The Human Embryonic Stem Cells Transcriptome: How Much Do We Know?

Shubha Vij[1], Lai ZhenYang[1], Wong Pui Mun[1], Chak Li-Ling[1], Wang Yue[1], Chan Woon-Khiong[1,*] and Ariff Bongso[2]

[1]*Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543 and* [2]*Department of Obstetrics and Gynecology, National University of Singapore, National University Hospital, Singapore 119074*

Abstract: The establishment of human embryonic stem cell (hESC) lines in 1998 served to set the pace for understanding the molecular biology behind the two hallmark features of stem cells: self renewal and pluripotency. The excitement was generated in the hope that understanding the molecular biology of hESCs would provide a good model for studying early human development, disease and drug discovery and also hold the promise for providing a cure for degenerative human diseases. In spite of the large number of studies, the molecular basis of pluripotency has remained a matter of intrigue ever since the embryonic stem cells (ESCs) were first identified. A considerable percentage of these studies have been transcriptome-based. Interestingly, significant differences are seen not only between mouse and human ESC transcriptomes but also amongst the hESC studies. Nevertheless, a key set of pluripotency genes seem to be common, reinforcing the utility of transcriptome-based approaches in identifying the molecular basis of pluripotency in hESCs.

INTRODUCTION

Embryonic stem cells (ESCs) are derived from the inner cell mass of embryos at the blastocyst stage (4-5 days post-fertilization). Mouse embryonic stem cells (mESCs) were reported for the first time in [1], while the first human embryonic stem cell (hESC) line was reported in [2]. The derivation of ESCs has brought a lot of excitement due to their hallmark features of self renewal and pluripotency. The ability to proliferate in an undifferentiated state for a prolonged period of time *in vitro* and the capability to differentiate into cells of different lineages [2] makes them distinct even from adult stem cells.

Most of our present knowledge of the pluripotent state has been contributed by studies on mESCs. A considerable number of studies have pinpointed *Nanog*, *Pou5f1* and *Sox2* as the key genes involved in the maintenance of pluripotency [3-6]. Out of these, *Pou5f1* which codes for a POU-domain transcription factor can be easily called the most well characterized ESC-specific gene. Its expression is necessary not only for pluripotency but also governs the three cellular fates following differentiation [7-8]. In a major breakthrough study last year, it was shown that only four genes (*Pou5f1*, *Sox2*, *c-Myc* and *Klf4*) were required to generate induced pluripotent stem cells (iPS) from adult mouse fibroblasts [9]. Further studies by various groups showed that iPS cells injected into mouse blastocysts were able to produce all tissue types with one study going on to successfully produce mice entirely from iPS cells [10-12]. Recently, Yamanaka's group was able to reproduce their work in humans, marking another major landmark in ESC research [13]. An independent group has also shown that a different combination of four factors (*POU5F1*, *NANOG*, *SOX2* and *LIN28*) [14] can be used to generate iPS from human fibroblasts. The expression levels of key transcription factors are also important in the maintenance of pluripotency. For example, overexpression of *Nanog* is sufficient to maintain the mESCs in the undifferentiated state even in the absence of leukaemia inhibitory factor (LIF) whereas its suppression leads to differentiation into the extra-embryonic endoderm [3,4]. On the other hand, overexpression of *Pou5f1* induces the mESCs to differentiate into the endoderm and mesoderm whereas its repression leads to differentiation into trophoectoderm [7].

Several studies have focused on comparing the mESC and hESC transcriptomes with the general conclusion that these two transcriptomes have limited overlap [15-17]. For instance, LIF-mediated JAK-STAT3 along with BMP4 signaling is required to keep the mESCs in the undifferentiated state [18]. However, in the case of humans, this role seems to be taken over by FGF2 [19]. This is further reflected by a higher preponderance of FGF receptors 1, 3, 4 in hESCs compared to mouse. Conversely, mESCs express high levels of LIFR along with JAK and STAT3 whereas LIFR and JAK have not been identified in hESCs [20,21]. Further, bone morphogenetic protein (BMP4) which is required to maintain the pluripotent state in mESCs induces trophoblast differentiation in hESCs [22]. A comparison of the key transcriptome studies of mESC and hESC revealed a commonality of only 13-55% transcripts, whereas, an overlap of 85-99% was seen amongst the hESC lines [23]. Thus, from the onset it became clear that though mESCs and hESCs do share a few common components, the differences are too large to be presumptuous of one another emphasizing the

*Address correspondence to this author at the Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, E-mail: dbscwk@nus.edu.sg

1874-0707/08

need for an extensive characterization of the hESC transcriptome. In this review, we summarize the outcome of hESC transcriptome analysis using a multitude of techniques such as microarray, SAGE, MPSS and EST analysis. We also offer an analysis as to why there is little overlap amongst the outcome of various studies and to what extent understanding the proteome can help resolve this disparity.

THE MANY WAYS OF UNRAVELING THE hESC TRANSCRIPTOME

The unlimited self renewal capacity coupled with their ability to differentiate into cells of the three germ layers and high plasticity has made ESCs the focus of intense research. Our current understanding of the molecular basis of pluripotency in hESCs stems largely from global profiling techniques such as DNA microarray [24], Serial Analysis of Gene Expression (SAGE) [25,26], Massively Parallel Signature Sequencing (MPSS) [17,27,28] and Expressed Sequence Tag (EST) analysis [21,28]. Amongst the various lines used for analysis, H1, H7, H9, H13 and H14 represent the most well studied human stem cell lines [29].

The pace was set by microarray studies to reveal the preliminary catalogue of stemness genes. In one such early experiment, the human ESC lines, human germ cell tumour cell lines, tumour samples, somatic cell lines and testicular tissue was compared revealing a higher overlap between the ESC and EC transcriptomes [20]. Microarray was also used for comparing the transcriptomes of hESC lines leading to the identification of 92 genes which could possibly represent the ESCs 'stemness' signature [30]. In a later analysis, the same group was able to show that >80% of these 92 genes were downregulated in embryoid bodies [31] providing confidence to the 'stemness' profile. A meta-analysis of the hESC transcriptome mainly comprising the results of microarray analysis was also reported in which comparisons were made with various fetal and adult tissues. One cluster each of hESC genes with high mitotic index genes, CNS genes along with a hESC-specific cluster could be recognized. In addition, there was one cluster which was overexpressed in hESCs but was also present in the majority of the tissues analyzed representing mostly the housekeeping genes [24].

SAGE like other tag-based approaches, has the advantage over microarray in its ability to identify novel transcripts. The first SAGE analysis of the hESC transcriptome was reported for two hESC lines, HES3 and HES4, representing different genetic and ethnic backgrounds [25]. The hESC transcriptome was found to be enriched in genes involved in DNA repair, stress response, apoptosis, cell cycle regulation as well as development [25]. In addition, a much higher level of expression was seen for genes involved in protein synthesis as well as mRNA processing. In fact, transcripts coding for ribosomal proteins were 4-8 times more abundant than in normal tissues. Comparison of HES3 and HES4 transcriptomes revealed an overall similarity in their profiles and the differences could be at least partially attributed to different gender backgrounds amongst other factors. Comparison of the hESC transcriptome with the 21 SAGE libraries from normal and cancer tissues revealed a set of ~200 upregulated transcripts including *POU5F1, SOX2, REX1, NANOG, LIN28* and *DNMT3B*. In addition, comparison with the mESC transcriptome revealed basic similarities such as the preponderance of metabolic enzymes, ribosomal proteins as well as cytoskeletal proteins. However, fundamental differences were found to exist in their regulatory pathways which obliterated a direct comparison [25]. Recently, longSAGE analysis was performed for nine human ESC lines altogether producing 2.5 million tags representing 379,645 unique tag sequences for analysis. Comparison of this ESC meta-library to 247 non-ESC libraries helped in identification of 20,047 tags unique to ESCs and also revealed a prevalence of RNA binding proteins in hESCs [26]. Though, SAGE allows for the identification of novel transcripts, its main drawback is its short length due to which it cannot be directly compared to the genome but only to a reference database such as SAGE map [32] or SAGE Genie [33]. Also, in many instances, tags either show hits to multiple loci or do not show any specific hit. Several strategies have been proposed to overcome this limitation including a reverse SAGE (rSAGE) method [34] which can convert these 'orphan' tags into useful information by generating longer 3′ cDNAs. Besides identification of novel genes, this approach proved useful in identification of natural antisense transcripts (NATs), novel introns and new splice variants of known transcripts.

Extensive EST analysis of the hESC transcriptome has also been reported [21,28]. In one such study, a total of 148,453 ESTs were generated from undifferentiated hESCs and three partially differentiated hESCs to yield 32,000 unique transcripts. The work sought to identify the differentially expressed genes especially with respect to LIF, FGF, WNT and NODAL signaling pathways [21]. While most of the LIF signaling components were not detected, all four FGF receptors were not only expressed but were also upregulated in undifferentiated ESCs. The undifferentiated ESCs also expressed most WNT pathway genes along with several agonists and antagonists of the NODAL pathway.

MPSS profiling has also been done for hESCs wherein three hESC lines (H1, H7 and H9) were pooled for the analysis. Three well known ESC-specific genes, *SOX*2, *DNMT3B* and *POU5F1* were present in the 200 most abundant signatures. Also, similar to the findings from hESC EST data, components of signaling pathways were detected but their inhibitors were also present, indicating the role of negative regulation in maintaining the undifferentiated state [27]. Comparison of MPSS data with a previous microarray analysis of six hESC lines reporting a set of 92 stemness genes revealed a ~95% overlap [27,30]. MPSS was also used for comparing human and mouse hESC transcriptomes. As in other analyses, only few genes were found to be similar including *POU5F1*, *SOX2*, *BMPR*, *NODAL*, *LEFTY*, *TERT* and *CRIPTO* and major differences in the LIF, TGFβ, WNT and FGF pathways were identified [17].

hESC TRANSCRIPTOME: THE SEARCH FOR STEMNESS GENES CONTINUES….

The transcriptome analyses of hESCs have largely focused on the belief that the 'stemness' signature can be defined by a small set of genes which should be enriched in all stem cell populations. However, equally interesting is the fact that there is very little consistency in the key 'stemness' genes identified by various groups. We sifted through the literature to understand whether these findings are a reflec-

tion of technical reasons or does it actually mean that no set of genes can actually represent the pluripotent state.

Embryonic stem cells eventually graduate into a diverse array of tissues required during normal human development. A subset of these which become adult stem cells exhibit self renewal and multi-lineage differentiation but do not exhibit the self renewal capacity or the pluripotent characteristic of ESCs [23]. To see if stem cells of different origins possess a similar signature, two studies attempted to compare the transcriptome of ESCs, haematopoietic stem cells (HSCs) and neurospheres. Interestingly, both studies reported ~200 genes representing the 'stemness' signature but only with a 3% overlap amongst them [35-37]. Similarly, even amongst hESCs, only seven genes were common to three independently published microarray datasets. An attempt to resolve the gap by conducting a re-analysis found 111 and 95 upregulated and downregulated genes, respectively, common to the three studies. Thus, even though the three studies were different in their starting material, data analyses, different versions of annotation databases used amongst other factors, a uniform method of data analysis seems to bridge the difference to an extent [15,20,30,38].

Most of the molecular profiling studies done in hESCs have been done with the aim of identifying genes implicated in pluripotency. However, few have focused on understanding the differences in gene expression patterns between different ESC lines. Such studies have shown that differences exist not only between different ESC lines but also between the same ESC lines from different labs. For instance, 25% variation in differentially expressed genes between H1 and BG01 [39] and 48% between HSF-1, HSF-6 and H9 [40] have been documented. In another such analysis, microarray was used to compare the transcriptome of seven hESC lines from two independent sources with all the lines subject to similar culture conditions [41]. The hESC transcriptomes were comparable including 8464 common transcripts of which 280 were hESC-specific. In spite of this, a line-specific profile could also be distinguished. Comparisons were also made with similar studies from different labs and an overlap of 30-93% was reported [15,20,30,40].

Recently, an attempt was made to create a universal expression atlas for ESCs [24]. A total of 38 transcriptome studies based on 28 different hESC lines were compiled and analyzed. They could identify a consensus hESC gene list (1,076 genes) and a consensus differentiation list (783 genes) based on the genes that were commonly overexpressed and downregulated, respectively, in at least three studies. Surprisingly, only one gene (*POU5F1*) was commonly overexpressed in all the studies and not even a single gene was common to the downregulation list. Some of these large variations, as they themselves point out can be attributed to differences in experimental design and data interpretation.

Thus, from the array of transcriptome studies, it is clear that though there are overlaps, vast differences exist between stem cell profiles at the cross-species level as well as within different lines from the same species. One clearly viable explanation is that the hESCs used in the various studies are presumed to be equivalent in terms of their pluripotent state. However, a major concern is that till now it is not possible to rigorously verify the pluripotent state in hESCs. This should be considered as one of the important factors causing limited overlap between transcriptome as well as proteome studies. Another major reason could be that the disparities are due to technical reasons and have more to do with the experimental design and data interpretation along with variable genetic backgrounds rather than to the actual *in vivo* condition. Some differences are also expected due to the different methodologies. For instance, the SAGE signature is dependent on the *Nla*III recognition site due to which, expression of ~1% of human genes present in RefSeq cannot be detected by SAGE [42]. Also, many identified transcripts do not match to annotated data further complicating the task of effective comparison. Another possibility is that the stem cell signature is not wholly unique and the key genes involved are expressed ubiquitously or possibly are expressed only transiently and hence difficult to detect. In fact the most logical way out of the problem is to analyze raw data from the different transcriptome profiling strategies and apply various statistical tests in order to generate a true representative hESC list. To analyze the basis of differences across various transcriptome data, we sought to compare our lab's SAGE data with the raw SAGE tag information from an independent source [26]. In order to compare the degree of similarity between two different hESC lines from the same lab in relation to the same hESC lines from different labs, we carried out the following two-way comparisons: HES3lab *vs* HES4lab [25] (Fig. **1a**), HES3Hirst *vs* HES4Hirst [26] (Fig. **1b**), HES3lab *vs* HES3Hirst (Fig. **1c**) and HES4lab *vs* HES4Hirst (Fig. **1d**). SAGE data of two different lines from the same lab (Figs. **1a**, **1b**) were clearly more closely related to each other than the same hESC lines from different labs (Figs. **1c**, **1d**). The first question which arose from the outcome of the analysis is whether the datasets from the two labs differed due to the differences in technique with our lab's data based on short SAGE and the latter based on LongSAGE. Alternatively, the difference could be attributed to the status of differentiation of the hESC lines. In order to address the first question, HES3lab was compared with short SAGE data for H9 (www.transcriptomes.org) (Fig. **1e**). The result of this comparison (Fig. **1e**) was definitely more similar than either Fig. (**1c** or **1d**) corroborating that the method used would have a strong bearing on the outcome of the analysis. Further, to find if the three different lab datasets were comparable in terms of quality, HES3lab, HES3Hirst and H9 were chosen for further analysis. All these three datasets were probed for the presence of 95 pluripotency genes common to $\geq$8 experiments and 75 differentiation genes common to $\geq$6 studies reported in the hESC transcriptome meta-analysis [24]. The HES3lab, H9, HES3Hirst showed an overlap of 66, 75 and 81 to the pluripotency dataset and 31, 34 and 46 to the differentiation dataset. Thus, the three datasets were comparable to an extent and the differences can be attributed to the depth of the analysis in each case.

A PROTEOME VIEW OF THE hESCs

To have a complete insight into the basis of stemness, the proteome should be investigated in parallel with the transcriptome. It is especially important to understand the proteome in terms of protein synthesis, transport, degradation, protein-protein interactions as well as post translational modifications (PTMs) since proteins are at the forefront of cellular functions. Further, there is evidence to show that the mRNA levels reflect only 40% of the changes occurring at

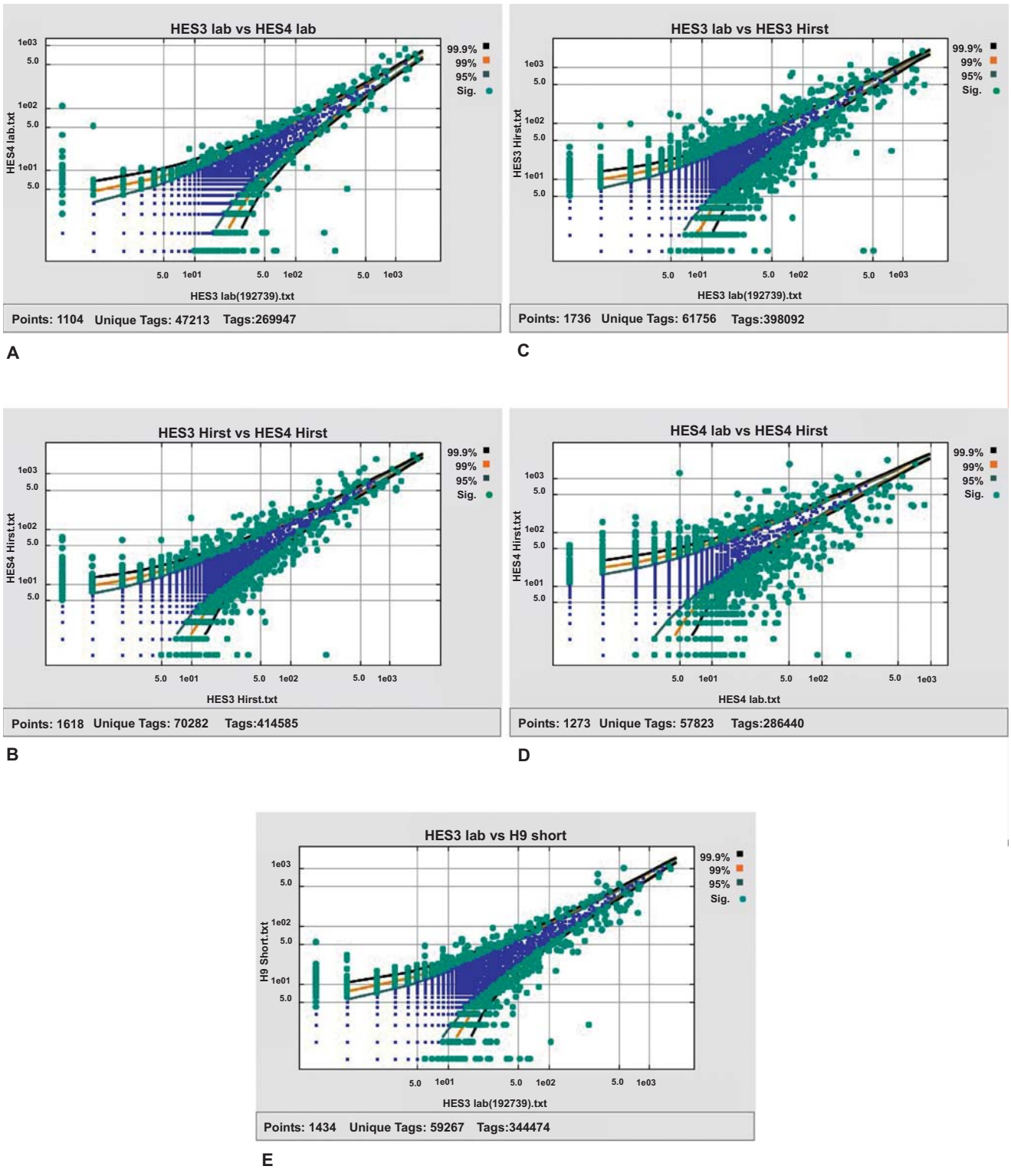

Fig. (1). Comparisons of different SAGE libraries using DiscoverySpace software [56]. Comparison of HES3 lab *vs* HES4 lab (**A**); HES3 Hirst *vs* HES4 Hirst (**B**); HES3 lab *vs* HES3 Hirst (**C**); HES4 lab *vs* HES4 Hirst (**D**) and HES3 lab *vs* short H9 (**E**). (HES3 Hirst and HES4 Hirst were originally LongSAGE libraries and hence converted to short SAGE for analyses).

the protein level [43,44]. Thus, proteome analysis could provide new insights into hESC biology and also help identify new markers for stem cells.

Proteomics has been given a boost by improved technologies both for detection as well as analysis and several independent groups have reported the mESC proteome [45-48]. A proteomic analysis of three hESC lines using 2-DE and MALDI TOF-TOF identified 685 protein spots of which the most abundant included chaperones, heat shock proteins, ubiquitin/proteasome and oxidative stress responsive proteins along with proteins involved in cell proliferation [49]. A similar association of the mESC transcriptome to stress

has been reported earlier [36]. In contrast to the mESC proteome where nuclear proteins were the maximum in number, cytoplasmic proteins are the biggest class in the hESC proteome. Further, the protein destination category including proteins involved in folding, targeting, modification and proteolysis represented the largest functional class rather than the metabolism category reported for the mESC proteome [47]. Also, comparison of the hESC proteome with the stemness proteins identified in the mESC proteome revealed an overlap of ~30%. Comparison of the most abundant proteins identified to those reported in a previous hESC transcriptome analysis [40] revealed an overlap of only seven proteins. A large-scale proteome comparison for mESCs and hESCs was reported last year in which a significant advantage was the use of the sensitive FT-ICR-MS/MS technique to compare the proteome profiles in the undifferentiated and differentiated ESCs in humans as well as mouse [50]. A subtractive approach was used to identify hESC/mESC proteins specific to the undifferentiated state. This dataset along with the proteins which were expressed >3 times more in undifferentiated ESCs gave a list of 730 human and 888 mouse proteins. Many of the ESC-specific proteins were found to be related to cell cycle progression. Cross-species differences were expected and were also found between mouse and humans. However, comparison of the mESC and hESC datasets did provide a list of 191 proteins common to both datasets and included known stem cell markers as well as novel proteins. The mESC proteome was also compared with the previously published mESC data to reveal a concordance of 89% [46] and 47% [47]. The higher overlap observed in the former dataset was probably due to the smaller scale of the data (218 proteins) in comparison to the latter study (1790 proteins). The specific function of the candidate proteins emerging out of various analyses can be deciphered through a more targeted approach. Recently, the protein interaction network of Nanog was studied to reveal a large number of nuclear factors as well as multiple co-repressor pathways providing novel insights into the maintenance of pluripotency [51]. Even though the proteome will give a better reflection of what is happening inside the cell, the technical limitations in proteome analysis persist resulting in proteomic data continually falling short of transcriptomic data in terms of magnitude. Thus both the platforms are needed in order to complement each other as well as to improve our understanding of ESCs.

PERSPECTIVE

ESC pluripotency and self renewal capability can be maintained indefinitely under artificial conditions, but a key question is how this ability is guarded in the *in vivo* condition, i.e. what is the role of pluripotency in normal development and when is it switched off for the sake of differentiation. In contrast to differentiated cells which express ~10-20% of their genes, ~30-60% genes are expressed in ESCs. Evidence exists for the ESC chromatin existing in an open state allowing for the expression of a diverse array of lineage-specific genes albeit at low levels. The presence of these genes might be necessary to receive the necessary cues from the microenvironment but this proposal remains to be validated [23]. Many hESC transcriptome analyses have focused on the search for those genes whose expression is upregulated and also those which become downregulated when differentiation occurs. This combined set could serve as a catalogue for assessing the state of the cell.

To a large extent, our current understanding of mESCs and hESCs has come through transcriptome analysis. Extensive molecular profiling of mouse and human ES cells have shown large differences between the two systems. Recently, epiblast derived stem cells have been reported in mouse and rats [52,53] which show a higher similarity to hESCs. This finding has lent support to a long held suspicion that the observed differences between hESCs and mESCs could represent altogether different stages of development and hence are not directly comparable. However, in spite of evident differences between mESCs and hESCs, some basic similarities exist such as the central role played by *Pou5f1*, *Nanog* and *Sox2* in maintaining the pluripotent state in mESCs as well as hESCs [4-7,54]. This implies the possibility that the expression levels and activity of a small group of transcription factors play a central role in guiding the fate of stem cells. The specific role of these transcription factors of course depends on the developmental stage of the cell along with the various epigenetic factors. To gain a deeper insight into the role of POU5F1, NANOG and SOX2 in regulation of pluripotency, an attempt was made to identify their downstream targets using ChIP-DNA microaray analysis [54]. Interestingly, POU5F1, SOX2 and NANOG were found to have a certain degree of overlap in the genes they regulate. In addition, they not only co-occupy promoters of their target genes but their binding sites were also found to be close to each other. Furthermore, many of the genes which are regulated by them are also transcription factors. POU5F1, SOX2 and NANOG were found to regulate 3%, 7% and 9% of the known protein coding genes in hESCs, respectively. Analysis of the 353 common set of genes regulated by POU5F1, NANOG and SOX2 showed that these three transcription factors mediate their function by positively regulating their own expression as well as of key signaling pathways in pluripotency and by repressing genes involved in differentiation.

The outcome of ESC transcriptome studies highlight that in addition to cross-species differences which are expected, disparities exist between the outcomes of hESC transcriptome analysis also. It remains to be solved whether the differences in gene expression patterns observed between the different hESCs could be attributed to culture conditions, genetic variation, epigenetic factors or difference in the pluripotent state amongst others or some of these differences may actually turn out to be of some biological significance. It is quite possible that the genes that are not detected through different transcriptome analyses may represent those which are dispensable for the pluripotent state. Thus, possibly the most abundant or differentially expressed proteins in ESCs which are also the targets in most of the studies could be the central players. However, the possibility exists that some other set of ubiquitous proteins are at the core of maintaining the stemness phenotype and have been ignored till now. However, transcriptome analysis has been able to throw in the candidate genes but a detailed characterization is needed especially of the key genes in understanding the basis of pluripotency in stem cells. Also, in addition to the key transcription factors, a large amount of evidence exists for the role of TGFβ, FGF4, WNT and Notch signaling in the regulation of pluripotency [55].

Due to their unique capability of self renewal and pluripotency, stem cells had given the promise of its immense potential in cell-based therapies ever since their discovery. However, several milestones need to be crossed before ESCs can get the green signal for human use. Some of the preliminary concerns include the immune response following hESC transplantation, the possibility of terminally differentiating cells reverting to their stemness phenotype and the possibility of stray hESCs which escaped lineage commitment amongst other factors. All these factors can be resolved once we have a complete understanding of the molecular biology of ESCs. This will also aid in crossing the milestone of clinical trials whose outcome needs to be known before stem cells can be brought to use. The generation of iPS cells in mouse as well as humans has also made a significant impact on our understanding of stem cell biology. However, the molecular basis of iPS cells needs to be fully understood before its potential can be fully realized. In the meantime, efforts should be increased to unveil the molecular basis of pluripotency so that the benefits of stem cell research can reach the millions of patients who think stem cells can be the magic potion which will give them the much needed new lease of life.

ACKNOWLEDGMENTS

The research work of our Laboratory is funded by Ministry of Education, Singapore (AcRF Grant, R-154-000-284-112) and Singapore Stem Cell Consortium (SSCC-06-02). Lai ZhenYang and Chak Li-Ling were supported by NUS postgraduate scholarship.

REFERENCES

[1] Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci USA 1981; 78: 7634-8.

[2] Thomson JA, Itskovitz-Eldor J, Shapiro SS, *et al.* Embryonic stem cell lines derived from human blastocysts. Science 1998; 282: 1145-7.

[3] Nichols J, Zevnik B, Anastassiadis K, *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 1998; 95: 379-91.

[4] Avilion AA, Nicolis SK, Pevny LH, *et al*. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 2003; 17: 126-40.

[5] Chambers I, Colby D, Robertson M, *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 2003; 113: 643-55.

[6] Mitsui K, Tokuzawa Y, Itoh H, *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 2003; 113: 31-42.

[7] Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 2000; 24: 372-6.

[8] 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: 225-35.

[9] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006; 126: 663-76.

[10] Maherali N, Sridharan R, Xie W, *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 2007; 1: 55-70.

[11] Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. Nature 2007; 448: 313-7.

[12] Wernig M, Meissner A, Foreman R, *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 2007; 448: 318-24.

[13] Takahashi K, Tanabe K, Ohnuki M, *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007; 131: 861-72.

[14] Yu J, Vodyanik MA, Smuga-Otto K, *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science 2007; DOI: 10.1126/science.1151526.

[15] Sato N, Sanjuan IM, Heke M, *et al*. Molecular signature of human embryonic stem cells and its comparison with the mouse. Dev Biol 2003; 260: 404-13.

[16] Ginis I, Luo Y, Miura T, Thies S, *et al.* Differences between human and mouse embryonic stem cells. Dev Biol 2004; 269: 360-80.

[17] Wei CL, Miura T, Robson P, *et al.* Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. Stem Cells 2005; 23: 166-85.

[18] Ying QL, Nichols J, Chambers I, Smith A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 2003; 115: 281-92.

[19] Xu C, Inokuma MS, Denham J, *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. Nat Biotechnol 2001; 19: 971-4.

[20] Sperger JM, Chen X, Draper JS, *et al.* Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Natl Acad Sci USA 2003; 100: 13350-5.

[21] Brandenberger R, Wei H, Zhang S, Lei S, *et al.* Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat Biotechnol 2004; 22: 707-16.

[22] Xu RH, Chen X, Li DS, *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nat Biotechnol 2002; 20: 1261-4.

[23] Eckfeldt CE, Mendenhall EM, Verfaillie CM. The molecular repertoire of the 'almighty' stem cell. Nat Rev Mol Cell Biol 2005; 6: 726-37.

[24] Assou S, Le Carrour T, Tondeur S, *et al.* A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. J Stem Cells 2007; 25: 961-73.

[25] 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.

[26] Hirst M, Delaney A, Rogers SA, *et al.* Long SAGE profiling of nine human embryonic stem cell lines. Genome Biol 2007; 8: R113.

[27] Brandenberger R, Khrebtukova I, Thies RS, *et al*. MPSS profiling of human embryonic stem cells. BMC Dev Biol 2004; 4: 10.

[28] Miura T, Luo Y, Khrebtukova I, *et al.* Monitoring early differentiation events in human embryonic stem cells by massively parallel signature sequencing and expressed sequence tag scan. Stem Cells Dev 2004; 13: 694-715.

[29] Allegrucci C, Young LE. Differences between human embryonic stem cell lines. Hum Reprod Update 2007; 13: 103-20.

[30] Bhattacharya B, Miura T, Brandenberger R, *et al.* Gene expression in human embryonic stem cell lines: unique molecular signature. Blood 2004; 103: 2956-64.

[31] Bhattacharya B, Cai J, Luo Y, *et al.* Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiating embryoid bodies. BMC Dev Biol 2005; 5: 22.

[32] Lash AE, Tolstoshev CM, Wagner L, *et al.* SAGEmap: a public gene expression resource. Genome Res 2000; 10: 1051-60.

[33] Boon K, Osorio EC, Greenhut SF, *et al.* An anatomy of normal and malignant gene expression. Proc Natl Acad Sci USA 2002; 99: 11287-92.

[34] Richards M, Tan SP, Chan WK, Bongso A. Reverse serial analysis of gene expression (SAGE) characterization of orphan SAGE tags from human embryonic stem cells identifies the presence of novel transcripts and antisense transcription of key pluripotency genes. Stem Cells 2006; 24: 1162-73.

[35] Ivanova NB, Dimos JT, Schaniel C, *et al*. A stem cell molecular signature. Science 2002; 298: 601-4.

[36] Ramalho-Santos M, Yoon S, Matsuzaki Y, *et al*. "Stemness": transcriptional profiling of embryonic and adult stem cells. Science 2002; 298: 597-600.

[37] Fortunel NO, Hatzfeld JA, Monier MN, Hatzfeld A. Control of hematopoietic stem/progenitor cell fate by transforming growth factor-beta. Oncol Res 2003; 13: 445-53.

[38] Suárez-Fariñas M, Noggle S, Heke M, *et al*. Comparing independent microarray studies: the case of human embryonic stem cells. BMC Genomics 2005; 6: 99.

[39] Rao RR, Stice SL. Gene expression profiling of embryonic stem cells leads to greater understanding of pluripotency and early developmental events. Biol Reprod 2004; 71: 1772-8.

[40] Abeyta MJ, Clark AT, Rodriguez RT, *et al*. Unique gene expression signatures of independently-derived human embryonic stem cell lines. Hum Mol Genet 2004; 13: 601-8.

[41] Skottman H, Mikkola M, Lundin K, *et al*. Gene expression signatures of seven individual human embryonic stem cell lines. Stem Cells 2005; 23: 1343-56.

[42] Boon K, Osorio EC, Greenhut SF, *et al.* An anatomy of normal and malignant gene expression. Proc Natl Acad Sci USA 2002; 99: 11287-92.

[43] Ideker T, Thorsson V, Ranish JA, *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001; 292: 929-34.

[44] Tian Q, Feetham MC, Tao WA, *et al.* Proteomic analysis identifies that 14-3-3zeta interacts with beta-catenin and facilitates its activation by Akt. Proc Natl Acad Sci USA 2004; 101: 15370-5.

[45] Guo X, Ying W, Wan J, *et al.* Proteomic characterization of early-stage differentiation of mouse embryonic stem cells into neural cells induced by all-trans retinoic acid *in vitro*. Electrophoresis 2001; 22: 3067-75.

[46] Elliott ST, Crider DG, Garnham CP, *et al*. Two-dimensional gel electrophoresis database of murine R1 embryonic stem cells. Proteomics 2004; 4: 3813-32.

[47] Nagano K, Taoka M, Yamauchi Y, *et al.* Large-scale identification of proteins expressed in mouse embryonic stem cells. Proteomics 2005; 5: 1346-61.

[48] Puente LG, Borris DJ, Carrière JF, *et al*. Identification of candidate regulators of embryonic stem cell differentiation by comparative phosphoprotein affinity profiling. Mol Cell Proteomics 2006; 5: 57-67.

[49] Baharvand H, Hajheidari M, Ashtiani SK, Salekdeh GH. Proteomic signature of human embryonic stem cells. Proteomics 2006; 6: 3544-9.

[50] Van Hoof D, Passier R, Ward-Van Oostwaard D, *et al*. A quest for human and mouse embryonic stem cell-specific proteins. Mol Cell Proteomics 2006; 5: 1261-73.

[51] Wang J, Rao S, Chu J, Shen X, *et al*. A protein interaction network for pluripotency of embryonic stem cells. Nature 2006; 444: 364-8.

[52] Brons IG, Smithers LE, Trotter MW, *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 2007; 448: 191-5.

[53] Tesar PJ, Chenoweth JG, Brook FA, *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 2007; 448: 196-9.

[54] Boyer LA, Lee TI, Cole MF, *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 2005; 122: 947-56.

[55] Rho JY, Yu K, Han JS, *et al*. Transcriptional profiling of the developmentally important signalling pathways in human embryonic stem cells. Hum Reprod 2006; 21: 405-12.

[56] Robertson N, Oveisi-Fordorei M, Zuyderduyn SD, *et al*. DiscoverySpace: an interactive data analysis application. Genome Biol 2007; 8: R6.

Received: December 17, 2007 Revised: January 25, 2008 Accepted: January 25, 2008