

Accepted Manuscript

Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains

Lioudmila V. Sharova, Alexei A. Sharov, Yulan Piao, Nabeebi Shaik, Terry Sullivan, Colin L. Stewart, Brigid L.M. Hogan, Minoru S.H. Ko

PII: S0012-1606(07)00880-9
DOI: doi: [10.1016/j.ydbio.2007.05.004](https://doi.org/10.1016/j.ydbio.2007.05.004)
Reference: YDBIO 3249

To appear in: *Developmental Biology*

Received date: 2 November 2006
Revised date: 1 May 2007
Accepted date: 4 May 2007



Please cite this article as: Sharova, Lioudmila V., Sharov, Alexei A., Piao, Yulan, Shaik, Nabeebi, Sullivan, Terry, Stewart, Colin L., Hogan, Brigid L.M., Ko, Minoru S.H., Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains, *Developmental Biology* (2007), doi: [10.1016/j.ydbio.2007.05.004](https://doi.org/10.1016/j.ydbio.2007.05.004)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains

Lioudmila V. Sharova¹, Alexei A. Sharov¹, Yulan Piao¹, Nabeebi Shaik¹, Terry Sullivan², Colin L. Stewart², Brigid L.M. Hogan³, and Minoru S.H. Ko¹

¹ Laboratory of Genetics, National Institute on Aging, NIH, Baltimore, MD 21224, USA.

² Cancer and Developmental Biology Laboratory, National Cancer Institute, NIH, Frederick, MD 21702, USA.

³ Department of Cell Biology, Duke University Medical Center, Durham, NC 27710, USA.

Corresponding author

Minoru S.H. Ko

National Institute on Aging (NIA/NIH)

Laboratory of Genetics

333 Cassell Drive, Suite 3000

Baltimore, Maryland 21224

Phone 410-558-8359; Fax 410-558-8331

E-mail kom@mail.nih.gov

Running title: Gene expressions in ES and EG cells

Keywords: microarray, gene expression, embryonic stem cells, embryonic germ cells, LIF, retinoic acid, differentiation, ES cells, EG cells, mouse strains, C57BL/6 mouse, 129 mouse, pluripotency, stem cells

Abstract

Pluripotent stem cell lines with similar phenotypes can be derived from both blastocysts (embryonic stem cells, ESC) and primordial germ cells (embryonic germ cells, EGC). Here, we present a compendium DNA microarray analysis of multiple mouse ESCs and EGCs from different genetic backgrounds (strains 129 and C57BL/6) cultured under standard conditions and in differentiation-promoting conditions by the withdrawal of Leukemia Inhibitory Factor (LIF) or treatment with retinoic acid (RA). All pluripotent cell lines showed similar gene expression patterns, which separated them clearly from other tissue stem cells with lower developmental potency. Differences between pluripotent lines derived from different sources (ESC vs. EGC) were smaller than differences between lines derived from different mouse strains (129 vs. C57BL/6). Even in the differentiation-promoting conditions, these pluripotent cells showed the same general trends of gene expression changes regardless of their origin and genetic background. These data indicate that ESCs and EGCs are indistinguishable based on global gene expression patterns alone. On the other hand, a detailed comparison between a group of ESC lines and a group of EGC lines identified 20 signature genes whose average expression levels were consistently higher in ESC lines, and 84 signature genes whose average expression levels were consistently higher in EGC lines, irrespective of mouse strains. Similar analysis identified 250 signature genes whose average expression levels were consistently higher in a group of 129 cell lines, and 337 signature genes whose average expression levels were consistently higher in a group of C57BL/6 cell lines. Although none of the genes was exclusively expressed in either ESCs versus EGCs or 129 versus C57BL/6, in combination these signature genes provide a reliable separation and identification of each cell type. Differentiation-promoting conditions also revealed some minor differences between the cell lines. For example, in the presence of RA, EGCs showed a lower expression of muscle- and cardiac-related genes and a higher expression of gonad-related genes than ESCs. Taken together, the results provide a rich source of information about the similarities and differences between ESCs and EGCs as well as 129 lines and C57BL/6 lines. Such information will be crucial to our understanding of pluripotent stem cells. The results also underscore the importance of

studying multiple cell lines from different strains when making comparisons based on gene expression analysis.

Introduction

Embryonic Stem Cells (ESCs) and embryonic germ cells (EGCs) are prototypical pluripotential stem cells cultured *in vitro*. Mouse ESCs can be derived from the inner cell mass (ICM) of blastocysts (Evans and Kaufman, 1981; Martin, 1981), whereas mouse EGCs can be derived from primordial germ cells (Matsui et al., 1992; Resnick et al., 1992). Similar cells have also been derived from human ICM (Thomson et al., 1998) and PGCs (Shamblott et al., 1998), respectively. Both ESCs and EGCs possess two defining characters of pluripotent stem cells: self-renewal, i.e., ability to produce daughter cells with an identical phenotype, and pluripotency, i.e., ability to give rise to most differentiated cell types, including germ cells (Hadjantonakis and Papaioannou, 2001; Labosky et al., 1994; Loebel et al., 2003; Smith, 2001; Stewart et al., 1994). Furthermore, both ESC and EGC express the same pluripotency markers, such as *Pou5f1* (also known as *Oct3/4*, *Oct3*, *Oct4*), *Nanog*, *Zfp42* (also known as *Rex1*), and alkaline phosphatase (ALP). Thus, ESCs and EGCs are often considered as essentially equivalent cell types. Such similarity has recently led to a hypothesis that ESCs are also derived from PGC precursors that are originated from the ICM (Zwaka and Thomson, 2005).

There are, however, some known differences between ESC and EGC. For example, chimeras generated using some EGCs derived from postmigratory PGC from the genital ridge had fetal overgrowth and skeletal abnormalities (McLaren and Durcova-Hills, 2001; Tada et al., 1998). It has been reported that EGCs failed to differentiate in a co-culture with lung tissue, whereas ESC successfully differentiated in these conditions (Durcova-Hills et al., 2003). *In vitro*, EGCs differentiated more efficiently to neuronal cells and less efficiently to cardiac and skeletal muscle cells than ESC (Rohwedel et al., 1996). As far as

we know, there are no surface or other markers that can distinguish between mouse ESC and EGC lines. The only known molecular difference is the DNA methylation patterns of some imprinted genes.

Another important difference between ESCs and EGCs is related to how these pluripotent cells are derived. Pluripotent ESCs are derived from pluripotent ICM of blastocysts, whereas pluripotent EGCs are derived from unipotent (or sometimes called nullipotent (Donovan and de Miguel, 2003)) PGCs, which can differentiate only to sperm or oocytes. Therefore, ICM cells converting to ESCs need to acquire only the capacity to proliferate actively in culture for long periods without differentiation (self-renewal), but PGCs converting to EGCs need to acquire both self-renewal and pluripotency. Therefore, the changes that occur during the conversion of PGCs to EGCs are often compared to those seen after the transplantation of nuclei from differentiated cells into enucleated oocytes (“nuclear reprogramming”) (Donovan and de Miguel, 2003; Hochedlinger and Jaenisch, 2006) . Here, we ask whether these seemingly different mechanisms involved in deriving ESCs and EGCs are reflected in the molecular characteristics of the cells in culture.

Although the mouse ESCs have been studied in great detail by global gene expression profiling (e.g., Anisimov et al., 2002; Brambrink et al., 2006; Furusawa et al., 2006; Ivanova et al., 2002; Ramalho-Santos et al., 2002; Tanaka et al., 2002; Wei et al., 2005), comparisons between ESC and EGC have been limited to the analysis of EST frequency in one ESC line versus one EGC line (Sharov et al., 2003).

Therefore, the main question still remains open of whether ESCs and EGCs are distinct or equivalent cell types from the viewpoint of global gene expression patterns. As a first step to address this question, we have examined the global gene expression profiles of mouse ESCs and EGCs. To avoid detecting cell line-specific gene expression patterns, we examined multiple ESC lines and EGC lines, which have been established independently. Considering that gene expression profiles might be affected by genetic background, we tested cell lines derived from different mouse strains. Finally, we examined possible

differential responses of ESC and EGC lines to differentiation-promoting stimuli, namely the absence of LIF and the presence of retinoic acid (RA).

Results

I. Phenotype and cell cycle analysis of pluripotent cell lines

We first examined the phenotypes of six mouse ESC lines and 10 mouse EGC lines (Table 1) cultured for three days in three different conditions: with LIF (LIF+, the standard culture condition), without LIF (LIF-, the differentiation-promoting condition), and with LIF and RA (RA+, the differentiation-promoting condition) (see the Materials and Methods for details). In the standard culture conditions, all cell lines examined here showed similar morphology and growth (Fig. 1A). When these cells were cultured in the differentiation-promoting conditions, we observed three early indicators of differentiation: a decreased signal intensity of alkaline phosphatase (ALP) staining, the flattening of ESC and EGC colonies, and the emergence of distinct single cells at the edges of colonies. These early signs of differentiation were noted by day 2 with RA and by day 3 in the absence of LIF (Fig. 1A). By day 3, most cells exposed to RA were already well-differentiated, with a morphology resembling epithelial or endodermal cells. However, at day 3, the cells cultured without LIF did not show any obvious signs of differentiation into a particular cell lineage, although the colonies already showed flat morphology and low intensity or patchy ALP staining. In the presence of RA, cells with a differentiated phenotype were slightly more abundant in ESC lines than in EGC lines.

It is known that pluripotent cells have a unique cell cycle: most cells are found in the S phase of the cell cycle and the G1 phase is short (Savatier et al., 1996). In the presence of LIF, all cell lines examined here showed a distribution of 57.7% - 66.3% cells in S phase and 13.8% - 23.4% cells in G1 phase (Fig. 1B), as it has been reported for undifferentiated ESC (Savatier et al., 1996). As expected, upon differentiation

the proportion of cells in the G1 phase of the cell cycle increased while the proportion of cells in S phase decreased. However, these changes were relatively small, probably because cells were not fully differentiated after 3 days in the RA+ and LIF- conditions. The proportion of cells in G1 phase of the cell cycle was lower in EGC than in ESC (GLM, $p = 0.033$), especially in the RA+ condition.

2. Global expression profiling of pluripotent cell lines

Using a whole-genome NIA 44K oligo-DNA microarray, we obtained the global gene expression profiles of six mouse ESC lines and six mouse EGC lines cultured for three days in LIF+, LIF-, and RA+ conditions. To assess the expression profiles of these cells in a larger context, we first compared them with our previous microarray data obtained from trophoblast stem (TS) cells and neural stem/progenitor cells (NSC) (Aiba et al., 2006). Direct comparison was possible because both studies used the same type of microarrays with a largely overlapping set ($N = 20,088$) of 60-mer oligos, and both studies included the same ESC line 129.3, which was used as a common standard for data normalization. Results of Principal Component Analysis (PCA) of log-transformed gene expression values showed that the gene expression profiles of ESC and EGC lines were similar to each other, and were clearly separated as a single group from those of TS and NSC (Fig. 2).

Although ESCs and EGCs were inseparable in the above analysis, we wished to identify the differences among individual cell lines. First, we applied ANOVA statistics to the microarray data of all ESC and EGC lines in the standard LIF+ condition and found that 6998 genes had a significant difference in their expression among individual ESC and EGC lines (Table S1). The PCA of these genes revealed considerable variations among individual cell lines. We found that 129 cell lines and C57BL/6 cell lines were best separated along a linear combination of principal components 1 and 2 ($PC1+0.69\cdot PC2$), whereas ESCs and EGCs were separated along the PC3 axis (Fig. 3). However, the cell lines-to-cell lines variations within ESC and EGC groups were too large to draw clear-cut boundary between these categories (Fig. 3). For example, two EGC lines (TGC 8.5-5 and TGC 8-8) had gene expression similar to

the ESC line BL6.9, whereas other 4 EGC lines had a more distinct gene expression pattern. In general, the difference in gene expression patterns between ESCs and EGCs was smaller than that between mouse strains, because the former was represented only by the 3rd principal component (PC3), but the latter was represented by the first two principal components (PC1 and PC2). Therefore, we conclude that ESCs and EGCs are not distinct entities in relation to global gene expression.

We expected sex-specific differences of global gene expression profiles in ESCs and EGCs. However, we did not find any principal component that separated female cell lines (CC9, TGC-12.8, and EG-3) from male lines. Three Y-linked genes (*Eif2s3y*, *Uty*, and *Ddx3y*) were more highly expressed ($p < 0.001$) in male cell lines than in female cell lines, whereas four X-linked genes (*Xlr5d*, *Xlr5c*, and *1700013H16Rik*) were more highly expressed ($p < 0.04$, fold change > 2) in female cell lines. However, such a few sex-related genes did not satisfy the stringent statistical criteria by the false discovery rates (FDR; see Materials and Methods for the detail). Pluripotent cell lines have thus limited sex-related differences in gene expression. We also expected stage-specific differences in the expression profiles of EGCs. However, we did not detect any specific gene expression patterns that correlated to the stage of EGC source, e.g., pre-migratory PGC (8-8.5 dpc) or post-migratory PGC (12 dpc) stages.

3. Variability of gene expressions among pluripotent cell lines

The separation of individual pluripotent cell lines by the PCA reflects the differences in their global gene expression patterns among the cell lines. To investigate what genes contributed to these differences, we sought genes whose expression levels changed significantly between cell lines (FDR = 0.05, fold change > 1.5) and correlated ($\text{abs}(r) > 0.7$) to those principal components that separated cell lines based on their strain and origin (Fig. 3). We found four sets of genes (Table S2-S3): 580 genes were correlated positively to $\text{PC1} + 0.69 \cdot \text{PC2}$, and thus were generally more highly expressed in 129 cells than in C57BL/6 cells (“129-overexpressed genes”); 658 genes were correlated negatively to $\text{PC1} + 0.69 \cdot \text{PC2}$, and thus were generally more highly expressed in C57BL/6 cells than in 129 cells (“C57BL/6-overexpressed genes”); 207 genes were

correlated positively to the PC3, and thus were generally more highly expressed in EGC lines than ESC lines (“EGC-overexpressed genes”); and 35 genes were correlated negatively with PC3, and thus were more highly expressed in ESC lines than EGC lines (“ESC-overexpressed genes”).

First, we analyzed these four sets of genes from the viewpoint of their pluripotency. Previously, we have shown that NSCs and TS cells already commit to specific cell lineages and express many genes that are characteristic of differentiated cells (Aiba et al., 2006). From the PCA (Fig. 2), we extracted 1,719 genes that were negatively associated with PC1, which represented genes expressed more highly in pluripotent ESCs and EGCs than in the lineage-committed NSCs and TS cells (Table S4). These candidate “pluripotency-related” genes included *Pou5f1* (*Oct3/4*), *Nanog*, *Zfp42* (*Rex1*), *Nr0b1* (*Dax1*), *Tcl1*, *Dppa3* (*Stella*), *Dppa4*, *Klf2*, *Klf4*, *Jarid2*, *Jarid1b*, *Foxd3* (Anisimov et al., 2002; Boyer et al., 2005; Gordeeva et al., 2005; Hanna et al., 2002; Li et al., 2005; Mongan et al., 2006; Niakan et al., 2006; Player et al., 2006; Sharov et al., 2003; Yoshikawa et al., 2006). The pluripotency-related genes were more abundant in the 129-overexpressed gene list than the C57BL/6-overexpressed gene list (chi-square = 19.4; $p < 0.001$) (Fig. 3), but there was no significant difference in the proportion of pluripotency-related genes in EGC- and ESC-overexpressed gene lists (chi-square < 0.001 ; $p = 1$). These data suggest that ESC and EGC show no difference in the degree of pluripotency, whereas C57BL/6 lines have a greater tendency to lose their pluripotency than 129 cell lines, at least after three days in culture in the standard LIF+ condition. Genes that were differentially expressed between strains did not include the canonical pluripotency genes (*Pou5f1*, *Nanog*, and *Sox2*); however, the expression level of *Zfp42*, which was already high in C57BL/6 lines, was much higher in 129 cell lines (Table S2, Fig. 5B).

Second, we analyzed these four sets of genes based on gene ontology (GO) terms (Fig. 3; see Tables S5-S7 for the details). The 129-overexpressed genes were classified as being associated with cytokine production, amino acid metabolism, lipid transport, and actin cytoskeleton. Higher cytokine production may be related to the fact that 129-derived cell lines have faster proliferation rate than C57BL/6-derived

cell lines in the standard LIF+ culture condition (data not shown). They also included many genes with unknown functions. The C57BL/6-overexpressed genes were classified as being associated with oxidoreductase activity, steroid metabolism, lysosome, spliceosome, protein transport, apoptosis, ubiquitination activity, and GTPase activity. Because the number of ESC-overexpressed genes was very limited (N = 35), there were no statistically significant GO categories. The EGC-overexpressed genes were associated with tRNA aminoacylation, cytokinesis, translation, and ubiquitination activity. It is worth mentioning that genes involved in methyltransferase activity (*Mettl5*, *Smyd3*, *Dnmt3l*) were overexpressed in the EGCs, which may indicate that the status of chromatin modifications in EGCs is different from that in ESCs.

4. Signature genes that distinguish strains and origins of pluripotent cells

One of our original goals was to identify marker genes that can separate ESCs and EGCs. The comparison between a single ESC line and a single EGC line revealed many such differentially expressed genes. However, subsequent analysis showed that these genes were merely cell line-specific and did not show consistent differential expression between multiple ESCs and multiple EGCs (e.g., Supplemental Fig. S1).

To find signature genes that are most consistently different between ESC and EGC lines in all culture conditions, we grouped the cells according to their origin (e.g. ESC vs. EGC) and carried out the statistical analysis (ANOVA) between these groups (see Materials and Methods for the detail). We found 104 signature genes (20 genes for ESC and 84 genes for EGC), whose average expression levels can distinguish each group (Tables S8-S9). Although we did not find any gene that was exclusively expressed in ESC or EGC, these groups were separated by the average expression levels of signature genes from each set (Fig. 4A), except one replication for cell line TGC 8-7. The top genes from signature lists were chosen for qRT-PCR verification: 2 ESC-overexpressed genes (*Grb10*, *Id2*) and 3 EGC-overexpressed genes (*Peg3*, *Spp1*, *Snrpn*). qRT-PCR results confirmed microarray data with all genes showing differential expression across all tested 6 ESC and

10 EGC lines (Fig. 5A). Out of 20 ESC-signature genes, two (*Id1* and *Id2*) are known to be directly activated by *Bmp4* (Hollnagel et al., 1999), which may suggest that *Bmp4* signaling is suppressed in EGC lines. Among EGC signature genes some are associated with extracellular matrix (*Coll16a1*, *Coll1a1*, *Col8a1*, *Ecm1*, *Mfge8*, *Postn*, *Serpinf1*, *Serpinh1*, *Slit2*, *Spp1*, *Timp1*), which may be related to the migratory pathway of parental PGC cells (Pereda et al., 2006).

We also found 250 signature genes for cells derived from the 129 strain and 336 signature genes for cells derived from the C57BL/6 strain in all three culture conditions (Tables S10-S11). Cell lines derived from these different strains were well separated by the average expression levels of top 100 signature genes from each set (Fig. 4B). Differential expression of 7 signature genes *Prdx2*, *Bhlhb2*, *Zadh2*, *Inpp5d*, *Tcstv3*, *Spink3*, and *Zfp42* was confirmed using qRT-PCR (Fig. 5B). To test if these differences were related to the degree of cell differentiation, we compared the sets of signature genes with a set of 514 genes whose expressions were induced in ESCs and EGCs in the LIF- condition (Table S12). In the set of 514 genes, more C57BL/6-signature genes were represented than 129-signature genes (16.6%, $N = 56$ in the strain C57BL/6 versus 4.0%, $N = 10$ in the strain 129; chi-square = 22.9, $p < 0.001$, see Tables S10, S11), indicating that cell lines derived from the C57BL/6 strain had a higher tendency to differentiate than cell lines from the 129 strain. However, the majority of strain-specific signature genes were not related to cell differentiation.

We found a few interesting examples that strain-specific phenotypes of mice may be reflected in the strain-specific signature genes. First, oxidoreductase activity was the top GO category for C57BL/6-overexpressed genes (Fig. 3), 17 of which were represented in the signature genes (e.g., *Prdx2*, *Dhrs7*, *Zadh2*, *Egln3*, *Mod1*, *Nxn*, *Phgdh*, *Gsr*). This seems to be consistent with the report that C57BL/6 mice have higher resistance to oxidative stress and have a longer life span than other strains (Rebrin and Sohal, 2004; Wesselkamper et al., 2000; Zraika et al., 2006). Second, we noticed genes involved in alcohol metabolism (*Pkm2*, *Sgpp1*, *Slc37a4*, *Soat1*, *Vldlr*, *Zadh2*) in the C57BL/6 signature genes, which may be related to the well-known alcohol-

preferring behavior of C57BL/6 mice (Kelai et al., 2006). Third, we found that apoptosis-related genes (*Card4/Nod1* and *Pmaip1*) were at the top of the list in the signature genes for strain C57BL/6, followed by additional genes (*Atm*, *Cebpb*, and *Dap*). Indeed, the higher incidence of apoptosis in C57BL/6 strain has been reported during oocyte atresia (Canning et al., 2003) and ES cell differentiation (Ward et al., 2004).

Frequent testicular cancer is a well-known phenotype of the 129 mouse strain (Matin and Nadeau, 2005). *Dnd1* has recently been identified as a gene responsible for *Ter* mutation, which is the most potent testicular cancer modifier gene (Youngren et al., 2005). We therefore examined whether the strain-specific signature genes were cancer-related (based on PubMed and NCI/Biomax Cancer Gene Database, <http://ncicb.nci.nih.gov/NCICB/projects/cgdc>) and expressed in testis (based on EST Profile Viewer (Larsson et al., 2000) in UniGene database). We found that 6 such genes were over-expressed in the strain 129 (*Anxa4*, *Dmd*, *Eno1*, *Gbp1*, *Gpr56*, *Tlr2*); however, all of them were tumor-suppressors, and therefore, their high expression is unlikely to cause cancer. In contrast, we found that 16 such genes were under-expressed in the strain 129 (i.e., signature genes for C57BL/6), 7 of which were either tumor suppressors or negatively correlated with cancer (*Atm*, *Dok2*, *Hdac1*, *Herpud1*, *Pik3cd*, *Prdx2*, *Sfrp1*). A list of genes identified here may provide additional candidates for testicular cancer modifier genes in the 129 strain.

5. Differential expression of imprinted genes in ESCs and EGCs

Previous study has shown the differential DNA methylation of some imprinted genes between ESCs and EGCs (Durcova-Hills et al., 2001; Labosky et al., 1994; Tada et al., 1998). Out of 60 mammalian imprinted genes reported so far (Luedi et al., 2005), 6 genes (*H19*, *Kcnq1*, *Nnat*, *Peg3*, *Snrpn*, and *U2af1-rs1*) were more highly expressed in EGCs, whereas 2 genes (*Calcr*, *Grb10*) were more highly expressed in ESCs in the standard LIF+ culture condition (Table 2). *Grb10*, a growth suppressor (Smith et al., 2006), was also the top ESC signature gene and *Snrpn*, *U2af1-rs1*, *H19* were among the top EGC signature genes (Tables S8-S9). Differential expression of *Grb10*, *Peg3*, and *Snrpn* was tested and confirmed by qRT-PCR (Fig. 5A). Five

genes out of 7 imprinted genes that were over-expressed in EGC were mapped to chromosome 7 (Table 2). Interestingly, 8 imprinted genes showed significant ($p < 0.05$ and >1.5 fold change) differences in expressions between mouse strains at least in one of the differentiation-promoting conditions (Table 2). For example, *Peg3* and *Dcn* were more highly expressed in the C57BL/6 strain, whereas *Impact*, and *Zfp264* were more highly expressed in the 129 strain.

6. Responses of pluripotent cells to differentiation-promoting culture conditions: Global Trends

Although ESCs and EGCs were not separable by global gene expression patterns in the standard culture condition, we suspected that they might show more discernible differences when induced to differentiate. To test this idea, we carried out the expression profiling of ESC lines and EGC lines cultured for three days in LIF- and RA+ conditions. The PCA of the centered gene expression profiles (see Materials and Methods for the detail) clearly showed that both ESCs and EGCs responded in a similar manner to LIF withdrawal and RA treatments (Fig. 6). The first major component PC1 corresponded to the changes associated with the RA+ condition, whereas the second major component PC2 corresponded to the changes associated with the LIF- condition (Fig. 6). The data indicate that the RA+ condition caused more dramatic change of global gene expression patterns than the LIF- condition. In fact, the RA treatment caused a qualitatively different and much stronger response in gene expression (2613 genes with >1.5 -fold change) than the LIF- condition (937 genes with >1.5 -fold change). This is consistent with the phenotypical differences observed between the RA+ and LIF- conditions (Fig. 1A): The RA+ condition induced more active and faster differentiation than the LIF- condition. In fact, the expressions of some pluripotency-related genes (e.g., *Pou5f1*, *Sox2*, *Nanog*, *Aire*, *Otx2*, *Foxd3*, and *Tcea3*) were more sharply reduced in the RA+ condition than the LIF- condition. The addition of RA to the basal conditions directed cells to a wide variety of lineages (neural, epithelial, endoderm, and mesoderm) with a

bias towards neural differentiation. For example, genes induced selectively by RA included homeobox genes (e.g., *Esx1*, *Meis1*, *Hoxa1*, *Hoxa7*, *Hoxb1*, *Hoxa5*, *Hoxa2*, *Hoxc9*, *Hoxa10*, *Hoxb7*, *Hoxb8*), and genes involved in morphogenesis (e.g. *Tbx3*), actin cytoskeleton, cell adhesion, cell migration, extracellular matrix, and various types of differentiation (e.g., neural, muscular, vascular, kidney, and endodermal cells). The most highly induced genes were related to neural differentiation (*Pmp22*, *Mrg1*, and *Hoxa1*). Genes induced selectively by LIF withdrawal included fibroblast growth factors (*Fgf5*, *Fgf8*, and *Fgf15*) and transcription factors (*T*, *Eomes*, *Lmyc1/Mycl1*, *Sp8*, *Irf1*, *Gsc*, and *Sp5*) (Table S12), many of which are mesoderm markers (*T*, *Eomes*, *Fgf8*, *Fgf15*) (Loebel et al., 2003). Some pluripotency-related genes were suppressed preferentially in the LIF- condition (*Spp1*, *Klf4*, *Esgp/EG653016*, *Piwil2*, *Kit*), whereas other pluripotency-related genes decreased both in the LIF- and RA+ conditions (*Nr0b1*, *Tcll*, *Zfp459*, *Sall1*, *Gli1*, *Rest*, *Nr5a2*, *Lrrc2*, *Klf2*, *Klf5*, *Klf9*).

7. Responses of pluripotent cells to differentiation-promoting culture conditions: Origin- and strain-specific variations

Differentiation-promoting culture conditions generally reduced the differences in gene expression levels among pluripotent cell lines. For example, most ESC- and EGC-signature genes identified in the standard LIF+ culture condition did not show significant differences in their expression in the LIF- and RA+ conditions. Only 2 (10%) ESC-signature genes (*Grb10* and *Glo1*) and 15 (18%) EGC-signature genes (*Atf5*, *Ccng2*, *Itpr3*, *Mylpf*, *Pgm2*, *Pink1*, *Ralgds*, *Snf1lk*, *Snrpn*, *Snx6*, *Spp1*, *Tbc1d13*, *Tmem40*, *Trib3*, *U2af1-rs1*) showed persistent differential expressions in all three culture conditions (tables S8-S9), indicating that these genes are the most reliable separator between ESCs and EGCs. Similarly, many C57BL/6-signature genes and 129-signature genes showed no difference in expression in the LIF- and RA+ conditions. For example, *Fgf5* showed a high expression in C57BL/6 lines and a low expression in 129 lines in the standard LIF+ culture condition, but its expression became equally high in all cell lines in the LIF- condition and equally low in all cell lines in the RA+ condition. However, a higher proportion of strain-specific signature genes retained their expression differences in the differentiation-promoting

conditions: 75 genes (30%) for 129-signature genes and 91 genes (27%) for C57BL/6-signature genes (Tables S10-S11).

To further investigate the origin- and strain-specific variations in responses to differentiation-promoting culture conditions, we focused on a set of genes that were differentially expressed between 129 and C57BL/6 (Fig. 7A) or between ESC and EGC (Fig. 7B), only after RA induced differentiation. Genes that were overexpressed in the 129 cell lines after RA treatment were associated with immune response (*Lgals4*, *H2-B1*, *Ly6h*, *Cxcl10*, *Ccl3*, *H2-Q1*, *H2-Q5*, *H2-M10.2*), muscle differentiation (*Myocd*, *Myl7*), and cell cycle control (*Gas1*, *Fos*) (Fig. 7A). In contrast, genes that were overexpressed in the C57BL/6 cell lines in response to RA were associated with secretion (*Saa3*, *Saa1*), the inhibition of Wnt signaling (*Nkd1*), and the extracellular matrix (*Col2a1*, *Mmp9*).

Genes that were overexpressed in ESCs compared with EGCs in the RA+ condition included the following functional groups: (i) muscle development and actin cytoskeleton (*Myocd*, *Acta1*, *Myl7*, *Cgnl1*, *Flnb*, *Myl6*, *Tagln2*, and *Cald1*), (ii) regulation of transcription (*Creb3l3*, *Foxd1*, *Hipk3*, *Rarb*, *Setbp1*, *Meis1*, and *Bach2*), and (iii) cell junction structure (*Cgnl1*, *Flnb*, *Shrm/Shroom3*) (Fig. 7B). These results indicate that genes activated in ESC lines by the RA treatment are more related to muscle differentiation than those activated in EGC lines. Genes that were selectively activated in EGCs compared with ESCs in the RA+ condition are known to be expressed in ovary (*Obox6*, *Hoxc9*, *Mmp9*, *Saa1*), blood (*Hoxa9*), and intestine (*Cdx1*). *Obox6* is expressed uniquely in oocytes (Rajkovic et al., 2002), and *Hoxc9* and *Mmp9* are expressed in granulosa cells in the ovary (Huntriss et al., 2006; Ke et al., 2004). qRT-PCR analyses confirmed the overexpression of *Myocd*, *Myl7*, and *Plat* in ESC and *Obox6* in EGC in the RA+ condition (Fig. 5C). These data thus indicate that RA induced some gonad-related genes more highly in the EGC lines than the ESC lines.

Discussion

We have compared the gene expression profiles of pluripotent stem cell lines with different genetic backgrounds and derived from cells at different stages of development: the ICM of blastocyst (ESC) and primordial germ cells (EGC). In addition to the standard LIF+ culture conditions, we tested gene expression after LIF withdrawal and RA treatment. As far as we know, this is the first large-scale gene expression study of various types of pluripotent stem cell lines. This data set (Table S13) will thus be a valuable resource to the research community.

One of the main questions we wanted to address in this work was whether ESCs and EGCs can be distinguished based on global gene expression patterns. Both ESC and EGC are pluripotent stem cells and can differentiate into essentially all the cell types *in vivo* and *in vitro*. Therefore, it has been an important question whether the different origins of these cells indeed cause any differences in their phenotypes and gene expression regulation (Donovan and de Miguel, 2003; Durcova-Hills et al., 2003; Hochedlinger and Jaenisch, 2006; Rossant, 1993). Here, we have shown many similarities and differences between ESCs and EGCs as well as between C57BL/6 lines and 129 lines. From the perspective of global gene expression profiles, we have shown in this paper that ESCs and EGCs are indistinguishable. First of all, when the expression profiles of multiple ESCs and EGCs were compared with those of NSC and TS cells, all ESCs and EGCs were clustered as one group. Second, even when comparing only ESCs and EGCs, variations between individual cell lines were too great to clearly separate them into two distinct groups, and the differences between ESCs and EGCs were smaller than those between cells derived from different strains. Differences of global gene expression patterns measured by the PCA between one ESC line to another ESC line were larger than those between one ESC line and EGC line. It is important to point out that if we compared the global expression profiles of only a few cell lines, we would have erroneously concluded that ESC and EGC are separate entities and identified marker genes specifically expressed in either ESC or EGC.

Although we expected that our study would reveal differences between ESCs and EGCs when they were induced to differentiate, we failed to find such an effect in global gene expression patterns. In both ESCs and EGCs, the RA treatment caused a much stronger response in gene expression than the LIF withdrawal. This is consistent with the phenotypical differences observed between RA+ and LIF- conditions. The addition of RA affected the expressions of genes involved in a wide variety of lineages (neural, epithelial, endoderm, and mesoderm) with a bias towards neural differentiation, whereas upon LIF withdrawal changes were biased towards mesoderm lineage.

Considering possible differences between the nuclear reprogramming process of ICM cells converting to ESCs and that of PGCs converting to EGCs, the remarkable similarity of gene expression patterns between ESCs and EGCs was unexpected. The work presented here suggests that these reprogramming events occur to the level where the global gene expression patterns of EGCs are inseparable from those of ESCs. Alternatively, this similarity between ESCs and EGCs may simply support the hypothesis that ESCs are also derived from PGC precursors that are originated from the ICM (Zwaka and Thomson, 2005).

Although large variations of gene expression profiles between cell lines within each category (e.g., ESCs) make it difficult to separate pluripotent cells to distinct groups as discussed above, it is possible to identify signature genes by first grouping the cells according to their categories and then by carrying out the statistical analysis to identify genes differentially expressed between groups. Using this approach, we identified 84 EGC signature genes and 20 ESC signature genes, which showed the most consistent and differential expression between EGC and ESC lines, irrespective of mouse strains. Although none of these genes was exclusively expressed in either EGCs or ESCs, in combination they provide a reasonable separation between these two cell types. These signature genes included several imprinted genes: *Grb10* was overexpressed in ESC lines and *Snrpn*, *U2af1-rs1*, *H19*, and *Nnat* were overexpressed in EGC lines. In total, 13 imprinted genes were differentially expressed between ESC and EGC at least in one culture

condition. Imprinted genes are thus one of the few features that reflect the origin of the pluripotent cell lines. However, imprinted genes constitute only a minor portion (~5%) of ESC- and EGC-signature genes and cannot account for the majority of differences between ESCs and EGCs. Although EGC signature genes identified in the standard LIF+ culture condition did not include germ line-specific genes, some ovary-related genes were upregulated in these cells in the RA+ condition. The level of induction of genes related to muscular differentiation was lower in EGCs compared with ESCs. As a result, EGCs might be more suitable starting material for generation of germ cells in culture and would be a poor choice in muscle development experiments. The most significant differences in induction of differentiation-related genes between EGC and ESC were observed for muscle-related genes in the RA+ condition (Fig. 8). It was reported previously that mouse EGC line (EG-1) undergoing differentiation in the embryoid body (EB) system has a limited ability to differentiate into muscle and cardiac lineages (Rohwedel et al., 1996). In addition, the proportion of cells in G1 phase of the cell cycle in EGC was lower than in ESC in the RA+ condition. These findings might indicate that EGCs were less differentiated in the RA+ condition than ESCs, which could be caused by either delayed response or lower sensitivity to differentiation stimuli.

Using similar approaches, we have also identified a set of genes that can distinguish C57BL/6 and 129 cells by their average expression levels. One notable difference between C57BL/6 cells and 129 cells is their relative pluripotency status at the day 3 of culture in the standard LIF+ condition. Cell lines derived from C57BL/6 strain have their gene expression already shifted towards more differentiated state, although they still have a high expression of canonical pluripotency markers (*Pou5f1*, *Nanog*, *Sox2*), maintain normal ESC-like phenotype (Fig. 1), and normal ESC-like cell cycle (Fig. 2). Interestingly, *Zfp42*, which is a well-known marker for pluripotent cells, belongs to 129-signature genes. Although *Zfp42* was expressed at relatively high level in C57BL/6 cell lines in the standard LIF+ condition compared to tissue-specific stem cells, its expression was up to 10 times higher in 129 cell lines than in C57BL/6 cell lines. We also noticed that the C57BL/6 cells tended to show the slight reduction of ALP

staining level and the slight flattening of the cell colonies at the day 3 in the standard LIF+ condition, compared with 129 cells (data not shown). This earlier shift towards differentiation may reflect cell's ability to respond to differentiating stimuli. However, this shift to differentiation is only a minor aspect of differences in gene expression between C57BL/6 and 129 strains, because the majority of strain-specific signature genes were not related to cell differentiation.

It is known that it is easier to derive ES cell lines from the 129 strain than most other mouse strains (Ledermann, 2000). ESCs derived from C57BL/6 are also known to have restricted differentiation ability in some conditions: for example in the medium without LIF and supplemented with RA they fail to differentiate in neurons and die within 2 weeks in culture (Ward et al., 2004). Their reported inclination to apoptosis matches well with our analysis (Fig. 3). They are also less efficient in producing chimeras and germ line transmission (Ward et al., 2004), although for practical applications higher breeding efficiency of the C57BL/6 strain seems to compensate these drawbacks (Seong et al., 2004). We have shown in this report that pluripotency-related genes are overexpressed in 129-derived cells compared with C57BL/6-derived cells, which might correlate with observations that 129-derived ES cells produce chimeras with high percentage of germ line transmission. We also found candidate genes that may be responsible for the difference in phenotypes between mouse strains. In particular we found genes that are likely to be associated with high resistance to oxidative stress, alcohol-preferring behavior, and high level of apoptosis in C57BL/6 mice as well as genes potentially associated with high prevalence of testicular cancer in the 129 strain.

The current study has examined multiple ESCs and EGCs, but the scope is still limited only to the subsets of all available ESCs and EGCs. It will be thus interesting to extend the study to other ESC and EGC lines. Moreover, it will also be interesting to examine other reported pluripotent stem cells such as multipotent adult progenitor cells (MAPCs; (Jiang et al., 2002), germline stem cells (GSCs; (Kanatsu-Shinohara et al., 2004), and multipotent adult germline stem cells (mpGSCs; (Guan et al., 2006) on the

same platform to compare with the current data sets. The ESC and EGC signature genes that we have identified in this work may have a practical use to distinguish between ESCs and EGCs as well as between 129 cells and C57BL/6 cells. Future studies on genes differentially expressed between 129 and C57BL/6 cells may also help to improve the germline transmission efficiency of C57BL/6 ES cells in mouse gene targeting strategy.

Materials and Methods

Cell Culture

We used 6 mouse ESC lines, including 1 line derived from strain C57BL/6 and 5 lines derived from strain 129, and 10 mouse EGC lines, including 8 TGC lines from strain C57BL/6 (established in the laboratory of B.L.M.H.) and 2 EG lines from strain 129 (established in the laboratory of C.L.S.) (Table 1). BL6.9 (aka MC2-B6) and 129.3 (aka MC1) ESC lines were purchased from The Transgenic Core Laboratory of the Johns Hopkins University School of Medicine (Baltimore, MD, USA). ES-D3-GL ESC line was purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). CC9.3.1 ESC line, which was originally generated by Dr. Allan Bradley, was a kind gift of Dr. Grant R. MacGregor (University of California, Irvine, CA, USA). EBRTcH3 and MG1.19 ESC lines were kind gifts of Dr. Hitoshi Niwa (RIKEN Center for Developmental Biology, Kobe, Japan). Three cell lines were female and the other were male; two ESC lines (EBRTcH3 and MG1.19) were genetically modified (Table 1). For experiments, both ESC and EGC were cultured for 2 passages on gelatin-coated plates in the presence of Leukemia Inhibitory Factor (LIF) in order to remove feeder cells. Cells were then transferred to gelatin-coated 6-well plates at the density of $1-2 \times 10^5$ cells/well ($1-2 \times 10^4$ cells/cm²) and cultured for 3 days in 3 different conditions: (1) complete ES medium: DMEM, 15% FBS; LIF (ESGRO, Chemicon, USA) 1000 U/ml; 1 mM sodium pyruvate; 0.1 mM NEAA, 2 mM glutamate, 0.1 mM beta-mercaptoethanol, and penicillin/streptomycin (50 U/50 µg per ml) (referred as “standard LIF+ condition”); (2) complete medium without LIF (referred as “LIF- condition”), and (3) complete medium with 1 µM RA (referred as

“RA+ condition”). Cells were cultured at 37 °C and 5% CO² condition and the culture medium was changed daily. Undifferentiated state of pluripotent cells was confirmed by ALP staining using Stem Cell Characterization kit (Chemicon, USA).

Cell cycle analysis

For cell cycle analysis, cells were plated on gelatin-coated 6-well plates at the density of 1x10⁴ cells/cm² (10⁵ cells/well). Cells were harvested at day 3, washed in PBS, fixed in 70% ethanol, treated with RNase A (1 mg/ml final concentration), and stained with propidium iodide (PI, Roche, 50 µg/ml final concentration). Cell cycle analysis was performed using FACScan. Cell cycle fit software (MacCycle-Phoenix Flow Systems) was used to estimate the percentages of cells in various stages of the cell cycle.

Microarray experiments

For microarray experiments we used all 6 ESC cell lines and 6 EGC lines in 3 culture conditions (LIF+, LIF-, and RA+) (See Table 1 for the list of cell lines). Experiments in the standard LIF+ condition were carried out with 2 to 3 biological replications, and experiments in the LIF- and RA+ conditions were done without replications. At the day 3, TrizolTM (1 ml/well; Invitrogen, USA) was added to the well and total RNAs were extracted using Phase lock gelTM columns (Eppendorf/Brinkman) according to the manufacturer's protocol. Total RNAs were precipitated with isopropanol, washed with 70% ethanol, and dissolved in DEPC-treated H₂O. 2.5 µg of total RNA samples were labeled with Cy3-CTP using a Low RNA Input Fluorescent Linear Amplification Kit (Agilent, USA). A reference target (Cy5-CTP-labeled) was prepared from the Universal Mouse Reference RNA (Stratagene, USA). Labeled targets were purified using an RNeasy Mini Kit (Qiagen, USA) according to the Agilent's protocol, quantitated by a NanoDrop scanning spectrophotometer (NanoDrop Technologies, USA), and hybridized to the NIA Mouse 44K Microarray v2.2 (whole genome 60-mer oligo; manufactured by Agilent Technologies, #014117) (Carter et al., 2005) according to the Agilent protocol (G4140-90030; Agilent 60-mer oligo microarray processing protocol - SSC Wash, v1.0). All hybridizations were carried out in the two color

protocol by combining one Cy3-CTP-labeled experimental target and Cy5-CTP-labeled reference target. Microarrays were scanned on an Agilent DNA Microarray Scanner, using standard settings, including automatic PMT adjustment.

Statistical analysis of microarrays

The data discussed in this publication have been deposited in NCBI Gene Expression Omnibus (GEO; Barrett et al., 2007; <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE5914. The data and analysis software is also available at the NIA ANOVA tool website (Sharov et al., 2005b; <http://lgsun.grc.nia.nih.gov/ANOVA/>). Differential gene expressions in various cell lines in the standard culture condition were analyzed using the NIA Array Analysis software (Sharov et al., 2005b), which implements ANOVA statistics with two additional methods to reduce the number of false positives: (1) small error variances were replaced with the average error variance estimated from 500 genes with similar signal intensity, and (2) false discovery rates ($FDR \leq 0.05$) were used to select genes with differential expression, instead of p-values (Benjamini and Hochberg, 1995). Mean center global adjustment was used to remove differences between 2 batches of replications, which were done with the same set of cell lines. Experiments without replication (LIF+, LIF-, and RA+ experiments that were done simultaneously) were analyzed with a 2-factor ANOVA (factorial design) with cell line ($n = 12$) and treatment ($n = 3$) as factors. The variance associated with interaction of factors was treated as error variance. This analysis may be biased towards the increased stringency, because some interaction terms may be biologically relevant, but the stringency was not compromised by these assumptions. Error variance adjustment and FDR estimation were done as described above.

Principal component analysis (PCA) of log-transformed gene expression levels was done using the SVD method within the NIA Array Analysis software (Sharov et al., 2005b) based on averages for each cell line. Individual replications were projected on the principal components and then ANOVA was used to

estimate standard errors. Genes contributing to each principal component were selected based on the following criteria: (1) significant differential expressions between cell lines ($FDR \leq 0.05$), (2) high correlation ($abs(r) \geq 0.7$), and (3) fold change in the direction of the principal component (fold change > 1.5). Because we wanted to separate the effects of cell lines from those of culture conditions (LIF+, LIF-, and RA+), we carried out PCA for the log-expression of genes in the standard (LIF+) condition, and for the centered log-expression obtained by subtracting cell line-specific average log-expression in all 3 culture conditions.

Signature genes for EGC and ECS lines and those for 129 and C57BL/6 strains were selected based on the ANOVA of grouped replications from all cell lines in the standard LIF+ culture condition ($FDR \leq 0.05$) with additional filtering based on: (1) significant differential expressions between all cell lines; (2) fold change > 1.5 ; and (3) the presence of proper gene symbols. Even if a gene fell into both EGC-ESC and 129-C57BL/6 categories, it was included in only one category with stronger effects. Genes that were differentially expressed between EGC and ESC (or between 129 and C57BL/6) only in the RA+ condition were selected based on the following criteria: (1) differential expressions in the RA+ condition was significant ($FDR \leq 0.05$, fold change > 1.5), and (2) the difference in ESC-EGC (or 129-C57BL/6) log-ratios between RA+ and LIF+ conditions was > 1.76 (1.5 fold change).

Analysis of selected genes by gene ontology (GO) terms was carried out using the hypergeometric distribution ($FDR = 0.1$) and enrichment ratio (> 1.5) as significance criteria with the NIA Mouse Gene Index (ver. mm7) software (Sharov et al., 2005a). Only non-redundant genes with gene symbols were used for analysis. Because many GO categories are redundant (i.e., contain similar lists of genes), we adjusted the calculation of FDR in the following manner. First, we identified all redundant pairs of GO terms that had a correlation of gene presence ≥ 0.7 . Second, in the list of GO terms ordered by increasing

p -values, p_i , estimated from the hypergeometric distribution, GO term i was considered redundant if it was redundant to at least one preceding term. Finally, the FDR was:

$$FDR_i = \frac{p_i \cdot N}{n_i},$$

where N is the total number of non-redundant GO terms, and n_i is the number of non-redundant GO terms from the start of the list till term i . This is a modification of the standard equation for FDR (Benjamini and Hochberg, 1995).

Real-time quantitative RT-PCR

Primer sets for quantitative reversetranscriptase PCR (qRT-PCR) were designed and tested for SYBR (Supplemental Table S14). Green chemistry using an established in-house protocol (Carter et al., 2003). Total RNA was used to prepare cDNA as described previously (Carter et al., 2003). Because the microarray targets were oligo(dT)-primed, all cDNA synthesis reactions were oligo(dT)-primed as well, and qRT-PCR primer sets were designed so that PCR-amplicons were matched to the region upstream of microarray 60-mer oligo probes when possible, or to the region within the 650 bp downstream. These steps were taken to minimize possible bias due to the different locations of amplicons and microarray oligos. Reactions were run on ABI 7900 HT Sequence Detection Systems using the default cycling program, and data were processed using SDS 2.2 software (Applied Biosystems).

Acknowledgements

We would like to thank Drs. Grant McGregor and Hitoshi Niwa for providing ES cell lines. We would also like to thank Ms. Christa M. Morris and Dr. Robert Wersto for performing the flow cytometry analysis, and Mr. Dawood Dudekula for help with bioinformatics analysis. The work was supported by the Intramural Research Program of National Institute on Aging, NIH.

Figure legends

Fig. 1. Phenotypes of representative pluripotent cell lines cultured on gelatin-coated dishes in the standard LIF+ condition, LIF- condition, and RA+ condition for 3 days. (A) Morphologies of cells stained with alkaline phosphatase (ALP). (B) Cell cycle structures analyzed by the Flow Cytometry.

Fig. 2. Principal component analysis (PCA) of global gene expression patterns in pluripotent cell lines (ESC and EGC), neural stem/progenitor cells (NSC), and trophoblast stem (TS) cells. Cells were plotted according to their coordinates on the principal component 1 (PC1) and the principal component 2 (PC2).

Fig. 3. Principal component analysis (PCA) of global gene expression patterns in pluripotent cell lines in the standard LIF+ culture condition. Cells were plotted according to their coordinates on the PC1+0.69·PC2 and PC3. These principal components (PCs) best separated pluripotent cells according to their strains (C57BL/6 vs. 129) and origins (EGC vs. ESC). Error bars were determined with ANOVA. Major Gene Ontology (GO) categories of genes that were overexpressed in each cell type are shown together with representative genes for transcription factors, growth factors, and hormones. Pie-charts show the proportion of “pluripotency-related genes” (see the text for the definition) in sets of genes that were positively or negatively correlated with each PC.

Fig. 4. Average expression levels of signature genes can separate pluripotent cells in the standard LIF+ culture condition according to their origins and strains. (A) Cells were plotted according to the average log-transformed expression levels of 20 ESC-signature genes and 84 EGC-signature genes. (B) Cells were plotted according to the average log-transformed expression levels of 100 129-signature genes and 100 C57BL/6-signature genes.

Fig. 5. Expression levels of selected genes measured by qRT-PCR. All 16 ESC and EGC cell lines were used to validate the microarray data obtained from the 12 subset lines. (A) Average expression levels of 6 selected genes in ESC and EGC cell lines in the standard LIF+ culture condition. The expression levels in EGC cell lines were normalized by those in ESC cell lines and represented as relative expression levels (fold difference). (B) Average expression levels of 10 selected genes in the cell lines derived from C57BL/6 and 129 mouse strains in the standard LIF+ culture condition. The expression levels in C57BL/6-derived lines were normalized by those in 129-derived lines and represented as relative expression levels (fold difference). (C) Average expression levels of 8 selected genes in ESC and EGC cell lines in the RA+ culture condition. The expression levels in EGC cell lines were normalized by those in ESC cell lines and represented as relative expression levels (fold difference).

Fig. 6. Principal component analysis (PCA) of global gene expression patterns of pluripotent cell lines in the standard LIF+, LIF-, and RA+ conditions. Gene expression levels were centered by subtracting cell line-specific average for all culture conditions. Cells were plotted according to their coordinates on the PC1 and PC2.

Fig. 7. Expression levels of a set of genes that selected for the differential expressions between 129 and C57BL/6 (A) or between ESC and EGC (B) only in the RA+ culture condition. Gene expression levels were centered by subtracting cell line-specific average for all culture conditions and were represented in the heatmap (low expressions in green and high expressions in red).

Supplementary materials

Fig. S1. Examples of genes that showed cell line-specific expressions, but did not represent cell strain or origin. Expression levels of four genes (Col6a3, Ctss, Lyz, and Ugt1a6a) were measured by qRT-PCR. If we studied only one of these ESCs or EGCs cells as representative of ESCs or EGCs, these genes would have been selected as ESC- or EGC-specifically expressed genes.

Table S1. Analysis of variance (ANOVA) of log-transformed gene expression levels in pluripotent cell lines in the standard LIF+ condition.

Table S2. A list of genes whose expressions were significantly correlated with $PC1+0.69\cdot PC2$ in Fig3. These genes represented mostly differences between mouse strains.

Table S3. A list of genes whose expressions were significantly correlated with PC3 in Fig. 3. These genes represented mostly differences between EGC and ESC lines.

Table S4. A list of genes whose expressions were correlated negatively with PC1 in Fig. 2. These genes can be considered “pluripotency-related genes”.

Table S5. GO annotations overrepresented in genes whose expressions were positively correlated with $PC1+0.69\cdot PC2$ in Fig. 3. These GO categories were associated mostly with the 129 strain.

Table S6. GO annotations overrepresented in genes whose expressions were negatively correlated with $PC1+0.69\cdot PC2$ in Fig. 3. These GO categories were associated mostly with the C57BL/6 strain.

Table S7. GO annotations overrepresented in genes whose expressions were positively correlated with PC3 in Fig. 3. These GO categories were associated mostly with EGC lines.

Table S8. A list of EGC-specific signature genes.

Table S9. A list of ESC-specific signature genes.

Table S10. A list of 129-specific signature genes.

Table S11. A list of C57BL/6-specific signature genes.

Table S12. Gene expression changes in differentiation-promoting culture conditions (LIF-, RA+).

Table S13. All the data of log-transformed (\log_{10}) gene expression levels normalized by signal intensities of Universal Mouse Reference RNAs.

Table S14. Primers used for quantitative RT-PCR.

References

- Aiba K, Sharov AA, Carter MG *et al.* 2006 Defining a developmental path to neural fate by global expression profiling of mouse embryonic stem cells and adult neural stem/progenitor cells. *Stem Cells* **24**, 889-95.
- Anisimov SV, Tarasov KV, Tweedie D *et al.* 2002 SAGE identification of gene transcripts with profiles unique to pluripotent mouse R1 embryonic stem cells. *Genomics* **79**, 169-76.
- Barrett T, Troup DB, Wilhite SE *et al.* 2007 NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* **35**, D760-5.
- Benjamini Y, Hochberg Y 1995 Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society, B* **57**, 289-300.
- Boyer LA, Lee TI, Cole MF *et al.* 2005 Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56.
- Brambrink T, Hochedlinger K, Bell G *et al.* 2006 ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc Natl Acad Sci U S A* **103**, 933-8.
- Camenisch G, Gruber M, Donoho G *et al.* 1996 A polyoma-based episomal vector efficiently expresses exogenous genes in mouse embryonic stem cells. *Nucleic Acids Res* **24**, 3707-13.
- Canning J, Takai Y, Tilly JL 2003 Evidence for genetic modifiers of ovarian follicular endowment and development from studies of five inbred mouse strains. *Endocrinology* **144**, 9-12.
- Carter MG, Hamatani T, Sharov AA *et al.* 2003 In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. *Genome Res* **13**, 1011-21.
- Carter MG, Sharov AA, VanBuren V *et al.* 2005 Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. *Genome Biol* **6**, R61.
- Doetschman TC, Eistetter H, Katz M *et al.* 1985 The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27-45.
- Donovan PJ, de Miguel MP 2003 Turning germ cells into stem cells. *Curr Opin Genet Dev* **13**, 463-71.
- Durcova-Hills G, Ainscough J, McLaren A 2001 Pluripotential stem cells derived from migrating primordial germ cells. *Differentiation* **68**, 220-6.
- Durcova-Hills G, Wianny F, Merriman J *et al.* 2003 Developmental fate of embryonic germ cells (EGCs), in vivo and in vitro. *Differentiation* **71**, 135-41.
- Evans MJ, Kaufman MH 1981 Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6.
- Furusawa T, Ikeda M, Inoue F *et al.* 2006 Gene Expression Profiling of Mouse Embryonic Stem Cell Subpopulations. *Biol Reprod.*
- Gordeeva O, Zinovieva R, Smirnova Y *et al.* 2005 Differentiation of embryonic stem cells after transplantation into peritoneal cavity of irradiated mice and expression of specific germ cell genes in pluripotent cells. *Transplant Proc* **37**, 295-8.
- Guan K, Nayernia K, Maier LS *et al.* 2006 Pluripotency of spermatogonial stem cells from adult mouse testis. *Nature* **440**, 1199-203.
- Hadjantonakis A, Papaioannou V 2001 The stem cells of early embryos. *Differentiation* **68**, 159-66.

- Hanna LA, Foreman RK, Tarasenko IA *et al.* 2002 Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* **16**, 2650-61.
- Hochedlinger K, Jaenisch R 2006 Nuclear reprogramming and pluripotency. *Nature* **441**, 1061-7.
- Hollnagel A, Oehlmann V, Heymer J *et al.* 1999 Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J Biol Chem* **274**, 19838-45.
- Huntriss J, Hinkins M, Picton HM 2006 cDNA cloning and expression of the human NOBOX gene in oocytes and ovarian follicles. *Mol Hum Reprod* **12**, 283-9.
- Ivanova NB, Dimos JT, Schaniel C *et al.* 2002 A stem cell molecular signature. *Science* **298**, 601-4.
- Jiang Y, Jahagirdar BN, Reinhardt RL *et al.* 2002 Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* **418**, 41-9.
- Kanatsu-Shinohara M, Inoue K, Lee J *et al.* 2004 Generation of pluripotent stem cells from neonatal mouse testis. *Cell* **119**, 1001-12.
- Ke FC, Chuang LC, Lee MT *et al.* 2004 The modulatory role of transforming growth factor beta1 and androstenedione on follicle-stimulating hormone-induced gelatinase secretion and steroidogenesis in rat granulosa cells. *Biol Reprod* **70**, 1292-8.
- Kelai S, Hanoun N, Aufrere G *et al.* 2006 Cannabinoid-serotonin interactions in alcohol-preferring vs. alcohol-avoiding mice. *J Neurochem* **99**, 308-20.
- Kim GH, Lee DH, Kim HW *et al.* 2005 Relationships between the expressions of CDX1 and CDX2 mRNA and clinicopathologic features in colorectal cancers. *Korean J Intern Med* **20**, 317-24.
- Labosky PA, Barlow DP, Hogan BL 1994 Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* **120**, 3197-204.
- Larsson M, Stahl S, Uhlen M *et al.* 2000 Expression profile viewer (ExProView): a software tool for transcriptome analysis. *Genomics* **63**, 341-53.
- Ledermann B 2000 Embryonic stem cells and gene targeting. *Exp Physiol* **85**, 603-13.
- Li Y, McClintick J, Zhong L *et al.* 2005 Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. *Blood* **105**, 635-7.
- Loebel DA, Watson CM, De Young RA *et al.* 2003 Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev Biol* **264**, 1-14.
- Luedi PP, Hartemink AJ, Jirtle RL 2005 Genome-wide prediction of imprinted murine genes. *Genome Res* **15**, 875-84.
- Martin GR 1981 Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8.
- Masui S, Shimosato D, Toyooka Y *et al.* 2005 An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic Acids Res* **33**, e43.
- Matin A, Nadeau JH 2005 Search for testicular cancer gene hits dead-end. *Cell Cycle* **4**, 1136-8.
- Matsui Y, Zsebo K, Hogan BL 1992 Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841-7.
- McLaren A, Durcova-Hills G 2001 Germ cells and pluripotent stem cells in the mouse. *Reprod Fertil Dev* **13**, 661-4.

- Mongan NP, Martin KM, Gudas LJ 2006 The putative human stem cell marker, Rex-1 (Zfp42): Structural classification and expression in normal human epithelial and carcinoma cell cultures. *Mol Carcinog*.
- Niakan KK, Davis EC, Clipsham RC *et al.* 2006 Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. *Mol Genet Metab* **88**, 261-71.
- Pereda J, Zorn T, Soto-Suazo M 2006 Migration of human and mouse primordial germ cells and colonization of the developing ovary: an ultrastructural and cytochemical study. *Microsc Res Tech* **69**, 386-95.
- Player A, Wang Y, Bhattacharya B *et al.* 2006 Comparisons between transcriptional regulation and RNA expression in human embryonic stem cell lines. *Stem Cells Dev* **15**, 315-23.
- Rajkovic A, Yan C, Yan W *et al.* 2002 Obox, a family of homeobox genes preferentially expressed in germ cells. *Genomics* **79**, 711-7.
- Ramalho-Santos M, Yoon S, Matsuzaki Y *et al.* 2002 "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science* **298**, 597-600.
- Rebrin I, Sohal RS 2004 Comparison of thiol redox state of mitochondria and homogenates of various tissues between two strains of mice with different longevities. *Exp Gerontol* **39**, 1513-9.
- Resnick JL, Bixler LS, Cheng L *et al.* 1992 Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550-1.
- Rohwedel J, Sehlmeier U, Shan J *et al.* 1996 Primordial germ cell-derived mouse embryonic germ (EG) cells in vitro resemble undifferentiated stem cells with respect to differentiation capacity and cell cycle distribution. *Cell Biol Int* **20**, 579-87.
- Rossant J 1993 Immortal germ cells? *Curr Biol* **3**, 47-9.
- Sakhinia E, Faranghpour M, Liu Yin JA *et al.* 2005 Routine expression profiling of microarray gene signatures in acute leukaemia by real-time PCR of human bone marrow. *Br J Haematol* **130**, 233-48.
- Savatier P, Lapillonne H, van Grunsven LA *et al.* 1996 Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene* **12**, 309-22.
- Seong E, Saunders TL, Stewart CL *et al.* 2004 To knockout in 129 or in C57BL/6: that is the question. *Trends Genet* **20**, 59-62.
- Shamblott MJ, Axelman J, Wang S *et al.* 1998 Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci U S A* **95**, 13726-31.
- Sharov AA, Dudekula DB, Ko MS 2005a Genome-wide assembly and analysis of alternative transcripts in mouse. *Genome Res* **15**, 748-54.
- Sharov AA, Dudekula DB, Ko MS 2005b A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics* **21**, 2548-9.
- Sharov AA, Piao Y, Matoba R *et al.* 2003 Transcriptome analysis of mouse stem cells and early embryos. *PLoS Biol* **1**, E74.
- Sligh JE, Levy SE, Waymire KG *et al.* 2000 Maternal germ-line transmission of mutant mtDNAs from embryonic stem cell-derived chimeric mice. *Proc Natl Acad Sci U S A* **97**, 14461-6.
- Smith AG 2001 Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* **17**, 435-62.
- Smith FM, Garfield AS, Ward A 2006 Regulation of growth and metabolism by imprinted genes. *Cytogenet Genome Res* **113**, 279-91.

- Stewart CL, Gadi I, Bhatt H 1994 Stem cells from primordial germ cells can reenter the germ line. *Dev Biol* **161**, 626-8.
- Tada T, Tada M, Hilton K *et al.* 1998 Epigenotype switching of imprintable loci in embryonic germ cells. *Dev Genes Evol* **207**, 551-61.
- Tanaka TS, Kunath T, Kimber WL *et al.* 2002 Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. *Genome Res* **12**, 1921-8.
- Thomson JA, Itskovitz-Eldor J, Shapiro SS *et al.* 1998 Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-7.
- Ward CM, Barrow KM, Stern PL 2004 Significant variations in differentiation properties between independent mouse ES cell lines cultured under defined conditions. *Exp Cell Res* **293**, 229-38.
- Wei CL, Miura T, Robson P *et al.* 2005 Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cells* **23**, 166-85.
- Wesselkamper SC, Prows DR, Biswas P *et al.* 2000 Genetic susceptibility to irritant-induced acute lung injury in mice. *Am J Physiol Lung Cell Mol Physiol* **279**, L575-82.
- Yoshikawa T, Piao Y, Zhong J *et al.* 2006 High-throughput screen for genes predominantly expressed in the ICM of mouse blastocysts by whole mount in situ hybridization. *Gene Expr Patterns* **6**, 213-24.
- Youngren KK, Coveney D, Peng X *et al.* 2005 The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. *Nature* **435**, 360-4.
- Zraika S, Aston-Mourney K, Laybutt DR *et al.* 2006 The influence of genetic background on the induction of oxidative stress and impaired insulin secretion in mouse islets. *Diabetologia* **49**, 1254-63.
- Zwaka TP, Thomson JA 2005 A germ cell origin of embryonic stem cells? *Development* **132**, 227-33.

Table 1. Pluripotent cell lines used in the study

Name of cell line	Mouse strain	Cell type	Feeder dependency	Sex	Source	Comments and Reference
BL6.9	C57BL/6	ESC	+	M	JHU	aka MC2-B6
CC9.3.1	129SvEv- <i>Gpi1^c</i>	ESC	+	F	Grant MacGregor	(Sligh et al., 2000); developed by Allan Bradley
129.3	129SvEv-Tac	ESC	+	M	JHU	aka MC1
ES-D3-GL	129/SV+c/+p	ESC	+	M	ATCC	Catalog No SCRC-1003 (Doetschman et al., 1985)
EBRTcH3	129Ola	ESC	-	M	Hitoshi Niwa	Derived from EB3 (E14tg2a), genetically modified (Masui et al., 2005)
MG1.19	129Ola	ESC	-	M	Hitoshi Niwa	Derived from CCE, genetically modified (Camenisch et al., 1996)
TGC 8-5*	C57BL/6	EGC	+	M	Brigid Hogan	Derived from 8 dpc embryo; methylation of Igf2r region 2 (met/unmet, partial) (Labosky et al., 1994)
TGC 8-7	C57BL/6	EGC	+	F	Brigid Hogan	Derived from 8 dpc embryo; methylation of Igf2r region 2 (unmet) (Labosky et al., 1994)
TGC 8-8	C57BL/6	EGC	+	M	Brigid Hogan	Derived from 8 dpc embryo; methylation of Igf2r region 2 (unmet) (Labosky et al., 1994)
TGC 8-9*	C57BL/6	EGC	+	M	Brigid Hogan	Derived from 8 dpc embryo; methylation of Igf2r region 2 (unmet) (Labosky et al., 1994)
TGC 8-12*	C57BL/6	EGC	+	M	Brigid Hogan	Derived from 8 dpc embryo; methylation of Igf2r region 2 (met/unmet, somatic) (Labosky et al., 1994)
TGC 8.5-5	C57BL/6	EGC	+	M	Brigid Hogan	Derived from 8.5 dpc embryo; methylation of Igf2r region 2 (met/unmet, partial) (Labosky et al., 1994)
TGC 8.5-19*	C57BL/6	EGC	+	M	Brigid Hogan	Derived from 8.5 dpc embryo; methylation of Igf2r region 2 (unmet); germ line transmission (Labosky et al., 1994)
TGC 12-8	C57BL/6	EGC	+	M	Brigid Hogan	Derived from genital ridge of 12.5 dpc embryo; methylation of Igf2r region 2 (unmet)

						(Labosky et al., 1994)
EG-1	129Sv	EGC	+	M	Colin Stewart	Derived from 9 dpc embryo (Stewart et al., 1994)
EG-3	129Sv	EGC	+	F	Colin Stewart	Derived from 9 dpc embryo (Stewart et al., 1994)

*Samples used for qRT-PCR only. All other samples were used for microarray analysis.

Table 2. Imprinted genes differentially expressed in pluripotent cell lines*

Symbol	Annotation	Maternal/ paternal	Chr	Enriched	Log ratio difference			p value***		
					LIF+	LIF-	RA	LIF+	LIF-	RA
Grb10**	growth factor receptor bound protein 10	Maternal	11a2	ESC	-0.476	-0.498	-1.012	0.000	0.000	0.000
Calcr	calcitonin receptor	Maternal	6a1	ESC	-0.297	-0.334	-0.217	0.000	0.027	0.003
Ins1	insulin I	Paternal	19d2	ESC	-0.038	-0.103	-0.205	0.144	0.008	0.000
Igf2r	insulin-like growth factor 2 receptor	Maternal	17a2	ESC	0.067	-0.027	-0.376	0.053	0.599	0.000
Slc38a4	solute carrier family 38, member 4	Paternal	15f2	ESC	0.040	-0.095	-0.210	0.321	0.089	0.000
Pon3	paraoxonase 3	Maternal	6a1	ESC	-0.029	-0.017	-0.200	0.332	0.821	0.001
Snrpn	small nuclear ribonucleoprotein N	Paternal	7b5	EGC	0.292	0.356	0.535	0.000	0.000	0.000
U2af1-rs1	U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 1, related sequence 1	Paternal	11a3	EGC	0.217	0.272	0.218	0.000	0.000	0.000
Nnat	neuronatin	Paternal	2h2	EGC	0.240	0.203	0.283	0.001	0.042	0.025
Kcnq1	potassium voltage-gated channel, subfamily Q, member 1	Maternal	7f5	EGC	0.117	0.206	0.183	0.010	0.006	0.031
H19	H19 fetal liver mRNA	Maternal	7f5	EGC	0.241	0.224	0.412	0.000	0.165	0.035
Peg3	paternally expressed 3	Paternal	7a1	EGC	0.287	0.157	0.116	0.000	0.112	0.008
Ascl2	achaete-scute complex homolog-like 2 (Drosophila)	Maternal	7f5	EGC	0.061	0.170	0.261	0.267	0.065	0.030
Peg3	paternally expressed 3	Paternal	7a1	C57BL/6	-0.375	-0.214	-0.103	0.000	0.016	0.033
Dcn	decorin	Maternal	10c3	C57BL/6	-0.139	-0.193	-0.105	0.005	0.006	0.391
Asb4	ankyrin repeat and SC	Maternal	6a1	C57BL/6	-0.257	-0.198	-0.120	0.000	0.226	0.329
Impact	imprinted and ancient zinc finger protein	Paternal	18a2	129	0.285	0.299	0.267	0.000	0.011	0.035
Zfp264	264	Paternal	7a1	129	0.175	0.094	0.296	0.000	0.084	0.000
Calcr	calcitonin receptor	Maternal	6a1	129	0.085	0.374	0.159	0.341	0.010	0.071
Dlk1	delta-like 1 homolog (Drosophila)	Paternal	12f2	129	0.021	0.116	0.228	0.668	0.142	0.015
Gtl2	GTL2, imprinted maternally expressed untranslated mRNA	Maternal	12f2	129	0.332	0.262	0.268	0.022	0.323	0.266

*Genes were differentially expressed ($p < 0.05$ and > 1.5 fold change) at least in one culture condition.

**Bold face of a gene symbol indicates significant differences ($p < 0.05$) in all culture conditions.

***Bold face of p-value indicates a significant difference in expression ($p < 0.05$).

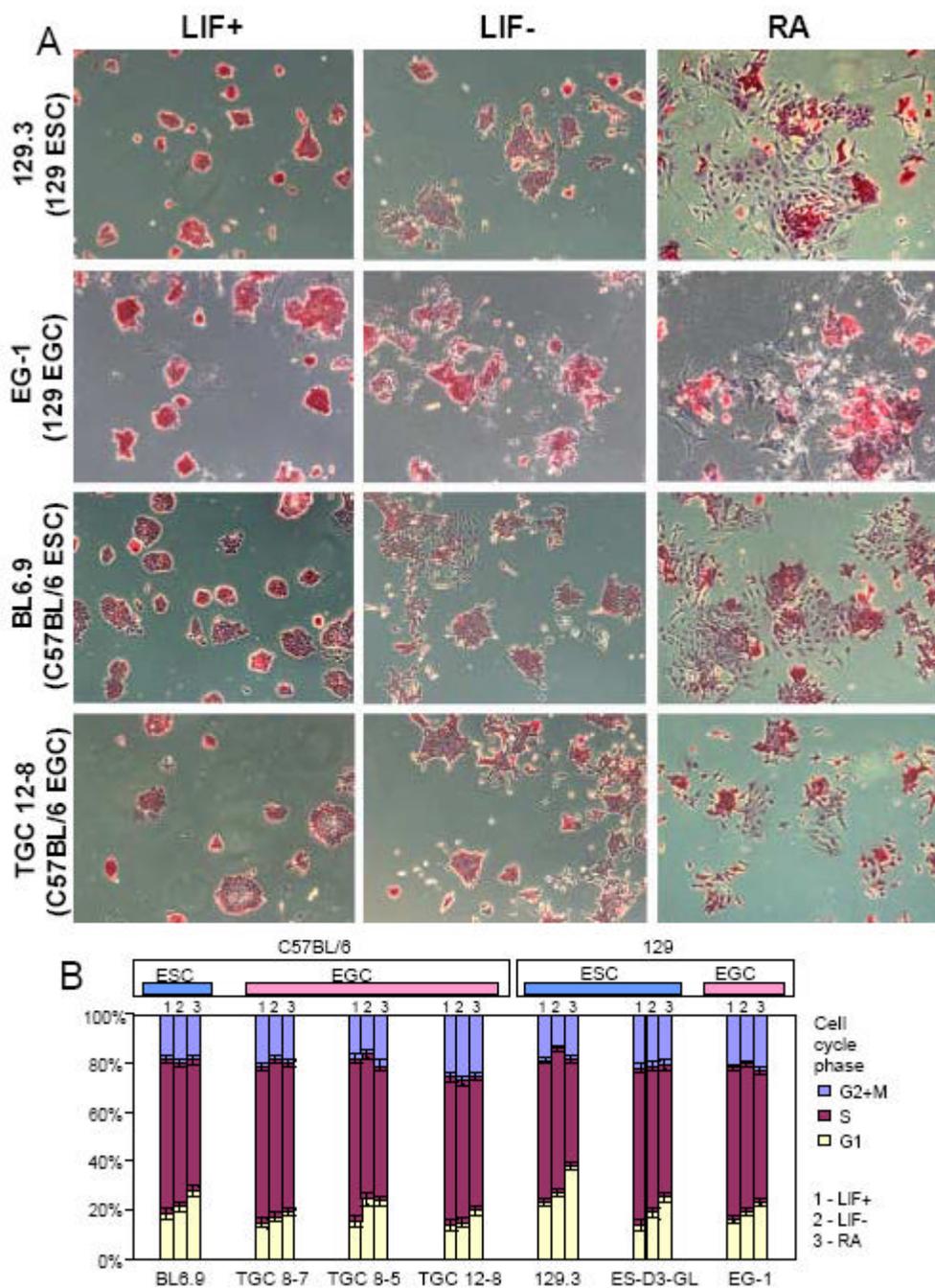
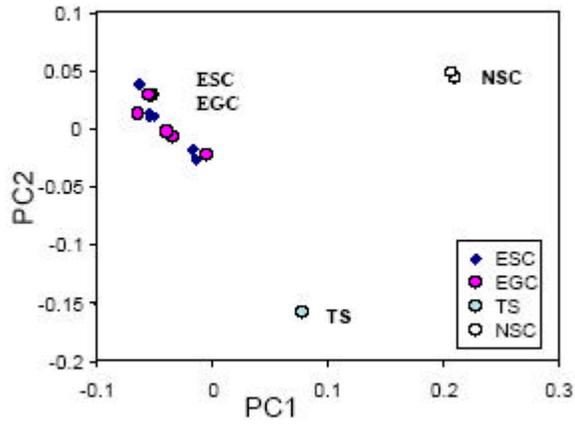
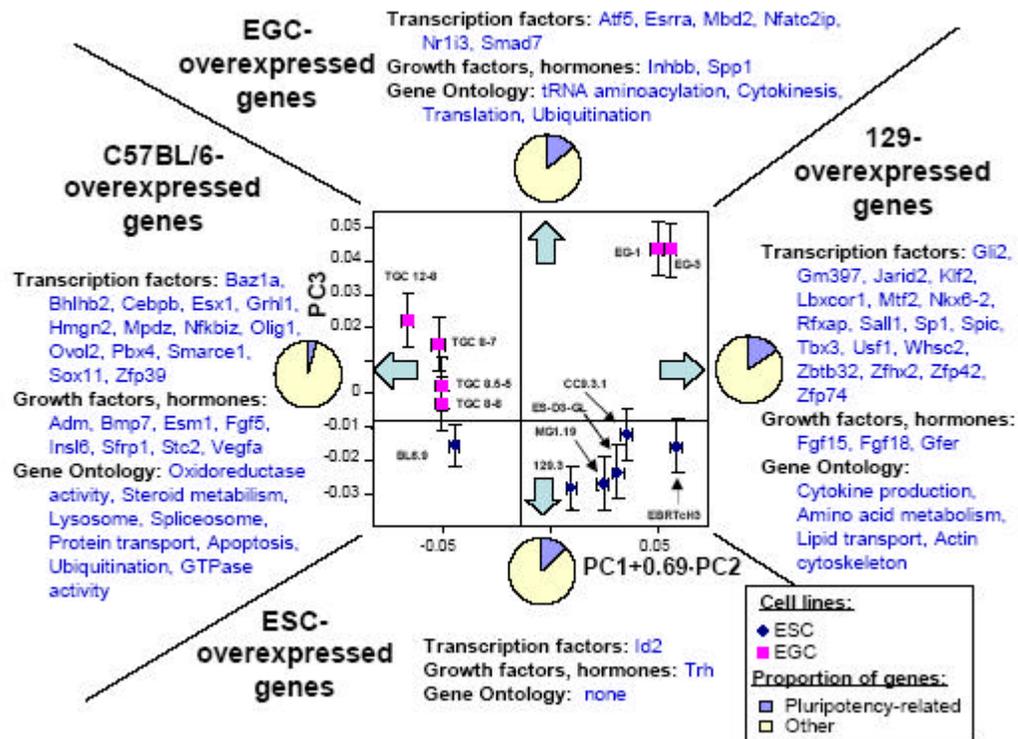
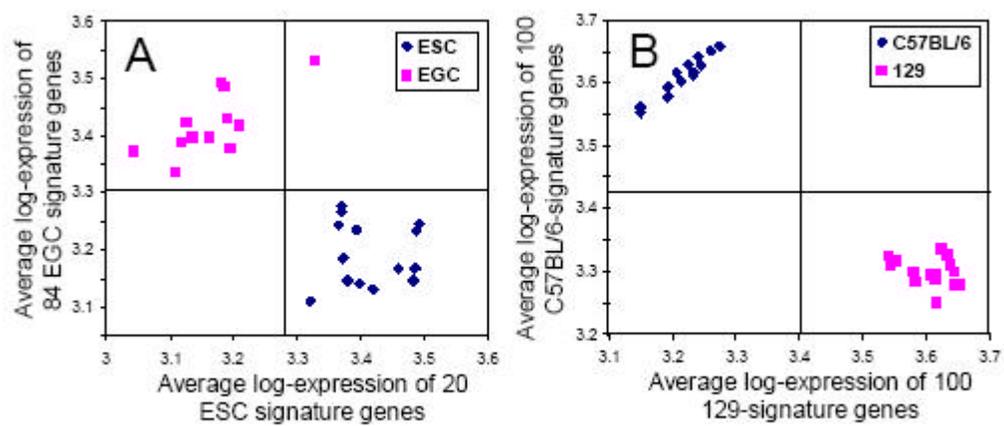


Fig. 1

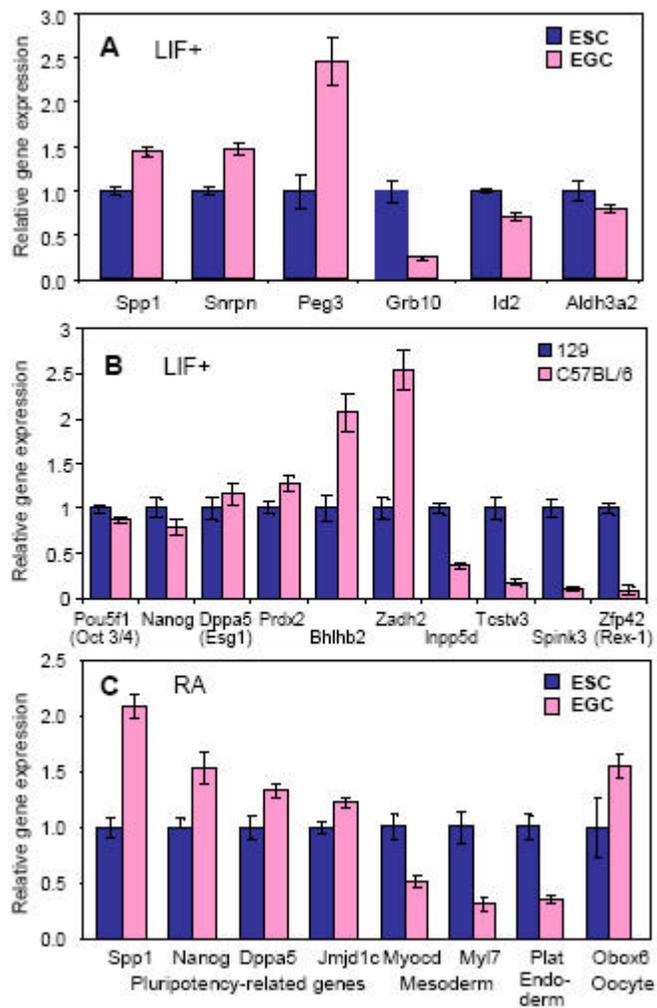


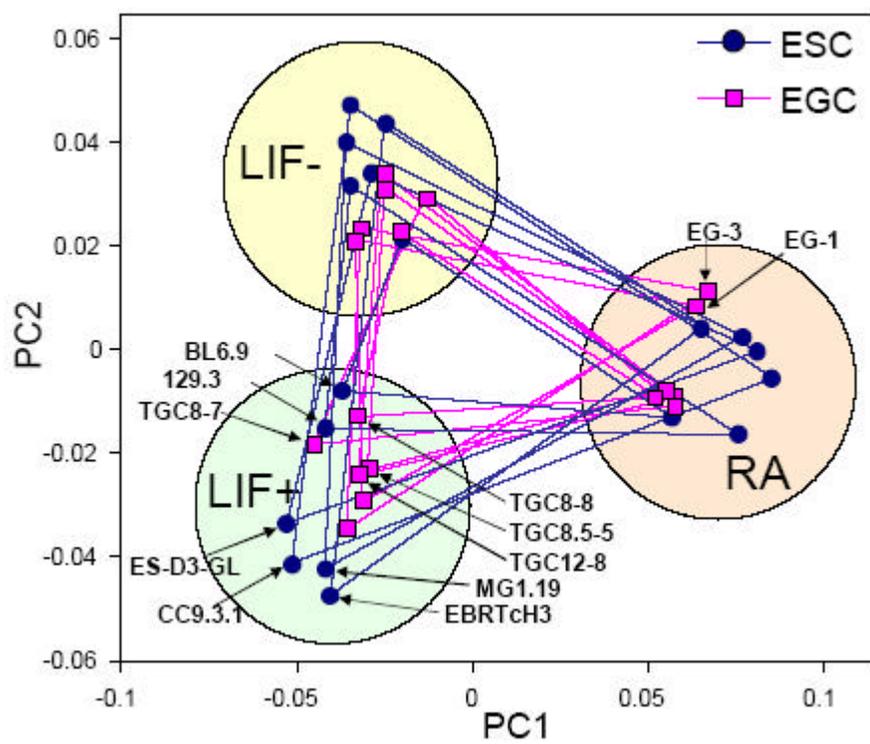
ACCEPTED MANUSCRIPT





ACCEPTED MANUSCRIPT





ACCEPTED

