Genome-wide gene expression analysis in mouse embryonic stem cells

JUAN SAINZ*[1,2,#], FERNANDO GARCÍA-ALCALDE[3,#], ARMANDO BLANCO[3] and ÁNGEL CONCHA[4]

[1]Genomic Oncology Area, GENYO (Pfizer-University of Granada-Andalusian Goverment Centre for Genomics and Oncological Research), [2]Hematology Department, University Hospital Virgen de las Nieves, [3]Department of Computer Science and Artificial Intelligence, University of Granada and [4] Department of Pathology, University Hospital Virgen de las Nieves, Granada, Spain

ABSTRACT Embryonic stem cell studies have generated great interest, due to their ability to form a wide variety of matured cells. However, there remains a poor understanding of mechanisms regulating the cell state of embryonic stem cells (ESCs) and of the genes they express during early differentiation. Gene expression analysis may be a valuable tool to elucidate either the molecular pathways involved in self-renewal and pluripotency, or early differentiation and to identify potential molecular therapy targets. The aim of this study was to characterize at the molecular level the undifferentiated mouse ESC state and the early development towards embryoid bodies. To attempt this issue, we performed CodeLink Mouse Uniset I 20K bioarrays in a well-characterized mouse ESC line, MES3, 3- and 7 day-old embryoid bodies and we compared our findings with those in adult tissue cells. Gene expression results were subsequently validated in a commercial stem cell line, CGR8 (ATCC). Significance Analysis of Microarrays (SAM) was used to identify statistically significant changes in microarray data. We identified 3664 genes expressed at significantly greater levels in MES3 stem cells than in adult tissue cells, which included 611 with 3-fold higher gene expression levels versus the adult cells. We also investigated the gene expression profile during early embryoid body formation, identifying 2040 and 2243 genes that were up-regulated in 3- and 7- day-old embryoid bodies, respectively. Our gene expression results in MES3 cells were partially confirmed in CGR8 cells, showing numerous genes that are expressed in both mouse stem cells. In conclusion, our results suggest that commonly expressed genes may be strong candidates for involvement in the maintenance of a pluripotent and undifferentiated phenotype and in early development.

KEY WORDS: *mouse stem cell line, gene expression profiling, early differentiation*

Embryonic stem cell (ESC) research has provided a deeper understanding of cell development and is a promising area in medicine, but it is in a preliminary stage (Lagasse *et al.*, 2001). Human ESCs have been used for a wide range of cell-based therapies, because they can potentially replace damaged cells in adult organisms (Korbling and Estrov, 2003). This replacement is possible because ESCs can propagate indefinitely in culture, maintaining a normal karyotype and undifferentiated state (Korbling and Estrov, 2003). This ability of ESCs to divide indefinitely provides an ideal system for the study of early development pathways and offers a potentially unlimited source of cells for organ transplantation and the development of novel therapies. In fact, therapies based on the use of human stem cells have already been proposed for regenerative medicine and tissue replacement after injury or disease. Human ESCs therefore appear to be an invaluable pre-clinical tool. However, utilization of the full therapeutic potential of ESCs requires an understanding of the signal pathways that control ESC development and regulate their specialization and proliferation for the repair of a specific tissue. It is also necessary to know the specific genes that drive ESCs to embryoid body (EB) formation, i.e., the genes involved in earlier cell differentiation and those responsible for ESC phenotypes and

---

***Address correspondence to:** Juan Sainz Pérez. Genomic Oncology area, Genyo - Pfizer-University of Granada-Andalusian Government Centre for Genomics and Oncological Research, 1st Floor, Office 4, Lab 6, Avda. de la Ilustración 114, 18007 Granada, Spain. Tel: +34-958-637-081. Fax: +34-958-637071. e-mail: juan.sainz@genyo.es **#Note:** These authors contributed equally to this paper.

behaviours. Few data are available on the genes and processes involved in ESC control and differentiation (Hailesellasse Sene *et al.*, 2007, Lee *et al.*, 2000). In recent years, several methods have been developed for the direct differentiation of ESCs into specific cell types (Kuo *et al.*, 2003, Lee *et al.*, 2000, Lumelsky *et al.*, 2001). Among these, differentiation of ESCs *via* EBs proved a suitable model to study cellular events during early development *in vitro* (Kramer *et al.*, 2006). Elucidation of molecular pathways that define mESC (mouse embryonic stem cells) pluripotency, self-renewal, differentiation and maturation to EBs are crucial for gaining a more comprehensive knowledge of ESC biology and behaviour. With the availability of a well-annotated mouse genome database (Waterston *et al.*, 2002), the gene-expression profiling of mESC lines may allow the identification of "stemness" genes and improve our understanding of the behaviour of mESCs, including the maintenance of the undifferentiated state and the mechanisms underlying self-renewal, early differentiation and development.

CodeLink™ is a microarray platform that allows the gene expression analysis of potentially thousands of transcriptomes using short oligonucleotide probes. CodeLink™ bioarrays are a well-established and powerful system that, in combination with bioinformatic tools (Bassett *et al.*, 1999, Howbrook *et al.*, 2003, Quackenbush, 2001), offers valuable insights into gene and molecular pathways involved in numerous biological events (Yoo *et al.*, 2009). Gene expression microarrays have been used to discover genes that are differentially expressed in developmental processes (Hemberger *et al.*, 2001, Kim *et al.*, 2001, Tanaka *et al.*, 2000). We hypothesized that genome-wide gene expression analysis of mESCs may facilitate discovery of the main genes involved in the maintenance of pluripotency and self-renewal and in the early differentiation process. The genes responsible for pluripotency and self-renewal would first be up-regulated in the undifferentiated state and then down-regulated during early cell differentiation and maturation. The genes involved in triggering early differentiation would also be initially down-regulated and then up-regulated during earlier development stages. To test this hypothesis, we performed a study to identify genes and hence cellular processes that may play a major role in the differentiation of mESCs to EB formation. We used the CodeLink™ microarray platform to assay (in duplicate) total RNA samples from a mESC line (MES3) and EBs. Measurements of gene expression were taken from biological duplicates in undifferentiated stem cells (day 0) and during differentiation at two time points (days 3 and 7) where pluripotency is already lost. Results were validated by replicating gene expression analyses in a second mESC line, CGR8. The analyses were expected to show genes involved in pluripotency and self-renewal of stem cells ("stemnes" genes) and the large expression changes during the initial period of differentiation facilitating the identification of the pathways underlying the early development process.

Results

Stem cell characterization

MES3 and CGR8 cell cultures were characterized by continuously performed *in vivo* visualizations to test whether the cells exhibited the properties of stem cells (Fig. 1A). The growing and subculturing continued over several months to ensure that cells were capable of long-term self-renewal. Cultures were examined microscopically to test their healthy and undifferentiated state. We also determined whether cells could be subcultured after freezing, thawing and replating. Finally, we tested whether they were pluripotent by allowing them to differentiate spontaneously in cell culture. For this purpose, the mESC lines were allowed to clump and form embryoid bodies. Evaluation of cell surface markers of an undifferentiated state showed that MES3 cells were positive for alkaline phosphatase (Fig. 1B) and *Ssea4*, another marker of undifferentiated mouse ESC lines (Fig. 1C). The undifferentiated state of the CGR8 cell line was analyzed by qRT-PCR to detect the expression level of undifferentiated cell markers previously shown to indicate undifferentiated ESCs or ESC self-renewal [e.g. *Nanog*, *Oct3/4* (also called *POU5f1*) and Rex-1]. For this purpose, we measured the gene expression levels of *Nanog*, *Oct3/4* and *Rex1* at days 0, 1, 2, 5, 15 and 20 (Fig. 2). As expected, *Nanog*, *Oct3/4* and *Rex1* were dramatically down-regulated after day 5 of differentiation. We also tested for the absence of early differentia-

tion markers, considering *Hnf4* and *Ttr* genes to be endoderm markers and *Fgf5* and *Wnt3* genes to be ectoderm markers. *Hnf4*, *Ttr* and *Fgf5* genes were down-regulated at baseline and showed a dramatic increase after day 5 of differentiation. These results showed a high degree of correlation with array data (Fig. 2) and confirmed the undifferentiated state of the mESCs and their progressive differentiation towards EBs.

Gene expression profiling of MES3 mESC line by microarrays

CodeLink UniSet Mouse I Expression Bioarray chips were assayed to compare the gene expression profiling of MES3 at day 0 (undifferentiated) with the expression of mURNA (reference for adult tissue cell expression). Duplicated hybridization experiments with
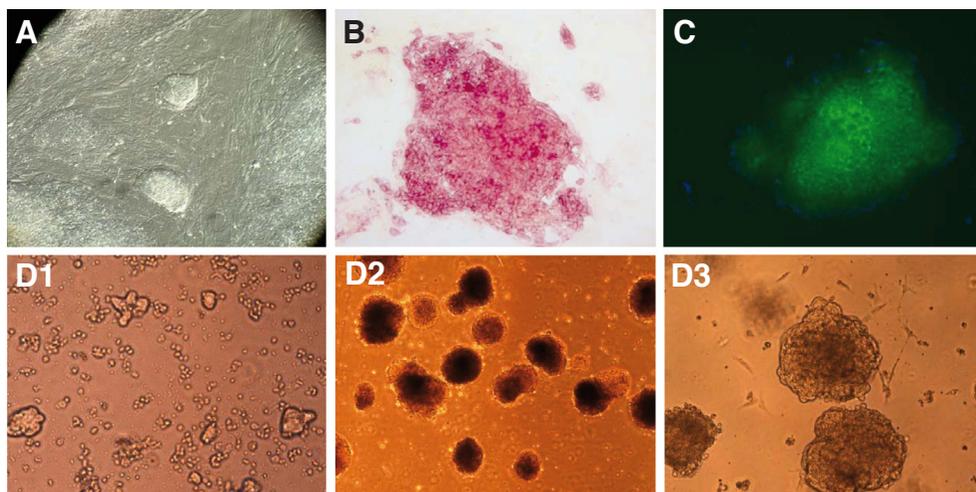
Fig. 1. Morphological characterization of mESCs and EBs. (A) *mESCs colonies growing on feeders;* **(B)** *mESCs staining for alkaline phosphatase activity;* **(C)** *mESCs immunodetection of* Ssea4 *surface marker;* **(D1-D3)** *representative images of embryonic bodies at days 1, 5 and 15, respectively.*
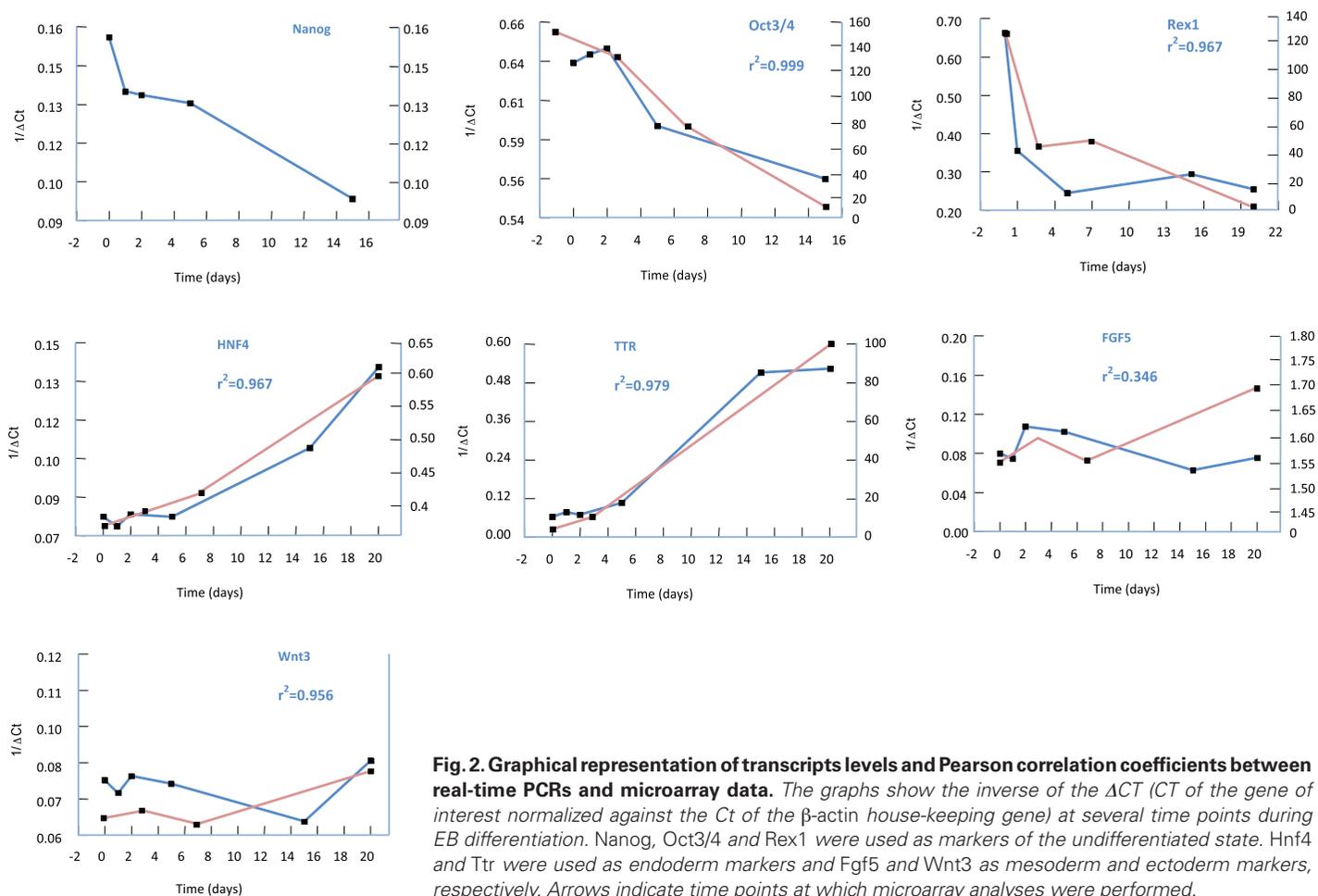
Fig. 2. Graphical representation of transcripts levels and Pearson correlation coefficients between real-time PCRs and microarray data. *The graphs show the inverse of the ΔCT (CT of the gene of interest normalized against the Ct of the* β-actin *house-keeping gene) at several time points during EB differentiation.* Nanog, Oct3/4 *and* Rex1 *were used as markers of the undifferentiated state.* Hnf4 *and* Ttr *were used as endoderm markers and* Fgf5 *and* Wnt3 *as mesoderm and ectoderm markers, respectively. Arrows indicate time points at which microarray analyses were performed.*

cRNA from both ESC lines showed a correlation coefficient >0.90, indicating high reproducibility (Fig. 3 A-D). Scatter plot analyses showed numerous differences between the MES3 and mURNA cell lines (Fig. 3 E-H). Gene expression analysis revealed that 5807 genes were differentially expressed by ≥ 2-fold versus mURNA. Out of these, 3664 genes were up-regulated, including 611 with a 3-fold higher expression than in the adult tissue cells. The remaining 2143 genes were down-regulated. Sup. Table 2 shows these representative genes with elevated (>4 fold) expression in MES3 ESCs. *Nanog*, *Rex-1* and *Oct3/4 (POU5f1)* were not detected in the MES3 microarrays, but their presence was confirmed by qRT-PCR (Fig. 2). Fig. 4A depicts the histogram of the distribution of GO terms significantly expressed in undifferentiated stem cells.

Gene expression profiling of 3 and 7 day-old embryoid bodies

CodeLink microarrays were hybridized using equal amounts of labelled cRNAs from undifferentiated ESCs and 3- and 7-day-old EBs. Undifferentiated mESCs were used as reference to detect gene expression differences between EBs at 3 and 7 days and adult tissue cells. Fig. 3 depicts a scatter plot of the normalized intensities of the genes in the hybridizations. The solid diagonal line is the linear regression of the intensity data; 20,000 genes were analyzed in each assay.

Gene expression analysis revealed that 8105 genes were differentially expressed in 3-day-old EBs in comparison to adult tissue cells (p<0.05): 6065 upregulated genes and 2040 down-regulated genes. Among the over-expressed genes, 624 showed ≥ 3-fold higher expression versus adult tissue cells. Sup. Table 3 shows the top 100 genes (ranked by fold change) with a changed expression in EBs at 3 days after spontaneous differentiation of mESCs. The set of 2040 genes that were down-regulated in the 3-day-old EBs included genes involved in amino acid biosynthesis and/or purine, carbohydrate, lipid and steroid metabolism and spermatogenesis (Sup. Table 3). Fig. 4B depicts the histogram of GO terms significantly over-represented in 3-day-old EBs.

The gene expression pattern found in 7-day-old EB mRNA revealed a high similarity with that in 3-day-old EBs. There were 9198 differentially expressed genes in 7-day-old EBs, of which 6995 were significantly up-regulated and 2243 significantly down-regulated. Sup. Table 4 shows the top 100 list (by fold change) of genes differentially expressed in 7-day-old EBs versus mURNA from adult tissue cells. Fig. 4B depicts a histogram of the GO terms significantly over-represented in 7-day-old EBs.

Cluster analysis to display overall features of early differentiation and embryoid body formation

Hierarchical clustering was used to display the differentially expressed genes (rows) between 3- and 7-day-old EBs (columns),
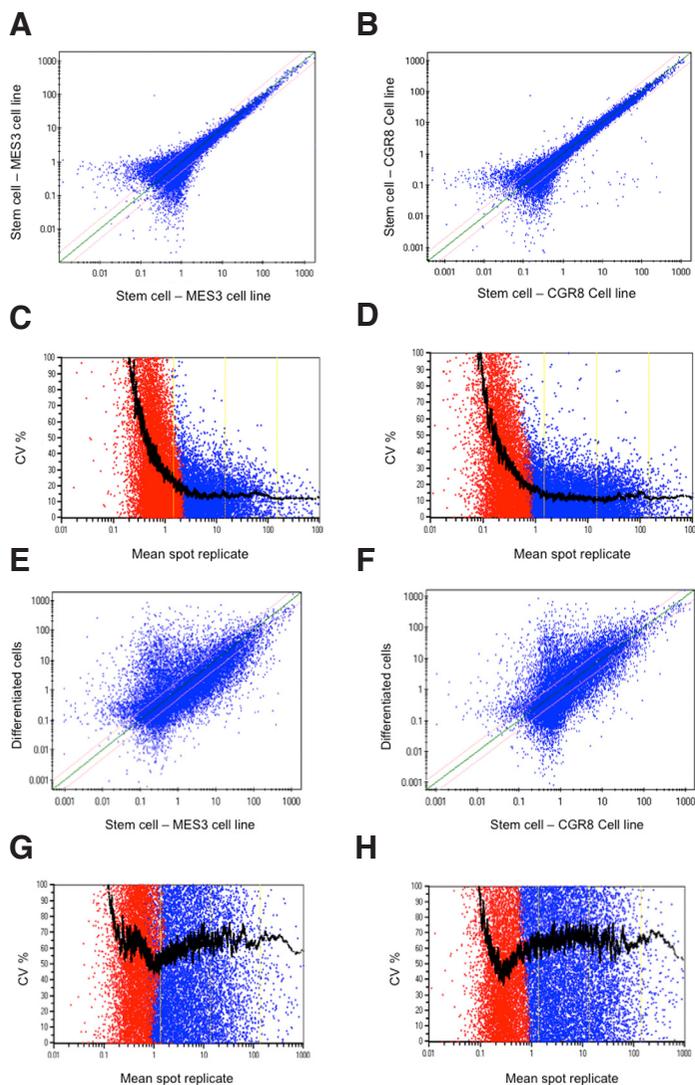
Fig. 3. Log-scale scatter plots of samples labelled by the CodeLink standard protocol. *Lines indicate a 2-fold difference.* *Ten micrograms of labelled cRNA loaded into microarray slides, showing all probe sets.* **(A)** *MES3 stem cell line.* **(B)** *CGR8 stem cell line.*

higher expression than in the adult tissue cells. The remaining 2143 genes were down-regulated. These results indicated that stem cells show a higher level of over-expressed genes that the adult cells and, therefore, those genes might be strong candidates for involvement in the maintenance of a pluripotent, undifferentiated phenotype. Comparison of the up-regulated genes with those previously described in other mESC lines showed a good agreement with expected expression patterns. Further analysis and organization of these differentially expressed genes yielded similar gene expression profiles to those of genes known to be differentially over-expressed in other mouse or human ES cell lines (Furusawa *et al.*, 2004, Phillips *et al.*, 2000, Sato *et al.*, 2003, Tanaka *et al.*, 2002). Among them, we identified the following functional groupings: (1) transcription factors/nucleic acid binding, e.g., *Top2A* (Sato *et al.*, 2003, Tanaka *et al.*, 2002), *Tek* (Phillips *et al.*, 2000, Sato *et al.*, 2003), *Ung* (Ivanova *et al.*, 2002, Kelly and Rizzino, 2000), *Klf3* (Ramalho-Santos *et al.*, 2002), *Klf4* (Nakatake *et al.*, 2006),

using half-square euclidean distance as similarity measure. Fig. 5 depicts the clustering of the differentially expressed genes. Comparison of the representation of functional groups of genes between the entire mouse genome and the set of differentially transcribed genes revealed whether clusters of genes of a specific functional group were enriched in the differentially expressed group. The identified clusters were annotated to common functional categories described in the Gene Ontology (GO) (Ashburner *et al.*, 2000). Genes in clusters belonged either to common metabolic pathways or cellular components, in which most of the genes encoded a protein (a gene product) that might have function as an enzymatic step in a cellular process or, alternatively, interacted directly with each other to form multimeric proteins or serve as ligands and receptors in signaling cascades. Based on the hierarchical clustering of our expression data, we drew a tree of early-development relatedness (Fig. 5). Our results revealed 6 distinct clusters with a diverse expression profile. Next, we made use of the FatiGO tool (Al-Shahrour *et al.*, 2004) in order to find significant associations between the genes in each cluster. Considering only the significant GO-terms with the p-value < 0.05 we found that, among 103 genes that were differentially expressed in 3days-EBs vs. 7d-EBs, genes participating in processes such as extracellular matrix and/ or collagen synthesis, protein transport, cell adhesion and tissue development were overrepresented (Fig. 6).
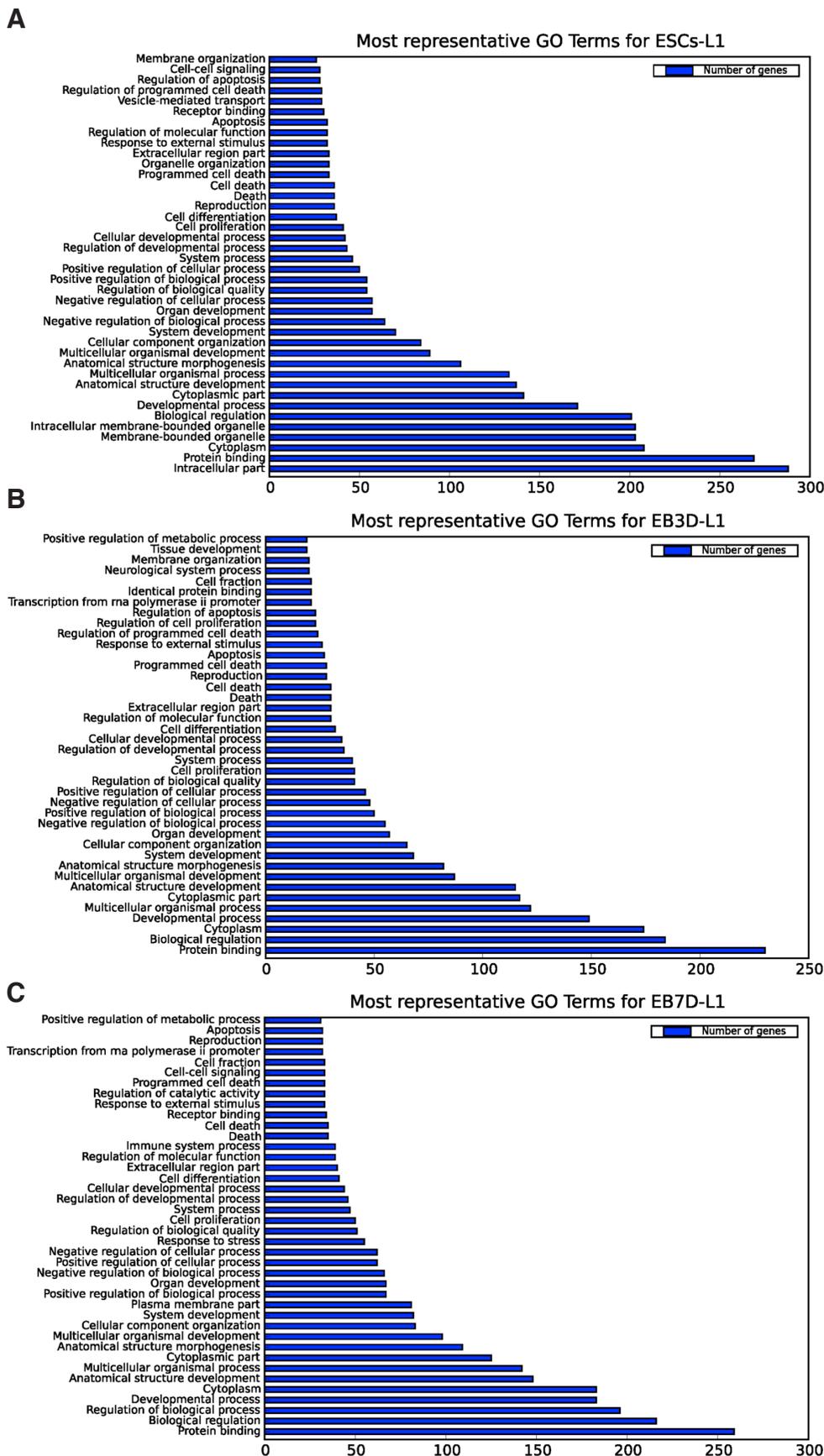
Gene expression analysis: validation study using CGR8 mESC line

As in the MES3 line, we performed microarrays to examine the variation at three different time points (0, 3 and 7 days of differentiation), using identical laboratory protocols. Comparison of gene expression levels between the two stem cell lines showed a large degree of replication as well as the expected differences. A total of 867 genes were differentially and commonly expressed in both mESC lines but not in adult tissue cells (Fig. 7A), and 1587 and 1791 genes were differentially and commonly expressed in 3 and 7 days-old EBs, respectively, in both mESCs but not in adult tissue cells (Fig. 7 B,C).

Discussion

The mechanisms regulating pluripotency, self-renewal and stem cell differentiation are far from being completely understood, and little is known about the genes involved in these complex processes. Microarrays for genome-wide gene expression profiling are important research tools for identifying potential biomarkers and candidate genes for the identification of diseases and other biological processes (Hemberger *et al.*, 2001, Kim *et al.*, 2001, Tanaka *et al.*, 2000). In this report, we performed the genome-wide gene expression profiling of MES3, a well-defined mouse embryonic stem cell line. We assayed CodeLink microarrays in undifferentiated MES3 cells and in EBs formed after 3 and 7 days of spontaneous differentiation and globally normalized the data by background subtraction. The quality of undifferentiated stem cells cultured in our laboratory was also confirmed by evaluating the expression of markers expressed by undifferentiated ES cells (Furusawa *et al.*, 2004, Jeong *et al.*, 2007, Zhao *et al.*, 2005).

Gene expression analysis identified 5807 genes as differentially expressed by ≥ 2-fold in MES3 cells versus adult cells. Among them, 3664 genes were up-regulated, including 611 with a 3-fold

A

Most representative GO Terms for ESCs-L1

Number of genes

Membrane organization
Cell-cell signaling
Regulation of apoptosis
Regulation of programmed cell death
Vesicle-mediated transport
Receptor binding
Apoptosis
Regulation of molecular function
Response to external stimulus
Extracellular region part
Organelle organization
Programmed cell death
Cell death
Death
Reproduction
Cell differentiation
Cell proliferation
Cellular developmental process
Regulation of developmental process
System process
Positive regulation of cellular process
Positive regulation of biological process
Regulation of biological quality
Negative regulation of cellular process
Organ development
Negative regulation of biological process
System development
Cellular component organization
Multicellular organismal development
Anatomical structure morphogenesis
Multicellular organismal process
Anatomical structure development
Cytoplasmic part
Developmental process
Biological regulation
Intracellular membrane-bounded organelle
Membrane-bounded organelle
Cytoplasm
Protein binding
Intracellular part

0   50   100   150   200   250   300

B

Most representative GO Terms for EB3D-L1

Number of genes

Positive regulation of metabolic process
Tissue development
Membrane organization
Neurological system process
Cell fraction
Identical protein binding
Transcription from rna polymerase ii promoter
Regulation of apoptosis
Regulation of cell proliferation
Regulation of programmed cell death
Response to external stimulus
Apoptosis
Programmed cell death
Reproduction
Cell death
Death
Extracellular region part
Regulation of molecular function
Cell differentiation
Cellular developmental process
Regulation of developmental process
System process
Cell proliferation
Regulation of biological quality
Positive regulation of cellular process
Negative regulation of cellular process
Positive regulation of biological process
Negative regulation of biological process
Organ development
Cellular component organization
System development
Anatomical structure morphogenesis
Multicellular organismal development
Anatomical structure development
Cytoplasmic part
Multicellular organismal process
Developmental process
Cytoplasm
Biological regulation
Protein binding

0   50   100   150   200   250

C

Most representative GO Terms for EB7D-L1

Number of genes

Positive regulation of metabolic process
Apoptosis
Reproduction
Transcription from ma polymerase ii promoter
Cell fraction
Cell-cell signaling
Programmed cell death
Regulation of catalytic activity
Response to external stimulus
Receptor binding
Cell death
Death
Immune system process
Regulation of molecular function
Extracellular region part
Cell differentiation
Cellular developmental process
Regulation of developmental process
System process
Cell proliferation
Regulation of biological quality
Response to stress
Negative regulation of cellular process
Positive regulation of cellular process
Negative regulation of biological process
Organ development
Positive regulation of biological process
Plasma membrane part
System development
Cellular component organization
Multicellular organismal development
Anatomical structure morphogenesis
Cytoplasmic part
Multicellular organismal process
Anatomical structure development
Cytoplasm
Developmental process
Regulation of biological process
Biological regulation
Protein binding

0   50   100   150   200   250   300

Klf5 (Ramalho-Santos *et al.*, 2002), *Klf9* (Ramalho-Santos *et al.*, 2002), *Smn1* (Liu *et al.*, 2007), *Ttf1* (Ivanova *et al.*, 2002, Kelly and Rizzino, 2000), *Sall1* (Sato *et al.*, 2003) and *FoxD3* (Abeyta *et al.*, 2004, Sperger *et al.*, 2003); (2) DNA repair / DNA and RNA-related genes, e.g., *Dnmt3a* (Tanaka *et al.*, 2002), *Msh2* (Ramalho-Santos *et al.*, 2002), *Hdac2* (Bhattacharya *et al.*, 2004), *Ddx21* and *Npm1* (Bhattacharya *et al.*, 2004); (Tanaka *et al.*) cell cycle / cell signalling / cell growth / cellular processes, e.g., *Ccdn1* (Abeyta *et al.*, 2004, Ivanova *et al.*, 2002, Kelly and Rizzino, 2000, Ramalho-Santos *et al.*, 2002, Sato *et al.*, 2003), *Ccne1* (Ivanova *et al.*, 2002, Kelly and Rizzino, 2000, Ramalho-Santos *et al.*, 2002), *Ccne2*, *Mcm3* (Sato *et al.*, 2003), (Bowman *et al.*, 2006, Rao and Stice, 2004, Venezia *et al.*, 2004), *Galanin* (Bhattacharya *et al.*, 2004), *Calumenin* (Bhattacharya *et al.*, 2004), *Pitx2* (Bhattacharya *et al.*, 2004, Ramalho-Santos *et al.*, 2002, Rao and Stice, 2004) and *Pttg1* (Bhattacharya *et al.*, 2004); (4) Metabolic pathways, e.g., *Mthfd2* (Bhattacharya *et al.*, 2004, Rao and Stice, 2004), *Tk1* (Bhattacharya *et al.*, 2004, Rao and Stice, 2004), *Serpinh1* (Bhattacharya *et al.*, 2004), *Cct8* (Bhattacharya *et al.*, 2004) and *Tubb4* [42]; (5) Growth factors / Receptors, e.g., *Epha1* (Sperger *et al.*, 2003); (6) Cell adhesion / Membrane proteins, e.g., *Lamr1* (Bhattacharya *et al.*, 2004, Ivanova *et al.*, 2002, Kelly and Rizzino, 2000, Rao and Stice, 2004), *Itga6* (Abeyta *et al.*, 2004, Ivanova *et al.*, 2002, Ramalho-Santos *et al.*, 2002, Rao and Stice, 2004) *Anxa2* (Kimura *et al.*, 2001); and (7) Miscellaneous, e.g., *Ppic* (Abeyta *et al.*, 2004, Ivanova *et al.*, 2002, Ramalho-Santos *et al.*, 2002, Rao and Stice, 2004), *Laptmb4* (Abeyta *et al.*, 2004, Ivanova *et al.*, 2002, Ramalho-Santos *et al.*, 2002, Rao and Stice, 2004) and *Arcn1* (Abeyta *et al.*, 2004, Ivanova *et al.*, 2002, Ramalho-Santos *et al.*, 2002, Rao and Stice, 2004). While many known markers of ES cells were detected in our arrays, other genes that

Fig. 4. Histograms showing the GO terms differentially expressed in undifferentiated stem cells and 3- and 7-day-old EBs. **(A)** *Histogram of stem cells;* **(B)** *Histogram of 3-day-old EBs;* **(C)** *Histogram of 7-day-old EBs.*

3-EB   7-EB
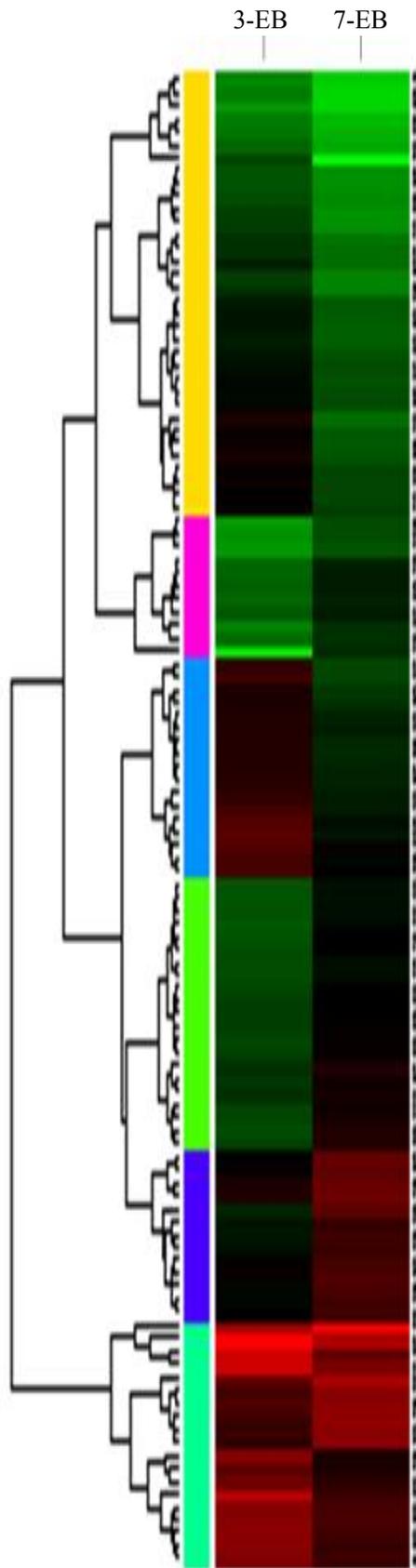
Gene Name	Description
Cluster 1 (22 genes)	
5031439A09Rik	RIKEN cDNA 5031439A09 gene (5031439A09Rik)
Col1a1	procollagen, type I, alpha 1 (Col1a1)
Shoc2	soc-2 (suppressor of clear) homolog (C elegans) (Shoc2)
Ctla2a	cytotoxic T lymphocyte-associated protein 2 alpha (Ctla2a)
Col3a1	procollagen, type III, alpha 1 (Col3a1)
Nrp1	neuropilin 1 (Nrp1)
Slc16a1	Mmusculus mRNA for monocarboxylate transporter
Pmp22	peripheral myelin protein (Pmp22)
F2r	coagulation factor II (thrombin) receptor (F2r)
2610001E17Rik	RIKEN cDNA 2610001E17 gene (2610001E17Rik)
Vldlr	very low density lipoprotein receptor (Vldlr)
Dlk1	delta-like 1 homolog (Drosophila) (Dlk1)
Hiatl1	RIKEN cDNA 5730414C17 gene (5730414C17Rik)
Slit2	slit homolog 2 (Drosophila) (Slit2)
Matr3	matrin 3, mRNA (cDNA clone MGC:28206 IMAGE:3989914)
Mrg1	myeloid ecotropic viral integration site-related gene 1 (Mrg1)
Hmga2	high mobility group AT-hook 2 (Hmga2)
Sfrs2	splicing factor Sc35 (Pr264) mRNA, partial cds, alternatively spliced
Tgfbi	transforming growth factor, beta induced (Tgfbi)
App	hippocampal amyloid precursor protein mRNA
D0H4S114	DNA segment, human D4S114 (D0H4S114)
Rps2	ui64g11x1 Sugano mouse liver mlia cDNA clone IMAGE:1887236 3' similar to gb:X81582 M.musculus mRNA for insulin-like growth factor binding (MOUSE).
Cluster 2 (18 genes)	
Nrk	Nik related kinase (Nrk)
Tbx2	T-box 2 (Tbx2)
Amot	angiomotin (Amot)
2810051F02Rik	RIKEN cDNA 2810051F02 gene (2810051F02Rik)
Calcb	calcitonin-related polypeptide, beta (Calcb)
Slco1a1	solute carrier organic anion transporter family, member 1a1 (Slco1a1)
Msx2	homeo box, msh-like 2 (Msx2)
Msr2	Fc receptor-like protein 2 immunoglobulin short isoform (Fcrh2) mRNA
Kdr	Mmusculus Flk-1 mRNA
BY759545.1	RIKEN 13 days embryo liver cDNA clone I920036M10 3'
1700018C11Rik	RIKEN cDNA 1700018C11 gene (1700018C11Rik)
Capn6	calpain 6 (Capn6)
Alcam	UI-M-BH3-asm-e-08-0-Uls1 NIH_BMAP_M_S4 cDNA clone UI-M-BH3-asm-e-08-0-UI 3'.
Msx1	homeo box, msh-like 1 (Msx1)
Atoh8	atonal homolog 8 (Drosophila) (Atoh8)
Pyy	peptide YY (Pyy)
Prss12	protease, serine, 12 neurotrypsin (motopsin) (Prss12)
W71826.1	me45a08r1 Soares mouse embryo NbME13.5 14.5 cDNA clone IMAGE:390422 5'.
Cluster 3 (18 genes)	
Hbb	mb84h05r1 Soares mouse p3NMF19.5 cDNA clone IMAGE:336153 5' similar to gb:V00497 HEMOGLOBIN BETA CHAIN (HUMAN); gb:J00413 Mouse beta-globin major gene (MOUSE).
Igfbp5	insulin-like growth factor binding protein 5 (Igfbp5), mRNA
8030466E21Rik	unc-5 homolog C (C elegans) (Unc5c)
Fgb	fibrinogen, B beta polypeptide (Fgb)
Tlx1	T-cell leukemia, homeobox 1 (Tlx1)
Hbb-bh1	AGENCOURT_10124687 NIH_MGC_134 cDNA clone IMAGE:6512013 5'
Krt2-1	keratin complex 2, basic, gene 1 (Krt2-1)
Tdo2	tryptophan 2,3-dioxygenase (Tdo2)
Hba-x	hemoglobin X, alpha-like embryonic chain in Hba complex (Hba-x)
Comt	ma46f02r1 Soares mouse p3NMF19.5 cDNA clone IMAGE:313755 5' similar to gb:X02162 APOLIPOPROTEIN A-I PRECURSOR (HUMAN);.
Rhox6	placenta specific homeobox 1 (Psx1)
Rspo3	thrombospondin, type I, domain 2 (Thsd2)
BQ032190.1	UI-1-CF0-apq-b-05-0-Uls1 NCI_CGAP_PITr1 cDNA clone UI-1-CF0-apq-b-05-0-UI 3'
Slc13a4	solute carrier family 13 (sodium/sulfate symporters), member 4 (Slc13a4)
Isl1	ISL1 transcription factor, LIM/homeodomain (islet 1) (Isl1)
Igf2	insulin-like growth factor 2 (Igf2)
Hapln1	cartilage link protein (Crtl1) mRNA
Myct1	myc target 1 (Myct1)
Cluster 4 (16 genes)	
Zfp207	RIKEN pooled tissues, spleen and thymus lymphocytes cDNA clone I6C0068G10 3'
Serpinb12	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 12 (Serpinb12)
Acsm1	butyryl Coenzyme A synthetase 1, mRNA (cDNA clone MGC:25954 IMAGE:4238964)
Olfr1403	olfactory receptor 1403 (Olfr1403)
Fabp1	fatty acid binding protein 1, liver (Fabp1)
Cyp1a2	cytochrome P450, family 1, subfamily a, polypeptide 2 (Cyp1a2)
Ulk2	RIKEN 0 day neonate lung cDNA clone E030037L18 3'
Olfr954	olfactory receptor 954 (Olfr954)
Ush3a	Usher syndrome 3A homolog (human) (Ush3a), mRNA
AI507531.1	vj76h02x1 Knowles Solter mouse blastocyst B1 cDNA clone IMAGE:934995 3'.
BY103204.1	RIKEN pooled tissues, adult spleen, etc cDNA clone K630149L12 5'
Slc12a4	RIKEN 0 day neonate kidney cDNA clone D630040N05 3'
Fgfbp1	fibroblast growth factor binding protein 1 (Fgfbp1)
AV240709.2	RIKEN 10 day neonate skin cDNA clone 4732442L07 3'
Slc15a4	RIKEN B16 F10Y cells cDNA clone G370092P18 3'
Chrnb2	12 days embryo spinal ganglion cDNA, RIKEN clone:D130070I21 product:cholinergic receptor, nicotinic, beta polypeptide 2 (neuronal), full insert sequence
Cluster 5 (16 genes)	
Gabrg1	RIKEN adult male hypothalamus cDNA clone A230092H19 3'
BB260391.1	RIKEN 7 days neonate cerebellum cDNA clone A730094F18 3'
Psmd11	adult male testis cDNA, RIKEN clone:1700089D09 product:unclassifiable, full insert sequence
Klhl21	10 day old male pancreas cDNA, RIKEN clone:1810045K06 product:hypothetical Galactose oxidase, central domain structure containing protein, full insert sequence
Map4k5	adult male testis cDNA, RIKEN clone:4931403G20 product:unknown EST, full insert sequence
Tm4sf4	transmembrane 4 superfamily member 4 (Tm4sf4)
Srgap2	UI-M-BH3-awc-c-12-0-Uls4 NIH_BMAP_M_S4 cDNA clone UI-M-BH3-awc-c-12-0-UI 3'.
BB668107.1	RIKEN 2 days neonate thymus thymic cells cDNA clone C920025J10 3'
BY755242.1	RIKEN RCB-1283 B16 melanoma cDNA cDNA clone G430107I07 3'
Rhbg	Rhesus blood group-associated B glycoprotein (Rhbg)
Dspg3	dermatan sulphate proteoglycan 3 (Dspg3)
2900024J01Rik	adult male hippocampus cDNA, RIKEN clone:2900024J01 product:unknown EST, full insert sequence
Ltb4r2	leukotriene B4 receptor 2 (Ltb4r2)
BB156955.1	RIKEN 16 days neonate thymus cDNA clone A130034B11 3', mRNA sequence
CA456872.1	AGENCOURT_10737396 NIH_MGC_152 cDNA clone IMAGE:6746735 5'
3000003F02Rik	RIKEN cDNA 3000003F02 gene (3000003F02Rik)
Cluster 6 (13 genes)	
Tcf15	transcription factor 15 (Tcf15)
AW047251.1	UI-M-BH1-amb-h-10-0-UI.s1 NIH_BMAP_M_S2 cDNA clone UI-M-BH1-amb-h-10-0-UI 3', mRNA sequence
Nnmt	nicotinamide N-methyltransferase (Nnmt)
Raly	(clones EcDNA1 and EcDNA4) RNA-binding protein mRNA sequence
Mmp3	matrix metalloproteinase 3 (Mmp3)
Ccl7	chemokine (C-C motif) ligand 7 (Ccl7)
Cbr3	carbonyl reductase 3 (Cbr3)
Ifit3	interferon-induced protein with tetratricopeptide repeats 3 (Ifit3)
Gdap1l1	ganglioside-induced differentiation-associated protein 1-like 1 (Gdap1l1)
Pou3f1	POU domain, class 3, transcription factor 1 (Pou3f1)
Mdfi	MyoD family inhibitor (Mdfi)
Ptpn11	protein tyrosine phosphatase SH-PTP2
Ltb	lymphotoxin B (Ltb)

Fig. 5. The hierarchical clustering depicted graphically in colour.

were more highly expressed in MES3 cell line than in adult cells have not been reported in previous gene expression studies of other mouse and human mESC lines. Taking these data together, we can hypothesize that the detailed analysis of these genes and known ESC markers may reveal new factors involved in the pluripotent and self-renewal state. However, the results should be interpreted with caution, since the ES cell expression profiling revealed a large percentage of uncharacterized genes.

Gene expression in differentiation of EBs from MES3 cells was monitored performing gene expression microarrays on mRNA isolated from 3- and 7-day-old EBs. Significantly more genes were expressed during the early development of ESCs than in the adult cells. As expected, most of the mESC markers were down-regulated during the initial period of differentiation and were significantly sub-expressed in 3- and 7-day-old EBs. When compared to adult tissue cells, gene expression analysis of 3-day-old EBs showed 8105 genes differentially expressed: 6065 up-regulated and 2040 down-regulated genes. The gene expression pattern found in 7-day-old EB mRNA was similar to that in 3-day-old EBs. There were 9198 differentially expressed genes in 7-day-old EBs, of which 6995 were significantly up-regulated and 2243 significantly down-regulated. The general direction of the expression profile was up-regulation in both 3- and 7-day-old EBs (vs. adult tissue cells and using ESC in undifferentiated state as reference). The fact that the majority of genes with a significant change in expression were up-regulated may help to elucidate candidate genes participating in the underlying signalling pathways that control the early differentiation and maturation of mESCs. However, it might be important to note that EBs are composed by many cell types that might be reflected by a higher gene expression pattern in comparison with mESCs.

The 3-day-EB cells contained genes related to growth and cell cycle control; *Mcm5* gene, which regulates cell growth and is involved in M/G1 and G2/M cell cycle check points and in the assembly of the pre-replicative complex (Kimura *et al.*, 2001); *Ccnd2,* related to cell cycle control and P53 signalling (>5-fold higher expression) (Furusawa *et al.*, 2006); *Prim1,* related to S/G1 cell cycle check point and DNA replication initiation (>5-fold higher expression); *Fst,* related to cell proliferation and growth (Skottman *et al.*, 2006); *Txndc1*, which participates in the induction of cell growth and proliferation through apoptosis inhibition (>5-fold higher expression) (Harkness *et al.*, 2008); and *Tk1*, related to nucleotide metabolism and DNA replication. A class of three related genes, *Lama1, Lamb1-1* and *Lamc1,* were also up-regulated at the start of differentiation. These genes are extracellular matrix receptors directly related to local cell adhesion and cell communication processes, and they encode laminin-111 (Li *et al.*, 2004, Li *et al.*, 2001), involved in the development of the ectoderm (Miner *et al.*, 2004, Scheele *et al.*, 2005). *Col4a2*, another extracellular matrix receptor, was also up-regulated, as were several genes associated with: amino acid metabolism, e.g., *Txndc12, Wars, Srm, Plod1* and *Plod2;* carbohydrate and/or lipid metabolisms, e.g., *Hs3st1*, *Hexa*, *Ggta1, Gale, Pla2g12b* and *Asah1*; and complement and coagulation cascades, e.g., *F3* and *Tfpi.*

The set of 2040 genes that were down-regulated in the 3-day-old EBs included genes related to n amino acid biosynthesis and/or purine metabolism (*Abp1, Nt5c1b* and *hpb1*), genes needed for carbohydrate utilization (*Idhc*, *Plcz1* and *Acsm1*) and genes related to lipid and steroid metabolism (*Hsd3b4, Cyp17a1, Cyp4b1,*

Cyp2e1, Cyp2f2 and *Cyp2d26*). Microarray analysis also revealed significant inactivation of a class of related genes *Spata3*, *Spata4*, *Spata16*, *Spata19*, *Spata20, Cst8, Cst12, Cst13, Svs5, Spag9, Spag4l, Tnp1* and *Tesp1*, which are associated with spermatogenesis, spermatogenesis cell apoptosis, testicular function and spermatocyte development. Interestingly, we also discovered a strong down-regulation in the expression of *Hils1* and *H1fnt* genes, which are essential for nuclear condensation during spermatogenesis (>6-fold change) through participation in the replacement of hystones with protamines (Sup. Table 3). Fig. 4B depicts the histogram of GO terms significantly overrepresented in 3-day-old EBs.

The subset of the most significantly upregulated genes in the 7-day-old EBs included genes involved in amino acid metabolism (*Plod1* and *Plod 2*), local cell adhesion and cell communication (*Lama1*, *Lamb1-1* and *Lamc1*), carbohydrate and/or lipid metabolism (*Hexa*, *Ggta1, Gale* and *Asah1*) and complement and coagulation cascades (*F3* and *Tfpi*). Interestingly, only a few genes had a higher expression than in 3-day-old EBs, suggesting that the proliferation and growth processes triggered during the formation of 3-day-old EBs remain activated after 7 days of spontaneous differentiation. Indeed, our microarray results indicate that several genes strengthen or complete the action of previously activated genes. Thus, *Mcm5 remained* activated in EBs at 7-days, when *Mcm2, Mcm3 and Mcm4* were also up-regulated. *Mcm2, Mcm3* and *Mcm4* belong to the same class of genes as *Mcm5*, i.e., Mcm family genes, and are involved in initiating replication through recruitment of DNA replication-related proteins. We can, therefore, hypothesise that first *Mcm5* and then *Mcm2, Mcm3* and *Mcm4* are key genes in cell proliferation promotion during early differentiation stages of ESCs to EBs. As expected, we found up-regulation of some other genes related to cell cycle and proliferation, including cyclin D2 (*ccnd2*) and cell division cycle 6 (*cdc6*), responsible for loading Mcm proteins onto DNA during DNA replication. We also recorded a strong up-regulation of genes related to cell growth, proliferation and differentiation: *Adamts9* gene, associated with the control of organ shape during early development (Clark *et al.*, 2000); *Cts1 and Ctgf*, genes, involved in cell migration and recruitment of mesenchymal stem cells (Luo *et al.*, 2004); *Pik3cb* gene, in the ErbB family of receptor tyrosine kinases (RTKs) involved in signalling pathways regulating different biological responses, including proliferation, differentiation, cell motility and survival (Citri and Yarden, 2006); *Ncl,* the main major nucleolar protein of growing eukaryotic cells and associated with the induction of chromatin decondensation by binding to histone H1 (Erard *et al.*, 1988); and *Loxl2,* related to the biogenesis of connective tissue (Smith-Mungo and Kagan, 1998). We also found some typical markers of undifferentiation such as *Fbxo15* and *Lin28*. The similarities found in the gene expression profile of 3- and 7-day-old EBs indicate that the gene expression analysis at shorter differentiation times would be needed to characterize the molecular mechanisms underlying ESC pluripotency and self-renewal. Our results are in concordance with those described by Palmqvist *et al.,* where they demonstrated that the main gene expression changes occur within the first 24 hours of spontaneous differentiation.

Because stem cell-derived assays have shown low confirmation rates and reproducibility (Evsikov and Solter, 2003, Fortunel *et al.*, 2003, Ivanova *et al.*, 2002), we tested the replicability of the gene expression results in a second mESC line, CGR8. A set of 867 genes were differentially and commonly expressed in both

Fig. 6. Significantly overrepresented GO terms for genes differentially expressed in 3 and 7-day-old EBs.

MES3 and CGR8 mESC lines but not in adult cells, and 1587 and 1791 genes were differentially and commonly expressed in 3 and 7 days-old EBs, respectively, in both mESCs but not in adult cells. Genes in these categories included various types of collagens or related proteins (*Col1a1, Col3a1* and *Tgfbi*) and other extracellular matrix components (*Matrin3* and *Mmp3*), proteins involved in neural and/or neuromuscular development (*Pou3f1, Nrp1, Pmp22, Nlk1, Slit2, App, Alcam, Atho8, Prss12* and *Gdap1l1*), blood coagulation factors (*F2r*), transporters (*Slc16a1* and *Vldlr*) and a variety of transcriptional regulating factors involved in tissue development and morphogenesis (*Pou3f1, Tcf15, Meis1, Hmga2, Tbx2, Msx1, Msx2, Kdr, Capn6, Mdfi* and *Ptpn11*).

The high variability found in previous gene expression studies of stem cells may result from the difficulty in standardizing culture conditions. The most feasible explanation for the variability is that identical culture conditions may have distinct effects on the growth of different ES cell lines. We grew our cell lines in the same laboratory under identical culture conditions, suggesting that our findings may reflect underlying biological differences between the cell lines rather than variations in the response to the culture environment.

In summary, our study provides gene expression profiling of
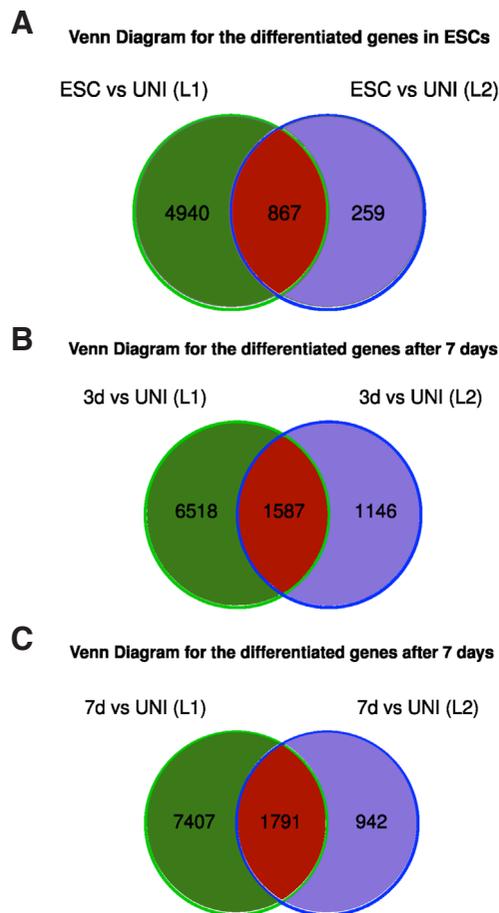
Fig. 7. Overlapping gene expression in MES3 and CGR8 murine ESCs. **(A)** *Venn diagram detailing shared and distinct gene expression between MES3 and CGR8 cells.* **(B)** *Venn diagram detailing shared and distinct gene expression between 3-day-old EBs in the two stem cell lines.* **(C)** *Venn diagram detailing shared and distinct gene expressions between 7-day-old EBs in the two stem cell lines.*

two mouse stem cell lines at baseline state and during their differentiation towards EB formation. Our findings were similar to previous reports in human and mouse stem cells, demonstrating that our approach yields comparable results to those obtained using a different technique. However, we also identified numerous other genes not been previously reported, which may offer novel insights into the behaviour of mouse ESC, the maintenance of their undifferentiated state and the mechanisms underlying their pluripotency, self-renewal and early differentiation.

Materials and Methods

Mouse embryonic stem cell cultures

MES3 mESC line was derived by whole blastocyst culture from the inner cell mass of a C57BLxCBA mouse embryo (Cortes *et al.*, 2008). Culture cells were maintained on Mitomicin C inactivated mouse embryonic fibroblasts (MEF) in knock-out DMEM (Gibco, Invitrogen) supplemented with 10% FBS (Gibco, Invitrogen), 2mM non-essential amino acids, 2 mM L-glutamine, 50 µg/mL Penn-Strep, 10 nM β-mercaptoethanol and 1000 units/mL LIF (ESGRO, Chemicon) (Fig. 1A). Culture medium was refreshed daily. Subculture was performed by mechanical excision and disruption of undifferentiated colonies, transferring these cell clusters to a new culture plate.

Generation of embryoid bodies

EBs were obtained by excising clusters of undifferentiated stem cells from growing colonies and transferring them to low-adhesion culture plates, where they were maintained in the same medium as the mESCs but without β-mercaptoethanol and LIF. EBs were harvested for extraction of total RNA at days 1, 2, 3, 5, 7, 9, 15 and 20 (Fig. 1D). Cell cultures were continuously tested to see whether cells exhibited ESC properties. Growing and subculturing continued for several months, ensuring that the cells were capable of long-term self-renewal. The healthy and undifferentiated state of the cells was tested by inspecting the cultures under a microscope. We also tested whether the cells could be subcultured after freezing, thawing and replating. Finally, we determined whether the mESCs were pluripotent by allowing them to differentiate spontaneously in cell culture to form EBs.

RNA isolation

Total RNA was isolated from well-characterized mouse ES cell lines (MES3 and CGR8) and 3- and 7-day-old EBs by using the Trizol® method (Invitrogen). All samples were treated with an RNase-free DNase set (Qiagen, Valencia, CA). The quality of total RNA was analyzed by standard agarose electrophoresis and capillary electrophoresis analysis using the Experion Automated Electrophoresis System (Bio-Rad Laboratories). All RNA samples had a 28S/18S ratio >1.8 and were tested in duplicate.

Total mouse universal RNA (mURNA) isolated from a collection of adult mouse tissues to represent a broad range of expressed genes (Ambion) was used as a universal reference control in the competitive hybridization.

Phenotypic characterization of cells

The undifferentiated state of cultures was established by using the AP detection kit (Chemicon) to stain colonies for alkaline phosphatase activity, following the manufacturer's instructions (Fig. 1B). Additionally, a specific monoclonal antibody was used to detect expression of Stage Specific Embryonic Antigen-4 (*Ssea-4*) by immunostaining (Fig. 1C). The quality of the MES3 and CGR8 cells cultured in our laboratory was also confirmed by evaluating the expression of several markers expressed by undifferentiated ESCs (Ben-Shushan *et al.*, 1998, Singh *et al.*, 2007, Thompson and Gudas, 2002).

We monitored the differentiation process during EBs culture by measuring the expression level of different markers, using real-time (RT)-PCR to evaluate relative changes in their intracellular concentration. *Nanog*, *Oct3/4* and *Rex1* were used as markers for the undifferentiated state; *Hnf4* and

Ttr as endoderm markers, and *Fgf5* and *Wnt3* as mesoderm and ectoderm markers respectively. Data were normalized with reference to β-*actin* levels.

cDNAs were obtained by using the Improm-II™ Reverse Transcription System (Promega) in the presence of 2 μg total RNA, 0.5 μg Oligo (Waterston *et al.*), 15 primer, 1U/μl RNase inhibitor and 3mM MgCl₂ in a final volume of 25 μl. After thermal inactivation of the enzyme at 70u61904C for 15 min, RT-PCR was performed with the SYBRGreen quantitec PCR master mix (Qiagen, Valencia) in the presence of 1 μl of the RT reaction and 1 μM of primers (Sup. Table 1). Expression levels of markers were determined at days 0, 1, 2, 5, 9, 15 and 20. Results were represented as $1/\Delta C_T$ versus time, where ΔC_T refers to normalized C_T calculated by subtracting the C_T of a control (β-actin) from the C_T of the gene of interest.

Microarray procedure: target preparation, hybridization, post-hybridization and scanning

DNA microarray analyses were performed using CodeLink Uniset Mouse I 20K Bioarray System (Amersham Biosciences/GE Healthcare, Piscataway, NJ, USA), which offers the expression profiling of ~20.000 well-annotated mouse genes in a single bioarray. CodeLink bioarray experiments were carried out following the recommended Amersham Biosciences CodeLink protocol (Ramakrishnan *et al.*, 2002). After the hybridation step, microarray slides were scanned in Cy5 (635 nm) channel using an Axon GenePix 4000B scanner (Axon Instruments, Foster City, CA) with 10 μM resolution. Scanned microarray images were exported as TIFF files to CodeLink™ Expression Analysis Software version 5.0 from Applied Microarrays.

Data preprocessing

Preprocessing of raw probe intensities is a key procedure in the analysis of gene expression microarray data. Briefly, background correction and normalization are needed to prevent from the effect of variations in experimental conditions. The raw intensity data for each probe was collected with CodeLink™ Expression Analysis Software version 5.0. For the preprocessing process we considered the raw data and computed them by means of the well-known R package limma. As we used one-color microarrays in our study, the preprocessing process consisted of three separated steps:

i) *Background correction of the raw data.* For this task we applied the normexp+offset method, computing the maximum likelihood by means of the saddle-point approximation introduced by Ritchie *et al.,* (Ritchie *et al.*, 2007). We selected an offset value of 50 in order to stabilize the log-ratio computations at low intensities. This methodology has been proved to provide much better results than the usual background subtraction.

ii) *Normalization process.* We normalized the background corrected data by applying the quantile normalization method. The quantile method is designed to make the distribution of intensities for each array in a set of arrays the same, and it has been shown to perform better than other related methods (Bolstad *et al.*, 2003).

iii) *Expression matrix generation.* For this matter, we averaged the replicates of each sample in order to obtain a single intensity value for each gene in each sample. Therefore, the obtained expression matrix is comprised of as many columns as samples in the experiment, and as many rows as genes in the array.

Gene expression analysis

As a first step in the analysis of the preprocessed data we identified the differentially expressed genes present in the experiments. These genes were identified by applying the *two*-class unpaired method from the program Significance Analysis of Microarrays (SAM) (Tusher *et al.*, 2001). SAM ranks genes based on a modified t-test statistic. Higher scores imply a more significant differential expression of the considered genes. This statistic improves traditional fold change or t-test approaches, as was stated in (Jones and Arvin, 2003, Smyth *et al.*, 2003). One of the main advantages of using SAM is that it estimates a global false discovery rate (FDR), defined as the fraction of genes that are likely to be incorrectly identified significantly regulated. Furthermore, SAM provides a parameter called delta that enables to vary the ratio of true positives to predicted false positives. For each experiment we selected a delta value that provided an optimal ratio of true positives to false positives, obtaining lists of significantly differentiated genes that were used for further analysis.

Cluster and gene ontology analysis

Functional information on the differentially expressed genes was obtained by examining their Gene Ontology (GO) annotations. Gene annotations were determined and compiled at several ontology levels. The Gene Ontology (GO) Consortium provides a structured, controlled vocabulary for describing roles of genes and gene products in organisms, which has become a *de facto* standard for gene description in the scientific community (Ashburner *et al.*, 2000). We used the GO data to explore the biological meaning of the genes selected in each experiment. GO annotations are widely used in microarray analysis in an attempt to reveal statistically significant biological processes or pathways among co-regulated genes. For our purposes, we used a recently published tool, csbl.go (Ovaska *et al.*, 2008), implemented as an R package. This tool facilitates the search for hypothetical gene functions in cells by computing GO similarities among genes and rapidly identifying and visualizing genes that share the same GO cluster. We established similarities among genes using the Resnik measure (Resnik, 1995) and obtained a hierarchical cluster from the resulting similarity matrix. We also analyzed sets of genes with their GO annotations to compute the most enriched GO terms by using the well-known FatiGo tool (Al-Shahrour *et al.*, 2004).

Validation of microarray data

Microarrays data from MES3 cell line were validated in a second mouse stem cell line, CGR8. This mouse stem cell line was obtained from American Type Culture Collection (ECACC, UK; http://www.atcc.org) for the validation of assay results. Additionally, expression data for specific biomarkers were validated by qPCR. In order to verify in part the differential expression of selected genes, we performed qPCR reactions in duplicate (50 μl) as previously reported (Furusawa *et al.*, 2004, Jeong *et al.*, 2007, Zhao *et al.*, 2005). Selected genes were highly expressed and exhibited previously demonstrated development-related differences (Furusawa *et al.*, 2004, Jeong *et al.*, 2007, Zhao *et al.*, 2005).

References

ABEYTA, M.J., CLARK, A.T., RODRIGUEZ, R.T., BODNAR, M.S., PERA, R.A. and FIRPO, M.T. (2004). Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum Mol Genet* 13: 601-608.

AL-SHAHROUR, F., DIAZ-URIARTE, R. and DOPAZO, J. (2004). FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* 20: 578-580.

ASHBURNER, M., BALL, C.A., BLAKE, J.A., BOTSTEIN, D., BUTLER, H., CHERRY, J.M., DAVIS, A.P., DOLINSKI, K., DWIGHT, S.S., EPPIG, J.T. *et al.,* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25-29.

BASSETT, D.E., JR., EISEN, M.B. and BOGUSKI, M.S. (1999). Gene expression informatics--it's all in your mine. *Nat Genet* 21: 51-55.

BEN-SHUSHAN, E., THOMPSON, J.R., GUDAS, L.J. and BERGMAN, Y. (1998). Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. *Mol Cell Biol* 18: 1866-1878.

BHATTACHARYA, B., MIURA, T., BRANDENBERGER, R., MEJIDO, J., LUO, Y., YANG, A.X., JOSHI, B.H., GINIS, I., THIES, R.S., AMIT, M. *et al.,* (2004). Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* 103: 2956-2964.

BOLSTAD, B.M., IRIZARRY, R.A., ASTRAND, M. and SPEED, T.P. (2003). A compari-

son of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19: 185-193.

BOWMAN, T.V., MCCOOEY, A.J., MERCHANT, A.A., RAMOS, C.A., FONSECA, P., POINDEXTER, A., BRADFUTE, S.B., OLIVEIRA, D.M., GREEN, R., ZHENG, Y. *et al.,* (2006). Differential mRNA processing in hematopoietic stem cells. *Stem Cells* 24: 662-670.

CITRI, A. and YARDEN, Y. (2006). EGF-ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol* 7: 505-516.

CLARK, M.E., KELNER, G.S., TURBEVILLE, L.A., BOYER, A., ARDEN, K.C. and MAKI, R.A. (2000). ADAMTS9, a novel member of the ADAM-TS/ metallospondin gene family. *Genomics* 67: 343-350.

CORTES, J.L., SANCHEZ, L., CATALINA, P., COBO, F., BUENO, C., MARTINEZ-RAMIREZ, A., BARROSO, A., CABRERA, C., LIGERO, G., MONTES, R. *et al.,* (2008). Whole-blastocyst culture followed by laser drilling technology enhances the efficiency of inner cell mass isolation and embryonic stem cell derivation from good- and poor-quality mouse embryos: new insights for derivation of human embryonic stem cell lines. *Stem Cells Dev* 17: 255-267.

ERARD, M.S., BELENGUER, P., CAIZERGUES-FERRER, M., PANTALONI, A. and AMALRIC, F. (1988). A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. *Eur J Biochem* 175: 525-530.

EVSIKOV, A.V. and SOLTER, D. (2003). Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* 302: 393; author reply 393.

FORTUNEL, N.O., OTU, H.H., NG, H.H., CHEN, J., MU, X., CHEVASSUT, T., LI, X., JOSEPH, M., BAILEY, C., HATZFELD, J.A. *et al.,* (2003). Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* 302: 393; author reply 393.

FURUSAWA, T., IKEDA, M., INOUE, F., OHKOSHI, K., HAMANO, T. and TOKUNAGA, T. (2006). Gene expression profiling of mouse embryonic stem cell subpopulations. *Biol Reprod* 75: 555-561.

FURUSAWA, T., OHKOSHI, K., HONDA, C., TAKAHASHI, S. and TOKUNAGA, T. (2004). Embryonic stem cells expressing both platelet endothelial cell adhesion molecule-1 and stage-specific embryonic antigen-1 differentiate predominantly into epiblast cells in a chimeric embryo. *Biol Reprod* 70: 1452-1457.

HAILESELLASSE SENE, K., PORTER, C.J., PALIDWOR, G., PEREZ-IRATXETA, C., MURO, E.M., CAMPBELL, P.A., RUDNICKI, M.A. and ANDRADE-NAVARRO, M.A. (2007). Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* 8: 85-105.

HARKNESS, L., CHRISTIANSEN, H., NEHLIN, J., BARINGTON, T., ANDERSEN, J.S. and KASSEM, M. (2008). Identification of a membrane proteomic signature for human embryonic stem cells independent of culture conditions. *Stem Cell Res* 1: 219-227.

HEMBERGER, M., CROSS, J.C., ROPERS, H.H., LEHRACH, H., FUNDELE, R. and HIMMELBAUER, H. (2001). UniGene cDNA array-based monitoring of transcriptome changes during mouse placental development. *Proc Natl Acad Sci USA* 98: 13126-13131.

HOWBROOK, D.N., VAN DER VALK, A.M., O'SHAUGHNESSY, M.C., SARKER, D.K., BAKER, S.C. and LLOYD, A.W. (2003). Developments in microarray technologies. *Drug Discov Today* 8: 642-651.

IVANOVA, N.B., DIMOS, J.T., SCHANIEL, C., HACKNEY, J.A., MOORE, K.A. and LEMISCHKA, I.R. (2002). A stem cell molecular signature. *Science* 298: 601-604.

JEONG, C.H., LEE, H.J., CHA, J.H., KIM, J.H., KIM, K.R., KIM, J.H., YOON, D.K. and KIM, K.W. (2007). Hypoxia-inducible factor-1 alpha inhibits self-renewal of mouse embryonic stem cells *in vitro* via negative regulation of the leukemia inhibitory factor-STAT3 pathway. *J Biol Chem* 282: 13672-13679.

JONES, J.O. and ARVIN, A.M. (2003). Microarray analysis of host cell gene transcription in response to varicella-zoster virus infection of human T cells and fibroblasts *in vitro* and SCIDhu skin xenografts in vivo. *J Virol* 77: 1268-1280.

KELLY, D.L. and RIZZINO, A. (2000). DNA microarray analyses of genes regulated during the differentiation of embryonic stem cells. *Mol Reprod Dev* 56: 113-123.

KIM, S.K., LUND, J., KIRALY, M., DUKE, K., JIANG, M., STUART, J.M., EIZINGER, A., WYLIE, B.N. and DAVIDSON, G.S. (2001). A gene expression map for Caenorhabditis elegans. *Science* 293: 2087-2092.

KIMURA, S.H., IKAWA, M., ITO, A., OKABE, M. and NOJIMA, H. (2001). Cyclin G1 is involved in G2/M arrest in response to DNA damage and in growth control after

damage recovery. *Oncogene* 20: 3290-3300.

KORBLING, M. and ESTROV, Z. (2003). Adult stem cells for tissue repair - a new therapeutic concept? *N Engl J Med* 349: 570-582.

KRAMER, J., STEINHOFF, J., KLINGER, M., FRICKE, L. and ROHWEDEL, J. (2006). Cells differentiated from mouse embryonic stem cells via embryoid bodies express renal marker molecules. *Differentiation* 74: 91-104.

KUO, H.C., PAU, K.Y., YEOMAN, R.R., MITALIPOV, S.M., OKANO, H. and WOLF, D.P. (2003). Differentiation of monkey embryonic stem cells into neural lineages. *Biol Reprod* 68: 1727-1735.

LAGASSE, E., SHIZURU, J.A., UCHIDA, N., TSUKAMOTO, A. and WEISSMAN, I.L. (2001). Toward regenerative medicine. *Immunity* 14: 425-436.

LEE, S.H., LUMELSKY, N., STUDER, L., AUERBACH, J.M. and MCKAY, R.D. (2000). Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. *Nat Biotechnol* 18: 675-679.

LI, L., ARMAN, E., EKBLOM, P., EDGAR, D., MURRAY, P. and LONAI, P. (2004). Distinct GATA6- and laminin-dependent mechanisms regulate endodermal and ectodermal embryonic stem cell fates. *Development* 131: 5277-5286.

LI, X., TALTS, U., TALTS, J.F., ARMAN, E., EKBLOM, P. and LONAI, P. (2001). Akt/PKB regulates laminin and collagen IV isotypes of the basement membrane. *Proc Natl Acad Sci USA* 98: 14416-14421.

LIU, X.L., YUAN, J.Y., ZHANG, J.W., ZHANG, X.H. and WANG, R.X. (2007). Differential gene expression in human hematopoietic stem cells specified toward erythroid, megakaryocytic, and granulocytic lineage. *J Leukoc Biol* 82: 986-1002.

LUMELSKY, N., BLONDEL, O., LAENG, P., VELASCO, I., RAVIN, R. and MCKAY, R. (2001). Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science* 292: 1389-1394.

LUO, Q., KANG, Q., SI, W., JIANG, W., PARK, J.K., PENG, Y., LI, X., LUU, H.H., LUO, J., MONTAG, A.G. *et al.,* (2004). Connective tissue growth factor (CTGF) is regulated by Wnt and bone morphogenetic proteins signaling in osteoblast differentiation of mesenchymal stem cells. *J Biol Chem* 279: 55958-55968.

MINER, J.H., LI, C., MUDD, J.L., GO, G. and SUTHERLAND, A.E. (2004). Compositional and structural requirements for laminin and basement membranes during mouse embryo implantation and gastrulation. *Development* 131: 2247-2256.

NAKATAKE, Y., FUKUI, N., IWAMATSU, Y., MASUI, S., TAKAHASHI, K., YAGI, R., YAGI, K., MIYAZAKI, J., MATOBA, R., KO, M.S. *et al.,* (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol* 26: 7772-7782.

OVASKA, K., LAAKSO, M. and HAUTANIEMI, S. (2008). Fast Gene Ontology based clustering for microarray experiments. *BioData Min* 1: 11-18.

PHILLIPS, R.L., ERNST, R.E., BRUNK, B., IVANOVA, N., MAHAN, M.A., DEANEHAN, J.K., MOORE, K.A., OVERTON, G.C. and LEMISCHKA, I.R. (2000). The genetic program of hematopoietic stem cells. *Science* 288: 1635-1640.

QUACKENBUSH, J. (2001). Computational analysis of microarray data. *Nat Rev Genet* 2: 418-427.

RAMAKRISHNAN, R., DORRIS, D., LUBLINSKY, A., NGUYEN, A., DOMANUS, M., PROKHOROVA, A., GIESER, L., TOUMA, E., LOCKNER, R., TATA, M. *et al.,* (2002). An assessment of Motorola CodeLink microarray performance for gene expression profiling applications. *Nucleic Acids Res* 30: e30-e41.

RAMALHO-SANTOS, M., YOON, S., MATSUZAKI, Y., MULLIGAN, R.C. and MELTON, D.A. (2002). "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science* 298: 597-600.

RAO, R.R. and STICE, S.L. (2004). Gene expression profiling of embryonic stem cells leads to greater understanding of pluripotency and early developmental events. *Biol Reprod* 71: 1772-1778.

RESNIK, P. (1995). Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In *International Joint Conference for Artificial Intelligence*. pp. 448-453.

RITCHIE, M.E., SILVER, J., OSHLACK, A., HOLMES, M., DIYAGAMA, D., HOLLOWAY, A. and SMYTH, G.K. (2007). A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 23: 2700-2707.

SATO, N., SANJUAN, I.M., HEKE, M., UCHIDA, M., NAEF, F. and BRIVANLOU, A.H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol* 260: 404-413.

SCHEELE, S., FALK, M., FRANZEN, A., ELLIN, F., FERLETTA, M., LONAI, P., ANDERSSON, B., TIMPL, R., FORSBERG, E. and EKBLOM, P. (2005). Laminin

alpha1 globular domains 4-5 induce fetal development but are not vital for embryonic basement membrane assembly. *Proc Natl Acad Sci USA* 102: 1502-1506.

SINGH, A.M., HAMAZAKI, T., HANKOWSKI, K.E. and TERADA, N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* 25: 2534-2542.

SKOTTMAN, H., STROMBERG, A.M., MATILAINEN, E., INZUNZA, J., HOVATTA, O. and LAHESMAA, R. (2006). Unique gene expression signature by human embryonic stem cells cultured under serum-free conditions correlates with their enhanced and prolonged growth in an undifferentiated stage. *Stem Cells* 24: 151-167.

SMITH-MUNGO, L.I. and KAGAN, H.M. (1998). Lysyl oxidase: properties, regulation and multiple functions in biology. *Matrix Biol* 16: 387-398.

SMYTH, G.K., YANG, Y.H. and SPEED, T. (2003). Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* 224: 111-136.

SPERGER, J.M., CHEN, X., DRAPER, J.S., ANTOSIEWICZ, J.E., CHON, C.H., JONES, S.B., BROOKS, J.D., ANDREWS, P.W., BROWN, P.O. and THOMSON, J.A. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci USA* 100: 13350-13355.

TANAKA, T.S., JARADAT, S.A., LIM, M.K., KARGUL, G.J., WANG, X., GRAHOVAC, M.J., PANTANO, S., SANO, Y., PIAO, Y., NAGARAJA, R. *et al.,* (2000). Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci USA* 97: 9127-9132.

TANAKA, T.S., KUNATH, T., KIMBER, W.L., JARADAT, S.A., STAGG, C.A., USUDA, M., YOKOTA, T., NIWA, H., ROSSANT, J. and KO, M.S. (2002). Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. *Genome Res* 12: 1921-1928.

THOMPSON, J.R. and GUDAS, L.J. (2002). Retinoic acid induces parietal endoderm but not primitive endoderm and visceral endoderm differentiation in F9 teratocarcinoma stem cells with a targeted deletion of the Rex-1 (Zfp-42) gene. *Mol Cell Endocrinol* 195: 119-133.

TUSHER, V.G., TIBSHIRANI, R. and CHU, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98: 5116-5121.

VENEZIA, T.A., MERCHANT, A.A., RAMOS, C.A., WHITEHOUSE, N.L., YOUNG, A.S., SHAW, C.A. and GOODELL, M.A. (2004). Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol* 2: e301-e312.

WATERSTON, R.H.LINDBLAD-TOH, K.BIRNEY, E.ROGERS, J.ABRIL, J.F.AGARWAL, P.AGARWALA, R.AINSCOUGH, R.ALEXANDERSSON, M.AN, P. *et al.,* (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562.

YOO, S.M., CHOI, J.H., LEE, S.Y. and YOO, N.C. (2009). Applications of DNA microarray in disease diagnostics. *J Microbiol Biotechnol* 19: 635-646.

ZHAO, R., WATT, A.J., LI, J., LUEBKE-WHEELER, J., MORRISEY, E.E. and DUNCAN, S.A. (2005). GATA6 is essential for embryonic development of the liver but dispensable for early heart formation. *Mol Cell Biol* 25: 2622-2631.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

Comparative study of mouse and human feeder cells for human embryonic stem cells
Livia Eiselleova, Iveta Peterkova, Jakub Neradil, Iva Slaninova, Ales Hamp and Petr Dvorak
Int. J. Dev. Biol. (2008) 52: 353-363

Mouse induced pluripotent stem cells
Eamon Geoghegan and Lucy Byrnes
Int. J. Dev. Biol. (2008) 52: 1015-1022

Analysis of parent-specific gene expression in early mouse embryos and embryonic stem cells using high-resolution two-dimensional electrophoresis of proteins
L Bowden, J Klose and W Reik
Int. J. Dev. Biol. (1996) 40: 499-506

Lithium influences differentiation and tissue-specific gene expression of mouse embryonic stem (ES) cells in vitro
M M Schmidt, K Guan and A M Wobus
Int. J. Dev. Biol. (2001) 45: 421-429

Myogenesis in cultures of uniparental mouse embryonic stem cells: differing patterns of expression of myogenic regulatory factors
L A McKarney, M L Overall and M Dziadek
Int. J. Dev. Biol. (1997) 41: 485-490