Published in final edited form as:
Genomics. 2007 January ; 89(1): 22–35. doi:10.1016/j.ygeno.2006.09.010.

Cross Species Transcriptional Profiles Establish a Functional Portrait of Embryonic Stem Cells

Yu Sun[1], **Huai Li**[1], **Ying Liu**[2], **Soojung Shin**[2], **Mark P. Mattson**[3], **Mahendra S. Rao**[2,4,*], and **Ming Zhan**[1,*]

1 *Bioinformatics Unit, Research Resources Branch, National Institute on Aging, NIH, Baltimore, MD, 21224, USA*

2 *The CRL, Invitrogen Corp, 1620 Faraday Ave, Carlsbad, CA 92008, USA*

3 *Laboratory of Neurosciences, National Institute on Aging, NIH, Baltimore, MD, 21224, USA*

4 *Neurosciences Program, Johns Hopkins University School of Medicine, Baltimore, MD 21224, USA*

Abstract

An understanding of the regulatory mechanisms responsible for pluripotency in embryonic stem cells (ESCs) is critical for realizing their potential in medicine and science. Significant similarities exist among ESCs harvested from different species, yet major differences have also been observed. Here, by cross-species analysis of a large set of functional categories and all transcription factors and growth factors, we revealed conserved and divergent functional landscapes underlining fundamental and species-specific mechanisms that regulate ESC development. Global transcriptional trends derived from all expressed genes, instead of differentially expressed genes alone, were examined, allowing for a higher discriminating power in the functional portrait. We demonstrate that cross-species correlation of transcriptional changes that occur upon ESC differentiation is a powerful predictor of ESC-important biological pathways and functional cores within a pathway. Hundreds of functional modules, as defined by *Gene Ontology*, were associated with conserved expression patterns but bear no overt relationship to ESC development, suggestive of new mechanisms critical to ESC pluripotency. Yet other functional modules were not conserved; instead, they were significantly up-regulated in ESCs of either species, suggestive of species-specific regulation. The comparison of ESCs across species and between human ESCs and embryonal carcinoma stem cells (ECCs) suggest that while pluripotency as an essential function in multicellular organisms is conserved through evolution, mechanisms primed for differentiation are less conserved and contribute substantially to the differences among stem cells derived from different tissues or species. Our findings establish a basis for defining the “stemness” properties of ESCs from the perspective of functional conservation and variation. The data and analyses resulting from this study provide a framework for new hypotheses and research directions, and a public resource for functional genomics of ESCs.

Keywords

embryonic stem cell; embryonic bodies; comparative transcriptomics; functional genomics; LIF; FGF; Nodal; Wnt; BMP; TGF; β

* To whom correspondence should be addressed: National Institute on Aging, NIH, 333 Cassall Drive, Baltimore, MD 21224, Tel: (410)-558- 8373 FAX: (410)-558- 8674, E-mail: zhanmi@mail.nih.gov; rao@invitrogen.com.

INTRODUCTION

Embryonic stem cells (ESCs) are pluripotent cells with indefinite replication potential and the ability to differentiate into all types of cells. As demonstrated in various *in vitro* studies, ESCs hold great promise for regenerative medicine and science [1;2;3;4;5]. Understanding the regulatory mechanism responsible for pluripotency in ESCs is of central importance towards realizing their therapeutic and scientific potential. Previous investigations examining differentially expressed genes during ESC differentiation have begun to elucidate the pathways and networks controlling ESC self-renewal and differentiation [6;7;8;9]. However, gene expression analyses focused primarily on fold changes often overlook relatively low, yet concerted changes in expression levels of genes; these changes in coordinated gene expression, despite being subtle, may dramatically alter the flux of pathways and their overall active or inactive states [10;11;12]. Moreover, functional profiling by differentially expressed genes is often limited to a relatively small number of functional modules that the genes are associated with. A global trend analysis of gene expression by employing all expressed genes allows for a greater discriminative power in functional profiling, even when the transcriptional changes on individual genes are rather small [10;11;12;13]. However, no large scale functional profiling has been conducted on ESCs with global patterns of gene expression.

The first ESC line was derived from the inner cell mass (ICM) of pre-implantation mouse embryos [14;15]. Since then, ESCs have been isolated from various organisms including human [16;17;18]. While significant similarities exist among ESCs harvested from different species, major differences have also been reported [19;20;21]. Important biological processes or pathways, like important sequence elements in the genome, are often evolutionarily conserved [19;22;23]. Comparative transcriptomics analysis has been successful in identifying functional modules conserved through evolution [10;22;24]. The observed conservation and variation among ESCs suggest that cross-species gene expression analysis and functional profiling may help distinguish between fundamental and species-specific mechanisms controlling ESC development and define the core “stemness” properties.

In this study, we employed a global gene expression analysis for a large scale cross-species functional profiling of ESCs. We systematically surveyed all biological processes, molecular functions, and cellular components determined by the Gene Ontology (GO), pathways defined by Kyoto Encyclopedia of Genes and Genomes (KEGG), and transcription factors and growth factors. In each functional module, we examined all expressed genes, instead of differentially expressed genes alone, to explore not only major but also subtle yet important transcriptional changes that occur upon ESC differentiation. Evolutionarily conserved and divergent functional modules were identified from human and mouse ESCs. The conserved modules, showing positive cross-species correlation of the expression pattern, represent core biological networks or mechanisms fundamental for ESC development. The divergent modules, showing negative or no correlation on the expression pattern, are suggestive of species-specific regulation in ESCs. The cross-species correlation of expression pattern is a powerful predictor of ESC-important biological pathways and functional cores within a pathway. Hundreds of functional modules, as defined by GO, were associated with conserved expression patterns but bear no overt relationship to ESC development, suggestive of new mechanisms critical to ESC pluripotency. The comparison of ESCs across species and between human ESCs and and embryonal carcinoma stem cells (ECCs) suggest that while pluripotency as an essential function in multicellular organisms is conserved through evolution, mechanisms primed for differentiation are shown to be less conserved and contribute substantially to the differences among stem cells derived from different tissues or species. The study establishes a basis to define the “stemness” properties by conserved and divergent functional landscapes of ESCs.

The data and analyses resulting from this study provide a framework for new hypotheses and research directions, and a public resource for functional genomics of ESCs.

MATERIALS AND METHODS

Human ESC culture

The human ESC lines used in this study included: BG01 (46 chromosomes with XY), BG01V (49 chromosomes with XXY, +12, +17), BG02 (46 chromosomes with XY), and BG03 (46 chromosomes with XX), which are registered with the NIH (http://stemcells.nih.gov/index.asp). Each cell line was provided as one sample and another sample was prepared by pooling different cell lines. Human ESCs were primarily cultured on mitotically inactivated mouse embryonic fibroblasts (MEF) or HS27 human fibroblast cells (HS27, ATCC) in DMEM/F12-Glutamax supplemented with 20% Knockout Serum Replacement, 0.1 mM nonessential amino acids, 100 μM β-mercaptoethanol, 50 μg/ml Pen-Strep (all from Invitrogen), and 4 ng/ml human recombinant basic fibroblast growth factor (PeproTech). Cultures were expanded by passaging clumps of ~10–100 cells. To remove contaminating feeder cells, hESCs were replaced on extracellular matrix of Matrigel (BD Biosciences) or fibronectin (Sigma) for at least three times before RNA extraction. During feeder-free culture, human ESC medium was pre-conditioned with inactivated MEF [25] for one day. Supplementary Figure S1 shows the cultured human ES cells. A detailed manual of the human ES cell described here is available online at the NIH stem cell web site (http://stemcells.nih.gov/index.asp).

Differentiation of ESC as embryoid bodies

For embryoid body formation, hESCs were harvested as clumps by collagenase (1 mg/ml, Invitrogen) and resupsended in DMEM/F12 supplemented with 15% FCS, 5% KSR, 2 mM L-Glutamine, 50 μg/ml Pen-Strep, 0.1 mM β-mercaptoethanol, and 0.1 mM non-essential amino acids. Medium was changed every other day and RNA was extracted at day 14 after EB formation. Supplementary Figure S1 shows the generated embryoid bodies.

NTera2 cell line

The NTera2 cell line was purchased from American Type Culture Collection (ATCC) and cultured in parallel with hESCs using protocols described in [26]. Briefly, NTera2 were cultured as monolayer on standard tissue culture plasticware in DMEM/F12 supplemented with 10% FCS and 2 mM L-glutamine. Cells were passaged every 3 to 5 days upon confluence. Differentiation of NTera2 cells were induced by co-culture with the mouse stromal cell line PA6 or with mitotically inactivated PA6 cells [26]. Two samples of differentiated NTera2 cells were prepared either with mitotically inactived or with active PA6 cells for this study. One sample was induced by the mouse stromal cell line PA6, another with mitotically inactivated PA6 cells. Since their expression patterns showed strong correlation ($r = 0.987$), they were used as replicates in the analysis. Supplementary Figure S1 shows the cultured undifferentiated and differentiated NTera2 cells.

RNA extraction, BeadArray preparation, and data processing

RNA was extracted using a standard TriZol (Invitrogen) method. The Illumina HumanRef-8 BeadArray was used in this study, which contained 23,584 probes, representing 20,692 genes recognized by RefSeq (http://ncbi.nih.gov/RefSeq/). Each gene or transcript was represented on the BeadArray by 3–10 oligonucleotides, each 50-base long. Arrays were scanned with the Illumina Beadarray Reader confocal scanner according to the manufacturer's instructions. The Illumina BeadStudio software was used to perform background correction and generate absolute expression estimates. Details of the RNA amplification, labeling, and hybridization

steps are available from www.Illumina.com. Low expression values (< 50) were substituted with 50. If expression values of a gene were all less or equal to 50, it is removed from further analysis. The mean intensity of an individual probe was calculated across all arrays, normalized by the quantile method [27], and the $\log_2$ ratio of each value to this mean was calculated. When several probes corresponded to the same gene (*i.e*. if different probes had the same gene symbol or GenBank ID), a single probe was kept for the analysis.

Cross-species comparison of gene expression pattern

The mESC dataset used in this study was obtained from the GEO database (GSE2375) and contained expression profiles of the R1 cell line and differentiated EB counterpart. To compare gene expression pattern between different organisms, we implemented a method similar to that previously described [10]. The list of human and mouse orthologous genes was obtained from the Affymetrix human-mouse orthologs link and the Inparanoid database (http://inparanoid.cgb.ki.se/). From the hESC and mESC datasets, we retrieved 7041 human - mouse orthologous gene pairs. After removing transcripts with low signal levels, we obtained 5459 pairs of orthologous genes for our analysis. Orthologous gene pairs were grouped by their corresponding GO categories. For each group of orthologous genes, evolutionary conservation on expression pattern is assessed by the correlation of the log-transformed (base 2) relative change in expression (condition 1 divided by condition 2) of the orthologous genes under different conditions. The probability of obtaining a Pearson correlation (r) from a group of n pairs of orthologous genes is computed by treating $\frac{\sqrt{(n-2)} \times r}{\sqrt{(1-r^2)}}$ as coming from a t distribution with n-2 degree of freedom [28].

Identification of differential expressed genes and principle component analysis

Differentially expressed genes between ESCs and EBs were identified by the paired t test. Differentially expressed genes between undifferentiated and differentiated NTera2 cells were identified using the local pooled error (LPE) algorithm [29], which is particularly useful for dealing with data of small sample sizes. The analyses were performed using bioconductor (www.bioconductor.org).

Significantly enriched GO terms

The Fisher's exact test was conducted to calculate the hypergeometric probability of observing a GO term as enriched in each group of genes. In specific, the probability p that a GO term is significantly enriched in a group of genes was calculated as: $p=1-\sum_{i=0}^{k-1}\frac{\binom{A}{i}\binom{G-A}{n-1}}{\binom{G}{n}}$, where k is the number of genes in the group, G is the total number of genes, n is the number of genes in the group with a given GO term, and A is the total number of genes with a given GO term.

RESULTS

We determined gene expression profiles of multiple human embryonic stem cell (hESC) lines and their differentiated cell counterparts, embryoid bodies (EBs), using Illumina BeadArray. Supplementary Figure S1 shows the cultured hESC and hEB cells. The array contained 23,584 probes, representing 20,692 unique genes. We obtained the mouse embryonic stem cell (mESC) dataset from the GEO database (GSE2375). The mESC dataset contained expression profiles of the R1 cell line and differentiated EB counterpart. From the hESC and mESC datasets, we selected 5459 human - mouse orthologous gene pairs that show significant signal

levels for our analyses. The hESC samples had correlation coefficient values greater than 0.96 with each other, while the hEB samples had pair-wise correlation values greater than 0.92. The mESC and mEB samples had correlations greater than 0.99 and 0.93, respectively. Many differentially expressed genes have been identified in ESCs and found to be important to the self-renewal or differentiation [7; 8; 13; 30; 31; 32; 33; 34]. However, small or subtle changes in the expression can sometime dramatically alter the flux of pathways and their overall active or inactive states [10; 11; 12]. A less than 2-fold increase Oct3/4, for example, can lead to differentiation of mouse ESCs to primitive endoderm and mesoderm, while a repression of Oct3/4 results in loss of pluripotency and differentiation to trophectoderm [35]. It is important to capture such relatively small yet significant expressional changes, which are however often overlooked by analysis focusing primarily on fold changes. In this study, we addressed this problem by including all expressed genes, regardless of their relative levels of changes, and analyzing the global trend of expression pattern across different functional sets in human and mouse. First, we summarized the expression change of a given gene *i* as the ratio of average expression levels before and after differentiation (represented by x_i). For a functional module with *n* genes (e.g. 84 genes of the Wnt pathway), the expression pattern of the module is represented as a vector of relative changes in expression across the *n* genes (represented by $\mathbf{x} = [x_1, x_2, \ldots x_n]$). We then assessed the evolutionary conservation of the functional module on the expression pattern based on the human - mouse orthologous gene pairs by calculating the Pearson correlation (r) between the vectors $\mathbf{x}_h$ (expression profile vector in human) and $\mathbf{x}_m$ (expression profile vector in mouse). The significance level of the observed correlation, measured by the P-value, is dependent on both the correlation value (*r*) and the group size (*n*). Therefore, a large group of genes with a relatively low correlation in the expression pattern could still be significant, and the method would thus help reveal subtle or hidden changes and a global trend of conservation.

Our results showed that the global cross-species correlation on the expression pattern over all expressed genes was 0.301. The correlation was highly significant statistically ($P < 10^{-114}$). This indicates that despite profound differences between hESCs and mESCs, a certain number of orthologous genes behave similarly upon ESC differentiation between human and mouse. We then conducted a systematic survey across all functional modules defined by the GO and KEGG databases, and examined all transcription factors and growth factors to identify the categories which contributed significantly to the observed global conservation of expression patterns across species. Three different types of functional modules were identified through the survey: 1) the expression patterns were positively correlated between human and mouse ESCs ($P < 0.05$); 2) the expression patterns were not correlated ($P \geq 0.05$); and 3) the expression patterns were negatively correlated ($r < 0$, $P < 0.05$). The functional modules of the first type were considered to be evolutionarily conserved, while those in the second and third types were evolutionarily divergent. The results are summarized in Table 1 and details are provided in Supplementary Tables S1 and S2. Out of the 903 GO-defined biological processes to which all human - mouse orthologous genes from our datasets were associated, 43.1% (i.e. 389 biological processes) were evolutionarily conserved ($P < 0.05$) (Table 1; Supplementary Table S1a). The other 56.9% (514 processes) were evolutionarily divergent, showing either a negative (0.07%) or no (56.2%) cross-species correlation. The GO-defined molecular function and cellular component exhibited a similar distribution of conserved and divergent groups (Table 1; Supplementary Tables S1b and S1c). Among all KEGG-defined pathways to which the human - mouse orthologous genes were associated, 15 showed a positively correlated expression pattern, 6 showed a negative correlation, while the remaining 88 showed a noncorrelated expression pattern (Table 1; Supplementary Table S2). Figure 1 depicts expression patterns of some conserved and divergent functional modules which are important to ESC development. As shown by the heatmap in the figure, each column represents an orthologous gene expressed in hESCs (top row) and mESCs (bottom row). The expression fold-change of a gene is represented by different colors (green: down-regulated in ESCs, red:

up-regulated in ESCs, black: no change between ESCs and EBs). The conserved modules (Figure 1a) were associated with high correlation values (r), in which many orthologous genes showed expression in the same direction (e.g. either up-regulated in both species or down-regulation in both) with similar expression fold-changes in human and mouse. The divergent modules (Figure 1b) were associated with low or negative correlation values (r), in which most orthologous genes showed expression in opposite directions (e.g. up-regulated in one species and down-regulated in another) with dissimilar expression fold-changes between human and mouse. Detailed information of each module in Figure 1 is provided in Supplementary Table S3.

1. Conserved Functional Modules

All conserved functional modules identified in this study are listed in supplementary Table S1 for GO categories and Table S2 for KEGG pathways (denoted by [E] on the term names). The conserved biological processes (GO) were classified into several major categories, shown by the pie chart in Figure 2. As illustrated, 13.8% of the conserved biological process were involved in development, such as cellular morphogenesis (97 orthologous genes, $r = 0.325$, $P < 0.001$), embryonic development (87 orthologs, $r = 0.318$, $P < 0.005$), pattern specification (50 orthologs, $r = 0.57$, $P < 10^{-5}$), and determination of left/right symmetry (6 orthologs, $r = 0.79$, $P < 0.05$) (Figure 2). The conserved functional modules also included establishment and/or maintenance of chromatin architecture (1% of all conserved processes, including chromosome remodeling; 67 orthologs, $r = 0.505$, $P < 10^{-5}$), responses to stimulus (7.2%, including response to damaged DNA and unfolded proteins), cell proliferation and cell cycle (6.2%), signal transduction (4.1%), apoptosis (1.8%), and metabolism (44.1%, including DNA replication and packaging) (Figure 2). Fifteen pathways defined by KEGG were conserved (Supplementary Table S2), including TGF-β (52 orthologs, $r = 0.411$, $P < 0.005$), Wnt (84 orthologs, $r = 0.236$, $P < 0.05$), Nodal (8 orthologs, $r = 0.695$, $P < 0.05$), and integrin-mediated signaling pathways (18 orthologs, $r = 0.427$, $P < 0.05$). These and many other conserved biological process and pathways have been considered to be important for ESC development [36; 37; 38; 39]. The conserved biological processes and pathways identified in our study, together with transcriptionally conserved transcription factors and growth factors (described below), represent core biological networks and mechanisms that are fundamental for ESC self-renewal and differentiation in different species. Figure 1a shows the expression patterns of some conserved functional modules important for ESC development (detailed information of each module in the figure is given in Supplementary Table S3). Table 2a shows conserved signal transduction pathways and their expression in ESCs and EBs of human and mouse.

Because of rapid DNA replication and the absence of cell cycle checkpoints, ESCs require extra mechanisms more sensitive than those of somatic cells to ensure genomic stability and to prevent transmission of mutations to offspring cells [40]. This may be achieved by active repression of spontaneous mutations, as well as removal of damaged cells by apoptosis. We found that genes involved in the response to DNA-damaging stimuli and DNA repair had relatively low, yet significant cross-species correlation in their expression patterns ($r = 0.261$ and 0.269, respectively; $P < 0.01$ in both) (Figure 1), supporting the hypothesis that their evolutionary conservation is of fundamental importance to ESC self-renewal. In particular, the genes MLH1, MSH2, MSH3, MSH5, PMS2, and POLD3, involved in DNA mismatch repair and replicative fidelity, were highly conserved ($r = 0.912$, $P < 0.01$) and thus comprised the core functional elements in DNA repair processes of ESCs. Apoptosis showed a low but highly significant cross-species correlation on the expression pattern (orthologs = 198, $r = 0.313$, $P < 10^{-5}$) (Figure 1). Moreover, genes involved in the response to unfolded proteins and telomere maintenance showed high levels of conservation ($r = 0.413$ and 0.802, respectively, $P < 0.01$ in both). The results suggest that these biological processes critically contribute to ensuring rapid yet accurate DNA replication and chromosome integrity during ESC self-renewal.

Cell cycle control mechanisms in ES cells differ from those in somatic cells [41]. ES cells are primed for rapid cell proliferation. Both rodent and monkey ES cells achieve this by shortening the G1 phase and bypassing DNA damage G1 checkpoints [41]. Our examination of 304 human - mouse orthologous genes involved in cell cycle progression showed a significantly conserved expression pattern ($r = 0.486$, $P < 10^{-18}$). In particular, 18 genes involved in the G1/S transition of the mitotic cell cycle and 17 genes involved in the cell cycle arrest were highly conserved ($r = 0.623$ with $P < 0.005$, $r = 0.573$ with $P < 0.01$, respectively). Moreover, 171 additional genes involved in regulation of the cell cycle were also significantly conserved ($r = 0.51$, $P < 10^{-12}$) (Figure 1). The results suggest that these biological processes are fundamentally important to ESC self-renewal of different species.

The Nodal pathway is implicated in early patterning of the embryo, left-right axis specification, and ESC development [42;43;44;45]. The pathway was highly conserved ($r = 0.695$, $P < 0.05$) (Figure 1; Supplementary Table S3). Among the 9 human - mouse orthologous genes examined for this pathway, while SMAD2 and SMAD3 did not show significant transcriptional changes in both hESCs and mESCs, LEFTB and TDGF1 exhibited at least 2-fold up-regulation in both hESCs and mESCs. The coordinated up-regulation of LEFTB and TDGF1 was also observed by EST frequency analysis (Brandenberger, et al., 2004). LEFTB is an antagonist and TDGF1 acts as a co-receptor for Nodal in Nodal signaling. The conserved expression pattern, especially for the co-activators of Nodal, TDGF1 and LEFTB, indicates that the Nodal pathway is fundamental for ESC pluripotency in different species.

The Wnt pathway is active in determining the fate of stem cells and the self-renewal [46;47; 48;49]. The pathway exhibited a low yet significant level of cross-species correlation on the expression pattern ($r = 0.236$, $P < 0.05$) (Figure 1;Supplementary Table S3). Nineteen of the 84 orthologous genes examined for this pathway showed at least 1.5-fold expression changes upon ESC differentiation in both species. Among the 19 genes, 15 showed up- and down-regulation in the same direction. The 15 conserved genes included Wnt family members (WNT1, WNT2B, and WNT5B), frizzled receptors (FZD2, FZD6, and FZD8), the antagonist DKK1, and down stream members of the pathway (FRAT2, CTBP2, TP53, MYC, DVL3, RAC2, CCND2, and CCND3). The Wnt family members and frizzled receptors, as well as DKK1, FRAT2, CTBP2, DVL3, RAC2, CCND2, and CCND3, were down-regulated in both hESCs and mESCs, while TP53 and MYC were up-regulated in ESCs. The observed evolutionary conservation of the expression pattern suggests that Wnt signaling plays a fundamental role in maintaining ES cell pluripotency. However, it is important to note that 4 genes of the pathway showed transcriptional changes over 1.5 fold in opposite directions between hESCs and mESCs: CSNK2A2 (down-regulated in hESCs but up-regulated in mESCs), FZD9, SFRP1, and PRKCB1 (up-regulated in hESCs but down-regulated in mESCs). These genes represented divergent peaks on the overall conserved landscape of this pathway. This divergence may explain the differing results of the effects of Wnt on hESC self-renewal [50].

The BMP/TGF-β pathway is involved in regulating cell proliferation and differentiation, mesoderm formation and patterning, and ESC development [51;52;53;54;55]. The pathway showed a significant level of positive correlation on the expression pattern between hESCs and mESCs ($r = 0.411$, $P < 0.005$) (Figure 1;Supplementary Table S3). Eight of the 52 orthologous genes examined for this pathway showed at least 1.5-fold expression changes upon ESC differentiation, in which BMP4, BMP5, MADH6, ID2, and PITX2 were down-regulated in both human and mouse, while THBS2, MYC, and GDF3 were up-regulated in both species. This conserved expression pattern of ligands, intermediate signal transducers, and target genes suggests that BMP/TGF-β signaling is fundamentally important for ESC self-renewal and differentiation. However, there were 5 genes showing divergent expression patterns: INHBA, MADH7, THBS1, and MAPK1 were up-regulated in mESCs while down-regulated in hESCs;

INHBE was down-regulated in mESCs while up-regulated in hESCs. The pathway therefore presents a mosaic structure with conserved functional cores and divergent domains, similar to the Wnt pathway. This is consistent with recent reports that BMP supports mESC self-renewal but activin appears to do so in hESCs [56;57].

The Notch pathway helps to determine cell fate through cell-cell interactions [58] and is reportedly involved in hESC self-renewal [59]. Overall, the pathway showed a conserved expression pattern across species ($r = 0.451$, $P < 0.05$) (Figure 1; Supplementary Table S3). However, 12 of the 23 orthologous genes examined for this pathway showed less than 1.5 fold expression changes upon ESC differentiation in both human and mouse, which included the receptors NOTCH1 and NOTCH4, and the ligands DVL2 and JAG2. Only 3 genes exhibited more than 1.5-fold expression changes in both human and mouse and positively correlated expression pattern: DVL3 (down-regulated in ESCs), RBPSUH, and CTBP2 (both up-regulated in ESCs). The subtle yet conserved expression pattern seems important for this pathway's role in ESC self-renewal. The cross-species analysis is effective in exploring such moderate, though conserved transcriptional changes for functional understanding.

Integrins mediate cell adhesion to the extracellular matrix (ECM), control cell cycle progression, regulate WNT1 and BMP4 expression, and activate the mesodermal and neuroectodermal lineage for ESC differentiation [60;61]. We examined 18 human - mouse orthologous genes involved in the integrin-mediated signaling pathway. The pathway was conserved on the transcriptional changes during ESC differentiation ($r = 0.427$, $P < 0.05$) (Figure 1; Supplementary Table S3). The transcriptional conserved genes included integrin members (e.g. ITGAs, ITGBs) and intracellular signal transducers (e.g. CD47, GAB2, CIB1). Among the 7 genes encoding integrins, 6 were either slightly down-regulated in ESC or had little transcriptional changes, while ITGB5 was at least 1.5-fold up-regulated in hESC but not in mESC. GAB2 and CD47 were at least 1.5-fold down-regulated in both human and mouse ESCs. The results suggest that integrins-mediated signaling is mostly repressed in ESCs or activated in EBs and fundamental for ESC development in both human and mouse.

Many functional modules were conserved on the expression pattern but not differentially expressed, and thus not disclosed from previous studies based on differentially expressed genes alone. Totally 237 biological processes, 102 molecular functions, and 66 cellular components fell into this category of modules (Supplementary Table S1a-c, denoted with [E] without [A.h] or [B.h] on the term names). The modules included embryonic morphogenesis (orthologous pairs = 37, $r = 0.42$, $P = 0.005$), gene silencing (orthologs = 8, $r = 0.67$, $P = 0.035$), cell cycle arrest (orthologs = 17, $r = 0.57$, $P = 0.008$), G1/S transition of mitotic cell cycle (orthologs = 18, $r = 0.62$, $P = 0.003$), and aging (orthologs = 8, $r = 0.70$, $P = 0.026$), which are known to be important for ESCs. Many of the modules in this category however bear no overt relationship to ESC development. Further analysis of these modules may lead to discovery of new mechanisms critical for ESC development.

2. Divergent Functional Modules

Divergent functional modules included those showing negative or no correlation on the gene expression pattern between human and mouse ESCs. Figure 1b shows expression patterns of some divergent functional modules important for ESC development (detailed information of each module on the figure is given in Supplementary Table S3). Table 2b shows divergent signal transduction pathways and their expression in ESCs and EBs of human and mouse.

The functional modules with negatively correlated expression patterns represented extreme cases of divergent modules (Supplementary Tables S1 and S2 with [F] denoted on term names). One such module is that involved in cytokine-cytokine receptor interactions. The 69 human - mouse orthologous genes examined for this module showed a significant level of negative

correlation on the expression pattern across species ($r = -0.312$, $P < 0.005$) (Figure 1b; Supplementary Table S3). The genes CCL7, PDGFRA, IL1R1, CXCL1, and INHBA were down-regulated in hESCs but up-regulated in mESCs with at least 1.5-fold changes in both species, while the genes INHBE, CXCL12, and BMP2 were up-regulated in hESCs but down-regulated in mESCs with at least 1.5-fold changes in both species. Opposite expression patterns of cytokines in hESCs and mESCs were also observed in other studies [20]. The negatively correlated expression pattern of cytokine-cytokine receptor interactions is suggestive of unique and distinct responses of hESCs and mESCs to growth factors.

The majority of functional modules we examined, including biological process, molecular functions, and cellular components defined by GO, and pathways defined by KEGG, showed no correlation on the expression pattern across species (Table 1). From the non-conserved modules, we identified those significantly up- or down-regulated in hESCs and mESCs. The results are summarized in Table 1 and details are given in Supplementary Table S1 and S2 (denoted by either [A] or [B] without [E] on the term names). As shown in Table 1, among the 508 non-conserved biological processes (56.2% of the total examined), 135 processes had the genes significantly up-regulated and 113 processes significantly down-regulated in hESCs. The up- or down-regulated and nonconserved modules identified in this study are suggestive of species - specific mechanisms controlling ESC self-renewal and differentiation.

The LIF pathway was not conserved on the expression pattern, with non- significant cross-species correlation ($r = -0.256$, $P = 0.17$), but was significantly down-regulated in hESCs although little changed in mESCs (Figure 1b; Supplementary Table S3). This result is consistent with the fact that the pathway is not essential in human but is critical in mouse for maintaining ESC pluripotency [57; 62; 63]. Among the 16 orthologous genes examined for this pathway, LIF was expressed at a much lower level in hESCs than in mESCs, while IL11, STAT6, STAT5B, SOCS1, and SOCS2 showed highly different transcriptional changes between hESCs and mESCs.

Similarly, the FGF pathway was not conserved on the expression pattern, with non-significant cross-species correlation ($r = -0.172$, $P = 0.24$), but was significantly up-regulated in mESCs but not in hESCs (Figure 1; Supplementary Table S3). The pathway is implicated in rodent ICM development and maintaining undifferentiated ES cells in culture [64;65;66]. Among the 19 orthologous genes examined for this pathway, the MAPK inhibitors DUSP1, DUSP6, and DUSP9 were highly up-regulated (over 2 fold) in mESCs but down-regulated or unchanged in hESCs, thus highly diverged on the expression. The repressed expression of these genes in hESCs were also observed by a RT-PCR analysis [59]. The results suggest that the FGF pathway acts differently between human and mouse for ESC self-renewal.

The non-conserved biological processes cytokine secretion and hormone biosynthesis were significantly up-regulated in hESCs but not in mESCs (Supplementary Table S1a). In contrast, the non-conserved biological processes base-excision repair, cytokinesis, and embryonic cleavage were significantly up-regulated in mESCs while not in hESCs (Supplementary Table S1a). Similarly, the insulin signaling pathway, being not conserved ($r = 0.06$, orthologs = 73, $P = 0.310$), was significantly up-regulated in mESCs while down-regulated in hESCs (Table 2). Analysis of such functional modules facilitates our understanding of species-specific mechanisms regulating ESC development.

It should be noted that since the non-conserved modules identified in this study were based on human - mouse orthologous genes only, not all species-specific biological mechanisms involved in ESC development were disclosed. Many other species-specific regulations should be present and activated by non-homologous genes in human or mouse ESCs.

3. Conserved and Divergent Transcription Factors and Growth Factors

We examined expression patterns of 304 human - mouse orthologous genes that show transcription factor activity. Globally, the transcription factors presented a relatively low yet highly significant level of cross-species correlation on the expression pattern ($r = 0.378$, $P < 10^{-11}$), suggesting the presence of conserved core regulatory network in ESCs. Specifically, 199 of the transcription factors were conserved by showing up or down-regulation in the same direction between hESCs and mESCs, in which 48 genes had at least 1.5 fold transcriptional changes upon ESC differentiation in both species. The remaining 105 transcription factors were divergent, showing up- or down-regulation in opposite directions between hESCs and mESCs, in which 9 genes exhibited at least 1.5 fold transcriptional changes upon differentiation in either or both species. Table 3a lists the 48 transcription factors with conserved expression patterns and at least 1.5 fold transcriptional changes. Also shown in Table 3a are the P-values derived from differential expression analysis of these transcription factors in hESC and mESC. At $P < 0.05$, 18 genes showed significant differential expression in both hESCs and mESCs. These transcription factors, with some up-regulated in both hESCs and mESCs and others down-regulated in both species, are functionally essential and are likely key elements in the regulatory network controlling ESC self-renewal and differentiation. Among the 48 transcription factors with conserved behavior, 26 are implicated in development (Table 3a). Some are known to be important for ESC pluripotency, such as POU5F1, UTF1, MYC, and TP53, all of which were up-regulated in ESC, and HAND1 and GATA6, both of which were down-regulated (Table 3a). Table 3b lists the 9 transcription factors that showed divergent expression patterns with at least 1.5 fold expression changes. These transcription factors may play different or species-specific roles in ESC development. Some of the factors are members of important signaling pathways: STAT6 (no significant change in hESCs, up in mESCs) and STAT5B (up in hESCs, down in mESCs) are members of the JAK-STAT pathway involved in LIF signaling; CEBPB (no significant change in hESCs, up in mESCs) is involved in IL6 signaling.

Similarly, growth factors overall also showed a conserved expression pattern between hESCs and mESCs ($r = 0.375$, $P < 0.01$). Sixty-one orthologous growth factors were examined in this study. Twenty-one of the growth factors exhibited negative correlation on the expression pattern across species, with 8 of them showing at least 1.5 fold transcriptional changes (Table 4b). The patterns of the remaining 40 growth factors were conserved, with 10 factors showing at least 1.5 fold transcriptional changes upon differentiation in both species (Table 4a). Among the 10 growth factors, 7 are known to be involved in development and morphogenesis. BMP4, BMP5 (both down-regulated in ESCs), and GDF3 (up-regulated in ESCs) are critical members of the BMP signaling pathway. LEFTB and TDGF1 (both up-regulated in ESCs) participate in Nodal signaling. Six of the 10 growth factors showed significant differential expression in both hESCs and mESCs ($P < 0.05$, Table 4a), including BMP5, GDF3, TDGF1, LEFTB, SCGF, and FBS1.

Many of the conserved transcription factors and growth factors show no obvious relationship with ESC development (Tables 3,4). These factors provide a basis for new hypothesis about mechanisms underlying ESC development. Further studies on these genes using RT-PCR, mutation and knock-out, or computational analysis of pathway dynamics will help confirm their possible roles in ESC development.

Transcription factors, growth factors and their receptors, and other intermediate signal transducers form signal transduction pathways which control the transition and integration of environmental stimuli and help to determine the fate of ESCs. Table 2 lists signaling pathways with information of transcriptional changes in hESCs and mESCs and cross-species conservation, illustrating conserved and divergent mechanisms regulating ESC development.

4. Functional Profiling of Embryonal Carcinoma Stem Cells

To further explore "stemness" properties, we examined common and different functional features between hESCs and NTera2 cells. NTera2 is a human embryonal carcinoma stem cell (hECC) line with less differentiation potential than hESCs. Supplementary Figure S1 shows cultured NTera2 cells in the undifferentiated and differentiated states. From the determined expression profiles, 2818 up-regulated and 2924 down-regulated genes were identified (FDR < 0.05), listed in Supplementary Table S4. The up-regulated genes in undifferentiated NTera2 cells included markers characteristic of undifferentiated hESCs such as POUT5F1, TDGF1, DNMT3B, CD24, LIN28, CD9, PUM1, PUM2, GJA1, and TERF1. The down-regulated genes included HAND1, IGF2, and other markers known to be involved in differentiation. The enriched GO and KEGG categories in the up- and down-regulated genes are shown in Supplementary Tables S1 and S2 (denoted with [C] and [D], respectively, on the term names).

There were 238 biological processes (GO) that were enriched in the up-regulated genes. Among them, 124 processes (i.e. 52.1% of the total) were shared with the 229 biological processes that were up-regulated in hESCs. In the shared 124 processes, 63 were conserved on the expression pattern between hESCs and mESCs (Supplementary Table S1a). There were 98 biological processes that were enriched in the set of down-regulated genes. Among them, 11 processes (i.e. 11.2% of the total) were shared with the 68 down-regulated processes of hESCs (Supplementary Table S1). Therefore, more biological processes are in common between hESCs and NTera2 cells in the undifferentiated states (as expected) than in the differentiated state. A similar pattern was also observed with the GO-defined molecular functions and cellular components, and KEGG-defined pathways (Supplementary Tables S1 and S2). The results suggest that hESCs and hECCs are more similar in mechanisms activated for self-renewal, but less similar in mechanisms activated for differentiation (or suppressed for maintaining self-renewal). This observation underlines the fact that while hESCs and hECCs share many characteristics of pluripotency and express similar "stemness" markers, the two stem cell types differ in their differentiation potential. While hESCs can differentiate to cells forming mesoderm, endoderm, and ectoderm, NTera2 cells mainly differentiate into neuroectoderm cells [26; 67; 68]. It was observed that biological processes related to embryonic development were significantly down-regulated and thus repressed in hESCs, but not down-regulated in NTera2 undifferentiated cells. Moreover, about 30% of the down-regulated biological processes in hESCs were related to development, whereas only ~7% of the down-regulated processes in undifferentiated NTera2 cells were related to development (Supplementary Table S1a). The results highlight the fact that NTera2 cells possess less differentiation potential than ESCs.

The functional modules that are common between hESCs and hECCs in the up-regulated genes and are conserved on the expression pattern between hESCs and mESCs represent core molecular mechanisms that are activated for self-renewal in stem cells of different sources. To achieve indefinite self-renewal and maintain the pluripotent state, stem cells are primed for rapid and accurate DNA replication by shortening the cell cycle, relaxing G1/S checkpoints, strengthening DNA repair, enhancing the response to unfolded proteins, and activating apoptosis. In addition, metabolic processes related to nucleotide synthesis, DNA replication, and DNA packaging are also activated [40]. All these biological processes or related pathways were conserved on the expression pattern between hESCs and mESCs and common between hESCs and hECCs (Supplementary Tables S1 and S2, denoted by [E], [A], and [C] at the term names; Figure 2; Figure 1a). In addition, the telomere maintenance and aging related processes were also conserved and up-regulated in both hESCs and hECCs (Supplementary Tables S1a), in keeping with the fact that stem cells have a longer half-life than somatic cells.

NTera2 cells are aneuploid; it is possible that over-presentation of some genes in the genome may lead to their over-expression in NTera2 cells. This "gene loading" effect on transcriptional

pattern of NTera2 has not yet been fully explored in previous studies [69;70]. There are some practical difficulties for examining the correlation between gene expression levels and the cytogenetic profiles in NTera2 cells. First, the karyotype of NTera2 is not very well defined yet. Plaia et al. [71] showed that 48% of the NTera2 cells had 63 chromosomes, 24% had 62 chromosomes, and the rest had other numbers of chromosomes. Second, some of the chromosomes are not normal [71]. There are 12 marker chromosomes, including der(9)t(1;9)(q25;q34.3), del(1)(q25), der(13)t(11;13)(q13;q34), t(Xq1q), and eight others. At least two markers are found only in some cells. Therefore, a well defined copy number of each gene in NTera2 cells is not yet available for an analysis [71]. The impact of "gene loading" on the expression analysis result of this study remains to be determined.

DISCUSSION

In this study, we systematically surveyed a large set of functional modules, and examined all transcription factors and growth factors for cross-species functional profiling of ESCs. We examined global transcriptional trends derived from all expressed genes, instead of differentially expressed genes alone, for each functional module and obtained a great discriminating power in the functional profiling. Despite the fact that only human - mouse orthologous genes were used in this study, a number of functional modules were identified that were transcriptionally conserved and thus predicted to be essential for ESC development. This study thus generated a global and comprehensive functional portrait of ESCs, featured by conserved and divergent landscapes underlining fundamental and species-specific mechanisms that regulate ESC development. The data and analyses resulting from this study provide a framework for new hypotheses and research directions, and a public resource for functional genomics of ESCs.

We demonstrate that cross-species correlation of gene expression patterns is a powerful predictor of functional mechanisms that are fundamental for ESC development. Much of our current knowledge on ESC-relevant pathways and networks are supported or reaffirmed by this correlation analysis, justifying the prediction. In hESCs, among the 407 up- or down-regulated biological processes, 159 (27% of the total) were identified to be conserved on the expression pattern and thus deemed to be functionally essential. In mESCs, a similar portion of conserved modules were identified from up- or down-regulated biological processes. Our studies thus provide a focused list of important pathways or functional modules for in-depth investigation. Moreover, we demonstrate that functional modules, such as BMP/TDGF-β and Wnt pathways, may be a mosaic of submodules showing different evolutionary patterns. Cross-species correlation of gene expression allows for prediction of conserved and thus core functional submodules present within a pathway, facilitating the examination. Yet, many of the conserved biological processes, pathways, and transcription or growth factors identified in this study bear no overt relationship to ESC development. In addition, many of the conserved modules are not recognized from previous fold-change-based analyses, since only a small portion of genes in the modules showed significant transcriptional changes upon ESC differentiation. Further examination of these functional modules may lead to discovery of new mechanisms critical for ESC pluripotency.

While conserved functional modules represent biological mechanisms that are fundamental, the non-conserved modules identified by this study are suggestive of species-specific regulation in ESC development. The non-conserved nature of functional modules could be due to several factors: 1) functions carried out by these modules are redundant so that different organisms may use different modules to carry out the same function; 2) there exist duplicated genes in a module, and different sets of genes of the module are activated or repressed to achieve the same effect in different organisms; 3) not all orthologous genes in a module are included in calculating the correlation of expression pattern. More hypotheses are possible based on the

profiled ESC functional portrait. Analysis driven by these or other hypotheses on the non-conserved modules would shed light on our understanding of species-specific mechanisms regulating ESC development. Relatively a small portion of functional modules showed significant negative correlation (Table 1), indicating that a very few modules carry out completely opposite activities in hESC and mESC during their differentiation. The result suggests that human and mouse in general share many similar mechanisms that controll ESC development. Nevertheless, there may exist more negatively correlated modules, which are not annotated by GO or KEGG and thus not captured by our analysis.

The observed conservation and variation of functional modules at the gene transcriptional level are consistent with their evolutionary patterns at the gene structural level. Consistent with the conserved expression pattern of the Nodal pathway (Figure 1), Nodal and TDFG1 of the pathway are highly conserved in the gene and promoter structure between human and mouse [19], suggestive of conserved activation mechanisms of the pathway. Consistent with the divergent expression pattern of the FGF pathway (Figure 1), FGF4 of the pathway shows sequence divergence in the Sox2 - Oct3/4 co-binding site of an enhancer [19]. The Sox2 - Oct3/4 co-binding site is critical to the self-renewal of mESCs; the sequence divergence in the co-binding site makes the transcription of the gene less effective in humans than in rodents [65]. Furthermore, consistent with the divergent expression pattern of the LIF pathway (Figure 1), LIFR and IL6ST (GP130) of the pathway are highly divergent in the gene and promoter structure between human and mouse [19], highlighting the fact that LIF signaling is required in mouse but is not essential in human for maintaining ESC pluripotency [57;62;63]. Reportedly, the collaboration between LIF signaling and the BMP/TGF-β pathway helps to sustain pluripotency in mESCs [57]. Since the LIF pathway is diverged and not fundamental, the collaboration is likely not fundamental in human or other species either. Although the BMP/TGF-β pathway is transcriptionally conserved (Figure 1), BMP4 has different effects on mouse and human ESCs: it blocks ESC differentiation along the neuroectoderm default pathway in mouse [57] while it induces trophectroderm differentiation of ESCs in human [72]. It is possible that the different impacts of BMP signaling on ESC development in human and mouse is due to the diverged LIF signaling that interacts with BMP signaling.

We demonstrate that while pluripotency as an essential function in multicellular organisms is conserved through evolution, the mechanisms primed for differentiation are less conserved and contribute in a larger extent to the differences among stem cells derived from different tissues or species. Between different stem cell types (e.g. hESCs and hECCs), there are more similarities on the mechanisms activated for self-renewal than on those activated for differentiation. Within the same stem cell type, more up-regulated modules are conserved than down-regulated modules. Specifically, in hESCs, 43.7% up-regulated biological processes (105 out of 240) were conserved, while 32.3% down-regulated biological processes (54 out of 167) were conserved (Supplementary Table S1a). In mESCs, 37.2% up-regulated biological processes were conserved, while 19.5% down-regulated biological processes were conserved (Supplementary Table S1a). Because biological mechanisms primed for differentiation are evolutionarily less conserved, ESC lines derived from different species (e.g. hESC, mESC) or isolated from different tissues (e.g. ESCs, ECCs, etc) likely adapt distinctive mechanisms activated for differentiation.

It would be interesting to compare the conserved modules detected in ESCs with those detected in other type of stem cells or even somatic cells. Such comparisons would help us further distinguish those processes critical for stem cell self-renewal from those governing differentiation, and those underlying house-keeping functions.

In summary, this study described conserved and divergent functional landscapes of ESCs, profiled by global transcriptional trend across species in addition to differentially expressed

genes. The conserved modules on the landscape represent core biological networks or mechanisms fundamental for ESC development, while the divergent modules are suggestive of species-specific regulation. The cross-species correlation of gene expression pattern is a powerful predictor of important biological pathways, as well as functional cores within a pathway. The examination of hECC - NEera2 revealed functional basis of different differentiation potentials yet similar pluripotency in different kinds of stem cells. While pluripotency is conserved through evolution, the mechanisms primed for differentiation are less conserved and contribute in a larger extent to the differences among stem cells derived from different tissues or species. The functional portrait of ESCs profiled by this study provides a basis for defining the "stemness" property in terms of evolutionary conservation and variation and a framework for new hypotheses and research directions on ESCs. The findings and methods reported in this investigation are significant in advancing our understanding on stem cell biology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported, in part, by the Intramural Research Program of the National Institute on Aging, NIH.

References

1. Reubinoff BE, et al. Neural progenitors from human embryonic stem cells. Nat Biotechnol 2001;19:1134–40. [PubMed: 11731782]
2. Rambhatla L, Chiu CP, Kundu P, Peng Y, Carpenter MK. Generation of hepatocyte-like cells from human embryonic stem cells. Cell Transplant 2003;12:1–11. [PubMed: 12693659]
3. Kehat I, et al. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. J Clin Invest 2001;108:407–14. [PubMed: 11489934]
4. Kaufman DS, Hanson ET, Lewis RL, Auerbach R, Thomson JA. Hematopoietic colony-forming cells derived from human embryonic stem cells. Proc Natl Acad Sci U S A 2001;98:10716–21. [PubMed: 11535826]
5. Segev H, Fishman B, Ziskind A, Shulman M, Itskovitz-Eldor J. Differentiation of human embryonic stem cells into insulin-producing clusters. Stem Cells 2004;22:265–74. [PubMed: 15153604]
6. Bhattacharya B, et al. Gene expression in human embryonic stem cell lines: unique molecular signature. Blood 2004;103:2956–64. [PubMed: 15070671]
7. Brandenberger R, et al. Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat Biotechnol 2004;22:707–16. [PubMed: 15146197]
8. Miura T, et al. Monitoring early differentiation events in human embryonic stem cells by massively parallel signature sequencing and expressed sequence tag scan. Stem Cells Dev 2004;13:694–715. [PubMed: 15684837]
9. Liu Y, et al. Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonal carcinoma cells to develop base profiles of U.S. Federal government approved hESC lines. BMC Dev Biol 2006;6:20. [PubMed: 16672070]
10. McCarroll SA, et al. Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. Nat Genet 2004;36:197–204. [PubMed: 14730301]
11. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003;34:267–73. [PubMed: 12808457]
12. Hughes TR, et al. Functional discovery via a compendium of expression profiles. Cell 2000;102:109–26. [PubMed: 10929718]
13. Li H, et al. Transcriptome coexpression map of human embryonic stem cells. BMC Genomics 2006;7:103. [PubMed: 16670017]

14. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature 1981;292:154–6. [PubMed: 7242681]
15. Martin GR. 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 1981;78:7634–8. [PubMed: 6950406]
16. Carpenter MK, Rosler E, Rao MS. Characterization and differentiation of human embryonic stem cells. Cloning Stem Cells 2003;5:79–88. [PubMed: 12713704]
17. Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat Biotechnol 2000;18:399–404. [PubMed: 10748519]
18. Thomson JA, et al. Embryonic Stem Cell Lines Derived from Human Blastocysts. Science 1998;282:1145–1147. [PubMed: 9804556]
19. Zhan M, Miura T, Xu X, Rao MS. Conservation and variation of gene regulation in embryonic stem cells assessed by comparative genomics. Cell Biochem Biophys 2005;43:379–405. [PubMed: 16244364]
20. Ginis I, et al. Differences between human and mouse embryonic stem cells. Dev Biol 2004;269:360–80. [PubMed: 15110706]
21. Rao M. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. Dev Biol 2004;275:269–86. [PubMed: 15501218]
22. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science 2003;302:249–55. [PubMed: 12934013]
23. Bergmann S, Ihmels J, Barkai N. Similarities and differences in genome-wide expression data of six organisms. PLoS Biol 2004;2:E9. [PubMed: 14737187]
24. Ihmels J, Bergmann S, Berman J, Barkai N. Comparative gene expression analysis by differential clustering approach: application to the Candida albicans transcription program. PLoS Genet 2005;1:e39. [PubMed: 16470937]
25. Brimble SN, et al. Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. Stem Cells Dev 2004;13:585–97. [PubMed: 15684826]
26. Schwartz CM, et al. NTera2: a model system to study dopaminergic differentiation of human embryonic stem cells. Stem Cells Dev 2005;14:517–34. [PubMed: 16305337]
27. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–93. [PubMed: 12538238]
28. SAS. SAS OnlineDoc, Version 8. SAS Institute Inc.; Cary, NC.: 1999.
29. Jain N, et al. Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. Bioinformatics 2003;19:1945–51. [PubMed: 14555628]
30. Abeyta MJ, et al. Unique gene expression signatures of independently-derived human embryonic stem cell lines. Hum Mol Genet 2004;13:601–8. [PubMed: 14749348]
31. Skottman H, et al. Gene expression signatures of seven individual human embryonic stem cell lines. Stem Cells. 2005
32. Rao RR, et al. Comparative transcriptional profiling of two human embryonic stem cell lines. Biotechnol Bioeng 2004;88:273–86. [PubMed: 15493035]
33. Sperger JM, et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Natl Acad Sci U S A 2003;100:13350–5. [PubMed: 14595015]
34. Richards M, Tan SP, Tan JH, Chan WK, Bongso A. The transcriptome profile of human embryonic stem cells as defined by SAGE. Stem Cells 2004;22:51–64. [PubMed: 14688391]
35. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 2000;24:372–6. [PubMed: 10742100]
36. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. “Stemness”: Transcriptional Profiling of Embryonic and Adult Stem Cells. Science 2002;298:597–600. [PubMed: 12228720]
37. Cai J, Weiss ML, Rao MS. In search of “stemness”. Exp Hematol 2004;32:585–98. [PubMed: 15246154]
38. Ivanova NB, et al. A stem cell molecutar signature. 2002;298:601.

39. Sato N, et al. Molecular signature of human embryonic stem cells and its comparison with the mouse. 2003;260:404.
40. Burdon T, Smith A, Savatier P. Signalling, cell cycle and pluripotency in embryonic stem cells. Trends Cell Biol 2002;12:432–8. [PubMed: 12220864]
41. Fluckiger AC, et al. Cell-cycle features of primate embryonic stem cells. Stem Cells. 2005
42. Pfendler KC, Catuar CS, Meneses JJ, Pedersen RA. Overexpression of Nodal promotes differentiation of mouse embryonic stem cells into mesoderm and endoderm at the expense of neuroectoderm formation. Stem Cells Dev 2005;14:162–72. [PubMed: 15910242]
43. Besser D. Expression of nodal, lefty-a, and lefty-B in undifferentiated human embryonic stem cells requires activation of Smad2/3. J Biol Chem 2004;279:45076–84. [PubMed: 15308665]
44. Vallier L, Reynolds D, Pedersen RA. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. Dev Biol 2004;275:403–21. [PubMed: 15501227]
45. Schier AF, Shen MM. Nodal signalling in vertebrate development. Nature 2000;403:385. [PubMed: 10667782]
46. Tang K, et al. Wnt-1 promotes neuronal differentiation and inhibits gliogenesis in P19 cells. 2002;293:167.
47. Aubert J, Dunstan H, Chambers I, Smith A. Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. 2002;20:1240.
48. Ding S, et al. Synthetic small molecules that control stem cell fate. PNAS 2003;100:7632–7637. [PubMed: 12794184]
49. Reya T, et al. A role for Wnt signalling in self-renewal of haematopoietic stem cells. 2003;423:409.
50. Dravid G, et al. Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells. Stem Cells 2005;23:1489–501. [PubMed: 16002782]
51. Whitman M. Smads and early developmental signaling by the TGFbeta superfamily. Genes Dev 1998;12:2445–62. [PubMed: 9716398]
52. Massague J, Blain SW, Lo RS. TGFbeta signaling in growth control, cancer, and heritable disorders. Cell 2000;103:295–309. [PubMed: 11057902]
53. Goumans MJ, Mummery C. Functional analysis of the TGFbeta receptor/Smad pathway through gene ablation in mice. Int J Dev Biol 2000;44:253–65. [PubMed: 10853822]
54. Johansson BM, Wiles MV. Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development. Mol Cell Biol 1995;15:141–51. [PubMed: 7799920]
55. Winnier G, Blessing M, Labosky PA, Hogan BL. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev 1995;9:2105–16. [PubMed: 7657163]
56. Chambers I, Smith A. Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene 2004;23:7150–60. [PubMed: 15378075]
57. Ying QL, Nichols J, Chambers I, Smith A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 2003;115:281–92. [PubMed: 14636556]
58. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. Science 1999;284:770–6. [PubMed: 10221902]
59. Rho JY, et al. Transcriptional profiling of the developmentally important signalling pathways in human embryonic stem cells. Hum Reprod 2006;21:405–12. [PubMed: 16239319]
60. Czyz J, Wobus A. Embryonic stem cell differentiation: the role of extracellular factors. Differentiation 2001;68:167–74. [PubMed: 11776469]
61. Walker JL, Assoian RK. Integrin-dependent signal transduction regulating cyclin D1 expression and G1 phase cell cycle progression. Cancer Metastasis Rev 2005;24:383–93. [PubMed: 16258726]
62. Daheron L, et al. LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. Stem Cells 2004;22:770–8. [PubMed: 15342941]
63. Humphrey RK, et al. Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. Stem Cells 2004;22:522–30. [PubMed: 15277698]
64. Xu C, et al. Feeder-free growth of undifferentiated human embryonic stem cells. Nat Biotechnol 2001;19:971–4. [PubMed: 11581665]

65. Lamb KA, Rizzino A. Effects of differentiation on the transcriptional regulation of the FGF-4 gene: critical roles played by a distal enhancer. Mol Reprod Dev 1998;51:218–224. [PubMed: 9740330]

66. Goldin SN, Papaioannou VE. Paracrine action of FGF4 during perimplantation development maintains trophectoderm and primitive endoderm. Genesis 2003;1:40–47. [PubMed: 12748966]

67. Walsh J, Andrews PW. Expression of Wnt and Notch pathway genes in a pluripotent human embryonal carcinoma cell line and embryonic stem cell. Apmis 2003;111:197–210. 210–1. [PubMed: 12760378]

68. Pera MF, Cooper S, Mills J, Parrington JM. Isolation and characterization of a multipotent clone of human embryonal carcinoma cells. Differentiation 1989;42:10–23. [PubMed: 2559868]

69. Przyborski SA, Smith S, Wood A. Transcriptional profiling of neuronal differentiation by human embryonal carcinoma stem cells in vitro. Stem Cells 2003;21:459–71. [PubMed: 12832699]

70. Houldsworth J, Heath SC, Bosl GJ, Studer L, Chaganti RS. Expression profiling of lineage differentiation in pluripotential human embryonal carcinoma cells. Cell Growth Differ 2002;13:257–64. [PubMed: 12114215]

71. Plaia TW, et al. Characterization of a new NIH-registered variant human embryonic stem cell line, BG01V: a tool for human embryonic stem cell research. Stem Cells 2006;24:531–46. [PubMed: 16293579]

72. Xu RH, et al. BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nat Biotechnol 2002;20:1261–4. [PubMed: 12426580]

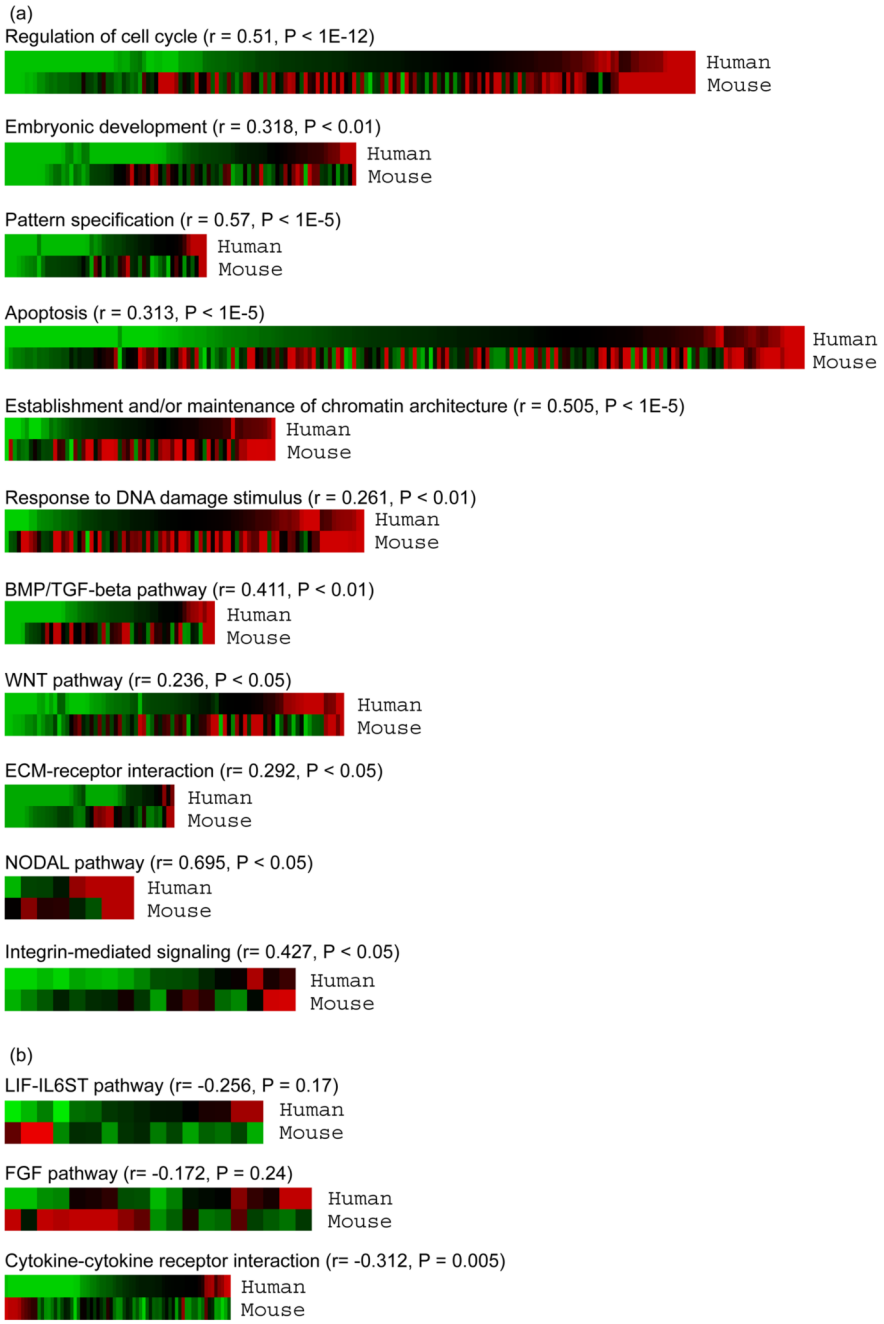

Figure 1.

Expression patterns of selected converged (**a**) and divergent (**b**) biological processes and pathways. On the headmap of each module, the column represents an orthologous gene expressed in hESC (top row) and in mESC (bottom row). The expression fold change of the gene is represented by different colors (green: down-regulated in ESC, red: up-regulated in ESC, black: no change). For conserved modules (**a**; with high r values), many orthologous genes showed expression in the same direction (e.g. either up-regulated in both species or down-regulation in both) with similar expression fold changes between human and mouse. For divergent modules (**b**; low or negative r values), most orthologous genes showed expression in opposite directions (e.g. up-regulated in one species and down-regulated in another) with

dissimilar expression fold changes between human and mouse. Details of the genes shown in each module along with the expression values and fold changes are provided in Supplementary Table S3.

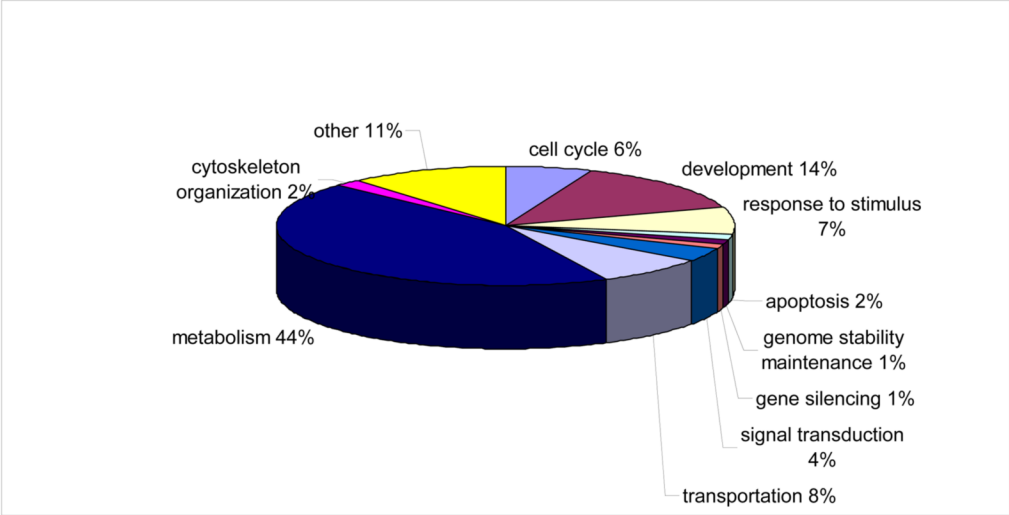

Figure 2.
Summary of conserved biological processes showing correlated expression patterns between human and mouse ESCs.

Table 1

Number of functional modules showing positive, negative, or no correlation on the transcriptional changes upon ESC differentiation between human and mouse ($P \leq 0.05$)

Cross-species transcriptional correlation	Gene group	Biological process (GO)	Molecular function (GO)	Cellular component (GO)	Pathway (KEGG)
positively correlated	Total	389	154	102	14
	up-regulated in hESCs	105	31	20	2
	down-regulated in hESCs	54	21	16	5
	up-regulated in mESCs	157	63	43	4
	down-regulated in mESCs	110	43	23	2
not correlated	Total	508	258	121	88
	Up-regulated in hESCs	135	67	28	4
	down-regulated in hESCs	113	85	39	10
	Up-regulated in mESCs	265	129	84	8
	down-regulated in mESCs	333	138	47	5
negatively correlated	Total	6	1	2	3

Table 2

Transcriptionally conserved (a) and diverged (b) signaling transduction pathways.

a)					
Pathway	**Correlation (r)**	**Ortholog pairs**	**P-value**	**Transcript changes**	
				HESC	**mESC**
TGF-beta signaling	0.41	52	0.001	Down	up & down
Wnt signaling	0.24	84	0.015	up & down	down
Notch signaling	0.45	23	0.016	Down	non
Nodal signaling	0.7	8	0.028	Up	up
ECM-receptor interaction	0.29	43	0.029	Down	down
Integrin-mediated signaling	0.43	18	0.039	Down	Down
b)					
Pathway	**Correlation (r)**	**Ortholog pairs**	**P-value**	**Transcript changes**	
				hESC	**MESC**
Cytokine-cytokine receptor interaction	−0.31	69	0.005	down	Non
Phosphatidylinositol signaling system	0.19	39	0.123	down	Non
JAK-Stat signaling	0.14	53	0.157	down	Non
Toll-like receptor Signaling	0.14	38	0.198	down	Non
Hedgehog signaling	0.17	25	0.204	down	Down
FGF signaling	−0.17	19	0.24	non	Up
LIF signaling	−0.26	16	0.169	non	down*
Insulin signaling	0.06	73	0.31	down	Up
MAPK signaling	0.03	128	0.355	down	Non
Calcium signaling	−0.01	75	0.448	down	Non
Neuroactive ligand-receptor interaction	−0.01	58	0.472	down	Down

* LIF pathway was down regulated in BG01, BG01V, BG02, but non-change in GB03.

Table 3

Transcriptionally conserved **(a)** and diverged **(b)** transcription factors. The transcription factors showed at least 1.5-fold up- or down-regulation in the same direction (conserved) or opposite directions (diverged) between human and mouse ES cells. P-values represent the significance of differential expression in hESCs or mESCs.

a)

Transcription Factors	P-value in hESC	P-value in mESC	Expression in hESC and mESC	Implicated functions
MXD4	0.0024	0.0175	Down	Cell proliferation
MYC	0.0555	0.0178	Up	
MYST2	0.0009	0.0062	Up	
TP53	0.0204	0.0008	Up	
ARNT2	0.1231	0.0302	Down	development
EVX1	0.2256	0.0112	Down	
GATA3	0.0150	0.0426	Down	
GATA6	0.0017	0.0561	Down	
HAND1	0.0033	0.0329	Down	
HEY1	0.0032	0.0118	Down	
HLX1	0.0588	0.0008	Down	
HOXB5	0.0014	0.0012	Down	
HOXB6	0.0043	0.0049	Down	
HOXB9	0.0578	0.0050	Down	
MAFB	0.0802	0.1325	Down	
MEF2C	0.0209	0.0551	Down	
MSX1	0.0030	0.0384	Down	
MSX2	0.0050	0.0933	Down	
PITX1	0.0121	0.0234	Down	
PITX2	0.0054	0.1141	Down	
RAX	0.0252	0.0004	Down	
SIX3	0.2344	0.0487	Down	
SIX5	0.0985	0.0163	Down	
SOX10	0.1450	0.0121	Down	
SOX11	0.2070	0.2287	Down	

a)

Transcription Factors	P-value in hESC	P-value in mESC	Expression in hESC and mESC	Implicated functions
TCF21	0.1014	0.0078	Down	
ZIC1	0.1209	0.3692	Down	
ZHX1	0.0013	0.0361	Down	
TGIF	0.0143	0.0128	Up	
POU5F1	0.0070	0.0328	Up	
ATBF1	0.0112	0.0086	Down	Others
CREB3L1	0.0644	0.1096	Down	
DSIPI	0.0423	0.1123	Down	
FOXA1	0.0273	0.0236	Down	
FOXF1	0.1172	0.0690	Down	
ISL1	0.0480	0.0877	Down	
NR2F1	0.1048	0.1984	Down	
SOX18	0.1016	0.0574	Down	
SOX7	0.0908	0.0705	Down	
SPI1	0.1408	0.0076	Down	
ZFHX4	0.0456	0.0730	Down	
ZNF42	0.3586	0.0206	Down	
MYB	0.0176	0.0600	Up	
MYCN	0.0032	0.0125	Up	
NFYB	0.0017	0.0132	Up	
POLR3K	0.0109	0.0519	Up	
POU2F1	0.0005	0.0870	Up	
UTF1	0.1303	0.0088	Up	

b)

Transcription factor	P-value in hESC	P-value in mESC	Expression in hESC	Expression in mESC	Implicated functions
CBFA2T2	0.0619	0.0195	Down	Up	Cell proliferation

a)

Transcription Factors	P-value in hESC	P-value in mESC	Expression in hESC and mESC	Implicated functions	
GTF2H1	0.0295	0.0076	Down	Up	
IRF1	0.0487	0.0778	Down	Up	
STAT5B	0.0030	0.0043	Up	Down	Signal transduction
STAT6	0.2758	0.0078	no-change	Up	
CEBPB	0.2703	0.0162	no-change	Up	Others
ZNF277	0.2153	0.0318	no-change	Up	
TPBG	0.0576	0.0237	Down	Up	
FOXA2	0.1128	0.0786	Up	Down	

Table 4

Transcriptionally conserved (**a**) and diverged (**b**) growth factors. The growth factors showed at least 1.5-fold up- or down-regulation in the same direction (conserved) or opposite directions (diverged) between human and mouse ES cells. P-values represent the significance of differential expression in hESCs or mESCs.

a)

Growth factors	P hESC –value	mESC P-value	Expression in ESC	Implicated functions
SCGF	0.0133	0.0468	Down	Cell proliferation
BMP4	0.0146	0.0600	Down	development/morphogenesis
BMP5	0.0067	0.0080	Down	
IGF2	0.0003	0.0563	Down	
MDK	0.0116	0.1218	Down	
GDF3	0.0039	0.0169	Up	
LEFTB	0.0089	0.0463	Up	
TDGF1	0.0021	0.0214	Up	
MST1	0.0543	0.0655	Down	
FBS1	0.0424	0.0233	Down	Others

b)

Growth factors	hESC P-value	mESC P-value	Expression in hESC	Expression in mESC	Implicated functions
AREG	0.0805	0.2423	Down	no-change	cell-cell signaling
CXCL12	0.0417	0.0188	Up	Down	
BMP2	0.0783	0.0231	Up	Down	development
FGF13	0.0212	0.0067	Up	Down	
DTR	0.0573	0.1424	Down	no-change	
INHBA	0.3262	0.0837	no-change	Up	
S100A6	0.0443	0.0243	Down	Up	

a)

Growth factors	P hESC –value	mESC P-value	Expression in ESC	Implicated functions	
				Down	Others
INHBE	0.2185	0.0598	no-change		