

Minireview

Chipping away at 'stemness'

April D Pyle*, Peter J Donovan* and Leslie F Lock†

Addresses: *Germ Cell and Stem Cell Group and †Genetic Models Group, Stem Cell Program, Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

Correspondence: Peter J Donovan. E-mail: pjdonovan@jhmi.edu. Leslie F Lock. E-mail: llock@jhmi.edu

Published: 29 July 2004

Genome Biology 2004, **5**:235

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/8/235>

© 2004 BioMed Central Ltd

Abstract

Global gene-expression analyses of human embryonic stem cells confirm the involvement of some known genes in stem-cell function and identify some new candidate regulators of stem-cell growth. Support remains elusive, however, for the concept of 'stemness' - a pattern of expression of genes that is common to all stem cells.

Embryonic stem (ES) cells are truly remarkable because of their ability both to renew themselves and to give rise to all the cells in the body. The ability to generate a wide range of differentiated cell types defines them as pluripotent. Human ES cells could be an incredible resource for the treatment of human disease. Differentiated cells derived from them could potentially be used to treat a wide variety of human conditions, including (but not limited to) heart disease, diabetes, stroke, Alzheimer's disease and Parkinson's disease. Several major hurdles remain to be overcome if such cells are to be used clinically, however. Stated simply, we know very little about their basic physiology or their true potential. What factors are required for their survival and proliferation? What factors can maintain them in a pluripotent state? What growth conditions affect their differentiation?

One way to address these questions is to analyze gene expression in ES cells. The goal is straightforward: if we can identify the full panoply of genes expressed in human ES cells and compare this with data from other stem-cell and non-stem-cell populations, it might be possible to define what makes ES cells unique. Such genes might be the ones that maintain ES cells in a self-renewing, pluripotent state. Knowledge of the genes expressed in ES cells could also have some very practical uses. For example, knowing that human ES cells express certain growth factor receptors could help in devising strategies to improve the growth of the cells in culture. Analyzing gene expression in human ES cells could

provide critical insights into the cell-surface receptors involved in growth control, cell-substrate adhesion and cell-cell adhesion in these and other cell types, and into intracellular signaling pathways involved in their basic physiology.

Several recent papers [1-5] have now reported global gene-expression analyses of a variety of human ES-cell lines (summarized in Table 1 and Figure 1). The datasets give important new insights into the basic physiology of these incredible cells. Several of the studies also compare the human ES-cell gene-expression datasets with published data from mouse ES cells [2,4,5] or with data derived from human embryonal carcinoma (EC) cells (the pluripotent stem cells derived from testicular tumors) or from seminomas (germ-cell-derived testicular tumor cells) [1]. Comparison of the genes that each study lists as candidates for involvement in self-renewal or pluripotency reveals both similarities and differences in gene-expression patterns among human ES-cell lines. The results include some tantalizing tidbits of information but also provide a cautionary tale for future research on ES cells.

The human ES-cell transcriptome

The data generated by the five groups [1-5] provide important confirmation of old ideas as well as providing support for some newly emerging concepts. For example, compelling evidence from studies carried out in mice [6] had previously

Table 1**Comparison of different methods and approaches for the analysis of human ES-cell gene expression**

	Sperger <i>et al.</i> [1]	Sato <i>et al.</i> [2]	Richards <i>et al.</i> [3]	Abeyta <i>et al.</i> [4]	Zeng <i>et al.</i> [5]
Human ES-cell lines used	H1, H7, H9, H13, H14	H1	HES3, HES4	H9, HSF-1, HSF-6	BG01, BG02
Culture conditions	MEFs	Matrigel	MEFs	MEFs	MEFs
Method of ES-cell isolation	Treatment with collagenase until colonies lifted off the MEFs	Treatment with dispase until cells were free of MEFs	Microdissection to free colonies of MEFs	Mechanical dissection of colonies from MEFs, then collagenase treatment	Trypsinization
Arrays used	Stanford microarrays	Affymetrix arrays (hU133A and mouse U74Av2)	SAGE	Affymetrix arrays (hU133A and hU133B)	Custom 16,659-spot 70-bp oligonucleotide array
Cells compared	hEC, hES and seminoma	hES and published mES [16]	hES and hES and additional SAGE libraries [24]	hES and hES versus published mES [16]	hES and hES versus mES [28]
Primary subtraction method	Somatic and cancer cell lines	Differentiated hES cells	None	None	Pooled human RNA
Software/analysis used	Significance analysis of microarrays (SAM) [25]	dChip and MAS 4.0 (Affymetrix)	Comparison of two SAGE resources with SAGE 2000 [26,27]	MAS 5.0 (Affymetrix)	Gene Pix (Axon Instruments)
Number of genes enriched in human ES cells	1,760	918	8,341	7,385	373
Candidate pluripotency genes*	565	227	192	76	92
Confirmation of gene expression using	RT-PCR	RT-PCR	RT-PCR	Quantitative RT-PCR	RT-PCR

*Candidate pluripotency genes are defined as genes that are found only in all pluripotent cell lines examined in each study. Abbreviations: hEC, human embryonic carcinoma cells; hES, human embryonic stem cells; MEFs, mouse embryonic fibroblasts; mES, mouse embryonic stem cells; RT-PCR, reverse-transcriptase-coupled PCR; SAGE, serial analysis of gene expression.

demonstrated a critical role for the POU domain transcription factor Oct4 (also called Pou5F1) in controlling the stability of the pluripotent state [6]. Encouragingly, all the studies [1-5] found high levels of expression of this factor in human ES cells. Also included in the list of genes enriched in human ES cells is the human homolog of the mouse *Nanog* gene. *Nanog* was recently found to be important for maintaining pluripotency in mouse ES cells [7,8]; deciphering the role of *NANOG* in human ES cells will therefore be of great interest. Interestingly, components of the leukemia inhibitory factor (LIF) pathway do not appear to be enriched in human ES cells, consistent with the observed lack of response by human ES cells to LIF, despite the fact that activation of the LIF pathway allows self renewal of mouse cells. Two additional enriched genes encode members of the transforming growth factor superfamily, LeftyA and LeftyB. These factors have been generally overlooked in stem-cell studies but now clearly warrant more attention. New insights into ES cell biology will emerge from analysis and functional testing of the candidates.

Comparison of the datasets from different human ES-cell lines suggests that each line has a unique gene-expression profile.

For example, the data from Abeyta *et al.* [4] show differences between lines: more than 20% of the approximately 14,000 genes found were expressed in only one of the three lines examined (see Figure 1d). What accounts for these differences in gene expression among human ES-cell lines? One possible explanation is genetic diversity. The genetic diversity present in the human population, represented to some degree in the available human ES cells, could affect gene-expression profiles among the cell lines. The expression of a gene can be affected by genetic polymorphisms in the gene itself or in modifier loci. The effects of modifier loci are well known to mouse geneticists and certainly act in human populations as well [9]. The growing list of available human ES-cell lines could be used to address the role that genetic diversity plays in regard to differential gene expression [10].

Genetic diversity may be part of the answer but is not likely to be the whole story. Differences in the purity of the cell populations used as the source of mRNA in the different studies could also account for some of the variation observed. The optimal conditions for maintaining human ES cells in an undifferentiated state are still not fully defined. For this

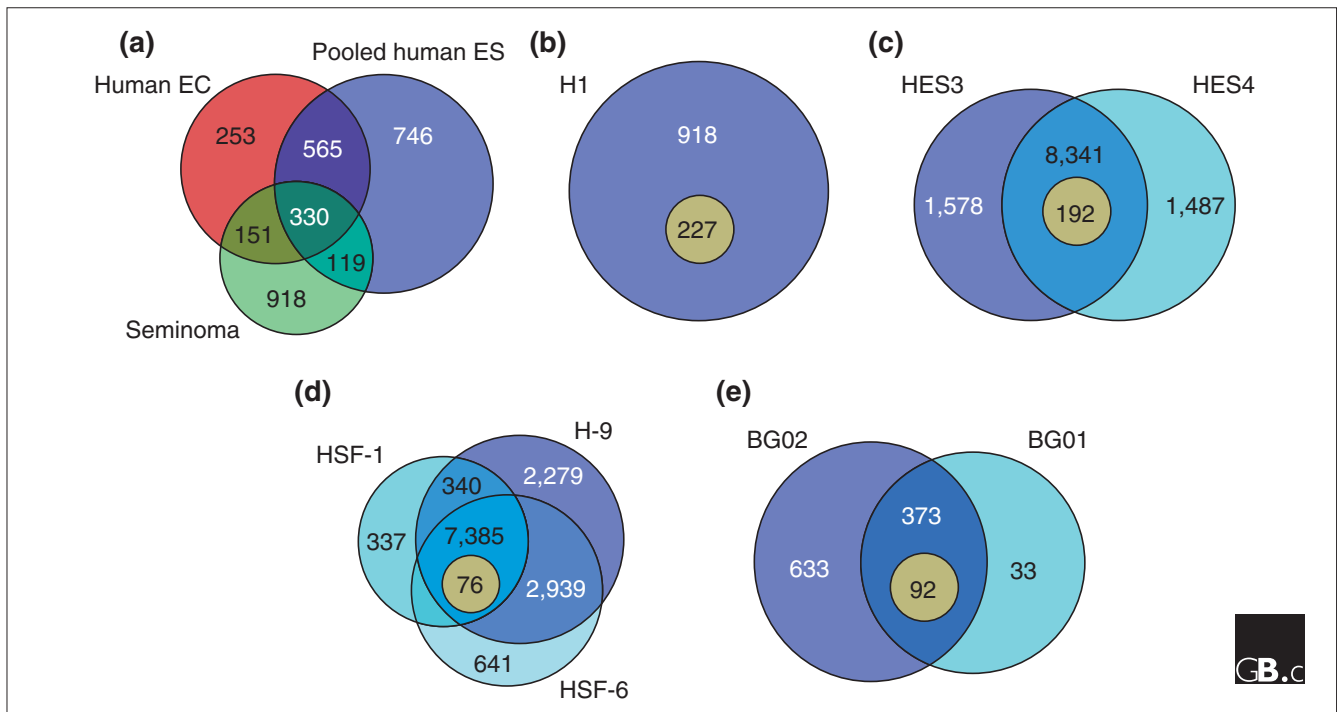


Figure 1

Comparison of the numbers of genes found to be enriched in human ES-cell lines in the five studies [1-5]. The circles each represent the number of genes found to be enriched in each cell line; the inner light circles represent genes shared with mouse cell lines (mouse genes not shared with human cell types are omitted). **(a)** Sperger *et al.* [1] compared five human ES-cell lines with seminoma and embryonal carcinoma (EC) cells and found 330 genes in common between them all. **(b)** Sato *et al.* [2] found 227 genes in common between human and mouse ES cells. **(c)** Richards *et al.* [3] found 192 genes that were upregulated in human ES cells compared with other human and mouse SAGE libraries. **(d)** Abeyta *et al.* [4] found 76 genes in common between three human ES-cell lines and mouse ES cells. **(e)** Zeng *et al.* [5] found 92 genes in common between two human ES-cell lines and mouse ES cells. In (b-d), the inner light circles represent the number of candidate pluripotency genes; in (a), candidate pluripotency genes are those shared by human ES and EC cells but not seminoma cells (565 genes).

reason, it seems possible that some of the variation in gene expression among human ES-cell lines may arise because the mRNA pools that were used contained not only ES-cell mRNA but also mRNA from differentiated derivatives. It is also worth noting that the behavior of the various cell lines can be quite different. Some lines are amenable to bulk culture, while others are not; some lines can be grown in feeder-independent culture, while others cannot; and some lines spontaneously differentiate, while others do not [11,12]. Culture conditions can affect gene imprinting, which in turn could have dramatic effects on gene expression. Variation in gene-expression patterns among human ES-cell lines could also reflect differences in the developmental stage represented by each line. If the cell lines are developmentally different - even slightly - then different gene-expression profiles could be observed. Indeed, *REX1*, an Oct4 target gene, and *FOXD3*, a gene required for development of pluripotent ES cells in mice, were enriched in some ES-cell lines but not others.

The stem-cell orchestra

One of the most challenging - yet crucial - unanswered questions regarding human ES cells is how they can be

maintained in a pluripotent state. By comparing the gene-expression profiles of pluripotent stem cells with those of other cell types, several of the groups [1-5] attempted to find genes whose expression could be unique to pluripotent stem cells; some have termed these the “stem-cell orchestra” [13]. One notable inclusion in the list of genes expressed by all human ES-cell lines is that encoding the transcription factor Oct4 mentioned above, consistent with its known role in maintaining pluripotency. Several other genes that have been associated with pluripotent cells are found in many of the datasets, including genes encoding Sox2, Nanog, Foxd3 and members of extracellular signaling pathway families including fibroblast (FGF), Wnt and bone morphogenetic protein (BMP).

All the studies [1-5] included lists of candidate genes that may play an important role in self-renewal and/or pluripotency. One perplexing observation is the lack of overlap among the gene lists. In part, this could result from the different subtraction methods used to removed genes expressed in non-ES cells from the datasets (see Table 1). Although there is unlikely to be one ‘correct’ method of subtraction, the differences in approach undoubtedly make comparison of results

complicated. Analysis is further complicated by the need to set a level of gene expression that is considered enriched or not enriched. In most of the studies [1,2,4,5], the authors chose to make the cut-off at two-fold to 3.5-fold; genes that show expression differences above these cut-offs were defined as enriched. It seems likely, however, that some of the most important players in the stem-cell orchestra may play a very soft tune. Indeed, studies have shown that changes in *Oct4* expression of less than two-fold can have profound effects on mouse ES-cell fate [14]. The arbitrary cut-off levels set by the authors [1,2,4,5] could exclude some of the key genes. Microarray analysis is also hampered by the 'lamp-post effect' - that is, only being able to see things where the light shines (microarray analysis excludes genes not present on the array). Thus, the studies that used microarray analysis [1,2,4,5] were able to identify only the genes with enriched expression in human ES cells that are present on the arrays. Importantly, one of the studies [3] used serial analysis of gene expression (SAGE). This technique avoids both the need for an arbitrary cut-off to define transcript enrichment and the lamp-post effect. It enables both quantitative characterization of gene expression and identification of novel splice variants, novel exons and novel genes [3,15].

A cautionary tale

In the last few years, the concept of 'stemness' has been introduced - the idea that different stem-cell populations, whether pluripotent or not, will share gene-expression signatures [16,17]. Can global gene-expression analysis provide new insights into this concept? It is difficult to answer this from the existing reports [1-5]. One idea that needs to be considered is that some of the genes that may be required to establish stemness may not be required to maintain that state. It also seems quite plausible that stem cells use pathways and molecules that are used elsewhere in non-stem-cell populations. A case in point is the *STAT3* gene, whose activity is sufficient for the maintenance of murine ES cells as stem cells. Activation of *STAT3* in mouse ES cells is brought about by LIF acting through gp130. This pathway is certainly not unique to stem cells, however, and plays important roles in many other aspects of development and adult homeostasis.

The recent gene-expression studies [1-5], nevertheless, provide a rich resource to gain important new insights into the physiology of human ES cells and provide an enormous amount of information that deserves our attention. But therein lies the difficulty: using microarray and SAGE approaches to examine gene expression is relatively easy, but deciphering gene function can be very difficult. How can gene function in human ES cells be determined? Altering gene expression by transfection or homologous recombination in human ES cells is technically difficult. Homologous recombination has been achieved in human ES cells but, to date,

there is only one published report of its use [18]. Moreover, some studies on gene function may require targeting of both alleles of a gene, a task that is wrought with difficulties even in mouse ES cells. Introducing genes into human ES cells with lentiviral vectors has been demonstrated and promises to be a workable technique [19-21]. RNA interference does work in human ES cells, but what is not yet clear is whether the effects are stable and whether true nulls (with no expression) are produced in addition to hypomorphs (with reduced expression) [2,23]. Nonetheless, such an approach may be useful for an initial examination of gene function.

Incremental advances in growing human ES cells in an undifferentiated state are likely to follow from the recent gene-expression studies [1-5]. The technology for manipulating the cells will improve, and this in turn will aid our ability to carry out analysis of gene function in human ES cells. Gradually we will begin to understand what gives these cells their unique properties, gene by gene. It may seem tedious, but consider the alternatives. If human ES cells had been discovered a decade earlier, some of the approaches now used would not have been available. The sequencing of the human genome has allowed us to examine the complexity of gene expression in a way that was inconceivable ten years ago. Discoveries made using additional genome-wide approaches may finally allow us to see beyond the light cast by microarrays and deeper into the mysteries of stemness.

Acknowledgements

We thank Robert W. Clouse III for help in preparing the figure and gratefully acknowledge support from the National Institutes of Health and Johns Hopkins University.

References

1. 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 USA* 2003, **100**:13350-13355.
2. 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.
3. 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.
4. 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.
5. Zeng X, Miura T, Luo Y, Bhattacharya B, Condie B, Chen J, Ginis I, Lyons I, Mejido J, Puri RK, et al.: **Properties of pluripotent human embryonic stem cells BG01 and BG02.** *Stem Cells* 2004, **22**:292-312.
6. 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.
7. 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.
8. 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.

9. Nadeau JH: **Modifier genes and protective alleles in humans and mice.** *Curr Opin Genet Dev* 2003, **13**:290-295.
10. Cowan CA, Klimanskaya I, McMahon J, Atienza J, Witmyer J, Zucker JP, Wang S, Morton CC, McMahon AP, Powers D, Melton DA: **Derivation of embryonic stem-cell lines from human blastocysts.** *N Engl J Med* 2004, **350**:1353-1356.
11. Richards M, Tan SP, Fong CY, Biswas A, Chan WK, Bongso A: **Comparative evaluation of various human feeders for prolonged undifferentiated growth of human embryonic stem cells.** *Stem Cells* 2003, **21**:546-556.
12. Vogel G: **Stem cells. Are any two cell lines the same?** *Science* 2002, **295**:1820.
13. Cavaleri F, Scholer HR: **Nanog: a new recruit to the embryonic stem cell orchestra.** *Cell* 2003, **113**:551-552.
14. 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.
15. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270**:484-487.
16. 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.
17. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR: **A stem cell molecular signature.** *Science* 2002, **298**:601-604.
18. Zwaka TP, Thomson JA: **Homologous recombination in human embryonic stem cells.** *Nat Biotechnol* 2003, **21**:319-321.
19. Pfeifer A, Ikawa M, Dayn Y, Verma IM: **Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos.** *Proc Natl Acad Sci USA* 2002, **99**:2140-2145.
20. Gropp M, Itsykson P, Singer O, Ben-Hur T, Reinhartz E, Galun E, Reubinoff BE: **Stable genetic modification of human embryonic stem cells by lentiviral vectors.** *Mol Ther* 2003, **7**:281-287.
21. Ma Y, Ramezani A, Lewis R, Hawley RG, Thomson JA: **High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors.** *Stem Cells* 2003, **21**:111-117.
22. Hay DC, Sutherland L, Clark J, Burdon T: **Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells.** *Stem Cells* 2004, **22**:225-235.
23. Vallier L, Rugg-Grunn PJ, Bouhon IA, Andersson FK, Sadler AJ, Pedersen RA: **Enhancing and diminishing gene function in human embryonic stem cells.** *Stem Cells* 2004, **22**:2-11.
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.
25. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
26. **CGAP SAGE Anatomic Viewer**
[<http://cgap.nci.nih.gov/SAGE/AnatomicViewer>]
27. **NCBI SAGEMAP** [<http://www.ncbi.nih.gov/SAGE/>]
28. Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, Joshi BH, Ginis I, Thies RS, Amit M, et al.: **Gene expression in human embryonic stem cell lines: unique molecular signature.** *Blood* 2004, **103**:2956-2964.