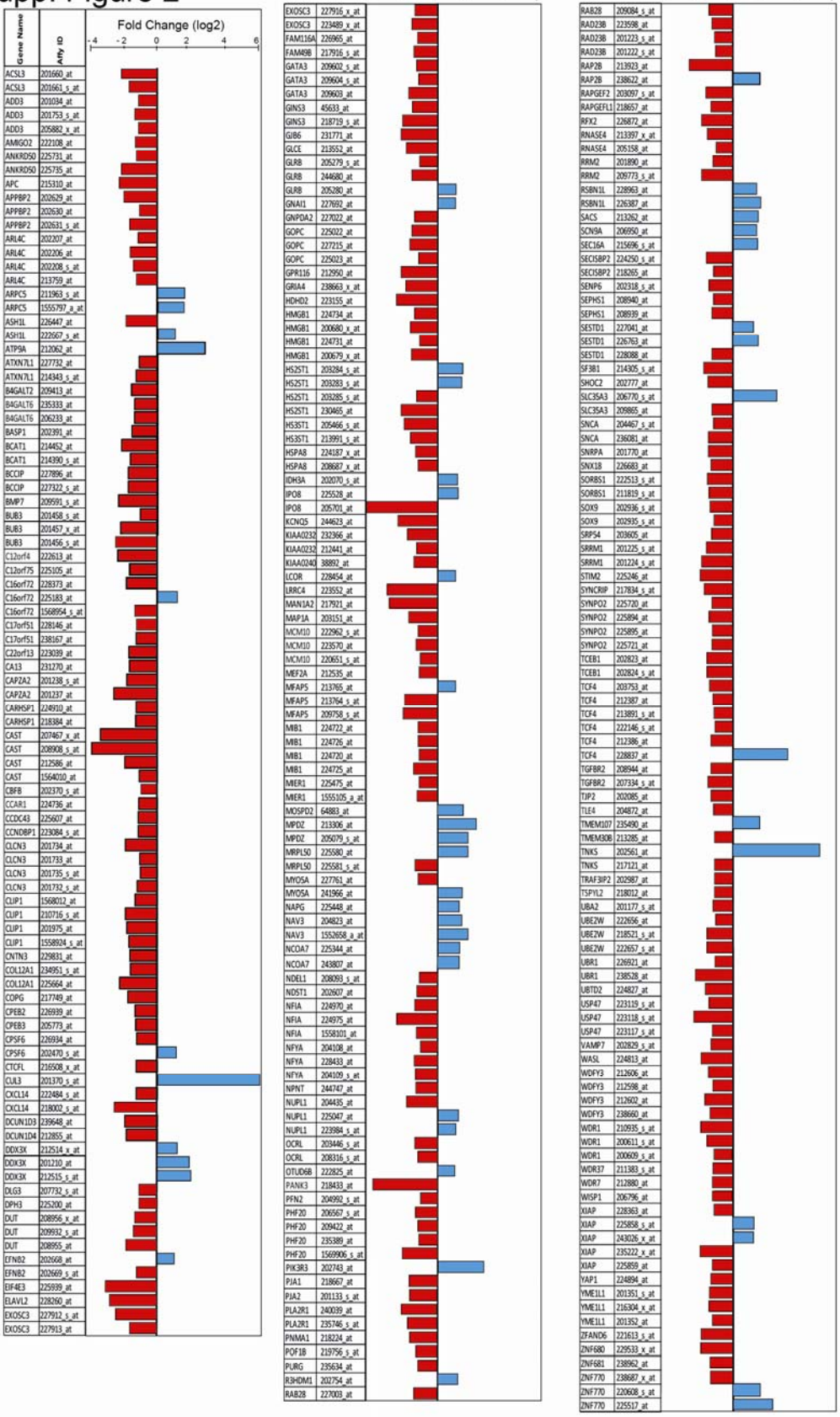


Figure S1

A, Representative levels of expression of CD24 and CD44 markers following transient transfection of either control mimic or miR106a-3p measured by FACS. B, 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).

Supp. Figure 2



729

Figure S2

The figure shows normalized expression ratio for the 40 genes targets identified by combining computational target prediction miRNA algorithm with Affymetrix expression profiling array (Log_2 fold change) in miR-106a-3p-transfected cells (*$P \leq 0.05$) for each Affymetrix probe. Positive or negative mean values indicate induction or repression of gene expression respectively ($P \leq 0.05$).

Supp. Figure 3

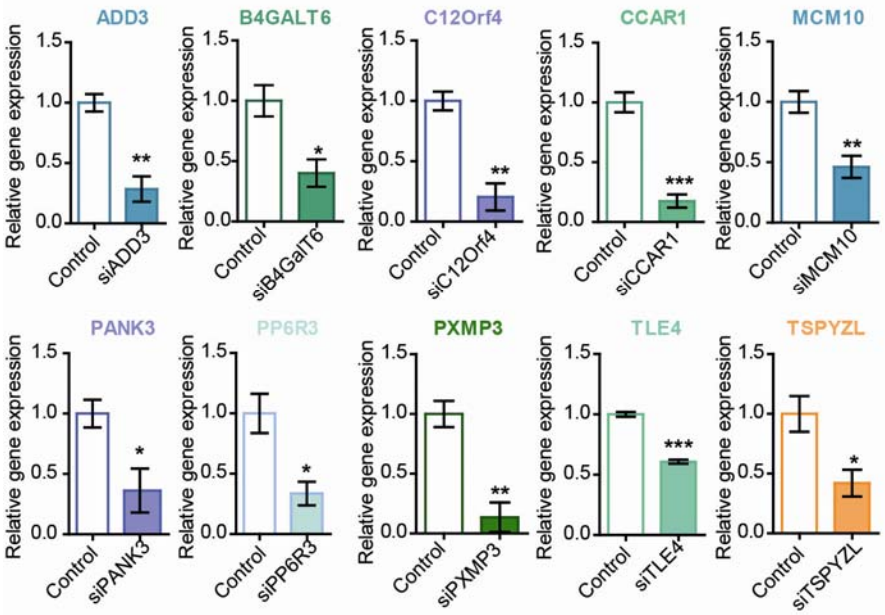


Figure S3

Efficiency of Gene Silencing. Verification of the efficiency of gene(s) silencing with specific human siRNA in HMEC cells. 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.

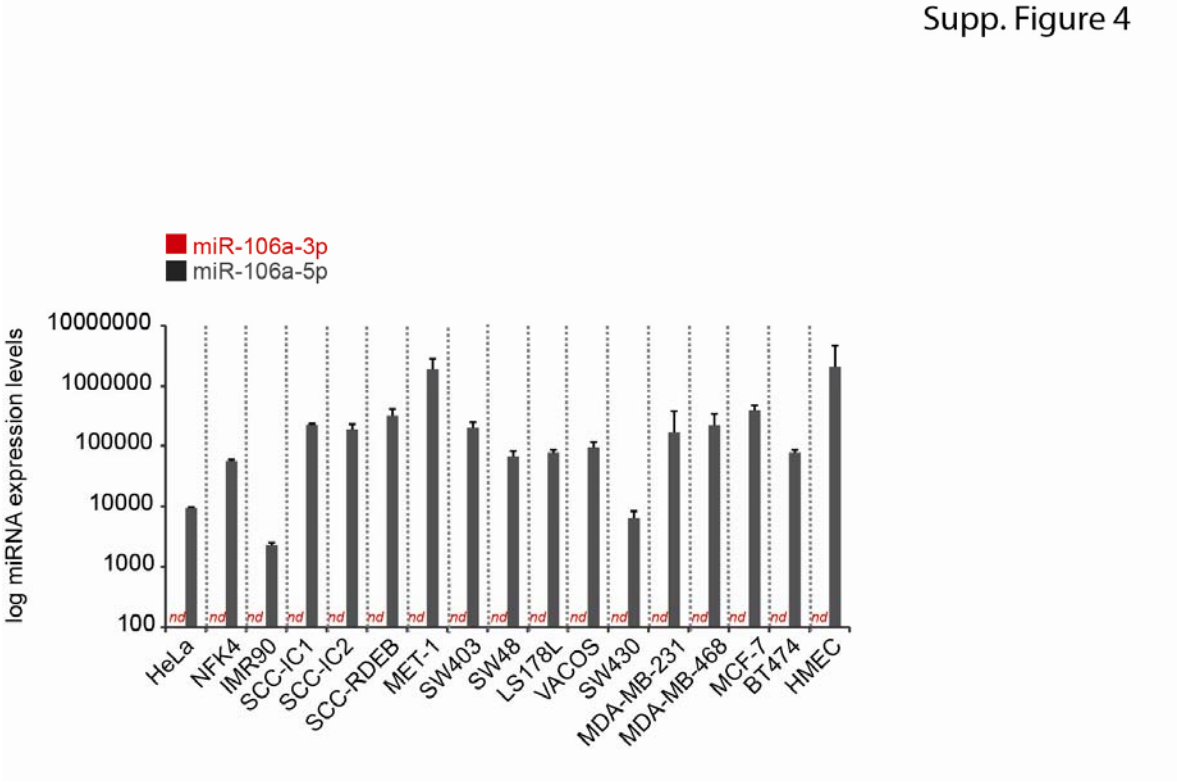


Figure S4

Relative miR-106a-3p and miR-106a-5p expression levels determined by RT-qPCR in a panel of 17 cell lines. *Nd* indicates not detectable. Means and standard errors are shown.

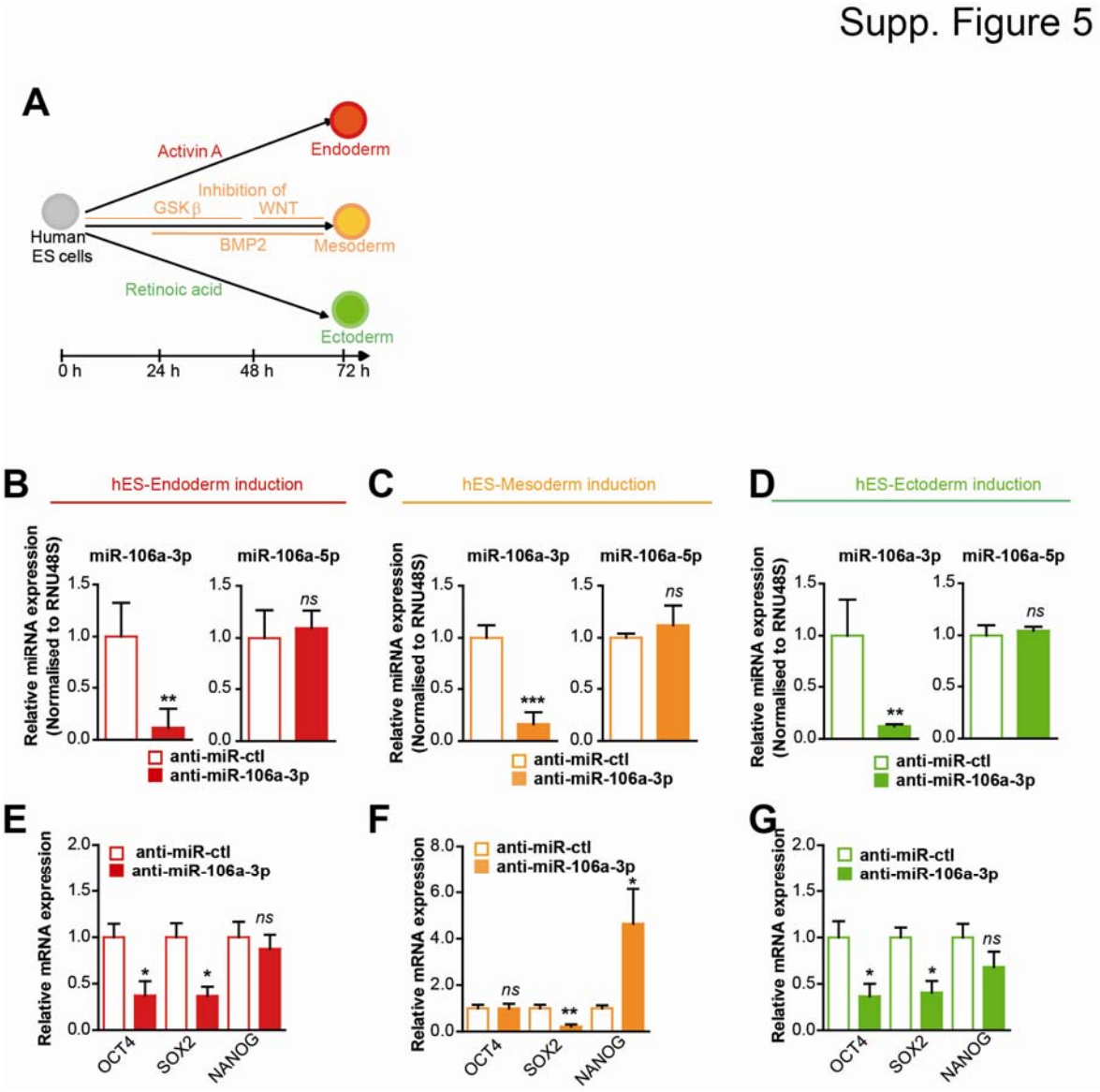


Figure S5

A, Schematic of the human ES cell differentiation system including timeline and key signaling pathways that are modulated. **B-D,** 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 (B), Mesoderm (C) and Ectoderm (D) differentiation. **E-G,** Relative mRNA expression levels of key regulators of pluripotency in hESCs transfected with control mimic or anti-miR106a-3p following induction of Endoderm (E), Mesoderm (F) and Ectoderm (G) 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.