Mol. Cells *31*, 315-326, April 30, 2011
DOI/10.1007/s10059-011-0039-1

Molecules and Cells

Gene Expression Profiles in CHA3 and CHA4 Human Embryonic Stem Cells and Embryoid Bodies

Sung-Hwan Moon[1,6], Sung-Whan Kim[2,6], Jong Soo Kim[3], Soon-Jung Park[4], Jeong Tae Do[3], Dong Ryul Lee[5], and Hyung-Min Chung[1,4,*]

The establishment of the first human embryonic stem cells (hESCs) in 1998 provided a unique tool for studying human development. Although several Western embryo-derived hESC lines are well characterized, the biological properties of Asian embryo-derived hESC lines remain unexamined. The aim of this study was to characterize Korean embryo-derived hESC lines and their differentiation potential. In this context, we conducted microarray-based differential gene expression analyses using two Korean embryo-derived hESC lines (CHA3 and CHA4) to identify undifferentiated and spontaneously differentiated (human embryoid body, or hEB) status. These two cell lines showed great similarity in gene expression. By comparing their expression patterns, we determined novel hESC-specific genes and transcriptomes that could serve as reliable hESC markers associated with the "stemness" phenotype. Additionally, we sought to identify hEB markers that could be used to determine the presence of differentiated cells in specific tissues, allowing for the purification of homogeneous cell populations or serving as indicators of hESC differentiation. Novel sets of 68 hESC-specific markers, 12 hESC-specific transcripts and 36 hEB markers were identified and shown by quantitative RT-PCR to be similarly expressed in CHA3- and CHA4-hESC lines, as compared to the Western embryo-derived H9-hESC line. Furthermore, our data analysis revealed that the cell cycle, urea cycle, p53 signaling, and metabolism of amino groups are significantly implicated in the regulation of hESC differentiation. These results provide another unique set of hESC/hEB markers and foster a better understanding of the molecular mechanisms underlying hESC biology. These results may thus facilitate studies of human developmental events and provide information regarding Korean embryo-derived hESCs, which could be used to determine differences in developmental events between human races.

INTRODUCTION

Since human embryonic stem cells (hESCs) were established in 1998, they have been widely regarded as a tool for studying human development (Thomson et al., 1998). Generally, hESCs exhibit unlimited self-renewal potential and pluripotency, meaning that they can give rise to all three germ layers: endoderm, ectoderm and mesoderm. This versatility becomes progressively limited as hESCs differentiate into tissue-specific lineages. *In vitro* differentiation of hESCs can induce the formation of human embryoid bodies (hEBs), which involves the three-dimensional aggregation of cells into morula-like structures if the cells are cultured in suspension media in the absence of basic fibroblast growth factor (bFGF) and mouse embryonic fibroblast (MEF) feeder cells (Bauwens et al., 2008; Itskovitz-Eldor et al., 2000). hEBs contain all three germ-layer populations and are therefore able to recapitulate the cellular differentiation that occurs during early embryogenesis (Fathi et al., 2009; Itskovitz-Eldor et al., 2000). For this reason, hEBs have been widely used *in vitro* to study the differentiation of hESCs. However, the heterogeneity of hEBs is detrimental to the reproducibility and synchronism of their differentiation (Bauwens et al., 2008).

Changes in molecular characteristics or mechanisms occurring during early developmental stages are still not fully understood. Before hESCs and hESC-derived cells of specific lineages may be used to understand human developmental events, insight into the serial processes regulating hESC differentiation is necessary. To date, more than 100 distinct hESC lines have been generated (Amit et al., 2000; Mitalipova et al., 2003; Reubinoff et al., 2000; Richards et al., 2004; Thomson et al., 1998). However, few hESCs are available for use in the characterization of the fundamental properties of this cell type or in culture condition-restricted analysis of common gene expression patterns during differentiation (Cai et al., 2006; Fathi et al., 2009). The studies that have been performed reported unique molecular signatures that could distinguish hESCs from differentiated progeny based on microarray analyses, expressed sequence tag scans, massively parallel signature sequences and serial

[1]CHA Bio & Diostech Co., Ltd., Seoul 135-081, Korea, [2]Regional Clinical Trial Center, Department of Cardiology, College of Medicine, Dong-A University, Busan 602-715, Korea, [3]Laboratory of Stem Cell and Developmental Biology, CHA Stem Cell Institute, CHA University, Seoul 135-907, Korea, [4]Stemcell Research Laboratory, Department of Developmental Biology, CHA University, Seoul 135-907, Korea, [5]Fertility Center, CHA Gangnam Medical Center, College of Medicine, CHA University, Seoul 135-081, Korea, [6]These authors contributed equally to this work.
*Correspondence: stemchung@gmail.com

Received July 19, 2010; revised December 31, 2010; accepted January 4, 2011; published online February 22, 2011

Keywords: differentiation, embryoid bodies, human embryonic stem cells, microarray, transcripts

analysis of gene expression (Bhattacharya et al., 2004; Mitalipova et al., 2003; Richards et al., 2004; Sato et al., 2003; Sperger et al., 2003; Wei et al., 2005). Until now, the majority of these data were generated from several Western embryo-derived hESC lines, while the biological properties of Asian embryo-derived hESC lines remained unexamined.

Recently, we have established and successfully maintained 29 Korean embryo-derived hESC lines (Lee et al., 2010). The aim of the current study was to investigate and characterize Korean-derived hESC lines and their differentiation potential. In this context, we conducted microarray-based differential gene expression analyses using two representative Korean embryo-derived hESC lines, CHA3 and CHA4, which we assessed for undifferentiated and spontaneously differentiated status (Ahn et al., 2006; Kim et al., 2007a). The latter was indicated by the presence of hEBs. We determined 68 hESC-specific markers, 12 hESC-specific transcriptomes and 36 hEB-specific markers by comparing published reports. Furthermore, we characterized hEB differentiation status by analysis of their gene expression patterns and functional classification. These markers could be used to assess the state of hESC differentiation and will be useful in the future characterization of Asian hESC lines.

MATERIALS AND METHODS

hESC culture and hEB formation

The maintenance and differentiation of undifferentiated hESC (CHA3 and CHA4 hESC lines) followed the procedures from our previous study (Kim et al., 2007a). In brief, undifferentiated CHA3- and CHA4-hESC line at passage 35-55 were grown on mitotically inactivated feeder cells in Dulbecco's modified Eagle's medium (DMEM)/F12 (50:50; GibcoBRL, USA) supplemented with 20% serum replacement (Gibco) and basic hESC medium components, including 1 mM L-glutamine (Gibco), 1% nonessential amino acids (Gibco), 100 mM β-mercaptoethanol (Gibco), and 4 ng/ml basic fibroblast growth factor (bFGF, Sigma, USA) (Park et al., 2003). The medium was changed every 24 h, and the hESC were transferred to new feeder cells every 5-7 days using dissecting pipettes. *In vitro* differentiation of hESC can induce the formation of hEB by culturing cells in suspension media in the absence of bFGF for 20 days (Bauwens et al., 2008; Itskovitz-Eldor et al., 2000).

Immunocytochemistry

To detect stemness markers in hESC, cultured cells were fixed with 4% paraformaldehyde for 20 min and permeabilized with 0.1% Triton X-100 in phosphate-buffered saline (Sigma) for 5 min. After treatment with 10% normal goat serum (Sigma) for 30 min at room temperature, the cells were incubated with Oct-4 against a pluripotency marker (Santa Cruz, USA) for 12 h at 4°C. After washing, bound primary antibodies were detected using rhodamine-conjugated goat anti-mouse immunoglobulin G antibody (Molecular Probes, USA) for 1 h at room temperature. The stained slides were mounted with a glycerol-based mounting solution containing 2.5% polyvinyl alcohol (Sigma). Images were analyzed using fluorescence microscopy (ECLIPSE TE2000, Nikon, Japan).

Karyotype analysis

Chromosome analysis was performed according to standard methods with minor modifications (Dutrillaux and Viegas-Pequignot, 1981). Three days after replating, hESC were incubated with 100 μl of colcemid (Gibco) for 3 h at 37°C in a CO_2 incubator and then trypsinized. After treating the cells with a hypotonic solution (1% citrate buffer), lysed cells were fixed in a methanol:glacial acetic acid (3:1) solution. G-banding was detected to identify chromosomes. Karyotype analysis was performed at the genetic analysis laboratory at the CHA General Hospital (Kim et al., 2007b).

Teratoma formation and immunohistochemical analyses

All animals received humane care in compliance with the "Guide for the Care and Use of Laboratory Animals" (National Institutes of Health publication No. 85-23, revised 1996). First, 3×10^6 undifferentiated hESC were injected into the dorsal part of 6-week-old non-obese diabetic/severe combined immunodeficiency (NOD/SCID) mice (The Jackson Laboratory, USA). The resulting teratomas were removed 12 weeks later. Samples from the teratomas were paraffin-embedded and serially sectioned (5 μm) with a microtome (Leica Microsystems, Germany). Sectioned slides were examined by hematoxylin and eosin and Alcian blue staining to analyze the three germ-layer lineages derived from the injected hESC in the teratomas. Images were analyzed using an inverted microscope (ECLIPSE 90i, Nikon).

Isolation of total RNA

Total RNA was isolated using the RNeasy Protect Mini Kit (Qiagen, USA). Pelleted cells were re-suspended and homogenized by passing the lysate through a 20-gauge needle fitted to a syringe five times. The samples were then processed following the manufacturer's instructions. In the final step, RNA was eluted with 50 μl of RNase-free water by centrifugation for 1 min at 10,000 rpm. RNA quality was analyzed on an RNA chip using a Bioanalyzer 2100 (Agilent, USA). Array-tested RNA was confirmed to have an A_{260}/A_{280} ratio of greater than 1.8 in a spectrophotometer (NanoDrop, Thermo Scientific, USA).

Applied Biosystems expression arrays

Applied Biosystems Human Genome Survey Arrays were used to analyze the transcriptional profiles of RNA samples (metastatic and non-metastatic) following acute morphine exposure. The Applied Biosystems Human Genome Survey Array contains 31,700 genes and 60-mer oligonucleotide RNA probes, which represent a set of 27,868 individual human genes and more than 1,000 control probes. Sequences used to design the microarray probes were obtained from curated transcripts from the Celera Genomics Human Genome Database (www.celera-discoverysystem.com), reference sequence transcripts that were structurally curated from the LocusLink public database (http://ncbi.nlm.nih.bov/LocusLink/refseq.html), high-quality cDNA sequences from the Mammalian Gene Collection (MGC) (http://mgc.nci.nih.gov) and transcripts that were experimentally validated by Applied Biosystems. The 60-mer probes were synthesized using standard phosphoramidite chemistry and solid-phase synthesis, and the quality of these peptides was confirmed by mass spectrometry. The probes were deposited and covalently bound onto a nylon substrate (2.5 × 3 inches) that was backed onto a glass slide by contact spotting with a feature diameter of 180 μm and space of > 45 μm between each feature. A 24-mer internal control probe was co-spotted at every feature on the microarray. Digoxigenin-UTP-labeled cRNA was generated and linearly amplified from 1 μg of total RNA using an Applied Biosystems Chemiluminescent RT-IVT Labeling Kit v 1.0 following the manufacturer's protocol. Array hybridization (two arrays per sample), chemiluminescence detection, and image acquisition and analysis were performed using an Applied Biosystems Chemiluminescence Detection Kit and an Applied Biosystems 1700 Chemiluminescent Microarray Analyzer following the manufacturer's protocol. Briefly, each mi-

croarray was first pre-hybridized at 55°C for 1 h in hybridization buffer containing a blocking reagent. Next, 16 μg of labeled cRNA targets were fragmented into 100-400 bases by incubating them in fragmentation buffer at 60°C for 30 min. These fragments were then mixed with the internal control target (24-mer oligo labeled with LIZ® fluorescent dye) and hybridized to each pre-hybridized microarray in a volume of 1.5 ml at 55°C for 16 h. After hybridization, the arrays were washed with hybridization wash buffer and chemiluminescence rinse buffer. Enhanced chemiluminescent signals were generated by first incubating arrays with anti-digoxigenin Fab-fragment conjugated with alkaline phosphatase, which was enhanced with Chemiluminescence Enhancing Solution, followed by the addition of Chemiluminescence Substrate. Images were collected for each microarray using a 1700 analyzer that was equipped with a high-resolution, large-format CCD camera. These images included two "short" chemiluminescent images (5-second exposure) and two "long" chemiluminescent images (25-second exposure) for gene expression analysis, two fluorescent images for feature-finding and spot normalization and two QC images for spectrum cross-talk correction. Images were auto-gridded and the chemiluminescent signals quantified, corrected for background fluorescence and spot-normalized and spatially normalized.

Quantitative real-time PCR (qRT-PCR)

cDNA synthesis was performed with the High Capacity cDNA Archive Kit (Applied Biosystems, USA) following the manufacturer's instructions. The level of marker gene expression was determined using the ABI 7300 qRT-PCR system (Applied Biosystems). qRT-PCR Primers were listed in Supplementary Table S1. The level of gene expression was calculated by the comparative $\Delta\Delta Ct$ method.

Data analysis

Applied Biosystems Expression System software was used to extract assay signal and assay signal-to-noise (S/N) ratio values from the microarray images. Bad spots flagged by the software were removed from the analysis. Selected gene signal values were transformed using a logarithmic function and normalized by the quantile normalization method. Next, two-way analysis of variance (ANOVA) was applied to determine specific cell type and time effects (no interaction effect). If a time effect existed, the Tukey HSD test was applied to determine the difference between times. Statistical significance was adjusted using the Benjamini-Hochberg multiple-testing correction with false discovery rate (FDR). Hierarchical cluster analysis was performed using complete linkage and Euclidean distance as a measure of similarity, whereas the k-means cluster was performed using a differential method. Biological ontology-based analysis was performed using the Panther database (http://www.pantherdb.org). In addition, KEGG pathway analysis was performed with DAVID (http://david.abcc.ncifcrf.gov). All data analysis and visualization of differentially expressed genes were conducted using Array Assist® (Stratagene, USA) and R statistical language v. 2.4.1. Biological ontology-based analysis was performed using the Panther database (http://www.pantherdb.org).

RESULTS AND DISCUSSION

Characterization of the CHA3- and CHA4-hESC lines

Over the last decade of hESC studies, many researchers have used hESCs derived from Western embryos. Only a few of these cell lines, such as the H1 and H9 hESCs, are well characterized (Amit et al., 2000; Heins et al., 2004; Mitalipova et al., 2003; Reubinoff et al., 2000; Richards et al., 2004; Thomson et al., 1998; Zeng et al., 2004). Given that hESCs exhibit cell line-dependent differences (Osafune et al., 2008; Richards et al., 2004), we sought to characterize CHA3 and CHA4 hESC lines, derived from Korean embryos, by microarray (Figs. 1A and 1G). Both cell lines possess 46 chromosomes, including X and Y chromosomes (Figs. 1B and 1H). To characterize these hESC lines *in vitro*, we examined the expression of Oct-4, a pluripotency marker, by immunocytochemistry (Figs. 1C and 1I). Then, to investigate the pluripotency of these cell lines *in vivo*, we injected 3×10^6 undifferentiated CHA3 and CHA4 hESCs into the dorsal region of NOD/SCID mice, which were sacrificed 12 weeks later. Immunohistochemical analysis of the teratomas extracted from the mice revealed tissues representative of all three germ layers (Figs. 1D-1F and 1J-1L), suggesting that the CHA3 and CHA4 hESC lines possess general hESC characteristics, such as pluripotency, *in vitro* and *in vivo*.

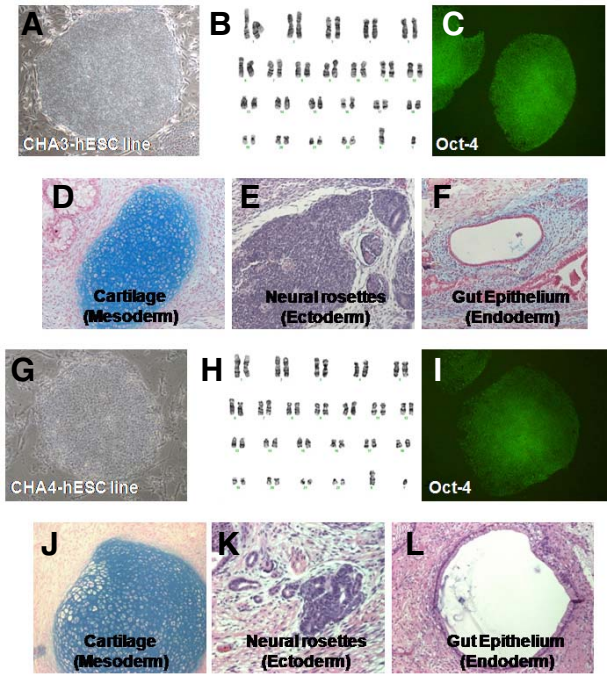

Fig. 1. Characterization of CHA3 and CHA4 hESC lines. (A) Morphology and (B) karyotype of CHA3 hESCs. (C) Pluripotent marker Oct-4 expression in CHA3 hESCs. (D-F) Differentiation of CHA3 hESCs *in vivo* into three germ layers by teratoma formation. (G) Morphology and (H) karyotype of CHA4 hESCs. (I) Pluripotent marker Oct-4 expression in CHA4 hESCs. (J-L) Differentiation of CHA4 hESCs *in vivo* into three germ layers by teratoma formation.

Evaluation of the reproducibility of the microarray results

Microarrays are a powerful tool that has been used to investigate global gene expression. Using this approach, several groups have reported molecular signatures of hESCs and differentiated cells, such as hEBs (Abeyta et al., 2004; Bhattacharya et al., 2004; Cai et al., 2006; Sato et al., 2003; Zeng et al., 2004). To investigate changes in gene expression during the differentiation of CHA3 and CHA4 hESCs, we used a spontaneous hEB-formation method, independent of cytokine induction (Itskovitz-Eldor et al., 2000). To obtain precise gene expression profile measurements between undifferentiated and

differentiated CHA3 and CHA4 hESCs, we assessed five serial developmental stages, including undifferentiated CHA3 and CHA4 hESCs. Recent reports have demonstrated that different hESC lines share similar expression of markers in the undifferentiated state but that the expression patterns change after differentiation (Adewumi et al., 2007; Bouchard and Lemmens, 2008; Kim et al., 2007c). Before analyzing expression profiles, we evaluated the reproducibility of our microarray results for the hEB stage. All data generated from the microarrays were highly reproducible, with no significant variation between cell lines or samples (data not shown). Scatter plot analysis, clustering analysis and gene annotation data revealed that the conventional hEB-formation protocol used in this study resulted in successful cell differentiation, characterized by multilayered cellular morphologies and a wider spectrum of cellular properties than previously reported (Cai et al., 2006). Therefore, we believe that our data provide important information regarding the gene expression profiles of CHA3 and CHA4 hESCs and hEBs.

Comparison of gene expression signatures during differentiation

To examine the gene expression pattern induced during the differentiation of CHA3 and CHA4 hESCs, we performed a pairwise comparison between undifferentiated CHA3 and CHA4 hESCs and hEBs differentiated for 5, 10, 15 or 20 days using microarray tools. Although two different hESC lines were studied, there was a large degree of similarity, as indicated by the inter-donor correlation coefficient (CC, 0.98) for the gene expression profiles of the two lines (Fig. 2A). These results indicate that the CHA3 and CHA4 hESC lines are similar and that differences between these cell lines do not affect gene expression. Additionally, pairwise comparison of the two CHA hESC lines and serially differentiated hEBs yielded a CC ranging from 0.89 to 0.93 (Fig. 2A), which was lower than inter-donor CC values obtained from each cell line. These data reflect the high reproducibility of repeated hybridization experiments as well as divergent gene expression during the differentiation of CHA3 and CHA4 hESCs.

Identification of differentially expressed genes

To identify differences in gene expression between the CHA3 and CHA4 hESC lines and serially differentiated hEB populations, the expression pattern of 33,155 probe sets was analyzed using a microarray. An FDR of less than 5% was chosen as the cutoff criterion for the analysis, and biased filtering was conducted using analysis of variance (ANOVA). Between the hESCs and day 20 hEBs, 4,297 and 4,610 genes were differentially expressed in the CHA3 and CHA4 hESCs, respectively. Unbiased hierarchical clustering results obtained from these samples using a microarray program indicated that the variation between donors was smaller than the differences between days in the differentiation of hEB.

Clustering analysis of the gene expression pattern obtained during CHA3 and CHA4 hESC differentiation

To analyze the relationship between undifferentiated hESCs and hEBs differentiated for various lengths of time, we compared the clusters identified from each differentiation stage. The comparison was performed on signature genes and was based on their expression and hierarchical clustering to reveal a pattern of differentially and similarly expressed genes in the CHA3 or CHA4 hESC lines (Fig. 2E and Supplementary Table S2). Of the comparison clusters, highly expressed genes in undifferentiated CHA3 and CHA4 hESCs showed a pattern similar to

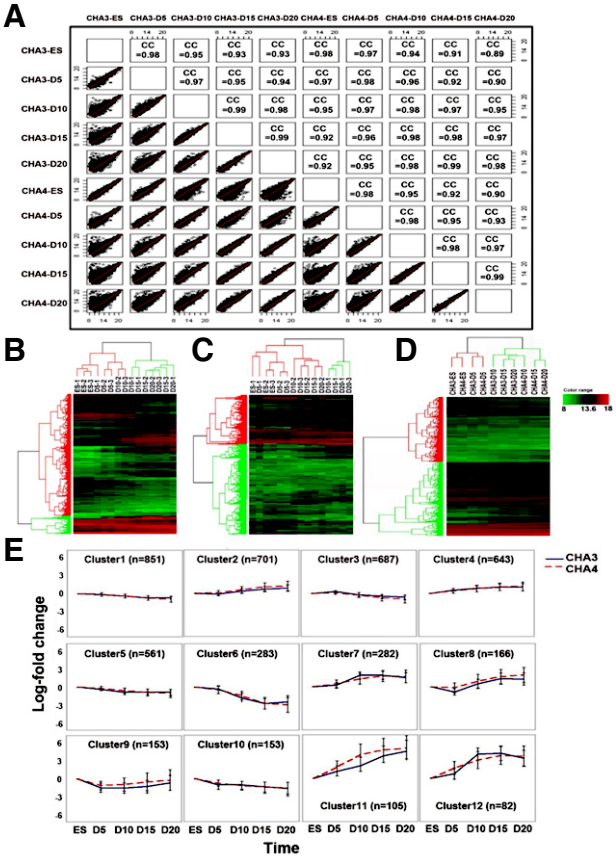

Fig. 2. Pairwise comparison, hierarchical clustering and k-means cluster analysis. (A) Scatter plots of average values for 16,143 probes expressed from undifferentiated hESC lines and serially differentiated hEB populations. The scatter plot results show similar CC values, which were greater than 0.89. CC, correlation coefficient. (B) Hierarchical clustering was analyzed in three independent experiments and filtered by ANOVA using a $P < 0.05$ cutoff to identify differences in gene expression in undifferentiated hESCs and serially differentiated hEBs from the CHA3 and (C) CHA4 hESC lines. (D) Hierarchical clustering of genes differentially expressed between CHA3 and CHA4 samples. (E) k-means cluster analysis of differentially regulated genes in CHA3 and CHA4 hESCs and their derivatives ($P < 0.05$). Twelve clusters were grouped based on their similar gene expression patterns between undifferentiated hESCs and differentiated hEBs. N, number of genes.

differentiated hESCs at cluster 6, which included *POU5F1* (*OCT4*), *SOX2* and *NANOG*, well-known markers of pluripotency and undifferentiated hESC status (Fig. 2E and Supplementary Fig. S1). The expression levels of several genes (INDO, DDX25, DPPA2 and DPPA4) in cluster 6 were also more highly expressed than in differentiated hEBs (Supplementary Fig. S2). In contrast, clusters 11 and 12 reflected tissue-specific genes associated with the development of the ectoderm, endoderm and mesoderm and included *AFP*, *SERPINA1*, *COL3A1*, *POSTN*, *DCN* and *STMN2* (Fig. 2E and Supplementary Fig. S3). These genes were highly expressed but did not show a similar pattern during the differentiation of the CHA3 or CHA4 hESC lines. Although all assays were performed under the same culture conditions, the CHA3 and CHA4 hESC lines appeared to present different differentiation potentials despite similar gene expression patterns in the undifferentiated state.

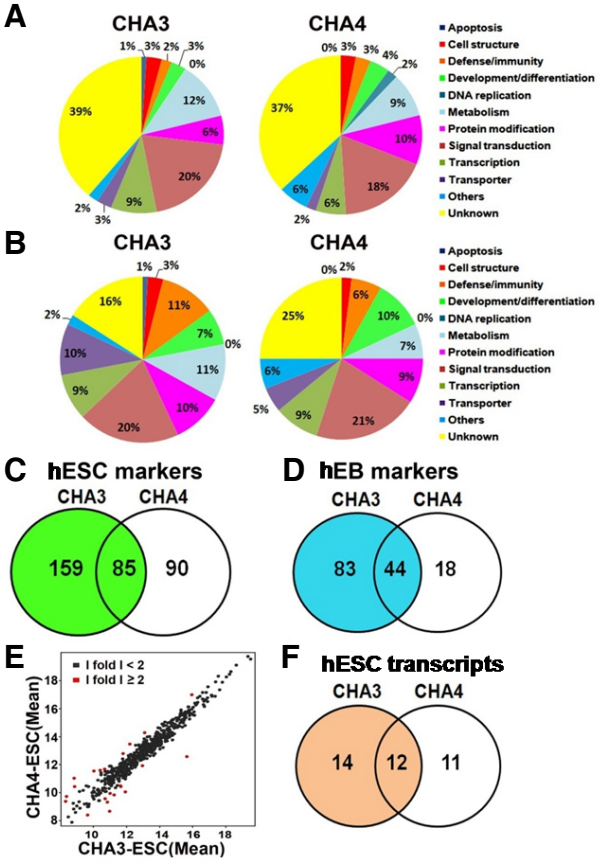

Fig. 3. Schematic representation of microarray hybridization experiments for CHA3 and CHA4 hESC lines. (A) Functional classification of the 100 most up-regulated genes in undifferentiated hESCs as compared to day 20 hEBs. (B) Functional classification of the 100 most significantly up-regulated genes in day 20 hEBs compared to undifferentiated hESCs. (C) Overlap of genes that were up-regulated more than four-fold in undifferentiated hESCs or (D) more than 20-fold in differentiated day 20 hEBs. (E) Scatter plot analysis of transcription factor genes in undifferentiated CHA3 and CHA4 hESCs. R^2 (CC) = 0.939. (F) Overlap of transcription factor genes that were down-regulated more than four-fold in day 20 differentiated hEBs compared to undifferentiated hESCs.

Functional classification of differentially expressed genes

To further examine the hallmarks of differential gene expression in each cell line, we classified the top 100 up-regulated genes based on the annotation results for undifferentiated CHA3 and CHA4 hESCs and day 20 hEBs. First, we knew that the up-regulated genes in undifferentiated CHA3 and CHA4 hESCs included many uncharacterized genes (Fig. 3A, 39% and 37% in undifferentiated CHA3 and CHA4 hESCs, respectively), and also knew that the up-regulated genes in differentiated CHA3 and CHA4 hESCs included many uncharacterized genes (Fig. 3B, 16% and 25% in day 20 CHA3 and CHA4 hEBs, respectively). The results indicate that many genes in this stage of hESC differentiation are not yet characterized and require further study. In contrast, well-defined gene groups, including those involved in development/differentiation, defense/immunity and transport, were up-regulated in the day 20 hEBs, as compared to undifferentiated CHA3 and CHA4 hESCs. Other groups, including transcription, metabolism, and protein modification genes, exhibited similar expression patterns in undifferentiated and differentiated CHA3 and CHA4 hESCs. This finding suggests that during CHA3 and CHA4 hESC differentiation, the cells mature into specialized types by the up-regulation of development- and differentiation-related genes.

Analysis of highly expressed genes in undifferentiated CHA3 and CHA4 hESCs

To investigate the genes specifically expressed in CHA3 and CHA4 hESCs and hEBs, we analyzed their gene expression profiles and examined the number of overlapping genes in undifferentiated CHA3 and CHA4 hESC lines and serially differentiated hEB samples. First, genes that were expressed four-fold greater in undifferentiated CHA3 and CHA4 hESCs than in hEBs were selected as hESC markers. Based on these selection criteria, 244 and 175 genes were found to be specifically expressed in the CHA3 and CHA4 hESC lines, respectively (Fig. 3C). The top 25 genes are listed in Supplementary Table S3. Among these genes, 85 genes overlapped between the undifferentiated CHA3 and CHA4 hESC lines (Fig. 3C and Table 1), including 17 well-characterized, hESC-specific genes, such as *NANOG*, *OCT4* and *SOX2*, as well as the recently reported genes *INDO*, *Lefty1*, *TDGF1*, *CRABP1*, *CHST4*, *CKMT1*, *C14orf115*, *GABRB3*, *ITGB1BP3*, *HESX1*, *DNMT3B*, *SCGB3A2*, *TERF1* and *TDGF1* (Beqqali et al., 2006; Bhattacharya et al., 2004; Cai et al., 2006; Richards et al., 2004; Sato et al., 2003) (Table 1). Until recently, *OCT4* and *NANOG* were used to gauge the state of hESCs because these two genes play a central role in the maintenance of hESC state (Mitsui et al., 2003; Zaehres et al., 2005). However, there are reports that these genes are expressed in mature tissues (Hart et al., 2004; Niwa et al., 2000; Tondreau et al., 2005), suggesting that these markers are not sufficient to determine the stage of hESC differentiation.

As shown in Table 1, *LEFTY1*, *INDO*, *GABRA5* and *SOX2* were more highly expressed in CHA3 and CHA4 hESCs than in hEBs. *LEFTY1*, which is the mouse homolog of *LEFTYB* and an inhibitory ligand belonging to the transforming growth factor (TGF)-beta super-family, is proposed to be the downstream target gene of *OCT3/4* (Niwa et al., 2000). The up-regulation of *LEFTY1* suggests that down-regulation of the TGF-beta pathway may play an important role in the maintenance of hESC status. Even if *INDO*, *GABRB3*, *DNMT3B* and *CKMT1* are up-regulated in undifferentiated CHA3 and CHA4 hESCs, as previously suggested (Beqqali et al., 2006; Cai et al., 2006), further verification is required to determine whether these genes regulate or are involved in the maintenance of the undifferentiated hESC stage. Other hESC markers identified, such as *C1orf182* and *C9orf61*, which encode hypothetical proteins, are similar to previously reported hESC markers (*C20orf1*, *C15orf15*, and *C20orf129*) (Bhattacharya et al., 2004).

Of the 85 specifically up-regulated genes during the undifferentiated stage, 17 were well characterized and an additional 68 of the hESC-specific markers were previously known genes, including *DPPA2*, *DPPA4*, and *DDX25/GRTH* (Table 1). These genes are from the same family as *DPPA5* and *DDX2*, which have been previously suggested as hESC markers (Bhattacharya et al., 2004; Cai et al., 2006) (Table 1). Dppa2 and Dppa4 may play a role in maintaining cellular potency in the inner cell mass (ICM) of pluripotent mouse and human cells and in the developing germ line of mice (Maldonado-Saldivia et al., 2007). Meanwhile, DDX25/GRTH is a multifunctional RNA helicase that is an essential post-transcriptional regulator of spermatid development and the completion of spermatogenesis (Tsai-Morris et al., 2004). DPPA2 and DDX25 are dramatically down-regulated in comparison with OCT4, SOX2 and

Table 1. List of 85 specifically up-regulated genes in undifferentiated CHA3 and CHA4 hESC lines

Gene symbol	Gene ID	Gene name	Reference
A2ML1	hCG1811680.2	Alpha-2-macroglobulin-like 1	
AASS	hCG33410.3	Aminoadipate-semialdehyde synthase	
ADCY2	hCG1749581.3	Adenylate cyclase 2 (brain)	
ADD2	hCG39485.3	Adducin 2 (beta)	
AK5	hCG1810787.1	Adenylate kinase 5	
ARTN	hCG25200.3	Artemin	
C14orf115	hCG1646085.2	Chromosome 14 open reading frame 115	Beqqali et al. (2006)
C1orf182	hCG1999359	Chromosome 1 open reading frame 182	
C9orf61	hCG30537.3	Chromosome 9 open reading frame 61	
CABYR	hCG37186.4	Calcium binding tyrosine-(Y)-phosphorylation regulated	
CACNA1G	hCG29633.3	Calcium channel, voltage-dependent, alpha 1G subunit	
CAMKV	hCG95981.4	CaM kinase-like vesicle-associated	
CDCA7L	hCG40002.4	Cell division cycle associated 7-like	
CHST4	hCG1779887.2	Carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4	Cai et al. (2006)
CKMT1	NM_020990.2	Creatine kinase, mitochondrial 1 (ubiquitous)	Cai et al. (2006)
CNTN1	hCG38261.3	Contactin 1	
CRABP1	hCG38931.2	Cellular retinoic acid binding protein 1	Bhattacharya et al. (2004)
CTGF	hCG22108.3	Connective tissue growth factor	
CXCL6	hCG16362.1	Chemokine (C-X-C motif) ligand 6	
DCAMKL1	hCG32199.3	Doublecortin and CaM kinase-like 1	
DDX25	hCG38912.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 25	
DEPDC2	hCG1810953.3	DEP domain containing 2	
DNMT3B	hCG37138.4	DNA (cytosine-5-)-methyltransferase 3 beta	Richards et al. (2004)
DPPA2	hCG1780948.2	Developmental pluripotency associated 2	
DPPA4	hCG27412.2	Developmental pluripotency associated 4	
FBXL16	hCG1779854.3	F-box and leucine-rich repeat protein 16	
FGF2	hCG37365.2	Fibroblast growth factor 2 (basic)	
FLJ12505	hCG37733.2	Hypothetical protein FLJ12505	
FLJ12684	hCG1746629.2	Hypothetical protein FLJ12684	
FLJ30707	hCG1646054.2	Hypothetical protein FLJ30707	
GABRA5	hCG1811407.1	Gamma-aminobutyric acid (GABA) A receptor, alpha 5	
GABRB3	hCG1811406.1	Gamma-aminobutyric acid (GABA) A receptor, beta 3	Cai et al. (2006)
GAP43	hCG2022854	Growth associated protein 43	
GPC4	hCG16065.3	Glypican 4	
GPR19	hCG2039474	G protein-coupled receptor 19	
GPR23	hCG19147.3	G protein-coupled receptor 23	
GRPR	hCG15741.3	Gastrin-releasing peptide receptor	
HESX1	hCG1640606.3	Homeo box (expressed in ES cells) 1	Richards et al. (2004)
INA	hCG24135.2	Internexin neuronal intermediate filament protein, alpha	
INDO	hCG27061.3	Indoleamine-pyrrole 2,3 dioxygenase	Beqqali et al. (2006)
INHBE	hCG1818161.1	Inhibin, beta E	
ITGB1BP3	hCG23523.3	Integrin beta 1 binding protein 3	Cai et al. (2006)
LEFTY1	hCG1737728.1	Left-right determination factor 1	Cai et al. (2006)
LOC168474	hCG1640107.4	hypothetical protein LOC168474	
LOC283174	hCG1820422.1	hypothetical protein LOC283174	
MDN1	hCG2036531	MDN1, midasin homolog (yeast)	
NALP4	hCG1733040.2	NACHT, leucine rich repeat and PYD containing 4	
NANOG	hCG1730824.3	Nanog homeobox	Mitsui et al. (2003)

(continued)

Gene symbol	Gene ID	Gene name	Reference
NAP1L2	hCG1641574.2	Nucleosome assembly protein 1-like 2	
NEF3	hCG16610.3	Neurofilament 3 (150kDa medium)	
NEFL	hCG16611.3	Neurofilament, light polypeptide 68kDa	
NELL2	hCG38302.3	NEL-like 2 (chicken)	
NMNAT2	hCG2039672	Nicotinamide nucleotide adenylyltransferase 2	
NMU	hCG20599.3	Neuromedin U	
NPTX2	hCG41329.3	Neuronal pentraxin II	
OLFM1	hCG20208.3	Olfactomedin 1	
OSBPL6	hCG43235.3	Oxysterol binding protein-like 6	
PCSK9	hCG33025.4	Proprotein convertase subtilisin/kexin type 9	
POU5F1	hCG25999.3	POU domain, class 5, transcription factor 1	Mitsui et al. (2003)
PTHB1	hCG1744365.1	Parathyroid hormone-responsive B1	
PTPRB	hCG25146.4	Protein tyrosine phosphatase, receptor type, B	
PTPRZ1	hCG33411.3	Protein tyrosine phosphatase, receptor-type, Z polypeptide 1	
RAB39B	hCG18045.3	RAB39B, member RAS oncogene family	
RARRES2	hCG15000.3	Retinoic acid receptor responder (tazarotene induced) 2	
RASL11B	hCG20672.3	RAS-like, family 11, member B	
RDH12	hCG1812051.1	Retinol dehydrogenase 12 (all-trans and 9-cis)	
RET	hCG22740.3	Ret proto-oncogene	
RNF182	hCG1647724.3	Ring finger protein 182	
SAMHD1	hCG17960.3	SAM domain and HD domain 1	
SCG3	hCG38751.4	Secretogranin III	
SCGB3A2	hCG1645642.2	Secretoglobin, family 3A, member 2	Cai et al. (2006)
SEPHS1	hCG2017606	Selenophosphate synthetase 1	
SLC10A4	hCG17411.2	Solute carrier family 10, member 4	
SLC7A3	hCG1990500	Solute carrier family 7, member 3	
SOX2	hCG2021649	SRY (sex determining region Y)-box 2	Richards et al. (2004)
SYT1	hCG2016754.1	Synaptotagmin I	
TAC1	hCG16982.3	Tachykinin, precursor 1	
TAF4B	hCG38478.3	TAF4b RNA polymerase II,	
TDGF1	hCG15320.3	Teratocarcinoma-derived growth factor 1	Sato et al. (2003)
TERF1	hCG1983637	Telomeric repeat binding factor (NIMA-interacting) 1	Sato et al. (2003)
TIMP4	hCG28316.3	TIMP metallopeptidase inhibitor 4	
TNFRSF8	hCG25063.3	Tumor necrosis factor receptor superfamily, member 8	
USP44	hCG1642341.3	Ubiquitin specific peptidase 44	
WIF1	hCG22745.3	WNT inhibitory factor 1	
ZIC3	hCG19782.2	Zic family member 3 heterotaxy 1	Lim et al. (2007)

NANOG in differentiating hEBs (Supplementary Figs. S1 and S2). Characterization of the expression and function of DPPA2 and DDX25 is likely to unveil new aspects of the regulation of pluripotency in hESCs.

Analysis of highly expressed genes in differentiated hEBs

To examine the differentiation potential of CHA3 and CHA4 hESCs, we investigated the genes that were highly expressed in differentiated hEBs. We first selected genes that were up-regulated at least 20 times more in the day 20 hEBs than in the undifferentiated CHA3 and CHA4 hESCs. In this context, 127 and 62 genes were detected in the CHA3- and CHA4-derived hEBs, respectively (Fig. 3D). Using our selection criteria, a total of 44 genes, including the well-known genes *AFP, CXCL14, DCN, GATA6, H19, IGFBP3, MGP and SERPINA1* (Cai et al., 2006), overlapped between the day 20 hEBs derived from both cell lines (Fig. 3D and Table 2), and we determined 36 candidate hEB-specific markers. Among them, SNAI2, also called SLUG, belongs to the Snail family of zinc-finger transcription factors that share an evolutionarily conserved role in mesoderm formation in invertebrates and vertebrates. SNAI2 specifically triggers epithelial-mesenchymal transitions and plays an important role in developmental processes (Perez-Mancera et al., 2007). Another marker, *TTR* is expressed in hepatocytes and the visceral yolk sac during embryogenesis (Costa et al., 1990). Meanwhile, fibrinogen beta chain and fibrinogen gamma chain are components of fibrinogen, a blood-borne glycoprotein composed of three pairs of non-identical polypeptide chains. Pro-

Table 2. List of 44 genes in common between the CHA3 and CHA4 hESC lines after 20 days of differentiation

Gene symbol	Gene ID	Gene name	Reference
AFP	hCG15197.2	Alpha-fetoprotein	Cai et al. (2006)
AGT	hCG14741.3	Angiotensinogen	
ALPK2	hCG16539.4	Alpha-kinase 2	
ANXA8	NM_001630.1	Annexin A8	
APOA1	hCG41332.2	Apolipoprotein A-I	
APOA2	hCG1766509.2	Apolipoprotein A-II	
APOB	hCG20898.2	Apolipoprotein B (including Ag(x) antigen)	
APOM	hCG43690.3	Apolipoprotein M	
ASS	hCG31245.3	Argininosuccinate synthetase	
CDH5	hCG26635.3	Cadherin 5, type 2, VE-cadherin (vascular epithelium)	
CGA	hCG33165.3	Glycoprotein hormones, alpha polypeptide	
COL3A1	hCG25172.3	Collagen, type III, alpha 1	
COL5A1	hCG2021790	Collagen, type V, alpha 1	
COL6A3	hCG23027.2	Collagen, type VI, alpha 3	
COLEC12	hCG38030.3	Collectin sub-family member 12	
CXCL14	hCG39697.3	Chemokine (C-X-C motif) ligand 14	Cai et al. (2006)
DCN	hCG24110.2	Decorin	Cai et al. (2006)
FGB	hCG2026446	Fibrinogen beta chain	
FGG	hCG28288.4	Fibrinogen gamma chain	
FLJ32115	hCG24421.3	Hypothetical protein FLJ32115	
FMOD	hCG24329.2	Fibromodulin	
GATA6	hCG37191.2	GATA binding protein 6	Cai et al. (2006)
GSTA2	hCG1779105.2	Glutathione S-transferase A2	
GUCY1A3	hCG28051.3	Guanylate cyclase 1, soluble, alpha 3	
H19	hCG1640777.2	H19, imprinted maternally expressed untranslated mRNA	Cai et al. (2006)
HOXB5	hCG29354.2	Homeo box B5	
IGFBP3	hCG1735376.2	Insulin-like growth factor binding protein 3	Cai et al. (2006)
LUM	hCG24108.2	Lumican	
MGP	hCG24416.2	Matrix Gla protein	Cai et al. (2006)
MTP	hCG37774.2	Microsomal triglyceride transfer protein	
PDZK1	hCG2004792.1	PDZ domain containing 1	
PITX2	hCG21422.3	Paired-like homeodomain transcription factor 2	
RELN	hCG18198.3	Reelin	
RSPO3	hCG1640838.3	R-spondin 3 homolog (Xenopus laevis)	
SERPINA1	hCG2029168.1	Serpin peptidase inhibitor, clade A, member 1	Cai et al. (2006)
SLC40A1	hCG24798.3	Solute carrier family 40 (iron-regulated transporter), member 1	
SLIT3	hCG36986.3	Slit homolog 3 (Drosophila)	
SLN	hCG39331.1	Sarcolipin	
SNAI2	hCG32298.3	Snail homolog 2 (Drosophila)	
STMN2	hCG20839.2	Stathmin-like 2	
TCF21	hCG20014.4	Transcription factor 21	
TTR	hCG25470.3	Transthyretin (prealbumin, amyloidosis type I)	
ZFHX1B	hCG2006491	Zinc finger homeobox 1b	
ZNF503	hCG23983.2	Zinc finger protein 503	

duced by the liver, fibrinogen helps to control blood loss by encouraging clotting (Mosesson et al., 2001) (Table 2).

Next, we analyzed the up-regulated genes associated with germ-layer specificity that are unique to the early development of CHA3- and CHA4-derived hEBs (Table 3). The expression of hEB markers associated with the development of the endoderm, ectoderm and mesoderm, such as *AFP*, *DCN*, *H19*, *SERPINA1*, *STMN2*, *COL3A1* and *MGP*, was 100-fold higher at day 20 of

Table 3. List of 28 genes connected with the early developmental stages of differentiating CHA3 and CHA4 hEBs

Classification	Gene ID	Gene symbol	Gene name	Group	Fold			
					D5/ES	D10/ES	D15/ES	D20/ES
Endoderm	hCG15197.2	*AFP*	Alpha-fetoprotein	CHA3	7.6	149.2	252.1	257.2
				CHA4	3.4	9.6	39.7	118.4
	hCG24110.2	*DCN*	Decorin	CHA3	11.4	86.5	190.7	349.5
				CHA4	1.9	7.1	11.9	16.8
	hCG37191.2	*GATA6*	GATA binding protein 6	CHA3	9.1	25.9	39.5	36.7
				CHA4	2.8	10.6	29.8	41.9
	hCG1640777.2	*H19*	H19, imprinted maternally expressed untranslated mRNA	CHA3	20.5	85.4	127.4	144.7
				CHA4	16.6	194.4	271.1	399.9
	hCG2029168.1	*SERPINA1*	Serpin peptidase inhibitor, clade A, member 1	CHA3	4.7	174.4	767.2	992.1
				CHA4	3.0	14.4	89.5	389.9
Ectoderm	hCG39012.2	*COBLL1*	COBL-like 1	CHA3	2.5	7.7	13.2	17.7
				CHA4	1.7	4.2	7.5	6.7
	hCG26636.3	*CDH11*	Cadherin 11	CHA3	5.3	13.8	20.7	19.5
				CHA4	2.1	10.2	7.4	5.9
	hCG29354.2	*HOXB5*	Homeo box B5	CHA3	4.5	17.5	24.9	28.2
				CHA4	1.6	8.0	15.9	20.3
	hCG18198.3	*RELN*	Reelin	CHA3	11.6	41.4	63.1	62.4
				CHA4	2.2	17.2	26.7	28.5
	hCG17980.4	*SEMA3C*	Sema domain, immunoglobulin domain, (semaphorin) 3C	CHA3	3.3	9.5	18.9	20.2
				CHA4	-1.1	9.4	14.8	8.9
	hCG36986.3	*SLIT3*	Slit homolog 3 (Drosophila)	CHA3	2.5	9.1	20.1	38.3
				CHA4	2.3	18.9	16.8	25.4
	hCG20839.2	*STMN2*	Stathmin-like 2	CHA3	7.8	32.3	89.3	203.4
				CHA4	1.7	39.2	36.2	37.5
Mesoderm	hCG18572.3	*AMOT*	Angiomotin	CHA3	3.0	5.4	10.6	9.9
				CHA4	2.5	15.4	15.9	9.4
	hCG25172.3	*COL3A1*	Collagen, type III, alpha 1	CHA3	18.4	119.3	225.1	299.6
				CHA4	12.7	107.0	302.4	393.0
	hCG2021790	*COL5A1*	Collagen, type V, alpha 1	CHA3	4.1	17.6	35.4	50.1
				CHA4	2.6	8.3	20.5	25.8
	hCG36799.2	*HAND1*	Heart and neural crest derivatives expressed 1	CHA3	6.9	18.2	18.2	16.9
				CHA4	7.4	20.0	20.0	18.6
	hCG39210.3	*HAPLN1*	Hyaluronan and proteoglycan link protein 1	CHA3	8.5	20.2	20.2	12.8
				CHA4	6.4	44.3	51.5	45.9
	hCG14618.3	*MBNL3*	Muscleblind-like 3 (Drosophila)	CHA3	3.3	6.3	9.0	9.2
				CHA4	1.4	9.9	13.0	7.6
	hCG24416.2	*MGP*	Matrix Gla protein	CHA3	3.4	35.6	74.4	202.1
				CHA4	4.6	7.0	17.9	55.3
	hCG39121.3	*MSX1*	Msh homeo box homolog 1 (Drosophila)	CHA3	4.0	10.3	11.0	6.3
				CHA4	2.1	4.1	10.0	9.7
	hCG18302.3	*MYL7*	Myosin, light polypeptide 7, regulatory	CHA3	4.6	12.9	15.2	8.1
				CHA4	2.3	3.3	13.6	13.3
	hCG27262.3	*PCSK6*	Proprotein convertase subtilisin/kexin type 6	CHA3	1.1	3.3	8.4	11.2
				CHA4	2.1	3.8	5.4	11.6
	hCG32814.3	*POSTN*	Periostin, osteoblast specific factor	CHA3	7.6	82.8	108.0	142.6
				CHA4	3.3	5.6	50.3	97.9
	hCG41400.2	*TGFBI*	Transforming growth factor, beta-induced	CHA3	2.7	4.0	11.5	16.1
				CHA4	1.3	15.8	22.3	9.3
Other	hCG24313.3	*LTBR*	Lymphotoxin beta receptor	CHA3	1.8	5.0	8.7	12.3
				CHA4	2.4	5.6	9.1	10.4
	hCG2043041	*PLAGL1*	Pleiomorphic adenoma gene-like 1	CHA3	2.7	12.1	29.8	34.6
				CHA4	2.4	3.3	15.3	27.9
	hCG1985832	*PDGFRB*	Platelet-derived growth factor receptor, beta polypeptide	CHA3	2.0	7.0	12.2	15.3
				CHA4	1.2	5.5	9.3	13.6
	hCG16030.2	*SDCCAG33*	Serologically defined colon cancer antigen 33	CHA3	3.8	15.9	26.5	34.9
				CHA4	1.0	4.2	8.5	12.2

Table 4. List of 12 transcripts down-regulated during hESC differentiation

Gene symbol	Gene ID	Gene name	Group	Fold			
				D5/ES	D10/ES	D15/ES	D20/ES
BCL11B	hCG1657560.3	B-cell CLL/lymphoma 11B (zinc finger protein)	CHA3	-1.3	-2.7	-6.4	-7.6
			CHA4	-1.4	-3.3	-5.4	-3.5
EGR1	hCG18777.1	Early growth response 1	CHA3	-6.1	-6.4	-11.7	-7.9
			CHA4	-13	-7.3	-7.8	-5.4
HESX1	hCG1640606.3	Homeo box (expressed in ES cells) 1	CHA3	-1	-3.3	-5.5	-7.8
			CHA4	-1.2	-2.6	-5.5	-6.2
NANOG	hCG1730824.3	Nanog homeobox	CHA3	-1.6	-2.2	-3.7	-5
			CHA4	-1.4	-3	-4	-4.5
NR5A2	hCG19256.3	Nuclear receptor subfamily 5, group A, member 2	CHA3	-5.5	-4.8	-2.7	-3.2
			CHA4	-1.4	-3	-4	-4.5
PHC1	hCG2003719.1	Polyhomeotic-like 1 (Drosophila)	CHA3	-1.5	-2	-4.5	-5.7
			CHA4	-1.1	-2.1	-3.6	-4
POU5F1	hCG25999.3	POU domain, class 5, transcription factor 1	CHA3	-1.3	-2.1	-4.8	-6.7
			CHA4	-1.1	-2.2	-4.4	-4.9
PRDM14	hCG18448.3	PR domain containing 14	CHA3	-1.5	-6	-13.2	-32.1
			CHA4	1	-3.6	-15	-6.8
SOX2	hCG2021649	SRY (sex determining region Y)-box 2	CHA3	-1.3	-2.5	-7.9	-14.3
			CHA4	-1.3	-2.7	-4.6	-7.3
TFAP2C	hCG37869.3	Transcription factor AP-2 gamma	CHA3	-2	-3.3	-7.2	-7.9
			CHA4	-1.6	-3.5	-5.7	-6.6
ZIC3	hCG19782.2	Zic family member 3 heterotaxy 1	CHA3	-1.1	-1.9	-4.5	-10.5
			CHA4	-1.1	-2.5	-5.4	-6.4
ZNF398	hCG16174.3	Zinc finger protein 398	CHA3	-1.1	-1.6	-3.2	-5.8
			CHA4	1.2	-1.7	-4.1	-2.5

hEB differentiation than before differentiation (Table 3). Because differentiated hESCs include many different cell types, detailing the gene expression profiles of differentiated hEBs could lead to the identification of markers involved in the directed differentiation of ectoderm, mesoderm or endoderm and could signal when specific pathways are activated during differentiation. Therefore, these markers could be used to indicate and define differentiated hEB status and to establish reproducible hEB-formation methods.

Analysis of the transcriptome in undifferentiated CHA3 and CHA4 hESCs

Generally, the transcriptome plays an important role in the functioning of all cells. We analyzed the expression pattern of transcriptionally activated genes between undifferentiated CHA3 and CHA4 hESCs using a pairwise comparison tool (Fig. 3E). The expression levels of distinct transcripts were highly correlated (CC = 0.936), with 12 transcripts down-regulated by more than 50% in the hEB stage as compared to the undifferentiated hESC stage of both cell lines (Fig. 3F and Table 4). These 12 transcripts, which included such hESC-specific transcripts as *POU5F1*, *SOX2* and *NANOG*, as well as such uncharacterized transcription factors as *ZIC3*, *HESX1*, *PRDM14*, *EGR1*, *TFAC2C* and *PHC1*, were expressed at four-fold higher levels during the undifferentiated stage than during the hEB stage. The high expression of *POU5F1*, *SOX2* and *NANOG* in hESCs has been previously reported (Mitsui et al., 2003; Niwa et al., 2000; Wei et al., 2005). Additional transcripts over-expressed in the undifferentiated CHA3 and CHA4 hESCs are listed in Supplementary Table S4.

Additionally, some transcripts (i.e., *ZIC3*, *HESX1*, *PRDM14*, *EGR1*, *TFAP2C*, and *PHC1*) were more highly up-regulated in CHA3 and CHA4 hESC lines than in differentiated hEBs. Recently, it was reported that *ZIC3* plays a key role in the maintenance of pluripotency by preventing endodermal lineage differentiation of hESCs (Lim et al., 2007). We also observed high expression of *HESX1*, a homeobox gene that regulates aspects of pituitary development, in both CHA3 and CHA4 hESCs, which is consistent with previous reports (Richards et al., 2004). *PRDM14*, which was up-regulated as well, is involved in histone methylation and regulates self-renewal by suppressing gene expression (Hu et al., 2005), while the *EGR1* is essential for cellular proliferation, growth and differentiation (Sukhatme et al., 1988). Meanwhile, *TFAP2C* plays an important role in the early development, morphogenesis and survival of mouse embryos (Luo et al., 2002); *PHC1* is a DNA-binding protein thought to be required for self-renewal and maintenance of the lineage compartment (Kim et al., 2008); and the orphan nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells (Heng et al., 2010). In sum, our data identify highly expressed and overlapping transcripts between the CHA3 and CHA4 hESC lines that may form and significantly affect transcriptional networks or regulate epigenetic modifications necessary for hESC maintenance. Future studies will determine whether these genes, which are associated with the self-renewal of hESCs, are unique stem cell markers.

Table 5. Classification of pathway-related genes whose expression significantly differed between hESCs and hEBs

KEGGID	KEGG pathway	Count	pvalue
hsa04110	Cell cycle	47	4.0E-06
hsa04115	p53 signaling pathway	28	5.5E-04
hsa00220	Urea cycle and metabolism of amino groups	16	5.6E-04
hsa05215	Prostate cancer	30	0.01
hsa00251	Glutamate metabolism	13	0.02
hsa00280	Valine, leucine and isoleucine degradation	17	0.02
hsa00520	Nucleotide sugars metabolism	5	0.02
hsa00903	Limonene and pinene degradation	12	0.03

Pathway analysis in CHA3 and CHA4 hESCs and hEBs

Finally, to determine the pathways associated with the pluripotency or differentiation of the CHA3 and CHA4 hESC lines, we examined the KEGG pathway database by assigning *p* values using the gene expression profiles of undifferentiated CHA3 and CHA4 hESCs and serially differentiated hEBs. As shown in Table 5, the cell cycle, prostate cancer and p53 signaling pathways contained 47, 30 and 28 of the genes identified in our study, respectively. Furthermore, several genes involved in metabolic processes were also identified (Table 5). The data suggest that these pathways may play crucial roles in the maintenance and differentiation of CHA3 and CHA4 hESCs. Our identification of differentially expressed genes supports previous reports and further demonstrates the importance of the cell cycle, p53 signaling, the urea cycle, and amino acid metabolism in CHA3 and CHA4 hESC differentiation. This observation also supports earlier findings suggesting that cell cycle control differs between hESCs and differentiated cells (Savatier et al., 1994; 1996). The proliferation of differentiated cells is controlled by the regulation of progression from the G1 to the S phase, and hESCs have a short G1 phase. Additionally, our results indicate that these cells exhibit different type of p53 signaling, consistent with an earlier report that p53 induces hESC differentiation by suppressing *NANOG* expression (Lin et al., 2005).

CONCLUSIONS

Safe and effective cell-based therapies require the characterization of CHA3 and CHA4 hESCs and hEBs and the targeted selection of tissue-specific progenitors. Our study identifies novel hESC-specific biomarkers, including transcriptomes, and hEB-specific markers in the newly established CHA3 and CHA4 hESC lines using genome-wide differential gene expression analysis. Our results provide a foundation for future studies of the molecular mechanisms governing “stemness,” the self-renewal properties of hESCs and hEB differentiation into three germ layers and tissue-specific lineages.

Note: Supplementary Information is available on the Molecules and Cells Website (www.molcells.org).

ACKNOWLEDGMENTS

This research was supported by a grant (10033642) from the Industry Sources Development Project funded by the Ministry of Knowledge Economy, Republic of Korea.

REFERENCES

Abeyta, M.J., Clark, A.T., Rodriguez, R.T., Bodnar, M.S., Pera, R.A., and Firpo, M.T. (2004). Unique gene expression signatures of independently-derived human embryonic stem cell lines. Hum. Mol. Genet. *13*, 601-608.

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S., Bevan, S., et al. (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat. Biotechnol. *25*, 803-816.

Ahn, S.E., Kim, S., Park, K.H., Moon, S.H., Lee, H.J., Kim, G.J., Lee, Y.J., Park, K.H., Cha, K.Y., and Chung, H.M. (2006). Primary bone-derived cells induce osteogenic differentiation without exogenous factors in human embryonic stem cells. Biochem. Biophys. Res. Commun. *340*, 403-408.

Amit, M., Carpenter, M.K., Inokuma, M.S., Chiu, C.P., Harris, C.P., Waknitz, M.A., Itskovitz-Eldor, J., and Thomson, J.A. (2000). Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. Dev. Biol. *227*, 271-278.

Bauwens, C.L., Peerani, R., Niebruegge, S., Woodhouse, K.A., Kumacheva, E., Husain, M., and Zandstra, P.W. (2008). Control of human embryonic stem cell colony and aggregate size heterogeneity influences differentiation trajectories. Stem Cells *26*, 2300-2310.

Beqqali, A., Kloots, J., Ward-van Oostwaard, D., Mummery, C., and Passier, R. (2006). Genome-wide transcriptional profiling of human embryonic stem cells differentiating to cardiomyocytes. Stem Cells *24*, 1956-1967.

Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., Luo, Y., Yang, A.X., Joshi, B.H., Ginis, I., Thies, R.S., Amit, M., et al. (2004). Gene expression in human embryonic stem cell lines: unique molecular signature. Blood *103*, 2956-2964.

Bouchard, R.A., and Lemmens, T. (2008). Privatizing biomedical research--a ‘third way’. Nat. Biotechnol. *26*, 31-36.

Cai, J., Chen, J., Liu, Y., Miura, T., Luo, Y., Loring, J.F., Freed, W.J., Rao, M.S., and Zeng, X. (2006). Assessing self-renewal and differentiation in human embryonic stem cell lines. Stem Cells *24*, 516-530.

Costa, R.H., Van Dyke, T.A., Yan, C., Kuo, F., and Darnell, J.E., Jr. (1990). Similarities in transthyretin gene expression and differences in transcription factors: liver and yolk sac compared to choroid plexus. Proc. Natl. Acad. Sci. USA *87*, 6589-6593.

Dutrillaux, B., and Viegas-Pequignot, E. (1981). High resolution R- and G-banding on the same preparation. Hum. Genet. *57*, 93-95.

Fathi, A., Pakzad, M., Taei, A., Brink, T.C., Pirhaji, L., Ruiz, G., Sharif Tabe Bordbar, M., Gourabi, H., Adjaye, J., Baharvand, H., et al. (2009). Comparative proteome and transcriptome analy-ses of embryonic stem cells during embryoid body-based differentiation. Proteomics *9*, 4859-4870.

Hart, A.H., Hartley, L., Ibrahim, M., and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. Dev. Dyn. *230*, 187-198.

Heins, N., Englund, M.C., Sjoblom, C., Dahl, U., Tonning, A., Bergh, C., Lindahl, A., Hanson, C., and Semb, H. (2004). Derivation, characterization, and differentiation of human embryonic stem cells. Stem Cells *22*, 367-376.

Heng, J.C., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., et al. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell *6*, 167-174.

Hu, M., Yao, J., Cai, L., Bachman, K.E., van den Brule, F., Velculescu, V., and Polyak, K. (2005). Distinct epigenetic changes in the stromal cells of breast cancers. Nat. Genet. *37*, 899-905.

Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., and Benvenisty, N. (2000). Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. Mol. Med. *6*, 88-95.

Kim, J., Moon, S.H., Lee, S.H., Lee, D.R., Koh, G.Y., and Chung, H.M. (2007a). Effective isolation and culture of endothelial cells in embryoid body differentiated from human embryonic stem cells. Stem Cells Dev. *16*, 269-280.

Kim, S., Kim, G.J., Miyoshi, H., Moon, S.H., Ahn, S.E., Lee, J.H.,

Lee, H.J., Cha, K.Y., and Chung, H.M. (2007b). Efficiency of the elongation factor-1alpha promoter in mammalian embryonic stem cells using lentiviral gene delivery systems. Stem Cells Dev. *16*, 537-545.

Kim, S.E., Kim, B.K., Gil, J.E., Kim, S.K., and Kim, J.H. (2007c). Comparative analysis of the developmental competence of three human embryonic stem cell lines *in vitro*. Mol. Cells *23*, 49-56.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.

Lee, J.E., Kang, M.S., Park, M.H., Shim, S.H., Yoon, T.K., Chung, H.M., and Lee, D.R. (2010). Evaluation of 28 human embryonic stem cell lines for use as unrelated donors in stem cell therapy: implications of HLA and ABO genotypes. Cell Transplant. *19*, 1383-1395.

Lim, L.S., Loh, Y.H., Zhang, W., Li, Y., Chen, X., Wang, Y., Bakre, M., Ng, H.H., and Stanton, L.W. (2007). Zic3 is required for maintenance of pluripotency in embryonic stem cells. Mol. Biol. Cell *18*, 1348-1358.

Lin, T., Chao, C., Saito, S., Mazur, S.J., Murphy, M.E., Appella, E., and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nat. Cell Biol. *7*, 165-171.

Luo, T., Matsuo-Takasaki, M., Thomas, M.L., Weeks, D.L., and Sargent, T.D. (2002). Transcription factor AP-2 is an essential and direct regulator of epidermal development in Xenopus. Dev. Biol. *245*, 136-144.

Maldonado-Saldivia, J., van den Bergen, J., Krouskos, M., Gilchrist, M., Lee, C., Li, R., Sinclair, A.H., Surani, M.A., and Western, P.S. (2007). Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. Stem Cells *25*, 19-28.

Mitalipova, M., Calhoun, J., Shin, S., Wininger, D., Schulz, T., Noggle, S., Venable, A., Lyons, I., Robins, A., and Stice, S. (2003). Human embryonic stem cell lines derived from discarded embryos. Stem Cells *21*, 521-526.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Mosesson, M.W., Siebenlist, K.R., and Meh, D.A. (2001). The structure and biological features of fibrinogen and fibrin. Ann. N Y Acad. Sci. *936*, 11-30.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372-376.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313-315.

Park, J.H., Kim, S.J., Oh, E.J., Moon, S.Y., Roh, S.I., Kim, C.G., and Yoon, H.S. (2003). Establishment and maintenance of human embryonic stem cells on STO, a permanently growing cell line. Biol. Reprod. *69*, 2007-2014.

Perez-Mancera, P.A., Bermejo-Rodriguez, C., Gonzalez-Herrero, I., Herranz, M., Flores, T., Jimenez, R., and Sanchez-Garcia, I. (2007). Adipose tissue mass is modulated by SLUG (SNAI2). Hum. Mol. Genet. *16*, 2972-2986.

Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. Nat. Biotechnol. *18*, 399-404.

Richards, M., Tan, S.P., Tan, J.H., Chan, W.K., and Bongso, A. (2004). The transcriptome profile of human embryonic stem cells as defined by SAGE. Stem Cells *22*, 51-64.

Sato, N., Sanjuan, I.M., Heke, M., Uchida, M., Naef, F., and Brivanlou, A.H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. Dev. Biol. *260*, 404-413.

Savatier, P., Huang, S., Szekely, L., Wiman, K.G., and Samarut, J. (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene *9*, 809-818.

Savatier, P., Lapillonne, H., van Grunsven, L.A., Rudkin, B.B., and Samarut, J. (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*, 309-322.

Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O., and Thomson, J.A. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc. Natl. Acad. Sci. USA *100*, 13350-13355.

Sukhatme, V.P., Cao, X.M., Chang, L.C., Tsai-Morris, C.H., Stamenkovich, D., Ferreira, P.C., Cohen, D.R., Edwards, S.A., Shows, T.B., Curran, T., et al. (1988). A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. Cell *53*, 37-43.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Tondreau, T., Meuleman, N., Delforge, A., Dejeneffe, M., Leroy, R., Massy, M., Mortier, C., Bron, D., and Lagneaux, L. (2005). Mesenchymal stem cells derived from CD133-positive cells in mobilized peripheral blood and cord blood: proliferation, Oct4 expression, and plasticity. Stem Cells *23*, 1105-1112.

Tsai-Morris, C.H., Sheng, Y., Lee, E., Lei, K.J., and Dufau, M.L. (2004). Gonadotropin-regulated testicular RNA helicase (GRTH/Ddx25) is essential for spermatid development and completion of spermatogenesis. Proc. Natl. Acad. Sci. USA *101*, 6373-6378.

Wei, C.L., Miura, T., Robson, P., Lim, S.K., Xu, X.Q., Lee, M.Y., Gupta, S., Stanton, L., Luo, Y., Schmitt, J., et al. (2005). Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. Stem Cells *23*, 166-185.

Zaehres, H., Lensch, M.W., Daheron, L., Stewart, S.A., Itskovitz-Eldor, J., and Daley, G.Q. (2005). High-efficiency RNA interference in human embryonic stem cells. Stem Cells *23*, 299-305.

Zeng, X., Miura, T., Luo, Y., Bhattacharya, B., Condie, B., Chen, J., Ginis, I., Lyons, I., Mejido, J., Puri, R.K., et al. (2004). Properties of pluripotent human embryonic stem cells BG01 and BG02. Stem Cells *22*, 292-312.