[Frontiers in Bioscience 13, 276-283, January 1, 2008]

# Genomic studies to explore self-renewal and differentiation properties of embryonic stem cells

**Ming Zhan**

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

## TABLE OF CONTENTS



## 1. ABSTRACT

Embryonic stem cells (ESCs) are pluripotent cells with an indefinite replication potential and an ability to differentiate into a variety of cell lineages, holding a great promise for regenerative medicine and biological research. The genome and transcriptome of ESCs have been extensively examined in order to decipher their self-renewal and differentiation mechanisms. Global transcriptional profiling allows the identification of genes expressed differentially or uniquely in ESCs and the elucidation of the molecular signatures. Comparative genomics and transcriptomics help to explore evolutionarily conserved and divergent transcriptional patterns and functional landscape of ESCs, and identify fundamental and species-specific mechanisms controlling the pluripotency. Chromosomal mapping of the transcriptome demonstrates the coexpression of neighboring genes along the chromosome and highlights their dynamic changes in response to ESC differentiation. Epigenetic analysis reveals methylation patterns and microRNA profiles unique to ESCs. In this article, various aspects of genomic and transcriptomic studies on ESCs are reviewed, and important findings regarding ESC self-renewal and differentiation are highlighted and discussed.

## 2. INTRODUCTION

Embryonic stem cells (ESCs) are pluripotent cells with an indefinite replication potential and an ability to differentiate into a variety of cell lineages. Due to these unique properties, ESCs hold an important therapeutic potential in regenerative medicine, particularly for the treatment of many currently incurable injuries or diseases such as spinal cord injury, diabetes, hematological, neurodegenerative and cardiovascular diseases (1-7). Moreover, recent evidence supports a long-standing notion that cancers are initiated and maintained by cancer stem cells that might have arisen from deregulated normal stem or progenitor cells (8). ESC studies may thus help to identify the molecular triggers and potential cures for various cancers. ESC differentiation is regulated by factors similar to those controlling germ layer formation and cell type specification. ESCs can thus serve as a model system to understand mechanisms controlling cell growth and differentiation, as well as early-stage development of human embryos, tissues, and organs (9-12).

A thorough understanding of the molecular factors and mechanisms that control the self-renewal and differentiation of ESCs is essential for realizing their

potential in medicine and science. The genome and transcriptome of ESCs have been studied extensively in recent years for identifying molecular mechanisms regulating ESC pluripotency. Genome-wide expressional profiling provides a basis to examine changes of genes and pathways in response to ESC differentiation. Comparative transcriptomics and genomics help to explore evolutionarily conserved and divergent transcriptional patterns and functional landscape of ESCs. Epigenetic analyses, proteomic studies, and microRNA surveys, as well as the application of new technologies such as chromatin immunoprecipitation have facilitated the discovery of regulatory mechanisms and deepened our understanding of the core properties of ESCs. Here, we review various aspects of genomic and transcriptomic studies in exploring self-renewal and differentiation properties of ESCs, with particular emphasis on human ESCs.

## 3. GENOME-WIDE EXPRESSION PROFILE

Genome-wide expressional profiling has been conducted on a variety of cell lines of ESCs using microarray, EST, SAGE, MPSS, and RT-PCR methods (13-26). These studies have generated a large volume of data to characterize the ESC transcriptome and to examine the changes of genes and pathways in response to ESC differentiation.

### 3.1. Molecular Signature

Comparative analyses of ESC expression profiles have been important for defining the molecular signature (16-20, 27). Some studies compared gene expression profiles in undifferentiated ESCs with those in differentiated cells, particularly the embryoid body (EB), which is the earliest stage of ESC differentiation in culture (21-24). Others compared transcript expression data of human ESCs (hESCs) with those of mouse ESCs (mESCs) (13, 16, 17, 20). By studying ESTs, Brandenberger *et al.* obtained 148,453 ESTs from undifferentiated hESCs and differentiated derivatives (21). Over 32,000 unique transcripts were identified as being expressed in undifferentiated hESCs. Among them, 532 were significantly up-regulated and 140 were significantly down-regulated. In studies using SAGE and MPSS methods, Richards *et al.* (19) and Miura *et al.* (23) identified 192 or 50 genes that were up-regulated in hESCs as compared with differentiated cells. Sharov *et al.* (25), on the other hand, identified potential marker genes characterizing mESCs by EST frequency analysis. Wei and colleagues (13) compared gene expression between human and mouse ESCs, and showed that LIFR was highly expressed in mouse but not in human, CD9 was highly expressed in human but not in mouse, and FGF2 was highly expressed in human but FGF4 was abundant in mouse. Importantly, some genes were often found in different expression datasets, such as OCT4, SOX2, NANOG, REX-1, UTF1, TERT, ABCG2, NODAL, TDGF1, LEFTB, BEX1, GATA4, and members of signaling pathways like FGF, WNT and BMP. Those genes, expressed exclusively or predominantly in ESCs and important for maintaining the pluripotency, were considered to be molecular signatures of ESCs (16, 22, 28, 29).

Among different expressional profile studies, the overlap is however small in the genes identified to be over-expressed in ESCs. For example, Invanoa *et al.* (15) and Ramalho-Santos *et al.* (14) independently identified more than 200 genes over-expressed in ESCs, but there are only six genes in common between the two gene lists, despite the fact that the two studies used identical microarray chips and the same cell types (30-32). Among the studies by Sperger *et al.* (18), Sato *et al.* (16), and Bhattacharya *et al.* (22), only seven genes were found to be shared out of a total of 2,226 genes. Genetic diversity among different cell lines, differences in cell culture conditions, and sampling issues may account for the differences among different gene expression data. There are also possible flaws in the methods employed for such comparative analysis. Efforts have been made to examine comparability of different hESC microarray studies and identify shared genes (14, 15, 33). Suarez-Farinas *et al.* (33) employed a different statistical method for a re-analysis of hESC array data from Bhattacharya *et al.* (22), Sato *et al.* (16) and Sperger *et al.* (18). This study indicates that comparing the lists of up-regulated genes is methodologically flawed. In such a comparison, a gene at the intersection of three lists must pass a statistical test three independent times, and the subsequent intersection's p-value is the cube of the lists' p-values, resulting in fewer genes found to be significant. By an integrated correlation analysis (34), a large set of common genes (111 up-regulated and 95 down-regulated genes) was identified from the three data sets, despite of the different ways that the experiments were conducted (33). The results were confirmed by real time RT-PCR (33). The study demonstrates that combined analyses of multiple experiments with proper statistical methods are helpful in reaching coherent conclusions.

### 3.2. Regulatory Pathways

Differential expression of genes between undifferentiated and differentiated states of ESCs highlights the dynamic change of transcriptional programs or pathways in response to ESC differentiation. Among the differentially expressed genes are various components of signaling pathways and transcriptional regulators that play key roles in ESC self-renewal and differentiation. Signaling pathways that are implicated from the ESC expression data include LIF, FGF, WNT, and NODAL pathways. The LIF pathway, for example, is required for rodent ESC self-renewal but not essential for human (35-40). Consistent with this pattern, gene expressional profiling by microarray, SAGE, MPSS, and EST showed no expression or under-expression of LIFR, IL6ST, and other components of LIF signaling in undifferentiated hESCs (18, 19, 21). Microarray and EST data also show the expression of FGF2 and all four FGF receptor transcripts in hESCs, suggesting that hESCs are responsive to FGF signaling (18, 21). Most WNT pathway genes are expressed in hESCs and EBs, while agonists and antagonists of this pathway show differential expression levels (21). This expression pattern suggests that hESCs are competent to respond to WNT signaling and that shifts in the balance of agonists and

antagonists occur as hESCs initiate differentiation (21). Moreover, the coordinated expression of members of the NODAL pathway, including agonists and antagonists (*e.g.* TDFG1, CER-1), suggests that this signaling pathway is important and tightly controlled for self-renewal and differentiation of hESCs (21).

### 3.3. Novel Genes and Splicing Variants

Studies on the ESC transcriptome have identified a number of expressed sequences and genes with unknown functions, and a diversity of splicing variants. The lack of expressed sequence data from hESCs prompted Brandenberger *et al.* to adopt the EST approach for expression profiling (21). Of the 32,764 unique transcripts identified, 16,990 (52%) were found not to match any UniGene cluster and 5,390 (16%) did not match any previously generated EST. Likewise, in a gene expression analysis using the MPSS method, Miura *et al.* estimated that approximately 25% of all detectable transcripts in hESCs were associated with unknown functions (23). On the other hand, Sharov *et al.* (25) discovered approximately 1,000 putative genes previously not identified by clustering ESTs from mESCs and other tissues. Novel transcripts or genes identified by these studies represent potential novel markers of ESCs and possible new mechanisms controlling ESC self-renewal and differentiation.

Pritsker *et al.* conducted a genome-wide survey of alternative splicing in human and mouse ESCs, using combined computational and experimental methods (41). EST sequences available from hESCs and mESCs were examined by sequence alignments. The computational analysis identified splicing variation in 413 (3.2%) of the genes expressed in mESCs and 1,163 (9.1%) of the genes expressed in hESCs. The results were confirmed by RT-PCR and sequencing. The study demonstrates that the frequency of alternative splicing is especially high in tissue-specific genes as compared to ubiquitous genes. Alternative splicing preferentially affects component genes of MAPK, WNT, apoptosis, cell cycle and other signaling pathways important to stem cells. Alternative splicing significantly increases the complexity of the ESC transcriptome, thus constituting an important aspect to address for a comprehensive characterization of ESCs.

## 4. COMPARATIVE TRANSCRIPTOMICS

Significant similarities exist among ESCs harvested from different species, yet major differences are also observed (29, 42). Important sequence elements in the genome, as well as important biological processes or pathways, are often evolutionarily conserved (43-45). The observed conservation and variation among ESCs suggest that comparative genomics or transcriptomics analyses may help distinguish between universal and species-specific mechanisms controlling ESC pluripotency. The evolutionary conservation on the genomic structure of ESC-critical genes and on the transcriptional changes during ESC differentiation is emerging as a powerful predictor of molecular factors and pathways important for ESC self-renewal and differentiation (46, 47).

### 4.1. Conserved and Divergent Functional Landscape

Sun *et al.* conducted a comparative transcriptomics study to examine a large set of functional categories and all transcription factors and growth factors (46). Global transcriptional changes derived from all expressed genes in ESCs and EBs, instead of changes in differentially expressed individual genes, were examined for a high discriminating power in functional profiling. The study identified many conserved functional categories showing positive correlations in the transcriptional changes that occur upon ESC differentiation, which represent fundamental molecular mechanisms regulation ESC pluripotency. The conserved biological processes include those involved in development, such as cellular morphogenesis, embryonic development, and pattern specification, and those involved in cell cycle, signal transduction, apoptosis, and establishment or maintenance of chromatin architecture (46). Transcriptional factors such as MYB, MYCN, NFYB, POLR3K, POU2F1, and UTF1, and growth factors such as GDF3, LEFTB, and TDFG1 also show conserved transcriptional changes in ESC differentiation. Conserved biological processes and pathways such as TGF-beta, WNT, NODAL, and integrin-mediated signaling are known to be important for ESC pluripotency. Yet, other conserved biological processes, pathways, and transcription or growth factors bear no overt relationship to ESC self-renewal and differentiation, which may represent new mechanisms controlling the pluripotency (46). On the other hand, many functional categories, transcriptional factors and growth factors were divergent, showing negative or no correlation in the expression patterns across species (46). The divergent functional categories, such as the LIF pathway, are suggestive of species-specific mechanisms controlling ESC self-renewal and differentiation (46).

### 4.2. Conserved and Divergent Coexpression Patterns

Gene coexpression analysis provides a systems-level context to evaluate transcriptional changes and network dynamics. Zhan and colleagues (unpublished) conducted a comparative study on gene coexpression from global networks to specific pathways for human and mouse ESCs. The analysis showed that coexpression networks of ESCs were scale-free and modular, containing a few highly connected genes (*i.e.* hub genes) that linked the rest of less connected genes to the system. Between human and mouse, the coexpression networks were highly divergent; the hub genes were mostly species-specific, and the overlap on dense areas between the networks was low. Tightly coexpressed gene complexes were identified from cell cycle, DNA replication, the OCT4/SOX2/NANOG directed network, and in TGF-beta, WNT, JAK/STAT, and ATK/PTEN pathways. The gene complexes were either conserved or divergent across species, suggestive of fundamental or species-specific mechanisms of tight gene-gene interactions in regulating ESC pluripotency. The results of this study support much of current knowledge of ESC-related pathways and biological processes, and provide clues of new signature genes and novel mechanisms controlling ESC development.

## 5. COMPARATIVE GENOMICS

Zhan and colleagues performed a comparative genomics study in combination with gene expression analysis on pathways critical for ESC self-renewal and differentiation (47). Genes of these pathways were examined for their genomic conservation and variation in both coding and regulatory regions. On the regulatory regions including promoters and enhancers, transcriptional factor binding sites (TFBS) were further examined. The study indicates that only a small set of transcription factors are likely critical in maintaining ESC pluripotency. The LIF pathway is evolutionarily divergent from the genomic perspective as the pathway components such as LIFR, EHOX, ERAS and GP130 are not conserved on the gene and promoter structure (47). EHOX, for example, is present in the mouse genome, but no ortholog is present in the human genome. The closest human paralogs of EHOX are quite divergent and more importantly not expressed in hESCs. Likewise, ERAS, which plays a critical role in rodent ESC self-renewal, exists as a pseudogene (*i.e.* HRAS or RASP) in human. The pseudogene shows a high divergence on the non-coding regulatory region and nonsense deletion mutations in the coding region, and is not expressed in hESCs (47-49). These results underline the fact that LIF and LIF-mediated signaling are not crucial for maintaining pluripotency in hESCs, and human and mouse ESCs show differential requirement of LIF signaling (38-40). The pathways directed by OCT4, SOX2, NANOG, FGF, and NODAL, on the other hand, are conserved in both gene and promoter structure. This study suggests that the conserved OCT4/SOX2 synergistic action is an important activation mechanism in the NANOG, LIF, FGF, and OCT4 directed pathways (47). SOX2, FOXD3, and ESG1 may serve in a feedback loop in these pathways; these pathways regulate the expression of the feedback loop, which in turn regulates the OCT4/SOX2 synergistic activation of these pathways. The NODAL/TDFG1/CER-1 pathway is independent of the OCT4/SOX2 activation and not involved in the feedback regulation. No FGF response element (FRE) (50) is present in STAT3, NANOG, or NODAL pathway genes, and no STAT-like binding sites are present in FGF4. However, FRE is present on regulatory regions of FGF4, SOX2, FOXD3, and ESG1, suggesting a basis for FGF regulating these genes in a parallel pathway to maintain ESC self-renewal.

A list of OCT4, SOX2, and NANOG target genes has been determined by cross-species genomic analysis followed by chromatin immunoprecipitation and gene expression analysis (51-54). Genes that are activated by OCT4, SOX2, and NANOG in ESCs are mostly up-regulated and important for ESC self-renewal (51, 52). The activated genes include transcription factors STAT3, ZIC3, REST, and HESX1, as well as chromatin and histone modification factors SMARCAD1, MYST3, and SET. Genes that are repressed by OCT4, SOX2, and NANOG are mostly down-regulated and participate in embryo development (51, 52). The repressed genes include HAND1, HOXB1, PAX6, and ISL1. NANOG binds to the promoter regions of OCT4 and SOX2, as well as to its own promoter region, and OCT4 and SOX2 jointly bind to NANOG, OCT4, and SOX2 (51, 52). The subsequently formed self-regulatory loop may help to stabilize the expression levels of these genes, since the precise level of OCT4 is required to maintain ESC pluripotency (40, 51). The transcription factors OCT4, SOX2, and NANOG hence form a core regulatory circuitry and play a key role in determining the fate of ESCs (51). This regulatory circuitry is unique and has no counterpart in non-mammalian genomes (55).

Chickarmane *et al.* conducted a computational modeling on the OCT4/SOX2/NANOG-directed network to examine the transcriptional dynamics, self-regulation, and target gene regulation (55). The modeling, based on the Shea–Ackers formalism (56), identified a bistable switch between self-renewal and differentiation in ESCs. The ESC box can be switched ON by such signals as WNT in human and mouse ESCs and LIF in mouse ESCs. Suppression of NANOG signaling molecules such as p53 turns the switch OFF. The bistable switch arises due to positive feedback loops formed by self-regulation of these genes. The model predicts that increasing the binding strength of NANOG to OCT4 and SOX2, or increasing the basal transcriptional rate of NANOG, leads to an irreversible bistable switch, in which the switch remains ON even when the activating signal is removed. Hence, ESCs can maintain self-renewal without further input signals. Such computational modeling furthers our understanding of important factors and their interplay on ESC pluripotency and facilitates manipulation of ESCs for directed differentiation.

## 6. CHROMOSOMAL MAP OF TRANSCRIPTOME

By examining the coexpression of neighboring genes along the chromosome, Li *et al.* (57) reported a chromosomal map of the hESC transcriptome. The distribution of coexpression domains on the genome is not random: they are significantly enriched in chromosomes 8, 11, 16, 17, 19, and Y in hESCs, and 6, 11, 17, 19 and 20 in EBs. Coexpression domains are biologically significant, and some of them are particularly associated with functions important to ESCs (57). Many chromosome domains show differential coexpression between ESCs and EBs. Some of the differentially coexpressed domains contain genes that are critical for ESC pluripotency, such as SOX15, STAT3, UTF1, TLE1, and OCT4. The SOX15 domain, for example, shows coexpression in ESCs but not in EBs, and contains cytokine TNFSF1, the ephrin gene EFNB3, and genes involved in transcriptional regulation (POLR2A, ZBTB4, TP53, and FXR2),. The observation of differential coexpression at chromosomal domains suggests that transcriptional regulation interacts with genomic structure and chromatin architecture in hESC differentiation. Genomic organization thus affects gene expression in hESCs and is important for regulating ESC pluripotency.

## 7. EPIGENETIC REGULATION

Emerging evidence suggests that transcriptional changes that are associated with ESC self-renewal and differentiation involve the action of unique epigenetic programs (42, 58-64). Epigenetic mechanisms important to ESCs include processes such as DNA methylation, non-

coding RNA-mediated regulation, and histone modifications.

### 7.1. DNA Methylation

DNA methylation is a general indicator of silenced genome regions and is essential for the establishment of chromatin structure during development (61-63). Many ESC-specific genes have been examined for the epigenetic status, and correlations between ESC development and DNA methylation have been established (58-60, 64-67). For hESCs, Lagarkova *et al.* (59) examined DNA methylation on promoter and putative regulatory regions of genes critical to ESC development (*e.g.* OCT4, NANOG, and DPPA) and showed the impact of these modifications on gene expression. Bibikova *et al.* (60) conducted a more comprehensive assessment, which surveyed the DNA methylation status from 1536 CpG sites of 371 genes in 14 hESC lines and five other cell types. The results suggest that hESCs have a unique epigenetic signature contributing to their developmental potential, and the DNA methylation profiles can clearly distinguish hESCs from adult stem cells, cancer cells, normal human tissues, and other cell types. A subset of 49 CpG sites from 40 genes contributes to the differences among different cell types. Another set of 25 sites from 23 genes distinguishes hESCs from normal differentiated cells and can be used as biomarkers to monitor ESC differentiation (60).

### 7.2. MicroRNAs

The critical role of microRNAs in ESCs was first revealed by the analysis of dicer-1, which cleaves double stranded RNA into small segments of approximately 22 nucleotides as an essential enzyme in microRNA biogenesis (68-71). Dicer-deficient ESCs, which can not generate microRNAs, were shown to be defective in ESC differentiation (72). ESC-specific microRNAs have been identified from both human and mouse and shown to play a unique role in ESC self-renewal and early differentiation (73, 74). These microRNAs are expressed only in undifferentiated ESCs, and the expression is significantly reduced as ESCs differentiate into EBs and becomes undetectable in adult organs (73, 74). MicroRNAs preferably expressed in other tissues or organs are poorly or not expressed in ESCs (73, 74). Many ESC-specific microRNAs have related sequences and are organized into various gene clusters (73, 74). In human, for example, 36 microRNAs are identified as ESC-specific, and eight of them (miR-302b, miR-302b*, miR-302c, miR- 302c*, miR-302a, miR-302a*, miR-302d, and miR-367) are located within a 200-bp region on chromosome 4, while another four microRNAs (miR-371, miR-372, miR-373, and miR-373*) are mapped within a 1050-bp region on chromosome 19 (74). Some ESC-specific microRNA genes are conserved across different mammalian genomes (73, 74). The ESC-specific microRNA miR-301 has the homolog in not only mammals but also in Fugu (73). Thus, ESC-specific microRNAs are involved in key and conserved molecular mechanisms in a diversity of organisms.

## 8. CONCLUDING REMARKS

The gene expression in ESCs is carefully regulated and cells either maintain the pluripotent state by self-renewal or undergo differentiation. The self-renewal is under the control of ESC-specific genes and by active repression of cell type-specific genes. The differentiation is initiated by activating cell type specific pathways and repressing self-renewal signals. Various genomic and transcriptomic studies have facilitated the discovery of regulatory mechanisms and deepened our understanding of the core molecular properties of ESCs. Nonetheless, great challenges remain ahead as we seek to comprehend the accumulating information from different aspects of ESC research. Integrated approaches are required in analyzing various kinds of 'omics' data to distill coherent conclusions. Sophisticated bioinformatics algorithms and software await development in order to address the complex patterns and massive data derived from ESC studies.

## 9. ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Institute on Aging, National Institutes of Health. Critical readings and comments by colleagues in our group and institute are gratefully acknowledged.

## 10. REFERENCES

1. Reubinoff, B. E., P. Itsykson, T. Turetsky, M. F. Pera, E. Reinhartz, A. Itzik & T. Ben-Hur: Neural progenitors from human embryonic stem cells. *Nat Biotechnol*, 19, 1134-40 (2001)
2. Rambhatla, L., C. P. Chiu, P. Kundu, Y. Peng & M. K. Carpenter: Generation of hepatocyte-like cells from human embryonic stem cells. *Cell Transplant*, 12, 1-11 (2003)
3. Kehat, I., D. Kenyagin-Karsenti, M. Snir, H. Segev, M. Amit, A. Gepstein, E. Livne, O. Binah, J. Itskovitz-Eldor & L. Gepstein: Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J Clin Invest*, 108, 407-14 (2001)
4. Kaufman, D. S., E. T. Hanson, R. L. Lewis, R. Auerbach & J. A. Thomson: Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc Natl Acad Sci U S A*, 98, 10716-21 (2001)
5. Segev, H., B. Fishman, A. Ziskind, M. Shulman & J. Itskovitz-Eldor: Differentiation of human embryonic stem cells into insulin-producing clusters. *Stem Cells*, 22, 265-74 (2004)
6. Kim, J. H., J. M. Auerbach, J. A. Rodriguez-Gomez, I. Velasco, D. Gavin, N. Lumelsky, S. H. Lee, J. Nguyen, R. Sanchez-Pernaute, K. Bankiewicz & R. McKay: Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature*, 418, 50-6 (2002)
7. Kehat, I., L. Khimovich, O. Caspi, A. Gepstein, R. Shofti, G. Arbel, I. Huber, J. Satin, J. Itskovitz-Eldor & L. Gepstein: Electromechanical integration of cardiomyocytes derived from human embryonic stem cells. *Nat Biotechnol*, 22, 1282-9 (2004)
8. Beachy, P. A., S. S. Karhadkar & D. M. Berman: Tissue repair and stem cell renewal in carcinogenesis. *Nature*, 432, 324-31 (2004)

9. Burdon, T., A. Smith & P. Savatier: Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol*, 12, 432-8 (2002)
10. Loebel, D. A., C. M. Watson, R. A. De Young & P. P. Tam: Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev Biol*, 264, 1-14 (2003)
11. Tiedemann, H., M. Asashima, H. Grunz & W. Knochel: Pluripotent cells (stem cells) and their determination and differentiation in early vertebrate embryogenesis. *Dev Growth Differ*, 43, 469-502 (2001)
12. Boiani, M. & H. R. Scholer: Regulatory networks in embryo-derived pluripotent stem cells. *Nat Rev Mol Cell Biol*, 6, 872-84 (2005)
13. Wei, C. L., T. Miura, P. Robson, S. K. Lim, X. Q. Xu, M. Y. Lee, S. Gupta, L. Stanton, Y. Luo, J. Schmitt, S. Thies, W. Wang, I. Khrebtukova, D. Zhou, E. T. Liu, Y. J. Ruan, M. Rao & B. Lim: Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cells*, 23, 166-85 (2005)
14. Ramalho-Santos, M., S. Yoon, Y. Matsuzaki, R. C. Mulligan & D. A. Melton: "Stemness": Transcriptional Profiling of Embryonic and Adult Stem Cells. *Science*, 298, 597-600 (2002)
15. Ivanova, N. B., J. T. Dimos, C. Schaniel, J. A. Hackney, K. A. Moore & I. R. Lemischka: A stem cell molecutar signature, 298, 601 (2002)
16. Sato, N., I. M. Sanjuan, M. Heke, M. Uchida, F. Naef & A. H. Brivanlou: Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol*, 260, 404 (2003)
17. Zeng, X., T. Miura, Y. Luo, B. Bhattacharya, B. Condie, J. Chen, I. Ginis, I. Lyons, J. Mejido, R. K. Puri, M. S. Rao & W. J. Freed: Properties of pluripotent human embryonic stem cells BG01 and BG02. *Stem Cells*, 22, 292-312 (2004)
18. Sperger, J. M., X. Chen, J. S. Draper, J. E. Antosiewicz, C. H. Chon, S. B. Jones, J. D. Brooks, P. W. Andrews, P. O. Brown & J. A. Thomson: Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A*, 100, 13350-5 (2003)
19. Richards, M., S. P. Tan, J. H. Tan, W. K. Chan & A. Bongso: The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells*, 22, 51-64 (2004)
20. Abeyta, M. J., A. T. Clark, R. T. Rodriguez, M. S. Bodnar, R. A. Pera & M. T. Firpo: Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum Mol Genet*, 13, 601-8 (2004)
21. Brandenberger, R., H. Wei, S. Zhang, S. Lei, J. Murage, G. J. Fisk, Y. Li, C. Xu, R. Fang, K. Guegler, M. S. Rao, R. Mandalam, J. Lebkowski & L. W. Stanton: Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat Biotechnol*, 22, 707-16 (2004)
22. Bhattacharya, B., T. Miura, R. Brandenberger, J. Mejido, Y. Luo, A. X. Yang, B. H. Joshi, I. Ginis, R. S. Thies, M. Amit, I. Lyons, B. G. Condie, J. Itskovitz-Eldor, M. S. Rao & R. K. Puri: Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood*, 103, 2956-64 (2004)
23. Miura, T., Y. Luo, I. Khrebtukova, R. Brandenberger, D. Zhou, R. S. Thies, T. Vasicek, H. Young, J. Lebkowski, M. K. Carpenter & M. S. Rao: Monitoring early differentiation events in human embryonic stem cells by massively parallel signature sequencing and expressed sequence tag scan. *Stem Cells Dev*, 13, 694-715 (2004)
24. Liu, Y., S. Shin, X. Zeng, M. Zhan, R. Gonzalez, F. J. Mueller, C. M. Schwartz, H. Xue, H. Li, S. C. Baker, E. Chudin, D. L. Barker, T. K. McDaniel, S. Oeser, J. F. Loring, M. P. Mattson & M. S. Rao: Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonal carcinoma cells to develop base profiles of U.S. Federal government approved hESC lines. *BMC Dev Biol*, 6, 20 (2006)
25. Sharov, A. A., Y. Piao, R. Matoba, D. B. Dudekula, Y. Qian, V. VanBuren, G. Falco, P. R. Martin, C. A. Stagg, U. C. Bassey, Y. Wang, M. G. Carter, T. Hamatani, K. Aiba, H. Akutsu, L. Sharova, T. S. Tanaka, W. L. Kimber, T. Yoshikawa, S. A. Jaradat, S. Pantano, R. Nagaraja, K. R. Boheler, D. Taub, R. J. Hodes, D. L. Longo, D. Schlessinger, J. Keller, E. Klotz, G. Kelsoe, A. Umezawa, A. L. Vescovi, J. Rossant, T. Kunath, B. L. Hogan, A. Curci, M. D'Urso, J. Kelso, W. Hide & M. S. Ko: Transcriptome analysis of mouse stem cells and early embryos. *Plos Biology*, 1, E74 (2003)
26. Tanaka, T. S., T. Kunath, W. L. Kimber, S. A. Jaradat, C. A. Stagg, M. Usuda, T. Yokota, H. Niwa, J. Rossant & M. S. Ko: Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. *Genome Research*, 12, 1921-8 (2002)
27. Cavaleri, F. & H. R. Scholer: Nanog: a new recruit to the embryonic stem cell orchestra. *Cell*, 113, 551-2 (2003)
28. Brandenberger, R., I. Khrebtukova, R. S. Thies, T. Miura, C. Jingli, R. Puri, T. Vasicek, J. Lebkowski & M. Rao: MPSS profiling of human embryonic stem cells. *BMC Dev Biol*, 4, 10 (2004)
29. Ginis, I., Y. Luo, T. Miura, S. Thies, R. Brandenberger, S. Gerecht-Nir, M. Amit, A. Hoke, M. K. Carpenter, J. Itskovitz-Eldor & M. S. Rao: Differences between human and mouse embryonic stem cells. *Dev Biol*, 269, 360-80 (2004)
30. Fortunel, N. O., H. H. Otu, H. H. Ng, J. Chen, X. Mu, T. Chevassut, X. Li, M. Joseph, C. Bailey, J. A. Hatzfeld, A. Hatzfeld, F. Usta, V. B. Vega, P. M. Long, T. A. Libermann & B. Lim: Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science*, 302, 393; author reply 393 (2003)
31. Evsikov, A. V. & D. Solter: Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science*, 302, 393; author reply 393 (2003)
32. Mikkers, H. & J. Frisen: Deconstructing stemness. *Embo J*, 24, 2715-9 (2005)
33. Suarez-Farinas, M., S. Noggle, M. Heke, A. Hemmati-Brivanlou & M. O. Magnasco: Comparing independent microarray studies: the case of human embryonic stem cells. *BMC Genomics*, 6, 99 (2005)
34. Parmigiani, G., E. S. Garrett-Mayer, R. Anbazhagan & E. Gabrielson: A cross-study comparison of gene expression studies for the molecular classification of lung cancer. *Clin Cancer Res*, 10, 2922-7 (2004)

35. Sun, Y., H. Li, H. Yang, M. S. Rao & M. Zhan: Mechanisms controlling embryonic stem cell self-renewal and differentiation. *Crit Rev Eukaryot Gene Expr*, 16, 211-31 (2006)
36. Humphrey, R. K., G. M. Beattie, A. D. Lopez, N. Bucay, C. C. King, M. T. Firpo, S. Rose-John & A. Hayek: Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. *Stem Cells*, 22, 522-30 (2004)
37. Daheron, L., S. L. Opitz, H. Zaehres, W. M. Lensch, P. W. Andrews, J. Itskovitz-Eldor & G. Q. Daley: LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells*, 22, 770-8 (2004)
38. Burdon, T., I. Chambers, C. Stracey, H. Niwa & A. Smith: Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. *Cells Tissues Organs*, 165, 131-43 (1999)
39. Ishiwata, I., Y. Tokeida, M. Iguchi, C. Ishiwata, K. Kiguchi, S. Yasumoto, K. Sato, T. Tachibana, H. Hashimoto & H. Ishikawa: New approach for the establishment of mouse early embryonic stem cells and induction of their differentiation. *Hum Cell*, 14, 283-91 (2001)
40. Niwa, H.: Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct Funct*, 26, 137-48 (2001)
41. Pritsker, M., T. T. Doniger, L. C. Kramer, S. E. Westcot & I. R. Lemischka: Diversification of stem cell molecular repertoire by alternative splicing. *Proc Natl Acad Sci U S A*, 102, 14290-5 (2005)
42. Rao, M.: Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev Biol*, 275, 269-86 (2004)
43. Stuart, J. M., E. Segal, D. Koller & S. K. Kim: A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302, 249-55 (2003)
44. Bergmann, S., J. Ihmels & N. Barkai: Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol*, 2, E9 (2004)
45. McCarroll, S. A., C. T. Murphy, S. Zou, S. D. Pletcher, C. S. Chin, Y. N. Jan, C. Kenyon, C. I. Bargmann & H. Li: Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat Genet*, 36, 197-204 (2004)
46. Sun, Y., H. Li, Y. Liu, S. Shin, M. P. Mattson, M. S. Rao & M. Zhan: Cross-species transcriptional profiles establish a functional portrait of embryonic stem cells. *Genomics*, 89, 22-35 (2007)
47. Zhan, M., T. Miura, X. Xu & M. S. Rao: Conservation and variation of gene regulation in embryonic stem cells assessed by comparative genomics. *Cell Biochem Biophys*, 43, 379-405 (2005)
48. Takahashi, K., K. Mitsui & S. Yamanaka: Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature*, 423, 541-5 (2003)
49. Miyoshi, J., M. Kagimoto, E. Soeda & Y. Sakaki: The human c-Ha-ras2 is a processed pseudogene inactivated by numerous base substitutions. *Nucleic Acids Res*, 12, 1821-8 (1984)
50. Haremaki, T., Y. Tanaka, I. Hongo, M. Yuge & H. Okamoto: Integration of multiple signal transducing pathways on Fgf response elements of the Xenopus caudal homologue Xcad3. *Development*, 130, 4907-17 (2003)
51. Boyer, L. A., T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch & R. A. Young: Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122, 947-56 (2005)
52. Loh, Y. H., Q. Wu, J. L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K. W. Sung, C. W. Lee, X. D. Zhao, K. P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C. L. Wei, Y. Ruan, B. Lim & H. H. Ng: The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet*, 38, 431-40 (2006)
53. Chambers, I., D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie & A. Smith: Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113, 643-55 (2003)
54. Mitsui, K., Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda & S. Yamanaka: The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 113, 631-42 (2003)
55. Chickarmane, V., C. Troein, U. A. Nuber, H. M. Sauro & C. Peterson: Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol*, 2, e123 (2006)
56. Buchler, N. E., U. Gerland & T. Hwa: On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*, 100, 5136-41 (2003)
57. Li, H., Y. Liu, S. Shin, Y. Sun, J. F. Loring, M. P. Mattson, M. S. Rao & M. Zhan: Transcriptome coexpression map of human embryonic stem cells. *BMC Genomics*, 7, 103 (2006)
58. Hattori, N., K. Nishino, Y. G. Ko, N. Hattori, J. Ohgane, S. Tanaka & K. Shiota: Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J Biol Chem*, 279, 17063-9 (2004)
59. Lagarkova, M. A., P. Y. Volchkov, A. V. Lyakisheva, E. S. Philonenko & S. L. Kiselev: Diverse epigenetic profile of novel human embryonic stem cell lines. *Cell Cycle*, 5, 416-20 (2006)
60. Bibikova, M., E. Chudin, B. Wu, L. Zhou, E. W. Garcia, Y. Liu, S. Shin, T. W. Plaia, J. M. Auerbach, D. E. Arking, R. Gonzalez, J. Crook, B. Davidson, T. C. Schulz, A. Robins, A. Khanna, P. Sartipy, J. Hyllner, P. Vanguri, S. Savant-Bhonsale, A. K. Smith, A. Chakravarti, A. Maitra, M. Rao, D. L. Barker, J. F. Loring & J. B. Fan: Human embryonic stem cells have a unique epigenetic signature. *Genome Res*, 16, 1075-83 (2006)
61. Lande-Diner, L. & H. Cedar: Silence of the genes--mechanisms of long-term repression. *Nat Rev Genet*, 6, 648-54 (2005)
62. Lorincz, M. C., D. R. Dickerson, M. Schmitt & M. Groudine: Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol*, 11, 1068-75 (2004)
63. Hashimshony, T., J. Zhang, I. Keshet, M. Bustin & H. Cedar: The role of DNA methylation in setting up chromatin structure during development. *Nat Genet*, 34, 187-92 (2003)
64. Shiota, K., Y. Kogo, J. Ohgane, T. Imamura, A. Urano, K. Nishino, S. Tanaka & N. Hattori: Epigenetic marks by

DNA methylation specific to stem, germ and somatic cells in mice. *Genes Cells*, 7, 961-9 (2002)
65. Allegrucci, C., A. Thurston, E. Lucas & L. Young: Epigenetics and the germline. *Reproduction*, 129, 137-49 (2005)
66. Keller, G.: Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev*, 19, 1129-55 (2005)
67. Maitra, A., D. E. Arking, N. Shivapurkar, M. Ikeda, V. Stastny, K. Kassauei, G. Sui, D. J. Cutler, Y. Liu, S. N. Brimble, K. Noaksson, J. Hyllner, T. C. Schulz, X. Zeng, W. J. Freed, J. Crook, S. Abraham, A. Colman, P. Sartipy, S. Matsui, M. Carpenter, A. F. Gazdar, M. Rao & A. Chakravarti: Genomic alterations in cultured human embryonic stem cells. *Nat Genet*, 37, 1099-103 (2005)
68. Bartel, D. P.: MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116, 281-97 (2004)
69. Lee, Y. S., K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer & R. W. Carthew: Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, 117, 69-81 (2004)
70. Jiang, F., X. Ye, X. Liu, L. Fincher, D. McKearin & Q. Liu: Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes Dev*, 19, 1674-9 (2005)
71. Saito, K., A. Ishizuka, H. Siomi & M. C. Siomi: Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biol*, 3, e235 (2005)
72. Kanellopoulou, C., S. A. Muljo, A. L. Kung, S. Ganesan, R. Drapkin, T. Jenuwein, D. M. Livingston & K. Rajewsky: Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev*, 19, 489-501 (2005)
73. Houbaviy, H. B., M. F. Murray & P. A. Sharp: Embryonic stem cell-specific MicroRNAs. *Dev Cell*, 5, 351-8 (2003)
74. Suh, M. R., Y. Lee, J. Y. Kim, S. K. Kim, S. H. Moon, J. Y. Lee, K. Y. Cha, H. M. Chung, H. S. Yoon, S. Y. Moon, V. N. Kim & K. S. Kim: Human embryonic stem cells express a unique set of microRNAs. *Dev Biol*, 270, 488-98 (2004)

**Abbreviations:** ESC: embryonic stem cell; EB: embryoid body; EST: expressed sequence tag; SAGE: serial analysis of gene expression; MPSS: massively parallel signature sequencing

**Key Words**: Embryonic Stem Cells, Pluripotency, Transcriptome, Comparative Genomics, Review

**Send correspondence to**: Dr. Ming Zhan, Bioinformatics Unit, Research Resources Branch, National Institute on Aging, NIH, 333 Cassall Drive, Baltimore, MD, USA, Tel: 410-558-8373, Fax: 410-558-8674, E-mail: zhanmi@mail.nih.gov

**http://www.bioscience.org/current/vol13.htm**