

# Gene Expression Signatures Defining Fundamental Biological Processes in Pluripotent, Early, and Late Differentiated Embryonic Stem Cells

John Antonydas Gaspar,<sup>1</sup> Michael Xavier Doss,<sup>2</sup> Johannes Winkler,<sup>3</sup> Vilas Wagh,<sup>1</sup> Jürgen Hescheler,<sup>1</sup> Raivo Kolde,<sup>4,5</sup> Jaak Vilo,<sup>4,5</sup> Herbert Schulz,<sup>6</sup> and Agapios Sachinidis<sup>1</sup>

Investigating the molecular mechanisms controlling the *in vivo* developmental program postembryogenesis is challenging and time consuming. However, the developmental program can be partly recapitulated *in vitro* by the use of cultured embryonic stem cells (ESCs). Similar to the totipotent cells of the inner cell mass, gene expression and morphological changes in cultured ESCs occur hierarchically during their differentiation, with epiblast cells developing first, followed by germ layers and finally somatic cells. Combination of high throughput -omics technologies with murine ESCs offers an alternative approach for studying developmental processes toward organ-specific cell phenotypes. We have made an attempt to understand differentiation networks controlling embryogenesis *in vivo* using a time kinetic, by identifying molecules defining fundamental biological processes in the pluripotent state as well as in early and the late differentiation stages of ESCs. Our microarray data of the differentiation of the ESCs clearly demonstrate that the most critical early differentiation processes occur at days 2 and 3 of differentiation. Besides monitoring well-annotated markers pertinent to both self-renewal and potency (capacity to differentiate to different cell lineage), we have identified candidate molecules for relevant signaling pathways. These molecules can be further investigated in gain and loss-of-function studies to elucidate their role for pluripotency and differentiation. As an example, siRNA knockdown of *MageB16*, a gene highly expressed in the pluripotent state, has proven its influence in inducing differentiation when its function is repressed.

## Introduction

EMBRYONIC STEM CELLS (ESCs) are derived from the inner cell mass (ICM) of preimplantation embryos [1] and can be cultured and propagated indefinitely in the pluripotent state. ESCs can be hierarchically differentiated into specialized cells of the germ layers ectoderm, endoderm, and mesoderm and, in later stages, into somatic cell types (for reviews see Refs. [2–4]). Functional genomics technologies, such as “Omics,” chip-based high-throughput assays, and gene silencing, so far have contributed to identify key molecules for the pluripotent state. Key transcription factors required for maintaining pluripotency of stem cells include *Pou5f1*, *Nanog*, and *Sox2* and other factors (for reviews see Ref. [2]). These molecules are vital for maintaining the identity of the ICM

during mouse preimplantation development [3,5–7]. Molecules stimulated via the LIF (LIFR, JAK, and STAT3), Wnt (Wnt proteins, Fzd, LRP5/6, and GSK-3 $\beta$ ), and bone morphogenetic protein 4 (BMP4) (SMAD complexes, TGF- $\beta$ , and Id) signaling pathways with the key pluripotency transcription factors control other genes in the core pluripotency network [8] as well. Besides transcriptional regulation, epigenetic modifications, mediation of miRNAs, and activation of cell signaling and cell-to-cell communication do also participate in maintenance of pluripotency. Although a number of genes and signaling transduction networks have been identified to regulate pluripotency, the biological mechanisms behind stemness and pluripotency are still not completely understood. Identifying key molecules involved in pluripotency and differentiation toward germ layers and somatic cell types

<sup>1</sup>Center of Physiology and Pathophysiology, Institute of Neurophysiology, University of Cologne, Cologne, Germany.

<sup>2</sup>Stem Cell Center, Masonic Medical Research Laboratory, Utica, New York.

<sup>3</sup>Institute for Genetics, University of Cologne, Cologne, Germany.

<sup>4</sup>Institute of Computer Science, University of Tartu, Tartu, Estonia.

<sup>5</sup>Quretec Ltd., Tartu, Estonia.

<sup>6</sup>Max-Delbrueck-Center for Molecular Medicine—MDC, Berlin, Germany.

can be useful in understanding differentiation networks controlling embryogenesis *in vivo*. Embryogenesis can be partly recapitulated by the use of cultured ESCs *in vitro* [2,5]. Similar to the totipotent cells of the ICM, gene expression and morphological changes of the cultured ESCs occur hierarchically during differentiation, first developing epiblast cells and then germ layers and finally somatic cells. Hundreds of genes participating in different biological processes, such as differentiation, proliferation, apoptosis, intercellular communication, and cell–matrix interactions, control the differentiation of ESCs to cells to tissue-specific cells (for reviews see Refs. [2,5]). Such biological processes are essential for normal embryonic development *in vivo* (for review see Ref. [9]). *In vitro* models, such as ESCs, combined with appropriate experimental protocols could allow the identification of genes participating in differentiation processes *in vivo* [2,6,7]. Progressive spontaneous differentiation of ESCs in multicellular 3-dimensional embryoid bodies (EBs) that results in the formation of tissue-specific cells can be used to identify those multilineage-decisive genes. Formation of EBs [2,3,6] is a spontaneous anchorage-independent process. However, 3 main problems exist toward identifying prominent biological and signal transduction pathways for pluripotency and differentiation using the ESC model. (i) The ESC model is extremely sensitive to the experimental conditions and even small variations of culture techniques as applied by different investigators will frequently lead to different results. (ii) The onset and shut-down of key biological differentiation processes occur in a narrow time window. Missing of this time window results in failing to detect or correctly interpret key biological processes.

(iii) Comprehensive omics technologies, stringent statistical criteria, and bioinformatics analysis are necessary to get a deep view into the complex biological processes occurring at the different time points of the differentiation.

To overcome these complications, we used a previously developed optimized feeder-free culturing protocol for the CGR8 murine ESC line [2,6,7] and applied the “hanging drop” protocol for obtaining EBs [2,3,6]. To cover all the possible specific differentiation processes occurring in a narrow window time, RNA was isolated each day until day 7 and finally at day 10 (Fig. 1) for microarray hybridization. Stringent statistical criteria were applied to avoid false positive and false negative results.

## Materials and Methods

### ESC culture and EB formation

CGR8 ESCs (ECACC 95011018) were cultured without feeder cells in Glasgow minimum essential medium supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 100 U/mL LIF, and 50  $\mu$ M  $\beta$ -mercaptoethanol (ME) in 0.2% gelatine-coated flasks as previously described [10]. CGR8 ESCs (passage No. 8) were treated with trypsin and used for preparation of cell suspensions (25,000 cells/mL) in Iscove’s modified Dulbecco’s minimal essential medium (IMDM) (Invitrogen) supplemented with 20% FBS, 1% non essential amino acid (NEAA) (vol/vol), 2 mM L-glutamine, and 100  $\mu$ M  $\beta$ -ME. For the “hanging drop” method, 20  $\mu$ L drops of this ESC suspension were placed on the inner surface of the lid of a Petri dish (diameter: 10 cm; Greiner).

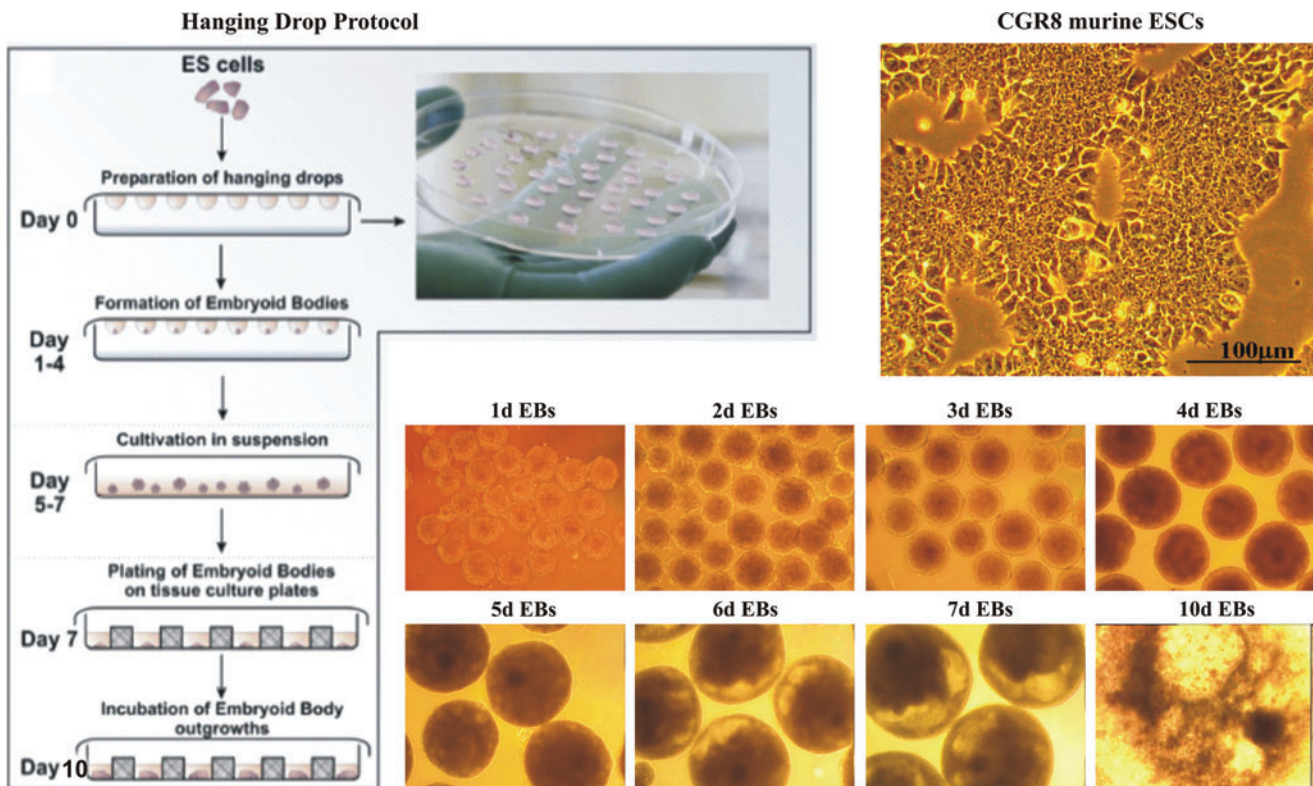


FIG. 1. Hanging drop protocol followed to generate embryoid bodies (EBs) to track the differentiation process from embryonic stem cells (ESCs) to 10-day old EBs. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

The Petri dishes that contained 5 mL phosphate-buffered saline (PBS) were closed with the lid and incubated under normal culture conditions. After days 1–7 and 10, all EBs were separately collected for RNA isolation. RNA isolation from the 10-day-old EBs was performed after the 7-day-old EBs were plated in cultured dishes and incubated in IMDM supplemented with 20% FBS, 100  $\mu$ M  $\beta$ -ME, 2 mM L-glutamine, and 1% NEAA (vol/vol) for 3 more days. Total RNAs from these time points were taken for the transcriptome study.

#### *RNA isolation and whole-genome gene expression profiling of ESCs and EBs of time kinetic study*

The experimental procedures followed for microarray hybridization have been described previously [11]. Microarray data were deposited in the EBI ArrayExpress database: [www.ebi.ac.uk/arrayexpress/experiments/E-TABM-672](http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-672).

#### *Quantitative real-time–polymerase chain reaction analysis*

Validation of the Affymetrix data was performed by quantitative real-time–polymerase chain reaction (qRT-PCR) analysis with ABI 7500 FAST Detection System (Applied Biosystems). One microgram of RNA from ES (day 0) and 4-, 7-, and 10-day-old EBs was reverse transcribed using SuperScript VILO cDNA synthesis kit (Invitrogen GmbH). qRT-PCR was performed in triplicates for every sample using SYBR-Green PCR Master Mix (QuantiFast qPCR Master Mix; Qiagen). Validation has been performed for *Pou5f1* (NM\_013633), *Nanog* (NM\_028016), *Mageb16* (NM\_001113734), *Ets* variant gene 5 (*Etv5*) (NM\_023794), ribonuclease P25 subunit (*Rpp25*) (NM\_133982), and  $\beta$ -*Actin* (NM\_007393) (for primer sequences see Supplementary Table S1; Supplementary Data are available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)). These genes were chosen in terms of their significance in relationship to pluripotency. All reactions were performed with the following conditions, beginning with Taq activation at 95°C for 5 min, thereafter 40 cycles of 10 s denaturation at 95°C, 30 s of combined annealing and extension at 60°C, and finally ending with a melting curve acquisition. The relative quantities of the target genes were normalized against relative quantities of  $\beta$ -actin keeping undifferentiated ESCs (day 0) as a reference sample. Fold-expression changes were calculated by raising to the power of the negative value of delta-delta Ct value ( $2^{-\Delta\Delta Ct}$ ) and the resulting values were plotted as log<sub>2</sub> values setting the expression value of the reference sample to 0.

#### *siRNA knockdown*

siRNAs for *Mageb16* [Mm\_ *Mageb16* (2410003)J06Rik\_5): TAGGGAGAAGGGAGCGTCTTA; Mm\_ *Mageb16* (2410003)J06Rik\_6): TACGGGCTAAAGCTGAACTA] were purchased from Qiagen and used at a concentration of 5 nM. Lipofectamine 2000 (Invitrogen) was used as siRNA transfection reagent as instructed by the manufacturer. CGR8 cells ( $7 \times 10^4$ ) seeded in 24-well plates with normal growth medium (containing LIF) were transfected the following day with siRNA-Lipofectamine complex. Negative transfection control labeled with Alexa Flour 488 provided by Qiagen (to monitor transfection efficiency with green fluorescence) was used as scrambled control; mock control (transfection agent

only) also was included. After 6 h of transfection, normal growth medium was added. Cells were harvested after 48 h for RNA and protein isolation. Knockdown efficiency was measured with qRT-PCR at mRNA expression level and with western blotting at protein expression level. Expression of the pluripotency markers *Pou5f1*, *Nanog*, and *Sox2* and early mesodermal, ectodermal, and endodermal markers *T* (*Brachyury*), *FGF5*, and *HNF4 $\alpha$* , respectively, was determined in untransfected ESCs and scrambled and *Mageb16*-knockdown samples using qRT-PCR. Nucleotide sequences of these primers are provided in Supplementary Table S1.

#### *Western blotting*

Separate knockdown experiment was set with *Mageb16* siRNA along with negative transfection control (scrambled) and untransfected ESC control following the same protocol mentioned previously in siRNA knockdown method. Cells ( $4 \times 10^5$ ) were seeded in 25 cm<sup>2</sup> flask and the same 5 nM siRNA concentration was used. Cells were collected after trypsinization and washed with cold PBS and lysed in 150 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl (pH 7.4), 1 mM phenylmethanesulfonyl fluoride, 1% Triton X-100, and 1% protease inhibitor cocktail (Sigma) at 4°C. Protein samples were quantified with the Bradford reagent (Sigma). From each sample 45  $\mu$ g of protein was separated using 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis gels and transferred to polyvinylidene fluoride membrane. Membranes were blocked with 5% bovine serum albumin at room temperature. Membranes were incubated with *Mageb16* polyclonal antibodies (Abgent Europe Ltd.) for overnight at 4°C on orbital shaking. After washing with PBST solution (PBS with 0.1% Tween-20), membranes were incubated with secondary antibodies and labeled with horseradish peroxidase for 1 h at room temperature. *Mageb16* was detected using enhanced chemiluminescence Western detection system (Fisher Scientific GmbH).

#### *Identifying significantly regulated transcripts related to pluripotency and self-renewal*

Array data were quantile normalized using the RMA implementation of the R Affy package [12,13]. Probe sets marginally expressed in all conditions were rejected using the MAS5 present call ( $P < 0.05$ ) [14]. Differential expression of the remaining 23,142 of 45,101 has been determined by the linear model implementation of the R Limma package [15] followed by a Benjamini Hochberg multiple testing correction (1% FDR). To determine specifically regulated pluripotency-associated transcripts, the expression levels of undifferentiated ESCs (day 0) were pairwise compared with the levels of differentiated 3- to 7-day-old EBs since many pluripotency markers, such as *Pou5f1*, do not show significantly varying expression up to day 3. Of particular interest were transcripts that show high expression in the undifferentiated state and low expression levels throughout differentiation. Size of change was stated with threshold value of fold change (FC) 2 in absolute scale.

To have a highly consistent list of significantly regulated transcripts, intersection of significant regulated genes from the CGR8 time course and from the transgenic experiment cluster of CGR8 was implemented. To obtain a significant list

of genes of the transgenic experiment cluster of CGR8, following transgenic cell lines of CGR8, such as  $\alpha$ -myosin heavy chain ( $\alpha$ -MHC) ESCs [9], BMP2 ESCs [16], T (Brachyury) ESCs [17],  $\alpha$ -smooth muscle actin 2 (Acta2) ESCs [18], and CD31 ESCs [19]), were taken for pairwise comparisons against their corresponding mixed (multilineage) and purified populations. Clear description of the strategy followed to obtain the significant list of genes in the transgenic experiment cluster of CGR8 mentioned previously is outlined in Fig. 2.

From the list obtained from the intersection study performed, only significant transcription factors were chosen to prove their homogeneous expression pattern both in undifferentiated and differentiated states other than in CGR8 cell lines. For that a time series of the mouse ESC lines, R1, J1, and V6.5, dataset of an 11-point time course (Embryonic Stem Cell and Embryoid Body Time course) study [20] made public in Gene Expression Omnibus, was chosen.

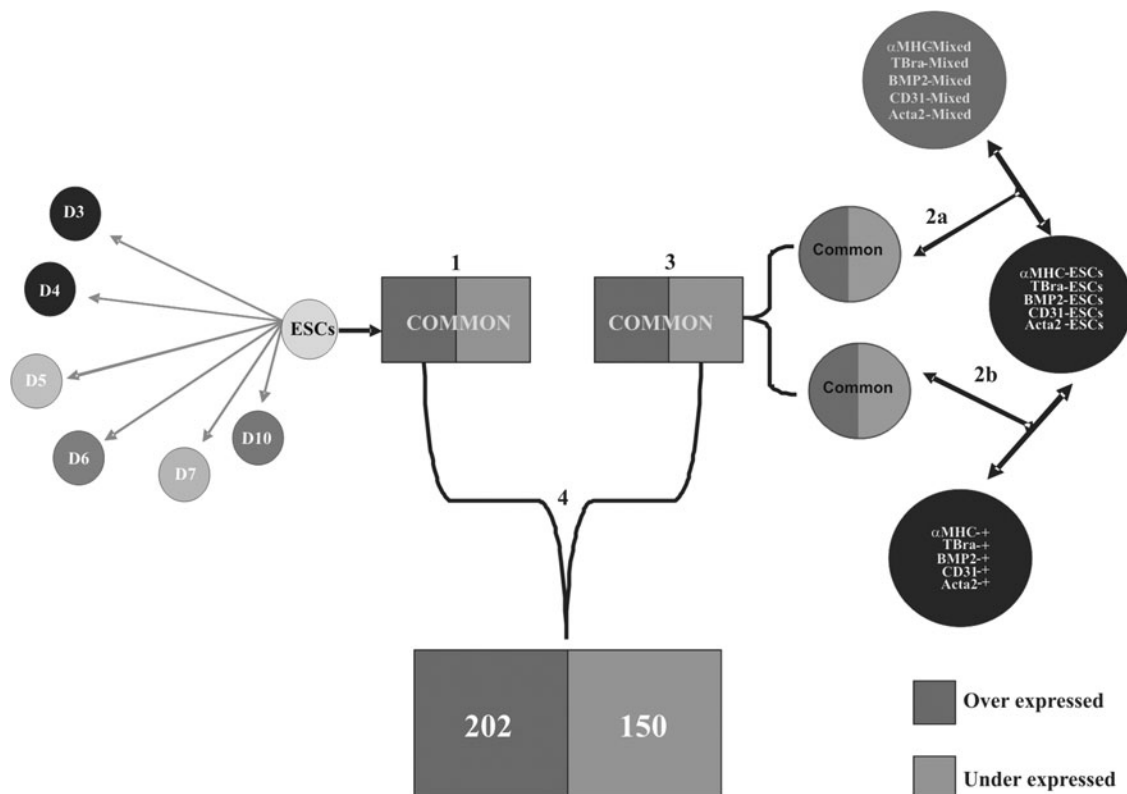
### Identifying significantly regulated transcripts in differentiating ESCs

To avoid biased interpretation of the large number of differentially expressed genes during differentiation of ESCs and to identify specific gene signature pathways, we used

the gene ontology–analysis of variance (GO-ANOVA) application of the Partek Genomic Suite version 6.4 for our analysis. For each annotation the application sets up an ANOVA with 3 factors: array, transcript, and as an additional parameter the time course to identify genes that follow a particular expression pattern. GO annotations with an FDR < 0.1% along the time course ( $n=5,772$  of 6,455) were selected as differentially expressed. The selection was additionally restricted to level 5 GO annotations ( $n=1,453$ ) to prevent overflowing overlap caused by the (multi-) hierarchical structure of GO.  $P$  values of consecutive points in time were  $-\log_{10}$  transformed and afterward normalized to a mean of zero and a standard deviation of one. Estimation of the number of K-means clusters using the Davies-Bouldin procedure leads to  $K=7$  groups.

### Principal component analysis

Principal component analysis (PCA) is performed using Partek Genomics Suite (Partek, Inc.) to visualize data variability in a multidimensional picture. Distance between the samples reflects the degree of dissimilarity between them. PCA mapping of the time kinetic data shows 54.9% of data variance in the first 3 principal components (PC1 33.6%, PC2 13.9%, and PC3 7.39%).



**FIG. 2.** Methodology applied to identify differentially expressed transcripts pertinent to pluripotency. Step 1: Obtained list of differentially expressed transcripts is commonly found in all pairwise statistical comparisons between ESCs (day 0) and day 3–7 and 10 EBs. Step 2: (a) Identified differentially expressed transcripts common to pairwise statistical comparisons between transgenic undifferentiated ESCs and their generated mixed populations. (b) Identified differentially expressed transcripts common to pairwise statistical comparisons between transgenic CGR8 ESC population and their generated pure populations. Step 3: Intersection-resulted list of lists from Step 2a and b. Step 4: Intersection-resulted list of lists from Step 1 and 3, and from these, 202 upregulated transcripts and 150 downregulated transcripts are identified after redundancy removal.

**Results and Discussion**

*Visualization of transcriptome alteration at differentiation process*

Early EB differentiation is characterized by fundamental changes of the transcriptome. PCA of the differentially expressed genes resulted in 3 characteristic principal components (PCs) that represent 54.9% of the dataset variance (Fig. 3A, B). The 100 most correlated probe sets in component loading of each PC [absolute ( $r$ ) > 0.8] have been visualized in a heatmap after hierarchical clustering to exemplify the characteristic of the PCs. PC1, PC2, and PC3 represent single, transient, and double-transient switches of the expression status, respectively (Fig. 3C). The most prominent switch time point in PC1 (33.6% of the dataset variance) is found between days 3 and 4; up to day 3 there is no significant change. Among the pluripotency-related core factors, *Pou5f1* has its constant expression up to day 3 and even some of the other pluripotency-related regulatory genes stay without significant changes in their expression profile.

*Significantly regulated transcripts pertinent to pluripotency*

Intersection of significant regulated genes in the CGR8 time course and CGR8 transgenic experiment cluster resulted in 202 upregulated transcripts and 150 downregulated transcripts after redundancy avoidance (Supplementary Tables S2 and S3).

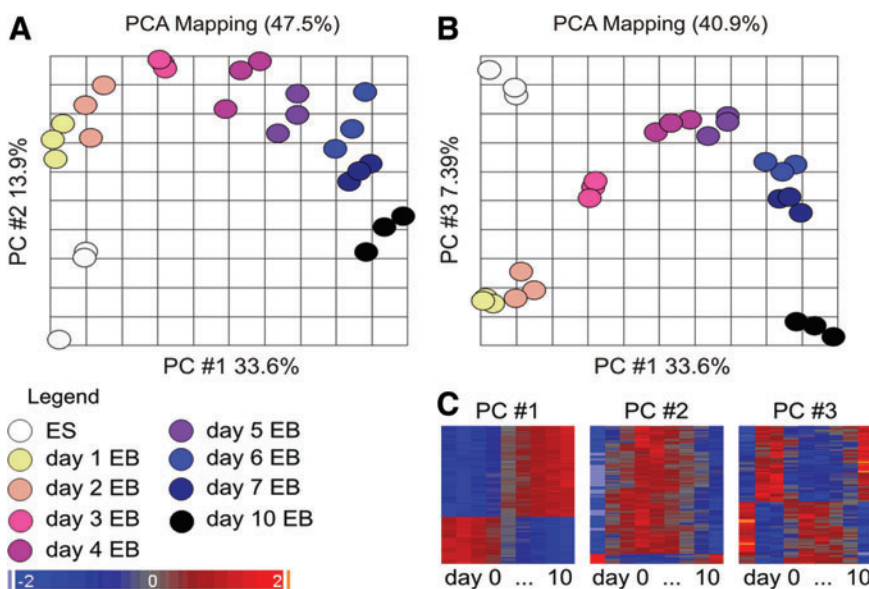
Among 202 high-expressed transcripts, there are 60 transcription factors (Supplementary Table S4) identified and so far reported pluripotency-associated markers are all present in these 60 genes. Among low-expressed genes, there are 26 transcription factors (Supplementary Table S5). Expression pattern wise, clear separation between high- and low-expressed transcription factors was observed (Fig. 4) in a time series of the mouse ESC lines R1, J1, and V6.5, when the pattern of expression of 54 significant transcription factors from the just outlined lists was visualized. Therefore, irrespective of different murine ESC lines, these identified

common transcription factors seem to have similar regulatory mechanism at pluripotency network.

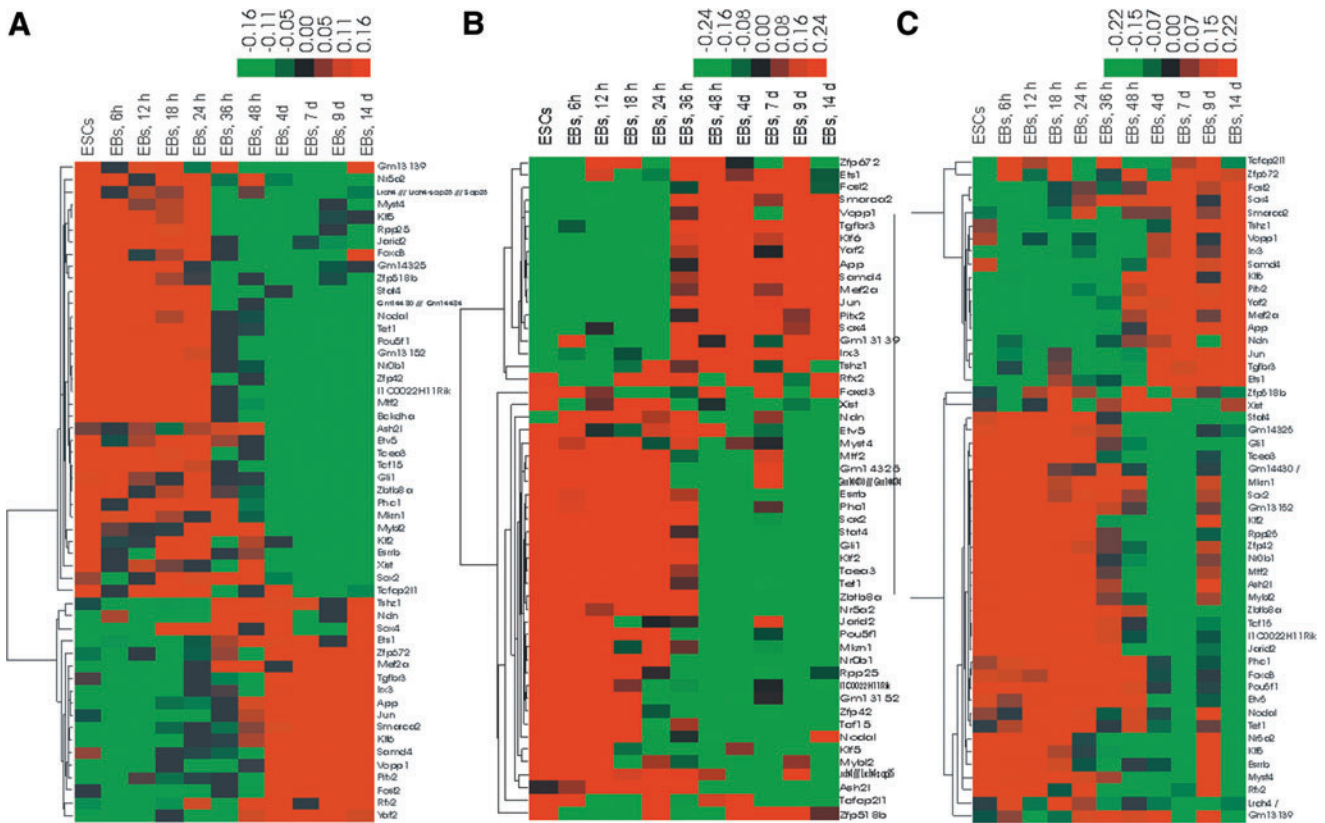
*Transcription factors prominently expressed in the pluripotent state*

Besides the well-known pluripotency-associated transcription factors, such as *Pou5f1*, *Nanog*, *Sox2*, Kruppel-like factor 4 (*Klf4*), other Klf factors—such as *Klf2*, *Klf5*, and *Utf1*—*Esrrb*, *Tcfcp2l1*, *Sall4*, *Dax1* (*Nr0b1*), *Stat4*, and *Zfp42*, few more regulatory genes (Table 1) were identified based on significant changes (downexpressed) of their expression level after the induction of differentiation. From this, only few transcription factors, which are neither discussed well with its pluripotency association nor taken for related genetic manipulation study, are reported in the following paragraph.

Among them, Polyhomeotic-like 1 (*Phc1*) is a core component of polycomb repressive complex 1 (*PRC1*), a polycomb group protein complex that is required for maintaining ESC pluripotency and plasticity during embryonic development [21]. *Jardi2* (jumonji, AT-rich interactive domain 2), a factor that negatively regulates cell proliferation, is a regulatory subunit of *PRC2*. It regulates Prc2-mediated H3k27 methylation and eventually modulates ESC pluripotency and somatic cell reprogramming due to methylation along with other subunits of *Prc2*, *MTF2*, and *esPRC2p48* [22]. *Rbpj* (recombination signal binding protein for immunoglobulin kappa J region) is involved in NOTCH signaling [23] and has been predicted to have an important role in maintaining pluripotency [24]. Forkhead box D3 (*Foxd3*) is essential for the development of endoderm, is expressed in ESCs, and is associated with early lineage-specific decisions with a coregulatory factor *Pou5f1* [25]. More recently, it has been shown that tet oncogene 1 (*Tet1*) along with *Tet2* is associated with the pluripotent state [26]. Regulatory factor X2 (*Rfx2*) that influences HLA class II expression is highly expressed in testis and is mentioned as *Pou5f1*-associated protein [27]. Also, it has been shown that depletion of developmental pluripotency associated 2 (*Dppa2*) initiated



**FIG. 3.** Principal component analysis (PCA) mapping of PCs 1 and 2 (A) and PCs 1 and 3 (B). PCA has been done using correlation as the dispersion matrix and normalized eigenvector scaling. The 100 best correlated probe sets to each PC have been heatmap visualized (C) after hierarchical average linkage clustering using the Euclidean distance function. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)



**FIG. 4.** Expression pattern of significant transcription factors during progressive differentiation of the ESCs (ESCs to 10-day old EBs). **(A)** R1 Embryonic Stem Cell and Embryoid Body Time Course. **(B)** J1 ESC and EB Time Course. **(C)** V6.5 Embryonic Stem Cell and Embryoid Body Time Course. An 11-point time course study is included as time course: 0 h, 6 h, 12 h, 18 h, 24 h, 36 h, 48 h, 4 days, 7 days, 9 days, and 14 days. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

differentiation and in parallel repressed proliferation in murine ESCs [28]. DNA methyltransferase 3-like (*Dnmt3l*) is required for DNA methylation in stem cells and is not expressed in differentiating cells [29]. Monocytic leukemia zinc-finger protein (MOZ) along with MORF (*MYST4*) safeguards identity of stem cell through maintaining epigenetic integrity [30]. In ESCs, Moz expression is regulated by the cascade of *Nanog*, *Pou5f1*, and *Sox2* [31]. Zinc finger and BTB domain containing 8a (*Zbtb8a*) is predicted to be one of the transcription factors having at least one preimplantation expressed sequence tag (EST) among many ESTs used to study the 7 stages of development from the unfertilized oocyte to the blastocyst stage [32]. Knockout studies of *Etv5* have shown its involvement in the maintenance of spermatogonial stem cells by regulating chemokine production from Sertoli cells [33]. *Etv5* is involved in the regulation of branching involved in mammary gland duct morphogenesis [34]. Transcription factor 15 (*Tcf15*) also known as Paraxis is involved in paraxial mesodermal development and regulates somite morphogenesis [35]. Tudor domain containing 12 (*Tdrd12*) also known as ECAT8 has been identified as one of the candidates of the LIF/Stat3-independent factor essential for pluripotent cells [36]. Makorin-ring finger protein (*Mkrfn1*) is a putative downstream gene of *Pou5f1* and is expected to play significant role at the maintenance of pluripotency [37]. Gli family zinc finger 1 (*Gli1*), a Hedgehog transcription factor, is involved in the positive regulation of cell proliferation, osteoblast differentiation, and in epidermal

cell differentiation [38]. *Rpp25* is involved in transcription and processing of tRNA and it also binds to H1 RNA in vitro [39]. DNA segment, Chr 1, Pasteur Institute 1 (*D1Pas1*) is expressed in germ cells of the testis and its expression is developmentally regulated [40]. However, its expression is much higher in ESCs compared with differentiated EBs. ATP-binding cassette, sub family B, member 1A (*Abcb1a*) along with other ABC transporters, such as *Abcg2*, *Abcb1b* and *Abcc1* is expressed both in early germ cells and latter in testis [41]. The mouse-specific Zinc finger proteins *zfp936*, *1700029I01Rik*, *zfp229*, and *zfp819* also are significantly overexpressed in the undifferentiated state. However, so far nothing is known about a potential functional role of these zinc finger proteins for the pluripotency of the ESCs.

#### Highly expressed new candidate genes are critical to the pluripotency network

The following group contains the highly upregulated genes (Table 2) that have not yet been associated with stemness or pluripotency properties. *Mageb16*, melanoma antigen family B, that still remains functionally non-annotated shows higher level of expression in the undifferentiated state. *Morcl* is located in the nucleus and is reported to be involved in spermatogenesis [42] and in "regulation of early meiotic prophase events involving chromosome synapsis or recombination" and even may perturb generation of male germ cells through meiosis [43]. Its role in cell cycle

TABLE 1. REGULATORY GENES (TRANSCRIPTION FACTORS) IDENTIFIED TO BE SIGNIFICANTLY UPREGULATED IN THE PLURIPOTENT STATE OF EMBRYONIC STEM CELLS

Gene symbol	FC (day 3)	FC (day 4)	FC (day 5)	FC (day 6)	FC (day 7)	FC (day 10)
<i>1700029I01Rik</i>	-9.45	-11.31	-8.75	-10.56	-12.21	-13.18
<i>2610305D13Rik</i>	-15.24	-29.45	-23.92	-21.56	-29.04	-9.65
<i>Abcb1a</i>	-38.44	-41.92	-36.75	-35.05	-35.98	-37.63
<i>Dnmt3l</i>	-27.73	-45.25	-49.29	-47.29	-46.42	-38.68
<i>D1Pas1</i>	-7.98	-8.72	-9.88	-11.49	-12.20	-12.34
<i>Dppa2</i>	-27.28	-51.27	-54.57	-50.56	-49.87	-37.27
<i>Enpp3</i>	-9.58	-12.55	-10.78	-7.36	-9.06	-10.34
<i>Etv5</i>	-2.07	-9.78	-7.52	-10.78	-10.70	-14.22
<i>Foxd3</i>	-2.39	-2.81	-2.73	-2.73	-2.97	-2.60
<i>Gli1</i>	-11.79	-12.30	-6.87	-8.63	-11.16	-12.64
<i>Irak3</i>	-3.33	-3.47	-3.55	-3.97	-2.60	-3.29
<i>Jarid2</i>	-2.91	-4.17	-5.62	-5.39	-4.69	-4.82
<i>Mkrr1</i>	-4.50	-6.11	-7.67	-6.36	-6.77	-4.14
<i>Mybl2</i>	-9.00	-9.19	-13.27	-16.11	-21.41	-37.01
<i>Myst4</i>	-7.06	-6.02	-4.03	-3.20	-3.34	-3.46
<i>Nkx6-3</i>	-4.17	-4.38	-4.47	-4.76	-4.29	-4.44
<i>Nr5a2</i>	-12.21	-10.93	-9.85	-6.15	-9.92	-10.06
<i>Phc1</i>	-2.83	-4.41	-5.43	-5.13	-5.74	-7.16
<i>Rbpj</i>	-3.61	-3.84	-4.00	-3.18	-2.41	-2.10
<i>Rfx2</i>	-3.63	-4.89	-3.41	-3.51	-4.14	-5.78
<i>Rpp25</i>	-14.42	-13.27	-16.34	-10.70	-10.34	-5.43
<i>Tcea3</i>	-2.60	-4.92	-4.56	-6.36	-6.77	-4.14
<i>Tcf15</i>	-2.58	-3.16	-3.23	-3.20	-2.93	-3.23
<i>Tdrd12</i>	-6.59	-9.19	-11.71	-11.88	-11.16	-12.64
<i>Zbtb8a</i>	-5.10	-7.67	-7.94	-7.78	-7.41	-5.98
<i>Zfp229</i>	-11.39	-6.96	-5.86	-6.15	-7.46	-11.39
<i>Zfp819</i>	-8.17	-10.20	-9.65	-10.20	-10.63	-9.99

FC values in absolute scale are provided in comparison with the time point day 0.

FC, fold change; Foxd3, Forkhead box D3; Rfx2, regulatory factor X2; Dppa2, developmental pluripotency associated 2; Dnmt3l, DNA methyltransferase 3-like; Zbtb8a, zinc finger and BTB domain containing 8a; Etv5, Ets variant gene 5; Tcf15, transcription factor 15; Tdrd12, tudor domain containing 12; Mkrr1, makorin-ring finger protein; Gli1, Gli family zinc finger 1; Rpp25, ribonuclease P25 subunit; D1Pas1, DNA segment, Chr 1, Pasteur Institute 1; Abcb1a, ATP-binding cassette, sub family B, member 1A.

regulation and proliferation of NT2 embryonal carcinoma stem cells has also been studied. *Cth*, also known as cystathionine gamma-lyase (CSE), is involved in the negative regulation of cell growth and proliferation. CSE-knockout mice show increased proliferation rates of smooth muscle cells [44]. Solute carrier family 7 (*Slc7a3*) is found in the plasma membrane and is involved in amino acid transportation and the regulation of Target of rapamycin signaling [45]. Heat shock transcription factor 2 binding protein (*Hsf2bp*) is expressed in human testis and its involvement in modulation of HSF2 activation in testis has been proposed [46]. Gametocyte specific factor 1-like gene (*Gtsf1l*), mapped at chromosome 2, is expressed at testis and is essential for spermatogenesis beyond the early meiotic phase [47]. Lipase member H (*Liph*), mapped at chromosome 3, involves in lipid catabolic process and is found in membrane. Cordon-bleu (*Cobl*) is expressed in the axial structure of the gastrulating mouse embryo [48] and is involved in the biological process neural tube closure. Proteasome subunit alpha type 8 (*Psm8*) is involved in the positive regulation of ubiquitin-protein ligase activity that regulates mitotic cell cycle. Its high level of expression in testis and in murine ESCs is recorded at BioGPS gene annotation portal [49]. Oocyte expressed protein homolog (*Ooep*) (dog) is involved in in utero embryonic development and is found in the cytosol. Its expression is higher in oocytes and has been reported to be

associated with *Ces5*, significantly regulated gene at murine ESCs [50]. Leucine-rich repeat containing 34 (*Lrrc34*) is highly expressed in murine ESCs and in testis. Pycard, containing PYD and CARD domain, is involved in signal transduction and regulation of the apoptosis pathway [51].

#### Correlation between mRNA detected by microarrays and qRT-PCR

Verification of the homogeneity of signal expression levels of few notable genes identified from microarray data analysis prominent at pluripotent state, such as *Mageb16* (NM\_001113734), *Etv5* (NM\_023794), and *RPP25* (NM\_133982) along with known markers *Pou5f1* (NM\_013633) and *Nanog* (NM\_028016), was performed with qRT-PCR. As indicated, in Fig. 5A and B, results of the Affymetrix analyses clearly correspond to the results obtained from the qRT-PCR. Therefore, the list of transcripts identified is reliable. As a proof of concept, *Mageb16*, prominently expressed at pluripotent state, was chosen for knockdown study to observe the expression of ESCs and germ layers' marker genes.

#### *Mageb16* knockdown activates differentiation even in the presence of LIF

From our study, it was identified that *Mageb16* mRNA expression decreases drastically with the initiation of

TABLE 2. TRANSCRIPTS HIGHLY UPREGULATED AT PLURIPOTENT STATE

Gene symbol	FC (day 3)	FC (day 4)	FC (day 5)	FC (day 6)	FC (day 7)	FC (day 10)
<i>Mageb16</i>	-24.19	-32.45	-32.60	-34.33	-33.75	-31.77
<i>Morc1</i>	-35.71	-43.38	-42.23	-43.68	-43.60	-43.50
<i>Cth</i>	-6.80	-26.81	-18.89	-21.59	-24.95	-28.03
<i>Slc7a3</i>	-10.91	-15.82	-15.94	-17.07	-18.21	-18.77
<i>Hsf2bp</i>	-19.84	-22.05	-23.00	-26.87	-24.42	-22.57
<i>Gtsf1l</i>	-16.37	-21.80	-20.28	-19.20	-18.74	-20.36
<i>Liph</i>	-16.53	-18.50	-21.43	-20.52	-17.07	-17.58
<i>Cobl</i>	-14.47	-17.78	-20.24	-21.07	-17.45	-17.12
<i>Psm8</i>	-11.52	-12.14	-12.60	-13.77	-12.98	-13.14
<i>Ooep</i>	-9.41	-10.94	-13.40	-10.99	-9.80	-8.64
<i>Lrrc34</i>	-5.84	-9.07	-8.40	-8.41	-9.61	-17.00
<i>Pycard</i>	-9.96	-14.41	-17.96	-17.14	-14.11	-12.41

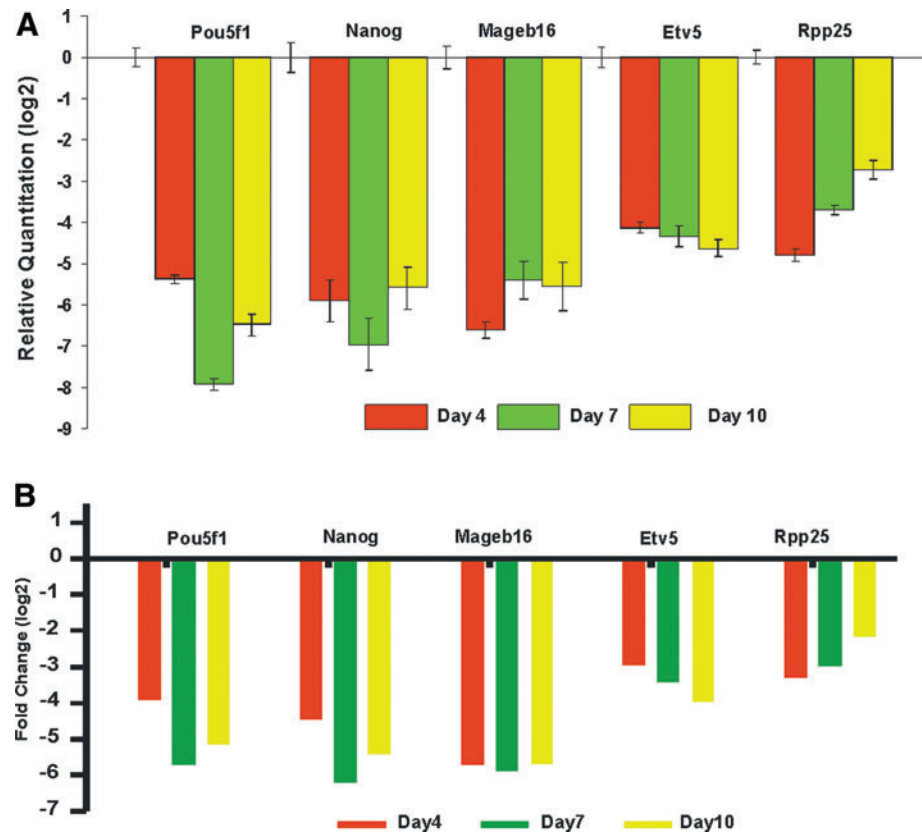
FC values in absolute scale are provided in comparison with the time point day 0.

*Slc7a3*, solute carrier family 7; *Hsf2bp*, heat shock transcription factor 2 binding protein; *Gtsf1l*, gametocyte specific factor 1-like gene; *Liph*, lipase member H; *Cobl*, Cordon-bleu; *Psm8*, proteasome subunit alpha type 8; *Ooep*, oocyte expressed protein homolog; *Lrrc34*, leucine-rich repeat containing 34.

differentiation and its expression level is very low in somatic cells derived from ESCs. Moreover, it was also reported to be one of the significant gene among 782 upregulated genes in ESCs determined by a combined study of RNA seq profile and microarray profiles [52]. Therefore, it was selected for siRNA knockdown to explore a possible relation to pluripotency. Approximately 70% of knockdown efficiency was achieved 48 h after transfection. Compared with untransfected (ESCs), scrambled, and mock control samples, only 30% of *Mageb16* mRNA expression was observed in siRNA-*Mageb16* targeted ESCs (Fig. 6A) and the same knockdown efficiency

was confirmed at protein expression level with western blotting (Fig. 6B). *Mageb16* siRNA-treated cells did not show significant expression changes of pluripotency markers, such as *Pou5f1*, *Nanog*, and *Sox2*. Interestingly, almost 2-fold overexpression of early meso-, ecto-, and endodermal lineage markers (*T Brachyury*, *FGF5*, and *HNF $\alpha$* ) was observed in the knockdown ESCs (Fig. 6C). These findings clearly show that *Mageb16* may also be essential for pluripotency via pathways that are independent of classic pluripotent genes, such as *Pou5f1*, *Nanog*, and *Sox2*. Even in the presence of LIF, under pluripotent conditions, knockdown of *Mageb16* resulted in an

FIG. 5. mRNA expression of *Pou5f1*, *Nanog*, *Mageb16*, Ets variant gene 5 (*Etv5*), and ribonuclease P25 subunit (*Rpp25*) derived from quantitative real-time-polymerase chain reaction (qRT-PCR) (A) and derived from microarrays (B). Relative quantitation is done compared with day 0 in both cases. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)





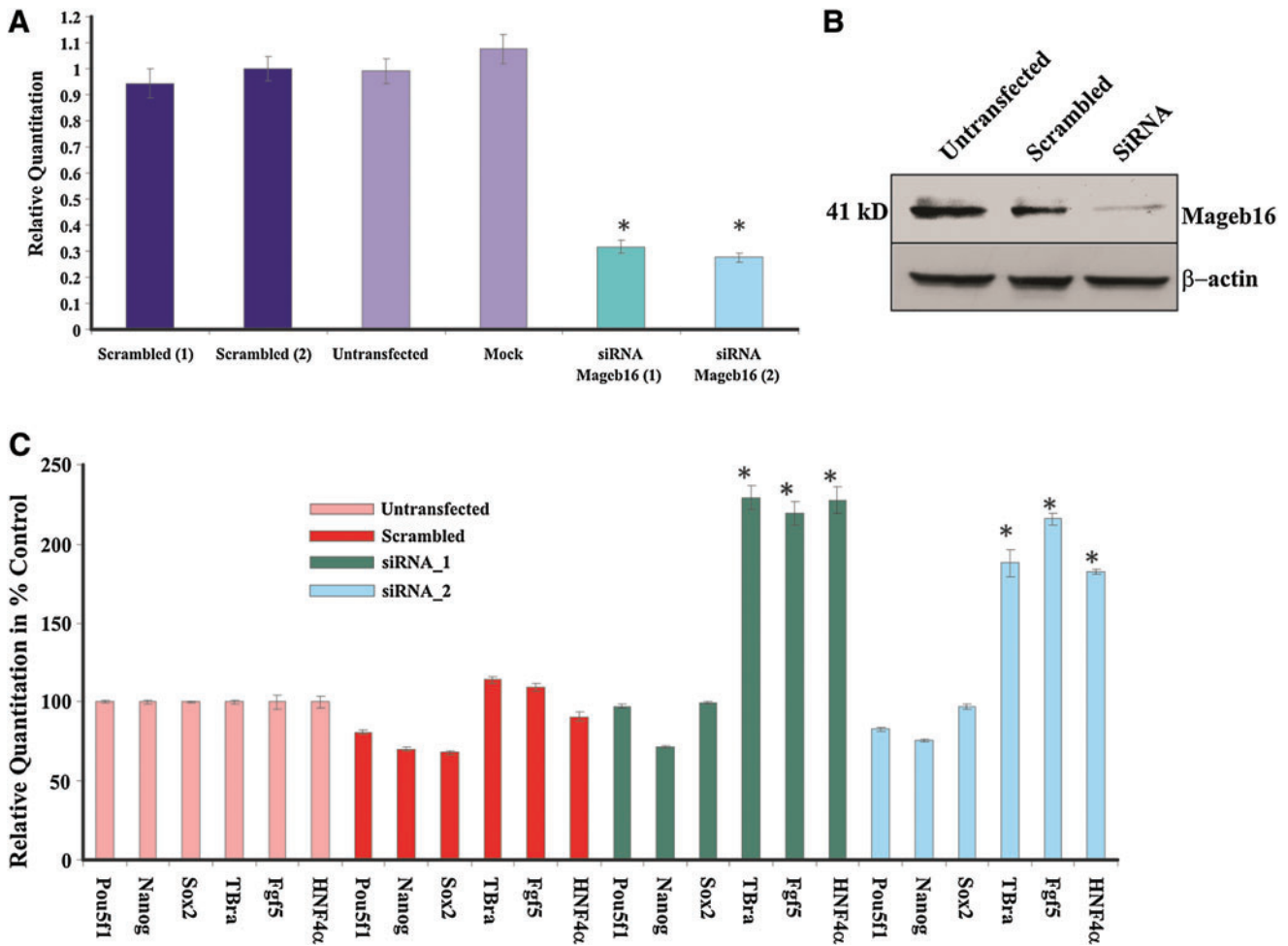


FIG. 6. Expression of specific pluripotent and germ layer gene markers in *Mageb16* knockdown ESCs mediated by 2 different siRNAs. (A) Knockdown efficiency of 2 different *Mageb16* siRNAs through transient transfection of undifferentiated ESCs determined by qRT-PCR [ $P < 0.05$  for siRNA (1) and (2) vs. scrambled (1), (2)], mock, and untransfected ESCs). (B) Western blot analysis showing protein expression level at untransfected, scrambled (1), and siRNA (1) samples. Protein levels were normalized to  $\beta$ -actin level. Compared with scrambled, the drastic reduction of MAGEB16 protein (Mol. weight=41 kDa) is noted in siRNA-mediated knockdown of *Mageb16*. (C) qRT-PCR expression levels expressed as fold changes of the pluripotent gene markers, *Pou5f1*, *Nanog*, and *Sox2*, of the germ layer markers *TBra* (mesodermal), *Fgf5* (ectodermal), and *HNF $\alpha$*  (endodermal). The expression values in untransfected ESCs were set as 100%. The expression and relative expression of pluripotency and early lineage markers were measured at scrambled as negative control and at 2 different *Mageb16* siRNAs (1 and 2).  $P < 0.05$  for *Tbra*, *Fgf5*, and *HNF4* siRNA (1) and (2) treated versus scrambled and untreated ESCs. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

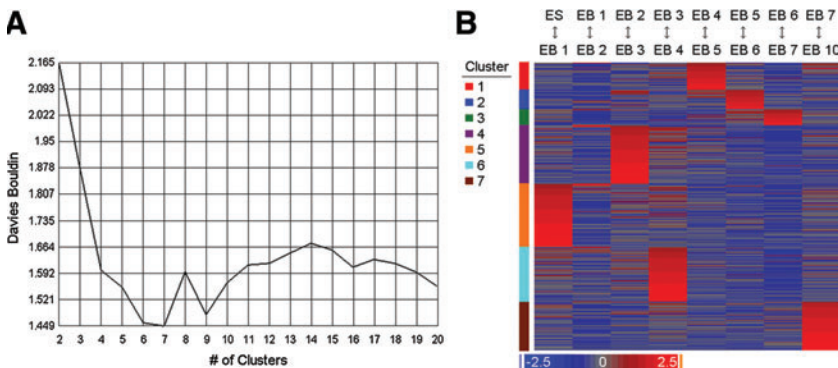


FIG. 7. K-means clustering of 1354 significant differentially expressed level 5 gene ontology annotations using the Euclidean distance function. (A) The Davies Bouldin K estimation shows an absolute minimum at  $K=7$ . (B) Clustering leads to clusters with differential expression between 1 specific consecutive time pair. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

upregulation of key genes required for the formation of the germ-layer-specific genes. Moreover, as shown in Fig. 5A, the expression pattern of *Mageb16* is similar to that of *Pou5f1* and *Nanog*. Therefore, the promoting actions of *Mageb16* toward pluripotency cannot be excluded. The relationship between *Mageb16* and the key pluripotency marker genes has to be elucidated. The subcellular localization of this protein is predicted by the Euk-mPloc 2.0 package [53] to be either in the nucleus or in the cytoplasm.

#### Identification of key differentiation gene clusters using K-means clustering pertinent to differentiation process

The K-means clustering (Fig. 7) using the Euclidean distance function resulted in straightforward and individual patterns of regulation between 7 of the 8 pairwise comparisons (Supplementary Table S6). The genes identified in these 7 clusters are all discussed here in terms of them enriching GO during the differentiation process. As shown in Supplementary Fig. S1, GO and probe set interconnections on this stage as given for cluster 1 (EB 4 to EB 5 differences) most probe sets are related to 1 GO annotation. Other probe sets are in central positions in the probes set to GO network interconnection. Probe sets have been selected as interconnectors if they (i) interconnect 4 or more GO annotations, (ii) are itself differentially expressed (FDR < 0.1%) with an LSD *P* value minimum at the time frame specified by the GO annotation cluster (LSD  $P < 10^{-3}$ ), and (iii) with significantly more interconnections in the specified cluster than in all other clusters together (Fisher's exact test FDR < 5%). Consequently, the filter cannot map the entire picture of CGR8 differentiation but the selected interconnectors are interesting due to their extraordinary position and fulfill the statistical criteria to be candidate key players in differentiation.

#### No prominent multilineage-related differentiation pathways noted in cluster 5 at ESCs versus day 1 EBs comparison

As shown in Supplementary Fig. S2 (part 1), 7 genes have been identified participating in 317 GO annotations in total. However, as indicated in the ArrayExpress Archive ([www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress), ID: E-TABM-672), size of change in expression (FC) between days 0 and 1 was relatively low except for 7 genes: *Prdx2*, *Nnat*, *Tia1*, and *Psmc9* (2- to 4-fold upregulated); *Stat5a*, *RenbP*, and *Hexb* (4- to 9-fold downregulated). From these results no significant conclusions can be made in terms of prominent differentiation pathways.

#### Early developmental processes are defined by the genes in cluster 4 (day 2 vs. day 3 EB comparison)

Interestingly, we identified in cluster 4 (day 2 vs. day 3 EBs) (Supplementary Fig. S2, part 1) 15 genes that are highly upregulated and interconnecting to 295 GOs participating mainly in early developmental processes such as the formation of the germ layers as well as on the development of derivative somatic cells. Among the genes are the key transcription factors *T* (Brachyury), *FoxC1*, *FoxC2*, *FoxA2*, *HoxB1*,

*LHX1*, and *Gata4*. Clearly, in day 3 EBs, transcripts participating in the early differentiation processes of all 3 germ layers are predominantly expressed. The *FOXC1*, *FOXC2*, and *FOXA2* genes belong to the forkhead family of transcription factors. *FOXC1* and *FOXC2* have been implicated in the regulation of many developmental processes, including foregut morphogenesis, proper lens formation, and hair follicle development (reviewed by Refs. [54–56]). *FOXC1* and *FOXC2* are expressed in the paraxial mesoderm and somites [57–59]. Knockout homozygous mutant mice for *Foxc1*<sup>-/-</sup>/*Foxc2*<sup>-/-</sup> lack expression of segmentation genes that are normally transcribed in anterior and posterior somites and somites fail to express the myogenic marker *MyoD* [60]. In summary several studies indicate that expression of *Foxc1* and *Foxc2* is crucial for the establishment of paraxial mesoderm, leading to skeletal myogenesis in the developing embryo [54].

*Foxa2* regulates the specification, neurogenesis, and differentiation of dopamine neurons in the midbrain floor plate (for review see Ref. [61]). Further, it has been demonstrated that *Foxa2* is also involved in the liver development in mouse [62]. *Foxa2* and *Gata4*, highly expressed, act also as markers for definitive endoderm [63].

*Lhx1* is expressed during embryogenesis in the mouse early in mesodermal tissue, and then later during urogenital kidney, liver, and nervous system development. Also, *Lhx1* is required for regulating the vertebrate head development and the nervous system, as well as the female reproductive tract development (for review see Ref. [64]). The prototype T-box transcription factor *T* (brachyury) is a canonical marker for early mesoderm [65]. *HoxB1* belongs to the homeobox family of genes encoding for a highly conserved family of transcription factors and are involved in developmental processes in all multicellular organisms. Altered segmental identity and abnormal migration of motor neurons in mice lack *Hoxb*. Lack of *Hoxbi* in *Hoxb1*-null mice demonstrated an important role for hindbrain patterning and development [66,67]. *Hoxb1* expression is not limited to the developing hindbrain but occurs also in the paraxial mesoderm and cervical somites [67] as well as in the development of the vertebrate skeleton [68]. *Hoxb1* controls cell fate specification and proliferative capacity of neural stem and progenitor cells in ESCs [69]. Taking into consideration all these findings it is obvious that the differentiation time point of days 2 to 3 defines the onset of the early differentiation processes.

#### Significant transcripts identified in day 3 versus day 4 EB comparisons, part of cluster 6

In cluster 6 (day 3 vs. day 4 EBs) (Supplementary Fig. S2, part 1), we identified only 3 genes, *TGF-β*, *Col1a1*, and *ATP7a*, that are highly upregulated and interconnecting to 277 GOs participating mainly to early developmental processes of the vasculature development, keratinocyte differentiation hematopoiesis (*TGF-β*), cartilage and epidermis development, bone trabecula formation, skeletal system morphogenesis, central nervous system neuron development, and limb morphogenesis. Based on mouse knockout and mutant analysis it has been demonstrated that *TGF-β* signaling plays an important role in the development of mesoderm formation, including the development of their organ derivatives such as heart, bone development, and keratin

development, as well as in endoderm formation, including organ derivatives such as liver and gastrointestinal tract development (for review see Ref. [70]). *COL1A1* encodes for the pro- $\alpha$ 1 chains of type I collagen, the fibril-forming collagen found in most connective tissues that is abundant in bone, cornea, dermis, and tendon (for review see Ref. [71]). The copper-translocating *ATP7A* is involved in the  $\text{Cu}^{2+}$  homeostasis and in the biosynthetic incorporation of Cu into copper-dependent enzymes that are essential for the secretory pathway, Cu detoxification via Cu efflux, and mutations in *ATP7* are associated with Menkes disease [72,73].

### *Insulin-like growth factor 1 is the prominent transcript in day 4 versus day 5 EB comparison (cluster 1)*

In cluster 1 (day 4 vs. day 5 EBs), only insulin-like growth factor 1 (*IGF1*) has been identified as highly upregulated and participating into 136 GOs (Supplementary Fig. S2, part 1). *IGF1* plays a central role in the growth and development of many tissues in the body and seems to be a key regulator of skeletal muscle development and repair [74]. Recently, mutations in human *IGF1* gene and transgenic animal studies demonstrate that *IGF1* plays an important role in the etiology of growth retardation [75]. *IGF1* network with 6 GO annotations is provided in Supplementary Fig. S2 (part 1).

### *Onset of cardiomyocyte differentiation and blood cell development*

In cluster 2 (day 5 vs. day 6 EBs), 8 highly expressed interconnecting genes participating to 101 GOs have been identified associated with contractile structure proteins of the heart *Myh6* (myosin, heavy chain 6, cardiac muscle), *Myl2* (myosin, light chain 2), *Ttn* (Titin), *Mybpc3* (myosin binding protein C, cardiac), and others (Supplementary Fig. S2, part 2). These data clearly show that cardiomyocyte differentiation started between days 5 and 6. In cluster 3, 4 expressed interconnecting genes participating in 79 GOs have been identified associated with blood cell development and development of the immune response system: *Ptprc*, *Igh-6*, *Fas*, and *Fcgr1g*.

### *Genes identified at day 7 versus day 10 comparison, part of cluster 7*

In cluster 7, 13 expressed interconnecting genes participating to 248 GOs have been identified associated with general biological processes that are not associated with developmental processes (Supplementary Fig. S2, part 2). Interestingly, among them we found that CD83 antigen is highly upregulated. *CD83* is a lectin that regulates the dendritic cells (DCs) [76] and the lymphocyte maturation and activation [77]. *Fcgr* (Fc $\gamma$  Receptors) is inflammatory immune system mediator protein located in the plasma membrane of leukocytes and eliciting the effects of immunoglobulin G antibodies on leukocytes resulting in an activation of the immune system [78]. Interestingly, *Fcgr3* and *Fcgr2b* are highly upregulated in between 7- and 10-day EBs.

Also, *Anxa1* has been found to be remarkably upregulated. *Anxa1* and other annexins, such as *AnxA6* and *AnxA8*, are  $\text{Ca}^{2+}$ -binding proteins that show differences in their subcellular localization and mode of action. It has been

shown that annexins are involved in the regulation of the EGF receptor activity via controlling the endocytic transport of the EGF receptor [79]. Recently, it has been shown that Annexin A1 plays a pivotal anti-inflammatory role in innate immunity and if it is absent inflammation is persisting. Therefore, Annexin A1 and its G protein-coupled receptor act as a target for anti-inflammatory therapeutics [80]. More recently, using *AnxA1*-KO mice it has been demonstrated that *AnxA1* has pro-angiogenic functions for vascular endothelial cell sprouting, wound healing, and tumor growth and metastasis [81].

From our findings we may conclude that the developmental/differentiation processes in ESCs are hierarchical and reflect the ongoing of these processes under in vivo conditions. Moreover, our study also contributes to the understanding of the mechanisms involved in the pluripotency and differentiation of ESCs. In this context, the candidate genes that are prominently expressed at the pluripotent state other than well-reported genes can be considered for genetic manipulation to elucidate their participation into pluripotency network. This strategy has been demonstrated by the knockdown of *Mageb16*. We also suggest that a consequential application of this strategy may significantly accelerate the generation of knowledge required for a stem-cell-based therapy and an effective drug discovery.

### Acknowledgments

This work was supported by a grant from the European Community (6th Framework Programme, Thematic Priority: Life sciences, genomics and biotechnology for health, Contract No: FunGenES LSHG-CT-2003-503494). We would like to thank Mr. James Mosedale, Abgent Europe Ltd., for supplying us with *Mageb16* polyclonal antibodies (sale order S11072703W).

### Author Disclosure Statement

No competing financial interests exist.

### References

- Evans MJ and MH Kaufman. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292:154–156.
- Niwa H. (2007). How is pluripotency determined and maintained? *Development* 134:635–646.
- Loebel DAF, CM Watson, A De Young and PPL Tam. (2003). Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev Biol* 264:1–14.
- Chambers I, D Colby, M Robertson, J Nichols, S Lee, S Tweedie and A Smith. (2003). Functional expression cloning of *Nanog*, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113:643–655.
- Avilion AA, SK Nicolis, LH Pevny, L Perez, N Vivian and R Lovell-Badge. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 17:126–140.
- Niwa H, J Miyazaki and AG Smith. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24:372–376.
- Scholer HR, S Ruppert, N Suzuki, K Chowdhury and P Gruss. (1990). New type of pou domain in germ line-specific protein Oct-4. *Nature* 344:435–439.

8. Okita K and S Yamanaka. (2006). Intracellular signaling pathways regulating pluripotency of embryonic stem cells. *Curr Stem Cell Res Ther* 1:103–111.
9. Doss MX, J Winkler, SH Chen, R Hippler-Altenburg, I Sotiriadou, M Halbach, K Pfannkuche, HM Liang, H Schulz, et al. (2007). Global transcriptome analysis of murine embryonic stem cell-derived cardiomyocytes. *Genome Biol* 8: R56.
10. Gissel C, C Voolstra, MX Doss, CI Koehler, J Winkler, J Hescheler and A Sachinidis. (2005). An optimized embryonic stem cell model for consistent gene expression and developmental studies—a fundamental study. *Thromb Haemost* 94:719–727.
11. Schulz H, R Kolde, P Adler, I Aksoy, K Anastassiadis, M Bader, N Billon, H Boeuf, PY Bourillot, et al. (2009). The FunGenES database: a genomics resource for mouse embryonic stem cell differentiation. *PLoS One* 4:e6804.
12. Gautier L, L Cope, BM Bolstad and RA Irizarry. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307–315.
13. Irizarry RA, BM Bolstad, F Collin, LM Cope, B Hobbs and TP Speed. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31:e15.
14. Pepper SD, EK Saunders, LE Edwards, CL Wilson and CJ Miller. (2007). The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics* 8:273.
15. Smyth GK. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3.
16. Doss MX, S Chen, J Winkler, R Hippler-Altenburg, M Odenthal, C Wickenhauser, S Balaraman, H Schulz, O Hummel, et al. (2007). Transcriptomic and phenotypic analysis of murine embryonic stem cell derived BMP2(+) lineage cells: an insight into mesodermal patterning. *Genome Biol* 8:R184.
17. Doss MX, V Wagh, H Schulz, M Kull, R Kolde, K Pfannkuche, T Nolden, H Himmelbauer, J Vilo, J Hescheler and A Sachinidis. (2010). Global transcriptomic analysis of murine embryonic stem cell-derived brachyury plus (T) cells. *Genes Cells* 15:209–228.
18. Potta SP, HM Liang, K Pfannkuche, J Winkler, SH Chen, MX Doss, K Obernier, N Kamisetti, H Schulz, et al. (2009). Functional characterization and transcriptome analysis of embryonic stem cell-derived contractile smooth muscle cells. *Hypertension* 53:196–204.
19. Mariappan D, J Winkler, SH Chen, H Schulz, J Hescheler and A Sachinidis. (2009). Transcriptional profiling of CD31(+) cells isolated from murine embryonic stem cells. *Genes Cells* 14:243–260.
20. Haileselasse SK, CJ Porter, G Palidwor, C Perez-Iratxeta, EM Muro, PA Campbell, MA Rudnicki and MA Andrade-Navarro. (2007). Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* 8:85.
21. Boyer LA, K Plath, J Zeitlinger, T Brambrink, LA Medeiros, TI Lee, SS Levine, M Wernig, A Tajonar, et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441:349–353.
22. Zhang Z, A Jones, CW Sun, C Li, CW Chang, HY Joo, QA Dai, MR Mysliwiec, LC Wu, et al. (2011). PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem Cells* 29:229–240.
23. Hori K, J Cholewa-Waclaw, Y Nakada, SM Glasgow, T Masui, RM Henke, H Wildner, B Martarelli, TM Beres, et al. (2008). A nonclassical bHLH-Rbpj transcription factor complex is required for specification of GABAergic neurons independent of Notch signaling. *Genes Dev* 22:166–178.
24. Mason MJ, GP Fan, K Plath, Q Zhou and S Horvath. (2009). Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* 10:327.
25. Guo Y, R Costa, H Ramsey, T Starnes, G Vance, K Robertson, M Kelley, R Reinbold, H Scholer and R Hromas. (2002). The embryonic stem cell transcription factors Oct4 and FoxD3 interact to regulate endodermal-specific promoter expression. *Proc Natl Acad Sci U S A* 99:3663–3667.
26. Walter J. (2011). An epigenetic Tet a Tet with pluripotency. *Cell Stem Cell* 8:121–122.
27. Pardo M, B Lang, L Yu, H Prosser, A Bradley, MM Babu and J Choudhary. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* 6:382–395.
28. Du JA, TJ Chen, X Zou, B Xiong and GX Lu. (2010). Dppa2 knockdown-induced differentiation and repressed proliferation of mouse embryonic stem cells (vol 147, pg 265, 2010). *J Biochem (Tokyo)* 147:929.
29. Ooi SKT, D Wolf, O Hartung, S Agarwal, GQ Daley, SP Goff and TH Bestor. (2010). Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics Chromatin* 3:17.
30. Yang XJ and M Ullah. (2007). MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene* 26:5408–5419.
31. Boyer LA, TI Lee, MF Cole, SE Johnstone, SS Levine, JR Zucker, MG Guenther, RM Kumar, HL Murray, et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122:947–956.
32. Guo GJ, M Huss, GQ Tong, CY Wang, LL Sun, ND Clarke and P Robson. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* 18:675–685.
33. Simon L, GC Ekman, T Garcia, K Carnes, Z Zhang, T Murphy, KM Murphy, RA Hess, PS Cooke and MC Hofmann. (2010). ETV5 regulates sertoli cell chemokines involved in mouse stem/progenitor spermatogonia maintenance. *Stem Cells* 28:1882–1892.
34. Chotteau-Lelievre A, R Montesano, J Soriano, P Soulie, X Desbiens and Y de Launoit. (2003). PEA3 transcription factors are expressed in tissues undergoing branching morphogenesis and promote formation of duct-like structures by mammary epithelial cells in vitro. *Dev Biol* 259:241–257.
35. Burgess R, A Rawls, D Brown, A Bradley and EN Olson. (1996). Requirement of the paraxis gene for somite formation and musculoskeletal patterning. *Nature* 384:570–573.
36. Mitsui K, Y Tokuzawa, H Itoh, K Segawa, M Murakami, K Takahashi, M Maruyama, M Maeda and S Yamanaka. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113:631–642.
37. Wang L, A Zheng, L Yi, CR Xu, MX Ding and HK Deng. (2004). Identification of potential nuclear reprogramming and differentiation factors by a novel selection method for cloning chromatin-binding proteins. *Biochem Biophys Res Commun* 325:302–307.
38. Vestergaard J, A Lind-Thomsen, MW Pedersen, HO Jarmer, M Bak, L Hasholt, N Tommerup, Z Tumer and LA Larsen. (2008). GLI1 is involved in cell cycle regulation and proliferation of NT2 embryonal carcinoma stem cells. *DNA Cell Biol* 27:251–256.

39. Reiner R, N Alfiya-Mor, M Berrebi-Demma, D Wesolowski, S Altman and N Jarrous. (2011). RNA binding properties of conserved protein subunits of human RNase P. *Nucleic Acids Res* 39:5704–5714.
40. Session DR, GS Lee and DJ Wolgemuth. (2001). Characterization of D1Pas1, a mouse autosomal homologue of the human AZFa region DBY, as a nuclear protein in spermatogenic cells. *Fertil Steril* 76:804–811.
41. Scaldaferrri ML, S Fera, L Grisanti, M Sanchez, M Stefanini, M De Felici and E Vicini. (2011). Identification of side population cells in mouse primordial germ cells and prenatal testis. *Int J Dev Biol* 55:209–214.
42. Inoue N, KD Hess, RW Moreadith, LL Richardson, MA Handel, ML Watson and AR Zinn. (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* 8:1201–1207.
43. Watson ML, AR Zinn, N Inoue, KD Hess, J Cobb, MA Handel, R Halaban, CC Duchene, GM Albright and RW Moreadith. (1998). Identification of morc (microrchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proc Natl Acad Sci U S A* 95:14361–14366.
44. Yang GD, LY Wu, S Bryan, N Khaper, S Mani and R Wang. (2010). Cystathionine gamma-lyase deficiency and overproliferation of smooth muscle cells. *Cardiovasc Res* 86:487–495.
45. Gao HJ, GY Wu, TE Spencer, GA Johnson and FW Bazer. (2009). Select nutrients in the ovine uterine lumen. III. Cationic amino acid transporters in the ovine uterus and perimplantation conceptuses. *Biol Reprod* 80:602–609.
46. Yoshima T, T Yura and H Yanagi. (1998). Novel testis-specific protein that interacts with heat shock factor 2. *Gene* 214:139–146.
47. Yoshimura T, S Toyoda, S Kuramochi-Miyagawa, T Miyazaki, S Miyazaki, F Tashiro, E Yamato, T Nakano and J Miyazaki. (2009). Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Dev Biol* 335:216–227.
48. Gasca S, DP Hill, J Klingensmith and J Rossant. (1995). Characterization of a gene trap insertion into a novel gene, cordon-bleu, expressed in axial structures of the gastrulating mouse embryo. *Dev Genet* 17:141–154.
49. Wu C, C Orozco, J Boyer, M Leglise, J Goodale, S Batalov, CL Hodge, J Haase, J Janes, JW Huss, III and AI Su. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10:R130.
50. Tashiro F, M Kanai-Azuma, S Miyazaki, M Kato, T Tanaka, S Toyoda, E Yamato, H Kawakami, T Miyazaki and JI Miyazaki. (2010). Maternal-effect gene *Ces5/Ooep/Moep19/Floped* is essential for oocyte cytoplasmic lattice formation and embryonic development at the maternal-zygotic stage transition. *Genes Cells* 15:813–828.
51. Mhyre AJ, AM Marcondes, EY Spaulding and HJ Deeg. (2009). Stroma-dependent apoptosis in clonal hematopoietic precursors correlates with expression of *PYCARD*. *Blood* 113:649–658.
52. Ouyang Z, Q Zhou and WH Wong. (2009). ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc Natl Acad Sci U S A* 106:21521–21526.
53. Chou KC and HB Shen. (2008). Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat Protoc* 3:153–162.
54. Savage J, A Voronova, V Mehta, F Sendi-Mukasa and IS Skerjanc. (2010). Canonical Wnt signaling regulates *Foxc1/2* expression in P19 cells. *Differentiation* 79:31–40.
55. Ang SL and J Rossant. (1994). *Hnf-3-beta* is essential for node and notochord formation in mouse development. *Cell* 78:561–574.
56. Hong HK, JK Noveroske, DJ Headon, T Liu, MS Sy, MJ Justice and A Chakravarti. (2001). The winged helix/forkhead transcription factor *Foxq1* regulates differentiation of hair in satin mice. *Genesis* 29:163–171.
57. Swiderski RE, RS Reiter, DY Nishimura, WLM Alward, JW Kalenak, CS Searby, EM Stone, VC Sheffield and JJC Lin. (1999). Expression of the *Mfl1* gene in developing mouse hearts: implication in the development of human congenital heart defects. *Dev Dyn* 216:16–27.
58. Winnier GE, L Hargett and BLM Hogan. (1997). The winged helix transcription factor *MFH1* is required for proliferation and patterning of paraxial mesoderm in the mouse embryo. *Genes Dev* 11:926–940.
59. Hiemisch H, AP Monaghan, G Schutz and KH Kaestner. (1998). Expression of the mouse *Fkh1/Mf1* and *Mfh1* genes in late gestation embryos is restricted to mesoderm derivatives. *Mech Dev* 73:129–132.
60. Kume T, KY Deng and BLM Hogan. (2000). Murine forkhead/winged helix genes *Foxc1* (*Mf1*) and *Foxc2* (*Mfh1*) are required for the early organogenesis of the kidney and urinary tract. *Development* 127:1387–1395.
61. Arenas E. (2008). *Foxa2*: the rise and fall of dopamine neurons. *Cell Stem Cell* 2:110–112.
62. Lee CS, JR Friedman, JT Fulmer and KH Kaestner. (2005). The initiation of liver development is dependent on *Foxa* transcription factors. *Nature* 435:944–947.
63. Mora-Castilla S, JR Tejedo, A Hmadcha, GM Cahuana, F Martin, B Soria and FJ Bedoya. (2010). Nitric oxide repression of *Nanog* promotes mouse embryonic stem cell differentiation. *Cell Death Differ* 17:1025–1033.
64. Hunter C and S Rhodes. (2005). LIM-homeodomain genes in mammalian development and human disease. *Mol Biol Rep* 32:67–77.
65. Wardle FC and VE Papaioannou. (2008). Teasing out T-box targets in early mesoderm. *Curr Opin Genet Dev* 18:418–425.
66. Studer M, A Lumsden, L ArizaMcNaughton, A Bradley and R Krumlauf. (1996). Altered segmental identity and abnormal migration of motor neurons in mice lacking *Hoxb-1*. *Nature* 384:630–634.
67. Studer M, A Gavalas, H Marshall, L Ariza-McNaughton, FM Rijli, P Chambon and R Krumlauf. (1998). Genetic interactions between *Hoxa1* and *Hoxb1* reveal new roles in regulation of early hindbrain patterning. *Development* 125:1025–1036.
68. Buckley RR and R Krumlauf. (2007). Contribution of *Hoxb1* to the development of the vertebrate skeleton. *Dev Biol* 306:883–893.
69. Gouti M and A Gavalas. (2008). *Hoxb1* controls cell fate specification and proliferative capacity of neural stem and progenitor cells. *Stem Cells* 26:1985–1997.
70. Kitisin K, T Saha, T Blake, N Golestaneh, M Deng, C Kim, Y Tang, K Shetty, B Mishra and L Mishra. (2007). Tgf-Beta signaling in development. *Sci Stke* 2007:cm1.
71. Basel D and RD Steiner. (2009). Osteogenesis imperfecta: recent findings shed new light on this once well-understood condition. *Genet Med* 11:375–385.

72. Veldhuis N, A Gaeth, R Pearson, K Gabriel and J Camakaris. (2009). The multi-layered regulation of copper translocating P-type ATPases. *Biometals* 22:177–190.
73. Lalioti V, G Muruais, Y Tsuchiya, D Pulido and IV Sandoval. (2009). Molecular mechanisms of copper homeostasis. *Front Biosci* 14:4878–4903.
74. Barton ER. (2006). The ABCs of IGF-I isoforms: impact on muscle hypertrophy and implications for repair. *Appl Physiol Nutr Metab* 31:791–797.
75. Klammt J, R Pfaffle, H Werner and W Kiess. (2008). IGF signaling defects as causes of growth failure and IUGR. *Trends Endocrinol Metab* 19:197–205.
76. Erbacher A, F Gieseke, R Handgretinger and I Muller. (2009). Dendritic cells: functional aspects of glycosylation and lectins. *Hum Immunol* 70:308–312.
77. Breloer M and B Fleischer. (2008). CD83 regulates lymphocyte maturation, activation and homeostasis. *Trends Immunol* 29:186–194.
78. Willcocks LC, KGC Smith and MR Clatworthy. (2009). Low-affinity Fc gamma receptors, autoimmunity and infection. *Expert Rev Mol Med* 11:e24.
79. Grewal T and C Enrich. (2009). Annexins—modulators of EGF receptor signalling and trafficking. *Cell Signal* 21: 847–858.
80. Perretti M and J Dalli. (2009). Exploiting the Annexin A1 pathway for the development of novel anti-inflammatory therapeutics. *Br J Pharmacol* 158:936–946.
81. Yi M and JE Schnitzer. (2009). Impaired tumor growth, metastasis, angiogenesis and wound healing in annexin A1-null mice. *Proc Natl Acad Sci U S A* 106:17886–17891.

Address correspondence to:

*Prof. Agapios Sachinidis  
Center of Physiology and Pathophysiology  
Institute of Neurophysiology  
University of Cologne  
Robert Koch Str. 39  
Cologne 50931  
Germany*

*E-mail: a.sachinidis@uni-koeln.de*

Received for publication November 13, 2011

Accepted after revision March 15, 2012

Prepublished on Liebert Instant Online March 15, 2012