

Minireview

Gene Expression Profiling of Embryonic Stem Cells Leads to Greater Understanding of Pluripotency and Early Developmental Events¹

Raj R. Rao and Steven L. Stice²

Rhodes Animal Science Center, University of Georgia, Athens, Georgia 30602-2771

ABSTRACT

Embryonic stem cells are characterized by their ability to propagate indefinitely in culture, maintaining a normal karyotype and their undifferentiated state. They have the potential of differentiating into any specialized cell type in the body. An understanding of the transcriptional profile related to pluripotency and early development is necessary to better tap their developmental potential and also maintain their undifferentiated phenotype. Currently, several techniques are in use to ascertain the gene expression profile of embryonic stem cells. This review summarizes the information generated using microarray and other approaches on the gene expression analyses of stem cells in both mouse and human cell lines. We also discuss specific approaches useful in future studies aimed at further deciphering the pluripotent nature of human embryonic stem cells.

developmental biology, early development, embryo, gene expression, gene regulation, pluripotency, stem cells

INTRODUCTION

Embryonic stem (ES) cells are undifferentiated cells that have piqued scientific curiosity primarily due to their inherent pluripotent nature. The isolation of human ES cells [1, 2] has generated enormous interest due to their ability to differentiate into derivatives of all three embryonic germ layers and form virtually any cell type in the body. Two major areas of study in current ES cell research include analyses and maintenance of pluripotency involving the continued culture of the cells in an undifferentiated state, and development of uniform and directed differentiation strategies for the production of different cell types of specific interest. Although proof of principle of human ES cells developing into many differentiated phenotypes has been demonstrated [3], successful in vitro differentiation will require a well-characterized starting pluripotent ES population.

¹Funding has been provided by the Georgia Tech/Emory Center (GTEC) for the Engineering of Living Tissues, an ERC Program of the National Science Foundation under award EEC-9731643.

²Correspondence: Steven L. Stice, Rhodes Animal Science Center, University of Georgia, 425 River Road, Athens, GA 30602-2771. FAX: 706 542 7925; e-mail: sstice@uga.edu

Received: 5 April 2004.

First decision: 20 April 2004.

Accepted: 6 May 2004.

© 2004 by the Society for the Study of Reproduction, Inc.

ISSN: 0006-3363. <http://www.biolreprod.org>

Transcriptional analyses of the pluripotent state of human ES cells will help uncover or further define signaling pathways and molecular mechanisms involved in the maintenance of the undifferentiated state and initial loss of pluripotency. A detailed understanding of these molecular mechanisms will thus be essential for developing human ES cells as in vitro model systems for studying embryonic development and for harnessing the differentiation potential that makes them highly attractive for cell-based therapies. The following review summarizes the methodologies used and the information gathered on the transcriptional analyses of mouse and human ES cells and discusses possible mechanisms and expression profiles that characterize the pluripotency state in these cell types.

TECHNIQUES UTILIZED TO CHARACTERIZE GENE EXPRESSION PROFILE OF STEM CELLS

Numerous techniques have been used for characterizing the gene expression profile of stem cells and this has been achieved typically in comparison with differentiated counterparts or with somatic or nonpluripotent cell types. These techniques include serial analysis of gene expression [4, 5], subtractive hybridization [6, 7], representational difference analysis [8, 9], cDNA microarray [10], and oligonucleotide microarray [11] technologies. These methodologies have been used to decipher genes that are specifically expressed in either ES cells, neural stem cells (NSCs), and/or hematopoietic stem cells (HSC). Although each technique has its own unique advantage, cDNA and oligonucleotide microarray technologies have gained precedence as they provide opportunities for comparison of global gene expression patterns between different populations. Specifically, they are proving to be a vital tool in the elucidation of the molecular mechanisms involved in the maintenance of pluripotency and initial differentiation events of human ES cells. Examples of commonly used cDNA arrays are GEArrays (SuperArray Bioscience Corporation, Frederick, MD) and the Atlas Array (BD Biosciences Clontech, Palo Alto, CA). These cDNA arrays are manufactured using cutting-edge, noncontact printing technology, with cDNAs deposited onto nylon membranes. An alternative approach involves custom-made microarray analysis using robotic systems to spot cDNAs or PCR-amplified fragments onto glass slides [12, 13] and offer the major advantage of using small reaction volumes in the hybridization procedures. A robust comparison of differential expression profiles is made possible through the use of dual-fluorescent dyes and differ-

| TECHNIQUE | CELL POPULATION | REFERENCE |
|---|------------------------|-----------|
| Serial Analysis of Gene Expression (SAGE) | mESC | [15] |
| | hHSC | [16] |
| | hESC | [17] |
| | hMSC | [18] |
| Representational Difference Analysis(RDA) | mNSC | [19] |
| Subtractive Hybridization | hHSC | [20] |
| Oligonucleotide Microarray | mESC | [21, 22] |
| | mESC, mHSC, mNSC | [23] |
| | mESC, mHSC, mNSC, hHSC | [24] |
| | mESC, mNSC, mRPC | [25] |
| | mESC, mNSC | [26] |
| | hESC | [27-30] |
| cDNA Microarray | mESC | [31, 32] |
| | hESC | [33, 34] |
| | mESC, mTSC | [35] |
| | rNSC | [36] |
| | mNSC, mESC, mHSC | [37] |
| | mNSC, mESC | [38] |
| | mHSC, mNSC | [39] |
| | mHSC | [40, 41] |

FIG. 1. Summary of techniques used to characterize differential gene expression in specific stem cell populations. m, Mouse; h, human; r, rat; ESC, embryonic stem cell; NSC, neural stem cell; HSC, hematopoietic stem cell; MSC, mesenchymal stem cell; TSC, trophoblast stem cell; RPC, retinal progenitor cell.

entially labeled probes from separate cell populations. The oligonucleotide GeneChip array system (Affymetrix, Santa Clara, CA) has gained popularity and is now widely used in multiple fields of biomedical research for a stringent quantitative analysis of gene expression. This system is a variation of custom-made microarrays on glass slides and utilizes photolithography and solid-phase chemistry to produce arrays with densely packed oligonucleotide probes [11, 14]. Figure 1 summarizes the different techniques and approaches for characterizing specific populations of stem cells.

TRANSCRIPTIONAL PROFILING OF MOUSE EMBRYONIC STEM (mES) CELLS USING MICROARRAYS

Elucidation of mES Cell-Specific Genes Using cDNA Microarrays

Subsequent to the isolation of mES cells [42], a handful of genes expressed in mES cells, including *Pou5f1* (*Oct-4*) [43, 44], *Nanog* [45, 46], *Sox2* [47], and leukemia inhibitory factor (*LIF*) [48], have been identified as critical to the maintenance of pluripotency. An initial study by Kelly and Rizzino [31] was very specific in its microarray analysis of only mES cells with the monitoring of the expression of 588 known regulatory genes using Clontech Atlas Mouse

cDNA expression arrays. Of the 588 genes, 292 were expressed in D3 mES and/or D3-differentiated cells with a comparative analysis indicating a downregulation of 18 genes (ES cell-enriched) by a factor of 2.5-fold or greater. Of specific interest among the mES cell-enriched genes were a number of transcription factors (e.g., *Oct-3*, *PEA-3*), growth factors (*FGF-4*), and cell-cycle regulators (*CCND1* and *E*) [31]. In a subsequent study, Tanaka and coworkers [49] used the NIA mouse 15K cDNA microarray to profile mES cells, extraembryonic-restricted trophoblast stem (TS) cells, and terminally differentiated mouse embryonic fibroblast (MEF) cells to identify expressed genes specific to each population [35]. Pairwise comparisons and further clustering revealed 124 ES-specific genes, 94 TS-specific genes, and 51 genes specific to both ES and TS cells (<http://www.genome.org/cgi/content/full/12/12/1921/DC1>). Key ES-cell-specific genes included transcription factors (*Oct3/4*, *Rex1*, *Zfp57*, *Dnmt3A*), cell cycle regulators (*Cdc10*, *cyclin B1*), matrix proteins (*Scamp1*), DNA replication (*Top2A*), signaling (*Fgf13*) and several uncharacterized genes that included embryonic stem cell-specific gene 1 (*Esg-1*). Specific evidence suggests that *Esg-1* exhibits gene expression patterns similar to that of *Oct3/4* with its expression restricted to pluripotent cell types [50]. Taken together, these two studies demonstrate the utility of cDNA

microarrays in deciphering differential expression of specific regulatory genes in mES cells.

Elucidation of mES Cell and Stemness Genes Using Oligonucleotide Arrays

The unique ability of stem cells to self-renew along with their pluripotent nature implies that all stem cells share a common genetic signature [23, 24]. Thus, the identification of common expressed genes should provide key insights into deciphering the mechanisms involved in the maintenance of the pluripotent state. Several studies have used Affymetrix oligonucleotide array technology applied to NSC, hematopoietic stem cells (HSC) in addition to mES cells to elucidate the list of genes that represented a stem cell molecular signature. Ivanova and coworkers [24] specifically focused on the gene expression profiles of fetal and adult HSCs and found that key HSC markers, such as *c-Kit*, *Tie1*, *Gata2* and *Tek* were expressed in many different populations of the hematopoietic hierarchy. It is important to note that 54% of the HSC-enriched genes observed were also previously identified by Phillips and coworkers [40], through a subtractive hybridization approach for HSC-specific gene products. To further characterize the common gene expression profile for diverse types of stem cells, comparison of the HSC-enriched genes were conducted with mES cells and NSCs; 283 genes were commonly expressed among all three populations with the key functional categories of transcriptional regulation, signal transduction, cell cycle regulation, RNA binding, and chromatin regulation. Examples of key genes identified as mES cell-specific include *Pou5f1*, *Gbx2*, *Fgf4*, *cyclin D1*, and *cyclin E1* (<http://www.sciencemag.org/cgi/content/full/1073823/DC1>). In an independent but similar study, Ramalho-Santos and coworkers [23], identified 216 genes found to be common among all three stem cell populations with the key functional categories of signaling, transcriptional regulation, cell-cycle regulation, DNA repair, and translational regulation. Examples of key genes that were identified as ES cell-specific include *Pou5f1*, *Rex1*, *Tdgf1/Cripto*, *Lefty1*, and *Fgf4* (<http://www.sciencemag.org/cgi/content/full/1072530/DC1>). Specific observations from these studies included an enrichment of genes involved in the JAK/STAT and TGF β pathways with both pathways shown to be important for self-renewal and early development in both mES cells [51], and HSCs [52]. These pathways often interact, indicating that a complex integration of genetic signals involved in different pathways could contribute to the nature of stemness. A brief review comparing these two studies provided further insight into the molecular mechanisms involved in stem cells and produced lists of genes relevant to stem cell biology [53]. Briefly, a comparison of the list of stemness genes identified by both the research groups revealed only 15 common genes. This lack of similarity has been attributed to significant differences in methodology, stem cell populations, and stringency in data analyses.

Another study compared three types of stem cells (ES cells, NSCs, and retinal progenitor cells) [25] and obtained a list of 385 genes common among all three cell types (http://giscompute.gis.a-star.edu.sg/suppdata_stemness). Additionally, a comparison of enriched genes obtained in this study with the two earlier studies [23, 24] revealed 332 ES cell-enriched genes and 236 NSC-enriched genes, while only 10 genes were commonly expressed in the two stem cell populations. Subsequently, a comparative study of the two earlier HSC datasets [23, 24] using identical compu-

tational techniques revealed a far greater overlap of 605 genes [54]. The similarities and differences among different stem cell populations can be interpreted in two ways: 1) there is still a significant random variation in stemness gene expression studies and/or 2) there is a core set of stem cell genes along with functionally important genes specific for each stem cell population.

Rather than seeking a common transcriptional profile between different stem cell populations, D'Amour and Gage [26] focused on elucidating differences between ES cells and multipotent NSCs. Using a transgenic approach, they purified multipotent NSCs and pluripotent ES cell populations based on Sox2-EGFP expression and generated direct genetic comparisons between the two stem cell populations. *Sox2* was used, as it is strongly associated with the pluripotent phenotype and is known to act cooperatively at several promoters with *Oct4* [55, 56]. Additionally, *Sox2* expression is characteristic of a neural tissue [57] and is a regulatory element expressed in both ES cells and neural progenitor populations [58]. A total of 112 genes was determined to be present uniquely in ES cells and differentially expressed at the 1.4-fold change level in a comparison between ES cells and NSC populations. Key genes were classified under the categories of transcriptional regulation (*Pou5f1*, *Rex1*, *Gbx2*, *Stat6*, *Nanog*), RNA binding (*Esg1*), and growth factors (*Fgf4*, *Tdgf1*, *LeftyA*). This study used a homogenous and uniform starting Sox2-defined NSC and ES cell population, revealing striking dissimilarities in the gene expression profiles of the two populations and also helped define key genes solely expressed in ES cells.

MICROARRAY ANALYSES OF HUMAN EMBRYONIC STEM CELLS

Ever since the isolation of human ES (hES) cells [1, 2] and a subsequent federal restriction on isolation of new cell lines (<http://stemcells.nih.gov/stemcell>), numerous research groups have adopted microarray approaches to characterize some of the NIH-approved cell lines. In this review, the nomenclature for the hES cells mentioned are those defined by the provider included in the NIH human embryonic stem cell registry (<http://stemcells.nih.gov/registry/index.asp>). Most of these studies have focused on elucidating the transcriptional profile of hES cells and for understanding initial differentiation events. Further transcriptional analyses and comparative profiling are essential to identify the key molecular components and mechanisms that are critical to the maintenance of the undifferentiated state and pluripotency.

Elucidation of hES Cell-Specific Genes

Utilizing the Affymetrix oligonucleotide array technology, Sato and coworkers [28] identified potential genes that could define molecular mechanisms related to pluripotency. In these studies, based on comparison of undifferentiated H1-ES cells with their differentiated counterparts, they obtained a set of 918 genes enriched in H1-ES cells. Key observations from these studies highlighted a role for components of the FGF, TGF β /BMP, and Wnt pathways in the maintenance of pluripotency. The most highly enriched genes included those known to be expressed in early embryogenesis, such as *Oct-3/4*, *LeftyA*, *LeftyB*, and *TDGF1*. Approximately 28% of the H1-ES-enriched genes corresponded to expressed sequence tags, indicating that a detailed analysis of these genes could reveal new factors that could contribute to the pluripotent state. A comparison of the 918 H1-ES-enriched genes with published mES datasets

[23] indicated an overlap of 227 genes with the key genes, including *Pou5f1*, *LeftyB*, and *TDGF1*. Detailed statistical analyses revealed that the probability of the observed overlap between the human and mouse datasets were far below the estimated probability for coincidental overlapping. This indicated that the observed overlapping genes between the human and mouse datasets were significant to the pluripotent state in both species. Several genes that are members of the TGF β /BMP signaling pathway were among the overlapping genes. These studies were conducted using only the H1 line, and the authors note the possibility that different human ES lines could have different transcriptional profiles.

To elucidate similarities and differences between transcriptional profiles of different cell lines, we conducted a direct stringent comparison between datasets obtained from two hES (H1 and BG01 [59]) cell lines (unpublished data). Toward further characterizing the pluripotent state, we used a HepG2 cell conditioned medium (MEDII) to generate another pluripotent population using BG01 cells [29]. MEDII [60, 61], has been known to induce early differentiation in mES cells, while retaining key pluripotent marker expression. The datasets from the three populations (H1-ES, BG01-ES, and BG01-MEDII) have been generated using the Affymetrix microarray technology. As expected, there was a large list of genes (8400) common among the three pluripotent populations (Fig. 2). To identify genes enriched in the pluripotent populations, we compared the three datasets against a dataset obtained from a H1-differentiated population (unpublished data). Our findings indicated that, based on stringent criteria for selection of enriched genes, a total of 133 genes overlapped between the three pluripotent populations with the key genes, including *Pou5f1*, *LeftyA*, *Sox2*, and *Fgf2*. Although many genes were unique to each pluripotent population, they all shared similarities based on key functional ontologies that can define pluripotency including signaling pathways, cell cycle regulation, and transcriptional regulation. A recent study by Abeyta and coworkers [27] comparing gene expression profiles of HSF-1, HSF-6, and H9 lines also found that the three populations overlapped by 7385 Affymetrix gene probes. Key genes expressed at significant levels in all three cell lines included *Pou5f1*, *Sox2*, *Rex1*, *TDGF1*, and *LeftyA*. Their studies also showed that, in spite of each stem cell line possessing a unique stem cell signature, they were all capable of contributing to multiple cell fates. Shades of pluripotency could thus exist between different stem cell lines, which could lead to either preferential or enhanced differentiation toward particular cell fates.

Using high-quality oligonucleotide glass arrays, Bhattacharya and coworkers [30] have examined gene expression in six (GE01, GE07, GE09, BG01, BG02, TE06) hES cell lines. Results revealed that all six cell lines expressed multiple markers of the undifferentiated state and shared significant homology in gene expression (<http://www.grc.nia.nih.gov/branches/lms/scbudata.htm>). Key among the 92 genes expressed in all six lines were *Nanog*, *GTCM-1*, *connexin 43*, *Pou5f1*, and *TDGF1*. However, when these results were compared with earlier published mES results [23, 24], only between 12 and 33 of the 92 hES genes were overexpressed in mES cells. The authors have attributed the limited overlaps to species differences and to potential variability in culture conditions between the different stem cell populations.

An alternate approach toward deciphering molecular mechanisms underlying pluripotency has used a compara-

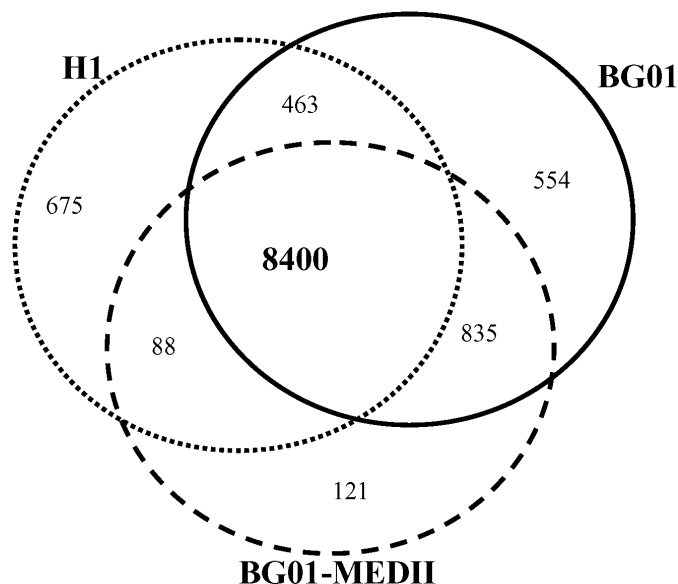


FIG. 2. Venn diagram illustrating intersection of genes from three pluripotent populations [H1(·); BG01(—); BG01-MEDII(-)] showing shared and unique gene expression. Only genes that received a present call as deciphered by the MAS 5.0 software detection algorithm were used for this comparison. Numbers are based on Affymetrix probe IDs. The region of overlap indicates the number of genes (8400) in all three populations.

tive analysis of the expression profiles of hES cell lines and human germ cell tumor (embryonal carcinoma-EC) lines. Using cDNA microarrays and hierarchical cluster analysis, Sperger and coworkers [33] have shown that five independent hES cell lines (H1.1, H7, H9, H13, and H14) clustered tightly together, thereby indicating major similarities. A total of 895 genes were expressed at significant levels in both hES and EC lines. Among the 25 most significant genes observed in hES cells were *Pou5f1*, *DNMT3B*, *CRABP1*, *SALL2*, and *GABRB3*. These studies also indicated that genes differentially expressed between hES and EC cells could constitute a list of genes that reflect an adaptation to tumor growth by suppression of differentiation. Through this adaptation, long-term survival and self-renewal mechanisms could be activated.

Elucidation of Genes Related to Initial Differentiation Events

Microarray analyses of hES cells have primarily focused on addressing issues related to the transcriptional profiles of different stem cell lines. Most of the experimental strategies adopted have involved a comparison of undifferentiated stem cells against their differentiated counterparts. A comparison of two closely related pluripotent populations could yield specific information on key genes involved in the maintenance or exit from the pluripotent state and also aid in the development of directed differentiation strategies. We have used HepG2 conditioned medium (MEDII) [60, 61], known to induce early differentiation in mES cells, while retaining pluripotent markers, to understand early differentiation events in hES cells [29]. Microarray data using the Affymetrix system showed that MEDII treatment of BG01 hES cells retained key pluripotent marker expression (*Pou5f1*, *Nanog*, *SSEA4*), while causing regulation of genes in the TGF β /nodal pathway. Results indicated that *TDGF1* (upregulated with MEDII treatment) and *LeftyA*, *Follistatin* (downregulated with MEDII treatment) were

FIG. 3. Representative genes highly expressed in embryonic stem cells. *, Italicized genes have also been observed in unpublished data involving a comparison of H1 and BG01 hES cell lines.

Transcription Factors/ Nucleic acid binding

POU5F1[23, 24, 26-31, 33, 34], *GBX2*[23, 24, 31], *TERF1*[28, 33], *NANOG* [28-30], *HOXA11*[23, 29], *PITX2*[23, 30], *UNG*[24, 31], *PEA3*[24, 31], *KLF2*[24, 31], *KLF3/5/9*[23], *TTF1*[24, 31], *JMJ*[28], *TOP2A*[28, 35], *AND1*[28], *SALL1*[28], *SALL2*[28, 33], *ZIC3*[28], *FOXD3*[27, 33], *SOX2*[26, 27, 30], *ZNF43*[30], *PSIP1*[28, 30], *RUVBL1*[23, 28], *MSH2*[23, 24, 27, 28], *ESG1*[26, 35], *REX1*[23, 26, 35]

Growth Factors/Receptors

FGF2[23, 24, 28], *FGF4*[24, 26, 31] *FGF13*[28, 35], *FGFR1/2/4*[28], *BMPR1A* [23, 28, 33], *FZD5*[28], *GABABR1*[23], *TGFB1*[23, 29], *GABABR3*[28, 33], *FZD7*[33], *GDF3*[26, 28, 30], *EPHA1*[33], *EPHA2/B4*[23, 26] , *EPHA4*[29]

Cell Cycle Regulators

CCNE1[23, 24, 31], *CCND1*[23, 24, 27, 28, 31], *CCNB1*[28, 30, 35] *CDK4*[23], *CDC2*[23, 30], *CHK2*[23, 28], *MCM3*[28],

Signaling Molecules/ Secreted Factors

TDGF1[26-30], *LEFTYA*[23, 26-30], *LEFTYB*[28, 30], *NODAL*[23, 27], *PATCHED2*[23, 26, 29], *WNT1*[23], *WNT5A*[29], *FST*[29, 30], *CRABPI*[30, 33]

Cell Adhesion/Membrane Proteins

PECAM[24, 31], *LAMR1*[24, 30, 31], *ITGA6*[23, 24, 27], *BYSL*[23, 24, 27], *GPC4*[28, 29, 33],

Miscellaneous

THY1[28], *FLJ10713*[23, 28, 33], *FLJ20105*[28], *FLJ10156*[28], *SILV*[28, 33], *DNMT3A*[35], *DNMT3B*[30, 33], *USP9X* [23, 24, 27, 28], *CGI30*[23, 24, 27, 28], *CGI32*[28], *PPIC*[23, 24, 27], *LAPTM4B*[23, 24, 27], *KIAA1018*[23, 24, 27], *KIAA0523*[28], *KIAA0922*[28], *ELOVL6*[23, 24, 27], *ARCN1*[23, 24, 27], *C20ORF1*[28, 30]

among the genes that were significantly differentially expressed between the two pluripotent populations. MEDII treatment of hES cells appeared to capture an event that has a gene-expression profile similar to primitive-streak stage of a developing mouse embryo.

Another early developmental event that was characterized by microarray analysis involved hES differentiation to trophoblast using BMP4 as an inducing agent [34]. Using cDNA microarrays, Xu and coworkers [34] analyzed genes differentially expressed between BMP4-treated and undifferentiated H1 hES cells. BMP4 influenced significant upregulation of key genes related to development of trophoblast and included *AP-2*, *MSX2*, *GATA2*, *GATA3*, and *CG β* and a significant downregulation of genes highly expressed in pluripotent cells such as *Oct-4* and *TERT*. Of particular importance is that germ-layer (ectoderm, endoderm, and mesoderm) -specific genes were not significantly elevated in the BMP4-treated cells. These initial studies have provided opportunities for identification of potential signals that could influence initial differentiation events and mechanisms involved in the exit from the pluripotent state.

CONCLUSIONS AND FUTURE DIRECTIONS

Gene expression profiling using microarray technologies provides an important basis for revealing the molecular mechanisms involved in pluripotency and initial differentiation events that involve embryonic stem cell populations. We have summarized the research efforts that have used microarray technologies to query gene expression in stem cells, with a focus on ES cells. These studies have helped identify similarities and differences between cell lines and aided in deciphering genes that could contribute to the maintenance of the pluripotent state. Figure 3 provides a list of representative genes that contribute to the pluripotent state, compiled from the many studies summarized. The major functional categories that define pluripotency relate to transcriptional regulation, signaling, and cell cycle regulation. Specifically, genes involved in FGF, TGF β , and Wnt signaling pathways have been implicated in early development and cell-fate decisions [62]. Recently, it has been shown that TGF β along with LIF and FGF2 support the propagation of undifferentiated hES cells in a feeder-free culture system [63]. Further, involvement of specific

components of the Wnt signaling pathway has also been shown to maintain the undifferentiated state in both mES and hES cells [64]. These studies have provided an example for the contribution of key signaling molecules for the maintenance of the undifferentiated state of hES cells. Continued characterization and functional testing of the genes obtained from microarray studies will further our understanding of the pluripotent state.

It is important to note that utilization of microarray technologies allows potential opportunities for comparison of datasets from different groups and this facilitates comparison of different stem cell lines. However, this can be achieved only through a direct comparison of stem cell populations that share biological similarities, using uniform and stringent statistical approaches. Current bioinformatics approaches involve data integration, using ontology, to query multiple data sources, to generate relevant biological information. From a scientist's perspective, challenges remain in formally defining and representing meaningful relationships (called semantic associations) and assessing the quality of results obtained. With the potential for experimental variations between different groups and generation of disparate data from different experiments, there is usefulness for semantic-based approaches [65] to analyzing gene expression data. Briefly, semantics involves a formal description of resources so that data obtained from different experimental methodologies can be easily understood and contribute toward a unique solution. These approaches will involve a combination of data generated from different experiments and knowledge stored in ontologies, to generate specific information. There is thus a potential for explaining important biological phenomena only if they are explained in the context of the vast amounts of data from different methodologies that is being made available. Specifically, from a stem cell standpoint, there is a major scope for semantic data analysis in establishing correlations between patterns of gene expression within different pluripotent populations from one cell line or between different cell lines. In conjunction with data generated from microarray and other experimental methodologies, semantic approaches can thus aid in unraveling the mystique of the undifferentiated pluripotent state and for the development of differentiation strategies to produce cell types of interest.

ACKNOWLEDGMENTS

We are grateful to Dr. Brivanlou for providing data (H1-ES) for alternate statistical reanalysis. We apologize for manuscripts that have not been cited here due to limitations in space. We appreciate proofreading of this document by Dr. Mary Ann Della-Fera.

REFERENCES

- Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science* 1998; 282:1145–1147.
- 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.
- Carpenter MK, Rosler E, Rao MS. Characterization and differentiation of human embryonic stem cells. *Cloning Stem Cells* 2003; 5:79–88.
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995; 270:484–487.
- Velculescu VE, Vogelstein B, Kinzler KW. Analysing uncharted transcriptomes with SAGE. *Trends Genet* 2000; 16:423–425.
- Duguid JR, Rohwer RG, Seed B. Isolation of cDNAs of scrapie-modulated RNAs by subtractive hybridization of a cDNA library. *Proc Natl Acad Sci U S A* 1988; 85:5738–5742.
- Sive HL, St John T. A simple subtractive hybridization technique employing photoactivatable biotin and phenol extraction. *Nucleic Acids Res* 1988; 16:10937.
- Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science* 1993; 259:946–951.
- Lisitsyn NA, Leach FS, Vogelstein B, Wigler MH. Detection of genetic loss in tumors by representational difference analysis. *Cold Spring Harb Symp Quant Biol* 1994; 59:585–587.
- Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Sampson R, Houlgatte R, Soularue P, Auffray C. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res* 1996; 6:492–503.
- Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996; 14:1675–1680.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270:467–470.
- Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* 1996; 93:10614–10619.
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999; 21:20–24.
- Anisimov SV, Tarasov KV, Tweedie D, Stern MD, Wobus AM, Boheler KR. SAGE identification of gene transcripts with profiles unique to pluripotent mouse R1 embryonic stem cells. *Genomics* 2002; 79:169–176.
- Zhou G, Chen J, Lee S, Clark T, Rowley JD, Wang SM. The pattern of gene expression in human CD34(+) stem/progenitor cells. *Proc Natl Acad Sci U S A* 2001; 98:13966–13971.
- 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.
- Tremain N, Korkko J, Ibberson D, Kopen GC, DiGirolamo C, Phinney DG. MicroSAGE analysis of 2,353 expressed genes in a single cell-derived colony of undifferentiated human mesenchymal stem cells reveals mRNAs of multiple cell lineages. *Stem Cells* 2001; 19:408–418.
- Geschwind DH, Ou J, Easterday MC, Dougherty JD, Jackson RL, Chen Z, Antoine H, Tersikh A, Weissman IL, Nelson SF, Kornblum HI. A genetic analysis of neural progenitor differentiation. *Neuron* 2001; 29:325–339.
- Graf L, Torok-Storb B. Identification of a novel DNA sequence differentially expressed between normal human CD34+CD38hi and CD34+CD38lo marrow cells. *Blood* 1995; 86:548–556.
- Aprelikova O, Pace AJ, Fang B, Koller BH, Liu ET. BRCA1 is a selective co-activator of 14-3-3 sigma gene transcription in mouse embryonic stem cells. *J Biol Chem* 2001; 276:25647–25650.
- Kriz V, Anneren C, Lai C, Karlsson J, Mares J, Welsh M. The SHB adapter protein is required for efficient multilineage differentiation of mouse embryonic stem cells. *Exp Cell Res* 2003; 286:40–56.
- 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.
- Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. A stem cell molecular signature. *Science* 2002; 298:601–604.
- Fortunel NO, Otu HH, Ng HH, Chen J, Mu X, Chevassut T, Li X, Joseph M, Bailey C, Hatzfeld JA, Hatzfeld A, Usta F, Vega VB, Long PM, Libermann TA, Lim B. Comment on "Stemness": transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature." *Science* 2003; 302(5644):393; author reply 393.
- D'Amour KA, Gage FH. Genetic and functional differences between multipotent neural and pluripotent embryonic stem cells. *Proc Natl Acad Sci U S A* 2003; 100(suppl 1):11866–11872.
- Abeyta MJ, Clark AT, Rodriguez RT, Bodnar MS, Pera RA, Firpo MT. Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum Mol Genet* 2004; 13:601–608.
- Sato N, Sanjuan IM, Heke M, Uchida M, Naef F, Brivanlou AH. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol* 2003; 260:404–413.
- Calhoun J, Rao RR, Warrenfeltz S, Rekaya R, Dalton S, McDonald J, Stice SL. Microarray analysis of early differentiation events in human embryonic stem cells. In: Program of the Annual Meeting of the International Society for Stem Cell Research; Washington, DC 2003; Abstract 108.
- Bhattacharya B, Miura T, Brandenburg R, Mejido J, Luo Y, Yang AX,

- Joshi BH, Irene G, Thies RS, Amit M, Lyons I, Condie BG, Iskovitz-Eldor J, Rao MS, Puri RK. Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* 2004 Apr 15; 103(8):2956–2964. Epub 2003 Dec 30.
31. Kelly DL, Rizzino A. DNA microarray analyses of genes regulated during the differentiation of embryonic stem cells. *Mol Reprod Dev* 2000; 56:113–123.
 32. Loring JF, Porter JG, Seilhammer J, Kaser MR, Wesselschmidt R. A gene expression profile of embryonic stem cells and embryonic stem cell-derived neurons. *Restor Neurol Neurosci* 2001; 18:81–88.
 33. Sperger JM, Chen X, Draper JS, Antosiewicz JE, Chon CH, Jones SB, Brooks JD, Andrews PW, Brown PO, Thomson JA. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A* 2003; 100:13350–13355.
 34. Xu RH, Chen X, Li DS, Li R, Addicks GC, Glennon C, Zwaka TP, Thomson JA. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol* 2002; 20:1261–1264.
 35. Tanaka TS, Kunath T, Kimber WL, Jaradat SA, Stagg CA, Usuda M, Yokota T, Niwa H, Rossant J, Ko MS. Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. *Genome Res* 2002; 12:1921–1928.
 36. Cai J, Cheng A, Luo Y, Lu C, Mattson MP, Rao MS, Furukawa K. Membrane properties of rat embryonic multipotent neural stem cells. *J Neurochem* 2004; 88:212–226.
 37. Easterday MC, Dougherty JD, Jackson RL, Ou J, Nakano I, Paucar AA, Roobini B, Dianati M, Irvin DK, Weissman IL, Terskikh AV, Geschwind DH, Kornblum HI. Neural progenitor genes. Germinal zone expression and analysis of genetic overlap in stem cell populations. *Dev Biol* 2003; 264:309–322.
 38. Luo Y, Cai J, Ginis I, Sun Y, Lee S, Yu SX, Hoke A, Rao M. Designing, testing, and validating a focused stem cell microarray for characterization of neural stem cells and progenitor cells. *Stem Cells* 2003; 21:575–587.
 39. Terskikh AV, Easterday MC, Li L, Hood L, Kornblum HI, Geschwind DH, Weissman IL. From hematopoiesis to neurogenesis: evidence of overlapping genetic programs. *Proc Natl Acad Sci U S A* 2001; 98:7934–7939.
 40. Phillips RL, Ernst RE, Brunk B, Ivanova N, Mahan MA, Deanehan JK, Moore KA, Overton GC, Lemischka IR. The genetic program of hematopoietic stem cells. *Science* 2000; 288:1635–1640.
 41. Park IK, He Y, Lin F, Laerum OD, Tian Q, Bumgarner R, Klug CA, Li K, Kuhr C, Doyle MJ, Xie T, Schummer M, Sun Y, Goldsmith A, Clarke MF, Weissman IL, Hood L, Li L. Differential gene expression profiling of adult murine hematopoietic stem cells. *Blood* 2002; 99:488–498.
 42. 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; 78:7634–7638.
 43. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998; 95:379–391.
 44. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 2000; 24:372–376.
 45. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003; 113:643–655.
 46. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003; 113:631–642.
 47. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003; 17:126–140.
 48. Smith AG, Nichols J, Robertson M, Rathjen PD. Differentiation inhibiting activity (DIA/LIF) and mouse development. *Dev Biol* 1992; 151:339–351.
 49. Tanaka TS, Jaradat SA, Lim MK, Kargul GJ, Wang X, Grahovac MJ, Pantano S, Sano Y, Piao Y, Nagaraja R, Doi H, Wood WH 3rd, Becker KG, Ko MS. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci U S A* 2000; 97:9127–9132.
 50. Pesce M, Wang X, Wolgemuth DJ, Scholer H. Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. *Mech Dev* 1998; 71:89–98.
 51. Smith AG. Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* 2001; 17:435–462.
 52. Fortunel NO, Hatzfeld A, Hatzfeld JA. Transforming growth factor-beta: pleiotropic role in the regulation of hematopoiesis. *Blood* 2000; 96:2022–2036.
 53. Burns CE, Zon LI. Portrait of a stem cell. *Dev Cell* 2002; 3:612–613.
 54. Ivanova N, Dimos JT, Schaniel C, Hackney JA, Moore KA, Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA, Lemischka . Response to comments on ‘Stemness’: transcriptional profiling of embryonic and adult stem cells” and “A stem cell molecular signature.” *Science* 2003; 302:393.
 55. Ambrosetti DC, Basilico C, Dailey L. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol* 1997; 17:6321–6329.
 56. Nishimoto M, Fukushima A, Okuda A, Muramatsu M. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol* 1999; 19:5453–5465.
 57. Wood HB, Episkopou V. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mech Dev* 1999; 86:197–201.
 58. Zappone MV, Galli R, Catena R, Meani N, De Biasi S, Mattei E, Tiveron C, Vescovi AL, Lovell-Badge R, Ottolenghi S, Nicolis SK. Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* 2000; 127:2367–2382.
 59. Mitalipova M, Calhoun J, Shin S, Winger D, Schulz T, Noggle S, Venable A, Lyons I, Robins A, Stice S. Human embryonic stem cell lines derived from discarded embryos. *Stem Cells* 2003; 21:521–526.
 60. Rathjen J, Lake JA, Bettess MD, Washington JM, Chapman G, Rathjen PD. Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J Cell Sci* 1999; 112(pt 5):601–612.
 61. Lake J, Rathjen J, Remiszewski J, Rathjen PD. Reversible programming of pluripotent cell differentiation. *J Cell Sci* 2000; 113(pt 3):555–566.
 62. Brivanlou AH, Darnell JE Jr. Signal transduction and the control of gene expression. *Science* 2002; 295:813–818.
 63. Amit M, Shariki C, Margulets V, Itskovitz-Eldor J. Feeder layer- and serum-free culture of human embryonic stem cells. *Biol Reprod* 2004; 70:837–845.
 64. Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* 2004; 10:55–63.
 65. Kochut K, Arnold J, Sheth A, Miller J, Kraemer E, Arpinar B, Cardoso J. Intelligen: a distributed workflow system for discovering protein-protein interactions. *Distributed and Parallel Databases* 2003; 13:43–72.