SUPPLEMENTAL INFORMATION

Generation of a drug-inducible reporter system to study cell reprogramming in human cells

Sergio Ruiz[1], Athanasia D. Panopoulos[1], Nuria Montserrat[2], Marie-Christine Multon[3], Aurélie Daury[3], Corinne Rocher[3], Emmanuel Spanakis[3], Erika M. Batchelder[1], Cécile Orsini[3], Jean-François Deleuze[3], and Juan Carlos Izpisua Belmonte[1,2*]

[1]Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037

[2]Center of Regenerative Medicine in Barcelona, Doctor Aiguader, 88, 08003 Barcelona, Spain

[3]Regenerative Medicine and Cell Therapy Platform, Sanofi R&D, Centre de Recherche de Vitry-Alfortville, 13, quai Jules Guesde - BP 14 -94403 Vitry-sur-Seine Cedex-France

RUNNING TITLE: Human reprogramming reporter system for screening purposes.

To whom correspondence should be addressed: Juan Carlos Izpisua Belmonte: Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037 (belmonte@salk.edu, izpisua@cmrb.edu, Tel: 858-453-4100, ext: 1130; Fax: 858-453-2573)

KEYWORDS: Reporter system, reprogramming, doxycycline, microRNAs.

ADDITIONAL EXPERIMENTAL PROCEDURES.

Plamids: pMX-OCT4, pMX-SOX2, pMX-KLF4, pMX-cMYC, FUdeltaGW-rtTA, pLVFUtetO-OCT4, pLVFUtetO-SOX2, pLVFUtetO-KLF4 and pLVFUtetO-cMYC were obtained from Addgene (plasmids 17217, 17218, 17219, 17220, 19780, 19778, 19779, 19777 and 19775 respectively). The lentiviral vectors encoding specific miRNAs (pMIRNA1) were purchased to SBY, System Biosciences.

Retroviral and lentiviral production: Moloney-based retroviruses were generated as described. Third generation lentiviral vectors (pLV-FU-tetO and pMIRNA) were co-transfected with packaging plasmids (pMDL, REV and VSVg, kindly provided by Dr. Oded Singer, Laboratory of Genetics Lab, The SALK Institute, La Jolla, CA) in 293T cells using Lipofectamine (Invitrogen). Supernatants were collected and passed through a 0.45 μM filter to remove cellular debris.

Immunostainings: Imnunodetection of pluripotent markers in hiPSCs or of differentiation-associated markers in embryoid bodies or teratomas were performed as described (Ruiz et al, 2011). Briefly, cells were fixed with 4% paraformaldehyde in PBS for 15 minutes, washed in PBS and incubated with 0.5% Triton-X100 in PBS for 10 minutes and blocked with 5 % normal donkey serum in 1% PBS-BSA for 1 hour at RT. Antibodies against E-CADHERIN (BD Transduction Laboratories, 610181), NANOG (Abcam, Ab21624), TRA-1-60 (Chemicon, MAB4360), AFP (DAKO, A0008), FoxA2 (R&D, AF2400), TUJ1 (Covance, MMS-435-P), Alpha-smooth muscle actin (ASMA) (Sigma, A5228), Alpha-sarcomeric actin (ASA) (Sigma, A2172) and GFAP (Dako, AB1980) diluted in 1% PBS-BSA were used for overnight incubation at 4ºC. After extensive washes with PBS, cells were incubated with secondary biotin-conjugated anti-rabbit antibody, AlexaFluor 488 or 568 (Invitrogen) anti-mouse, rabbit or goat antibodies where correspond for an additional 2 hours at room temperature. For the immunodetection of NANOG, cells were afterwards incubated with streptavidin-HRP (Vector) and a DAB substrate kit for peroxidase (Vector, SK-4100) was used to develop the staining. DAPI was used to visualize nuclei at a concentration of 10μg/ml in PBS. For immunofluorescence analysis of BrdU, cell cultures were incubated with BrdU at 10μM final concentration for one hour in fibroblast media. Cells were fixed with 4% formaldehyde in PBS for 15 minutes, washed in PBS and incubated with HCl 2N for one hour. After extensive washing with PBS, cells were incubated with 0.5% Triton-X100 in PBS for 10 minutes and blocked with 5 % normal donkey serum in 1% PBS-BSA for 1 hour at RT. Antibodies against BrdU (Axill, OBT0030) diluted in 1% PBS-BSA were used for overnight incubation followed by incubation with a secondary AlexaFluor 568 (Invitrogen) for an additional 2 hours at room temperature. DAPI was used to visualize nuclei at a concentration of 10μg/ml in PBS.

RNA isolation and real time-PCR analysis: Total RNA was obtained using Trizol Reagent (Invitrogen) according the manufacturer's recommendations. The reverse transcription (RT) step was performed with 2μg of total RNA using the RT Supermix M-MuLV kit (BioPioneer). 0.25 μl of the final reaction was used to quantify gene expression by real time PCR using the SYBR-Green PCR Master mix (Applied Biosystems) in the ViiA 7 Real-time PCR System (Applied Biosystems). Values of gene expression were normalized using GAPDH expression and data is shown as fold change relative to the value of the sample control. All the samples were done in triplicate. The list of the primers used for Real-time-PCR experiments is available upon request. For the miR arrays, total RNA (including small RNA) from the samples indicated in the table below was extracted from $3 \cdot 10^6$ cells by using RNeasy® Mini Kit (Qiagen) according to manufacturer's protocols. RNA was quantified on a NanoDrop 8000 spectrophotometer (Thermo Scientific). For microarray analysis, RNA quality was determined on a Bioanalyser 2100 (Agilent) and only RNA samples with an RNA integrity number (RIN) between 8 and 10 were used.

miRNA quantitation and data normalization: Megaplex profiling using human TaqMan Low Density miRNA Arrays (TLDA) (Applied Biosystems) was used to evaluate the expression of 667 miRs as described by the manufacturer. Briefly, 600 ng of total RNA were used in two Megaplex reverse transcription (RT) reactions containing each one, a pool of Megaplex RT primers designed to detect and quantify up to 380 miRNAs and controls in a single reaction. No prior miR preamplification step was needed. Each reverse transcribed product was mixed with 2X TaqMan Universal PCR Master Mix, without AmpErase UNG (Applied Biosystems) and loaded onto a TLDA v2.0 containing 380 TaqMan miRNAs assays. TLDAs were run on a 7900HT Thermocycler (Applied Biosystems) using Sequence Detection Systems (SDS) software version 2.3. Data analysis was performed using SDS RQ manager v1.2 (Applied Biosystems) which utilizes the delta-delta CT method. The threshold cycle (CT) is defined as the fractional cycle number at which the fluorescence exceeds the fixed threshold of 0.1. For each sample, the median of the CT values from detected signals was used to normalize the RNA input. A total of 26 miRs were identified as uniquely or strongly expressed in pluripotent cells compared to somatic cells (see table with the list of miRs).

Identification of hsa-miR-519a putative targets: We used bioinformatic predictions (DIANA-MICROT, MICRORNA.ORG, MIRDB, RNA22-HAS, TARGETSCAN, TARGETMINER) to identify putative targets of hsa-miR-519a. Based on the highest scores of predicted targets we selected 14 genes (C7ORF60, C10ORF140, CHD9, IER3IP1, PHF14, EPHA4, PCDHB4, ZNF238, HEG1, ZNFX1, MYT1L, HABP4, ZBTB4 and TGFβRII) for further analysis.

AP-staining: For the AP staining, cells were fixed in a solution of 4% paraformaldehyde in PBS for 20 minutes. After extensive washes in PBS, cells were incubated in NTMT solution (10 mM NaCl, 100 mM Tris-HCl (pH 9), 50 mM MgCl2 and 0.1% Tween-20) for 5 minutes and then in NTMT solution supplemented with NBT (Nitro-Blue Tetrazolium Chloride) and BCIP (5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt) in the dark until the staining was developed.

SUPPLEMENTARY FIGURE LEGENDS:

Figure S1: dFib-OCT4^{GFP} cells behave as human fibroblasts in terms of their reprogramming ability. **(A)** Schematic representation of the experiment performed. **(B)** GFP detection and NANOG immunofluorescence analysis of dFib-OCT4^{GFP} cells infected with retroviruses encoding the indicated factors. Scale bar: 100 μm.

Figure S2: Generation of dFib-OCT4^{GFP}–ind lines with doxycycline-inducible expression of different subsets of reprogramming factors. **(A)** Morphology of the dFib-OCT4^{GFP}–ind fibroblast-like cells (at passage 7) after embryoid body-mediated differentiation. Scale bar: 50 μm. **(B)** dFib-OCT4^{GFP}–indSKC and dFib-OCT4^{GFP}–ind-OSK cells were infected with retroviruses encoding OCT4 and cMYC respectively and either untreated or treated with 100 ng/ml of doxycycline for 24 hours. Real time PCR analyses were performed to detect transcript corresponding to the 4 reprogramming factors. Note that there is no expression from the factors delivered by retrovirus to generate originally the primary hiPSC lines. Data are shown as relative averages ± SD of two biological replicates analyzed in triplicate. **(C)** Cell cultures of the indicated dFib-OCT4^{GFP}-ind cells were infected with retroviruses encoding the missing reprogramming factor and treated with different doses of doxycycline for 18 days (*upper wells*). Uninfected cells either untreated or treated with doxycycline were used as negative controls of the experiment (*lower wells*). hiPSC colonies were detected by alkaline phosphatase staining.

Figure S3: Identification of miRs strongly expressed in pluripotent cells or downregulated in somatic cell types. **(A)** Samples plotted along the first and second principal components. Note that pluripotent cells group together. **(B)** Heat map for the 26 miRNAs selected due to their differential expression in pluripotent cells or strong downregulation in somatic cells. Red indicates high expression level; yellow, medium; blue, low expression level.

Figure S4: Expression of hsa-miR-519a does not influence the endogenous level of pluripotent stem cell markers. **(A)** Real-time PCR analysis to detect the levels of putative markers targeted by hsa-miR-519a performed with RNA obtained from BJ fibroblasts infected with the indicated miRs. RNA extraction was performed five days after infection. Only genes showing significant down-regulation are shown. See the complete list of genes evaluated in Additional Experimental Procedures. Data are shown as relative averages ± SD of two biological replicates analyzed in triplicate. **(B)** Real-time PCR analysis to detect the levels of the indicated endogenous stem cell markers with RNA obtained from BJ fibroblasts infected with three (OSK) or four (OSKC) reprogramming factors and the indicated miRs. RNA extraction was performed six days after infection. Data are shown as relative averages ± SD of two biological replicates analyzed in triplicate.

Table S1: *Samples profiled on miR arrays.*

Sample name	**Cell type**	**Origin**
CRL-2429	Neonatal fibroblasts	ATCC
CRL-2522	BJ foreskin fibroblasts	ATCC
CC-2511	Adult fibroblasts	Lonza
CC-2509	Neonatal fibroblasts	Lonza
CDMPRO	Melanocytes	Promocell
WA01	ESC H1 cell line	Wicell
WA07	ESC H7 cell line	Wicell
WA09	ESC H9 cell line	Wicell
hiPSC43A2	CRL-2429 fibroblasts reprogrammed with lentivirus STEMCCA	
hiPSC43B2	CRL-2429 fibroblasts reprogrammed with lentivirus STEMCCA	
hiPSC43ID6	CRL-2429 fibroblasts reprogrammed with lentivirus STEMCCA	
hiPSC57A5	CRL-2522 fibroblasts reprogrammed with lentivirus STEMCCA	
hiPSC57A7	CRL-2522 fibroblasts reprogrammed with lentivirus STEMCCA	
hiPSC57B7	CRL-2522 fibroblasts reprogrammed with lentivirus STEMCCA	
F1hiPSC4F	F1 fibroblasts reprogrammed with lentivirus STEMCCA	
CBhiPSC4F3	umbilical cord blood cells reprogrammed with retroviruses encoding OCT4, SOX2, KLF4 and cMYC	
CBhiPSC3F12	umbilical cord blood cells reprogrammed with retroviruses encoding OCT4, SOX2 and KLF4	
CBhiPSC2F3	umbilical cord blood cells reprogrammed with retroviruses encoding OCT4 and SOX2	

Table S2: *Table with the description of the 26 miRs specifically expressed in pluripotent cells or strongly downregulated in somatic cells.*

miRNA	***Similarity***	***PC1 +***	***Expr. Level somatic cells/pluripotent cells***	***Cluster chr (cluster)***	**miR family TargetScan**
hsa-miR-302a	1	0.98	-/+++	chr4 (302-367)	miR-106/302
hsa-miR-302b	0.98	0.97	-/++	chr4 (302-367)	miR-106/302
hsa-miR-302d	0.97	0.86	-/+++	chr4 (302-367)	miR-106/302
hsa-miR-302c	0.98	0.96	-/+++	chr4 (302-367)	miR-106/302
hsa-miR-372	0.98	0.97	-/++	chr19 (371-373)	miR-106/302
hsa-miR-520c-3p	0.96	0.95	-/++	ch19 (C19MC)	miR-106/302
hsa-miR-373	0.96	0.97	-/++	chr19 (371-373)	miR-106/302
hsa-miR-367	0.98	0.96	-/+++	chr4 (302)	miR-25/32/92ab/363/367
hsa-miR-363	0.88	0.93	-/++	chrX (363/92a-2/19b-2/20b/18b/106a)	miR-25/32/92ab/363/367
hsa-miR-371-3p	0.83	0.83	-/++	chr19 (371/372/373)	miR-371-3p
hsa-miR-200c	0.95	0.94	-/+++	chr12 (200c/141)	miR-200bc/429
hsa-miR-200b	0.93	0.9	-/++	chr1 (200b/200a/429)	miR-200bc/429
hsa-miR-429	0.94	0.91	-/++	chr1 (200b/200a/429)	miR-200bc/429
hsa-miR-141	0.96	0.98	-/++	chr12 (200c/141)	miR-141/200a
hsa-miR-519a	0.99	0.98	-/+	ch19 (C19MC)	miR-291b-3p/519a/519b-3p/519c-3p
hsa-miR-519b-3p	0.98	0.97	-/+	ch19 (C19MC)	
hsa-miR-512-3p	0.98	0.98	-/++	ch19 (C19MC) (512-1; 512-2)	miR-512-3p/1186
hsa-miR-517a	0.99	0.98	-/++	ch19 (C19MC)	miR-517ac
hsa-miR-517c	0.99	0.97	-/++	ch19 (C19MC)	miR-517ac
hsa-miR-518b	0.99	0.97	-/+	ch19 (C19MC)	miR-518a-3p/518bcf/518d-3p
hsa-miR-518e	0.98	0.98	-/+	ch19 (C19MC)	miR-518e
hsa-miR-520h	0.98	0.97	-/++	ch19 (C19MC)	miR-520gh
hsa-miR-520g	0.99	0.98	-/++	ch19 (C19MC)	miR-520gh
hsa-miR-519d	0.99	0.98	-/++	ch19 (C19MC)	miR-17-5p/20/93.mr/106/519.d
hsa-miR-20b	0.93	0.96	+/+++	chrX (363/92a-	miR-17-

				2/19b-2/20b/18b/106a)	5p/20/93.mr/106/519.d
hsa-miR-205	0.99	0.98	-/++	chr1 (no)	miR-205

Similarity score: To obtain the list of miRs enriched in pluripotent cells, miRs having similar expression profiles to hsa-miR-302a were identified using Pearson correlation (GeneSpring 11.5.1, Agilent) and principal component analysis (SPSS Statistics® version 19 software, IBM). In brief, using Ward clustering, the samples were classified into two main clusters: pluripotent samples and somatic cells. Then, 177 major miRNAs presenting a significant ($p<0.05$) difference between those two clusters by analysis of variance were selected and entered a principal components analysis. The first principal component separated the hiPSC lines from the differentiated cells and was, therefore, interpreted as the pluripotency component. Values in PC1 column correspond to the calculated coefficients of each miRNA in principal component 1; These results, combined with expression level comparisons (miRNAs expression levels are based on normalized Ct values, +++: high expression level; ++: medium expression level, +: low expression level, -: no expression), allowed the selection of 26 miRNAs. C19MC: chromosome 19 microRNA cluster. miR families are defined as described in TargetScan website (http://www.targetscan.org/).

FIGURE S1:

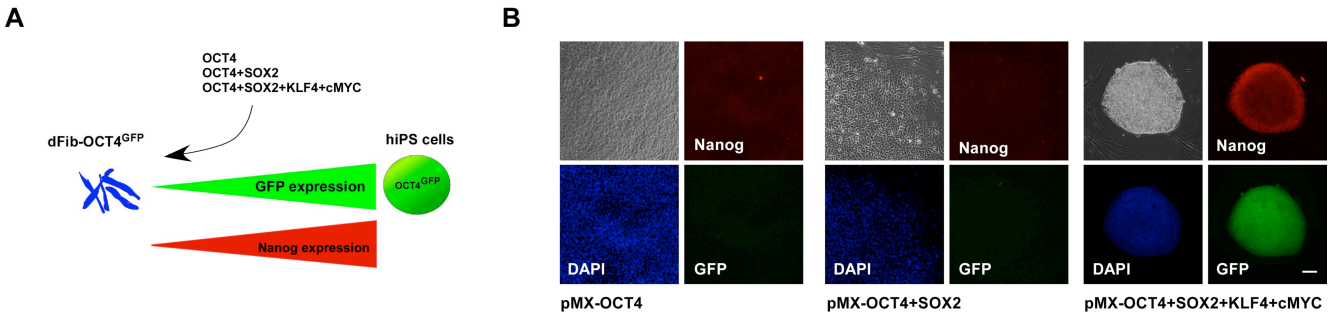

A
OCT4
OCT4+SOX2
OCT4+SOX2+KLF4+cMYC
dFib-OCT4GFP
hiPS cells
GFP expression
OCT4GFP
Nanog expression
B
Nanog
Nanog
Nanog
DAPI
GFP
DAPI
GFP
DAPI
GFP
pMX-OCT4
pMX-OCT4+SOX2
pMX-OCT4+SOX2+KLF4+cMYC

FIGURE S2:

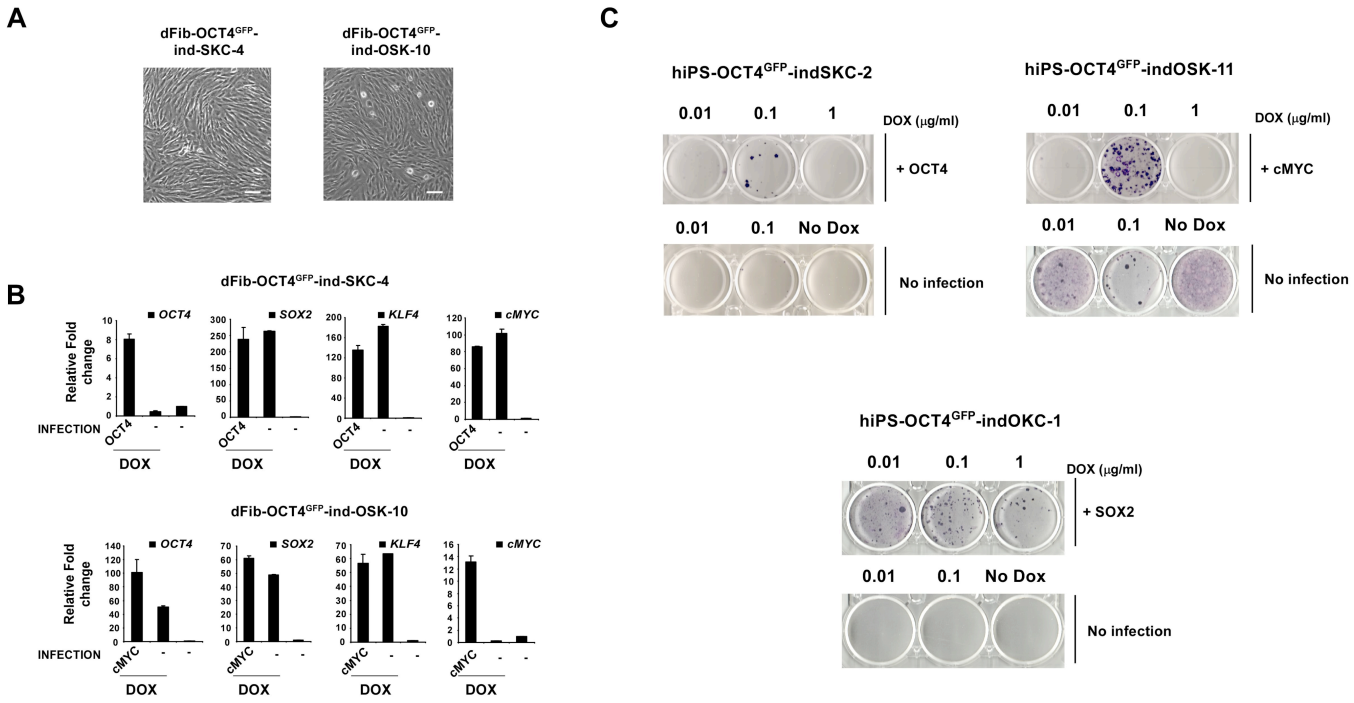

A
dFib-OCT4GFP-ind-SKC-4
dFib-OCT4GFP-ind-OSK-10
B
dFib-OCT4GFP-ind-SKC-4
OCT4
SOX2
KLF4
cMYC
Relative Fold change
INFECTION
DOX
dFib-OCT4GFP-ind-OSK-10
C
hiPS-OCT4GFP-indSKC-2
hiPS-OCT4GFP-indOSK-11
hiPS-OCT4GFP-indOKC-1
0.01
0.1
1
No Dox
DOX (μg/ml)
+ OCT4
+ cMYC
+ SOX2
No infection

FIGURE S3:

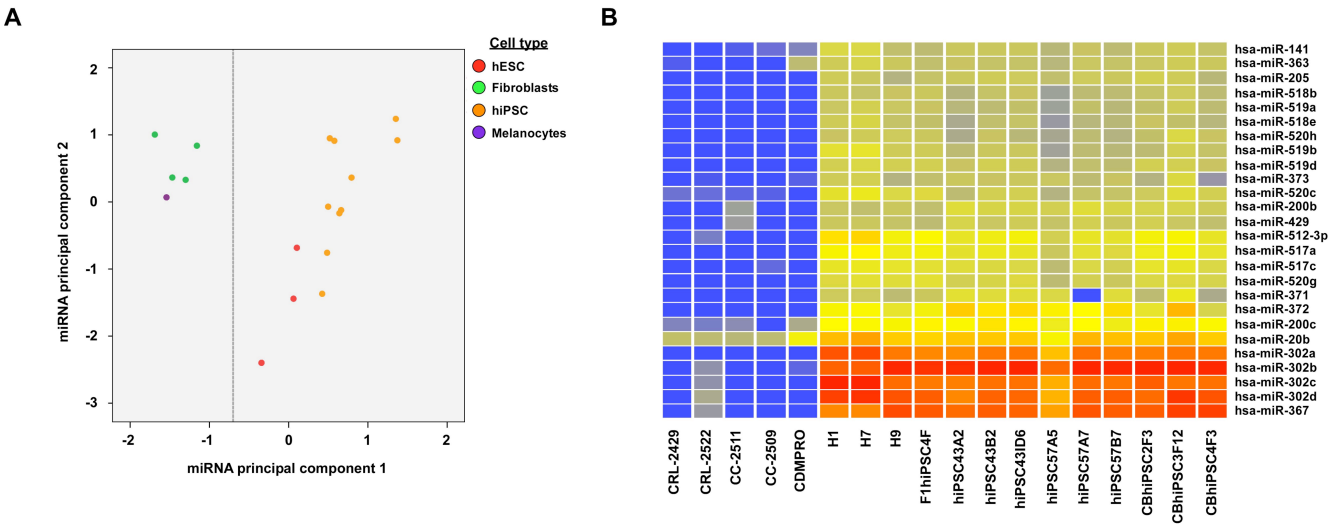
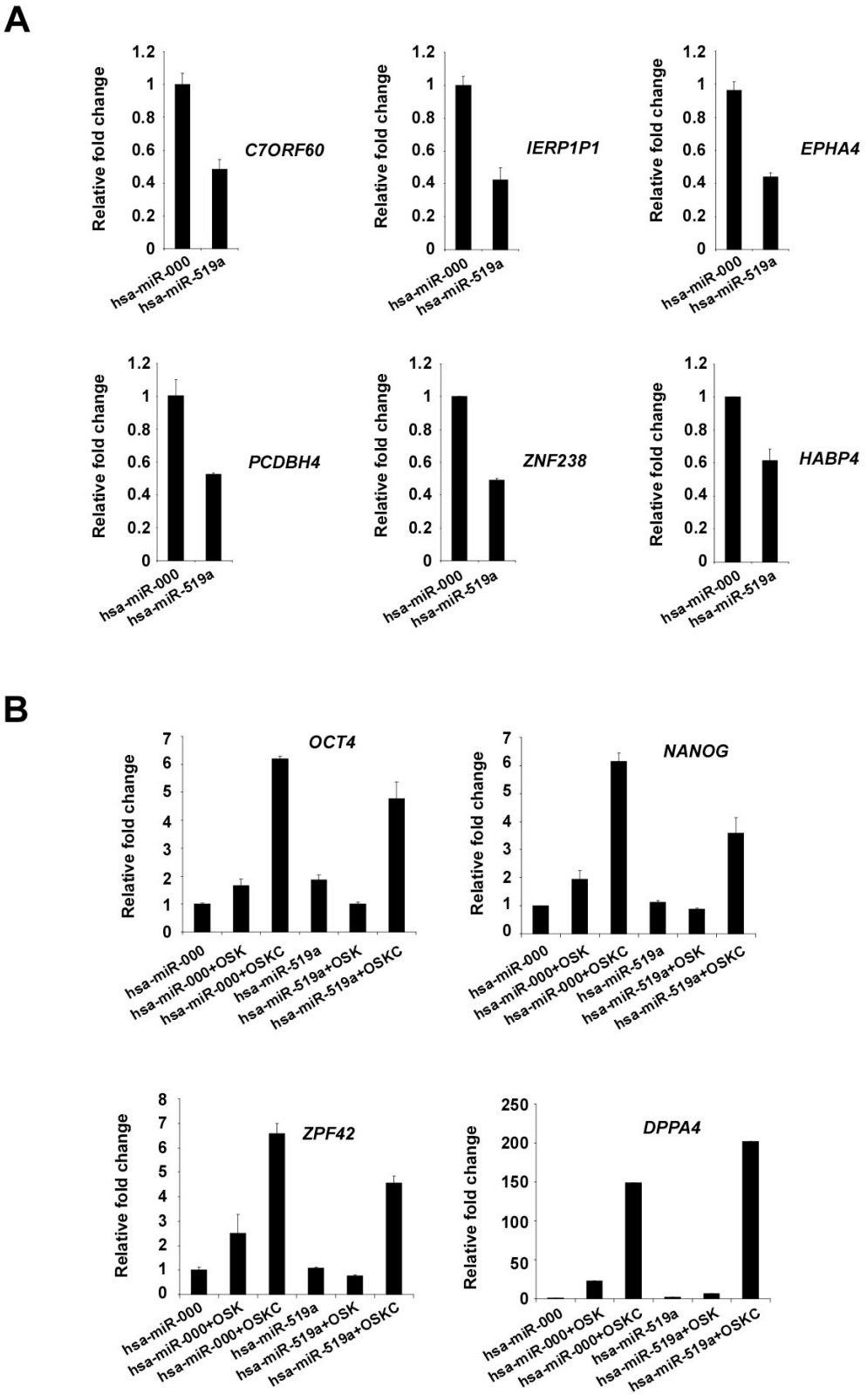

A
Cell type
hESC
Fibroblasts
hiPSC
Melanocytes
miRNA principal component 2
miRNA principal component 1
2
1
0
-1
-2
-3
-2
-1
0
1
2
B
hsa-miR-141
hsa-miR-363
hsa-miR-205
hsa-miR-518b
hsa-miR-519a
hsa-miR-518e
hsa-miR-520h
hsa-miR-519b
hsa-miR-519d
hsa-miR-373
hsa-miR-520c
hsa-miR-200b
hsa-miR-429
hsa-miR-512-3p
hsa-miR-517a
hsa-miR-517c
hsa-miR-520g
hsa-miR-371
hsa-miR-372
hsa-miR-200c
hsa-miR-20b
hsa-miR-302a
hsa-miR-302b
hsa-miR-302c
hsa-miR-302d
hsa-miR-367
CRL-2429
CRL-2522
CC-2511
CC-2509
CDMPRO
H1
H7
H9
F1hiPSC4F
hiPSC43A2
hiPSC43B2
hiPSC43ID6
hiPSC57A5
hiPSC57A7
hiPSC57B7
CBhiPSC2F3
CBhiPSC3F12
CBhiPSC4F3

FIGURE S4:


A
Relative fold change
C7ORF60
IERP1P1
EPHA4
PCDBH4
ZNF238
HABP4
hsa-miR-000
hsa-miR-519a
B
OCT4
NANOG
ZPF42
DPPA4
hsa-miR-000
hsa-miR-000+OSK
hsa-miR-000+OSKC
hsa-miR-519a
hsa-miR-519a+OSK
hsa-miR-519a+OSKC