Human Molecular Genetics, 2005, Vol. 14, No. 21 ***3129–3140***
doi:10.1093/hmg/ddi345
Advance Access published on September 13, 2005

Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells

Tariq Enver[1], Shamit Soneji[1], Chirag Joshi[1], John Brown[1], Francisco Iborra[1], Torben Orntoft[2], Thomas Thykjaer[2], Edna Maltby[3], Kath Smith[3], Raed Abu Dawud[4], Mark Jones[4], Maryam Matin[5], Paul Gokhale[4], Jonathan Draper[4] and Peter W. Andrews[4,*]

[1]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK, [2]Molecular Diagnostics Laboratory, Aarhus University Hospital, Skejby, Denmark, [3]North Trent Clinical Cytogenetics Service, Sheffield Children's Trust, Western Bank, Sheffield S10 2TH, UK, [4]Department of Biomedical Science, Centre for Stem Cell Biology, University of Sheffield, Sheffield, UK and [5]Department of Biology, Institute of Biotechnology and Tissue Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Received May 26, 2005; Revised August 1, 2005; Accepted September 7, 2005

Human embryonic stem cell (HESC) lines vary in their characteristics and behaviour not only because they are derived from genetically outbred populations, but also because they may undergo progressive adaptation upon long-term culture *in vitro*. Such adaptation may reflect selection of variants with altered propensity for survival and retention of an undifferentiated phenotype. Elucidating the mechanisms involved will be important for understanding normal self-renewal and commitment to differentiation and for validating the safety of HESC-based therapy. We have investigated this process of adaptation at the cellular and molecular levels through a comparison of early passage (normal) and late passage (adapted) sublines of a single HESC line, H7. To account for spontaneous differentiation that occurs in HESC cultures, we sorted cells for SSEA3, which marks undifferentiated HESC. We show that the gene expression programmes of the adapted cells partially reflected their aberrant karyotype, but also resulted from a failure in X-inactivation, emphasizing the importance in adaptation of karyotypically silent epigenetic changes. On the basis of growth potential, ability to re-initiate ES cultures and global transcription profiles, we propose a cellular differentiation hierarchy for maintenance cultures of HESC: normal SSEA3+ cells represent pluripotent stem cells. Normal SSEA3− cells have exited this compartment, but retain multilineage differentiation potential. However, adapted SSEA3+ and SSEA3− cells co-segregate within the stem cell territory, implying that adaptation reflects an alteration in the balance between self-renewal and differentiation. As this balance is also an essential feature of cancer, the mechanisms of culture adaptation may mirror those of oncogenesis and tumour progression.

INTRODUCTION

Cell identity and potency is ultimately a function of gene expression. Several studies have sought to gain insights into the mechanisms of self-renewal and differentiation in human embryonic stem cells (HESCs) through global gene expression profiling of the undifferentiated stem cells and comparison with their differentiated derivatives, the latter often in the form of haphazardly differentiated embryoid bodies (1–5). Others have attempted to analyse the transcriptome by quantifying the relative abundance of expressed RNA in undifferentiated HESC and to identify those genes thought to be involved in pluripotency (6,7). A number of specific genes that are characteristically expressed in undifferentiated HESC, and downregulated upon their differentiation, have been identified in each of these studies, notably *POU5F1*

*To whom correspondence should be addressed. Tel: +44 1142224173; Fax: +44 1142222399; Email: p.w.andrews@sheffield.ac.uk

(*Oct4*), *NANOG* and *TDGF*. In contrast, substantial differences are also evident between the various HESC lines and between different studies. In part, these differences might reflect the genetic heterogeneity of HESC lines derived, as they are, from a genetically diverse, outbred population (1,2).

However, in addition, these studies have treated cultures of HESC as homogeneous populations of cells, despite evidence to the contrary (8,9), potentially confounding the reliability of the transcripts identified as putative pluripotency-related genes. HESC cultures are often heterogeneous because they contain both the undifferentiated stem cells and the spontaneously arising differentiated derivatives. This heterogeneity can be addressed by fractionating ES cultures according to the expression of specific cell surface markers. Several such antigens have been proposed as markers of undifferentiated HESC, including the globoseries glycolipid antigens SSEA3 and SSEA4 and the keratan sulphate-associated antigens TRA-1-60, TRA-1-81 and GCTM2 (10–12). Studies of the expression patterns of these antigens in human embryonal carcinoma (EC) cells, the malignant counterparts of ES cells, and in human ES cells themselves (9,12–14) suggest that SSEA3 in particular might represent a sensitive marker of the most primitive state for human ES cells. On the basis of our past extensive experience with human EC and ES cells, we have consistently found that, in various differentiation protocols, the expression of SSEA3 is almost always lost most rapidly and before that of the other human EC/ES cell marker antigens.

A further problem is that HESC cultures may contain cells that have adapted to the conditions under which they are grown. Certainly, the initial outgrowth of HESC from an explanted blastocyst is likely to involve some degree of adaptation of the cells to the conditions of proliferation *in vitro*, for example, indicated by the simple fact of indefinite growth of HESC in culture compared with the limited existence of pluripotent cells of the inner cell mass and epiblast in the developing embryo. However, long-term maintenance in culture seems likely to select for further genetic changes that promote self-renewal and limit differentiation or apoptosis. They may involve gross karyotypic changes (15), but might also include other more subtle genetic or epigenetic changes or indeed reversible 'dynamic' changes in gene regulation or intermediary metabolism. Such progressive culture adaptation is likely to also be reflected in changes in the transcriptome of the cells at different passage levels.

Determining the cellular and molecular bases of heterogeneity due to differentiation and adaptation is important for understanding the mechanisms of ES cell self-renewal and differentiation and for devising improved culture conditions that minimize the selective advantage of variant cells and therefore help maintain genetically normal cells suitable for therapeutic applications. We have addressed both aspects of differentiation and adaptation in a study of gene expression patterns of a single HESC line, H7 (10), cultures of which we have sorted for expression of the surface antigen SSEA3 at a relatively early passage level when the cells were diploid (46, XX) and at a much later passage when the cells had acquired an additional copy of the long arm of chromosome 17 (15). We denote these diploid, early passage and aneuploid, late passage sublines as 'normal' and 'adapted', respectively. The use of sublines derived from a single HESC line obviates the effects of genotypic differences that exist between cell lines derived from different individuals and thus eliminates one potential source of variability.

Taken together, our results emphasize the importance of addressing the heterogeneity of ES cell cultures in any efforts to identify the key genes of the pluripotent state. Moreover, they provide a basis for identification of genes associated with adaptation and have allowed us to develop a hierarchical model of the commitment of HESC to differentiation, reminiscent of those developed for other stem cell systems.

RESULTS

ES cell lines

Normal H7 cells were confirmed to have a diploid 46, XX karyotype, whereas the adapted H7 cells had acquired an extra copy of chromosome 1 by the time of the present study, in addition to the chromosome 17q amplification described previously (15). The karyotype of the adapted cells was 47,XX, + 1,der(6)t(6;17)(q27;q1) (Supplementary Material). No other consistent structural chromosomal abnormalities affecting all cells were observed. The adapted H7 subline used in this study had been maintained in culture for over a year and more than 100 passages. It had a substantially greater population growth rate than the normal H7 subline (Fig. 1A), but still exhibited a capacity for extensive differentiation (Fig. 1B and C).

Transcriptional profiling

Three maintenance cultures of both normal and adapted H7 cells were separately labelled for expression of SSEA3 and sorted by flow cytometry. Although broadly similar, a slightly lower proportion of SSEA3(+) cells were found in the normal when compared with the adapted cultures [average SSEA3(+) cells: 64 and 83%, respectively] (Fig. 1D). Fractions corresponding to the brightest and dimmest cells were collected and processed for gene expression profiling using Affymetrix genechips. We also analysed the transcriptional profiles for three separate cultures of normal and adapted cells that had been allowed to differentiate extensively by growth as embryoid bodies followed by attachment and further outgrowth.

Signatures of pluripotency and self-renewal

A simple present/absent analysis of gene expression in the different samples is shown as Venn diagrams (Fig. 1E). By this analysis, 293 genes were uniquely expressed in normal SSEA3+ (N3+) cells compared with their SSEA3− (N3−) and differentiated (ND) counterparts; 212 genes were uniquely expressed in adapted SSEA3+ (A3+) cells compared with their SSEA3− (A3−) and differentiated (AD) counterparts. However, this stringent signature of pluripotency excludes genes such as *POU5F1* (*Oct4*), a key regulator of murine ES cell and HESC self-renewal (16–18), as it is expressed in all samples, albeit at significantly different levels. We therefore assessed differential gene expression in all normal and adapted cell samples using multiclass significance analysis

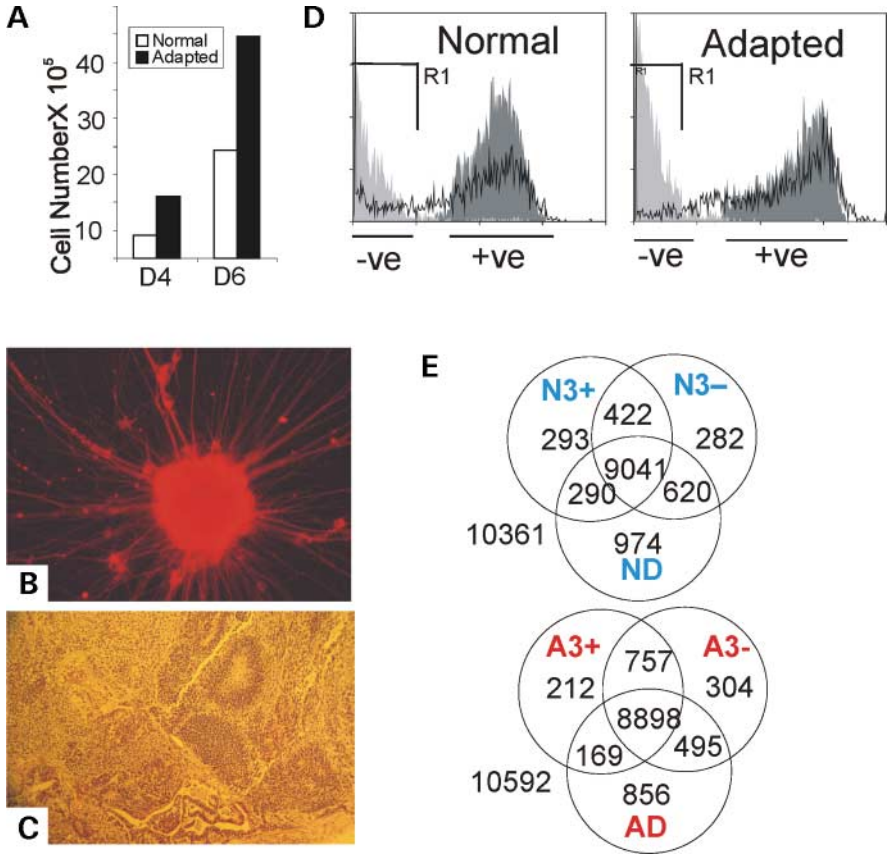
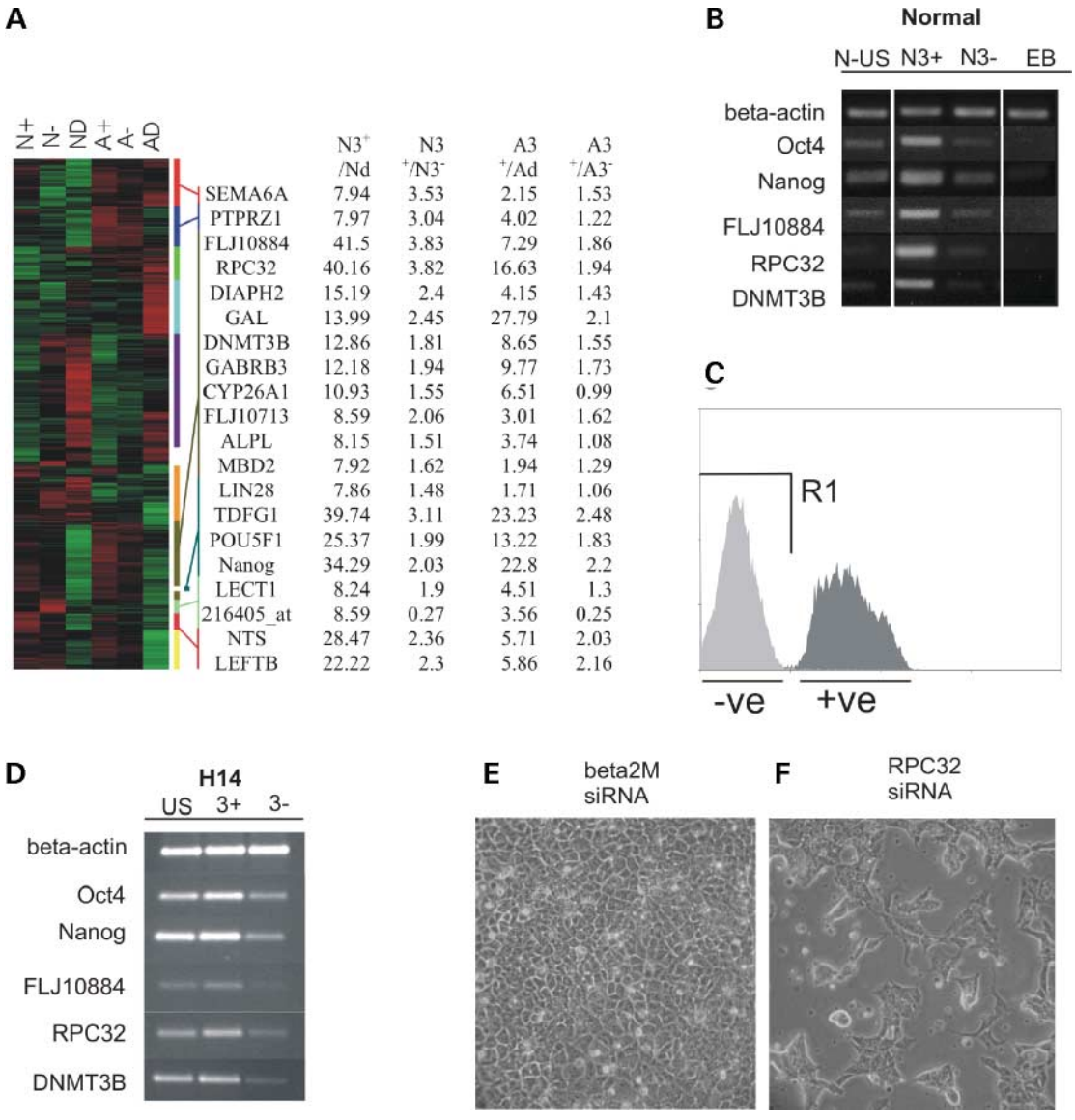


Figure 1. The adapted H7 HESC showed a substantially greater population growth rate than the normal H7 HESC, but retained the ability to differentiate and exhibited similar patterns of SSEA3 expression to the normal H7 HESC. (**A**) Growth of normal and adapted HESCs single cells seeded at 2×10^5 cells/25 cm^2 on feeders and counted 4 and 6 days after seeding. (**B**) Outgrowth of neurons from an embryoid body derived from adapted cells; the neurons are identified by immunostaining with a monoclonal antibody to TuJ1. (**C**) A teratoma xenograft of adapted HESC in a SCID mouse, showing neural rosettes and glandular structures. (**D**) Fluorescence histograms of SSEA3 expression in normal and adapted HESCs; R1 indicates the negative gate set following staining with the negative control antibody P3X63Ag. Similar distributions of SSEA3 expression were evident in both sublines, though with slightly more SSEA3− cells in the normal cultures (average from three separate sorts: normal H7, 64% SSEA3+ and adapted H7, 83% SSEA3+). Superimposed on these histograms are the re-analysed histograms (shaded) of subsets of cells sorted to be SSEA− and SSEA3+. (**E**) Venn diagrams showing the numbers of genes scored 'present' in different subsets of normal and adapted HESCs (N, normal HESC; A, adapted HESC; 3+, SSEA3+; 3−, SSEA3− and D, differentiated cells).

of microarrays (SAMs); the resulting genes were subjected to hierarchical clustering (Fig. 2A).

In total, 866 genes were more highly expressed in the N3+ when compared with the ND samples (Supplementary Material); 20 genes exhibiting the greatest changes in expression are shown together with their relative expression in the different HESC subsets (Fig. 2A). These include several genes previously implicated as candidate regulators or markers of the pluripotent state, such as *POU5F1*, *NANOG* and *TDGF1* (2,16,19). The differential expression of several of these genes in N3+, N3− and ND subsets was verified by reverse transcription (RT)–polymerase chain reaction (PCR) of RNA prepared from sorted populations of normal H7 cells independent of those used for the Affymetrix Array experiment (Fig. 2B). Note that these genes exhibited significant, although relatively modest (2-fold), differences in expression between the N3+ and N3− subsets; in all, 124 downregulated and 425 upregulated genes showed at least a 2-fold difference between the N3+ and N3− cells (Supplementary Material). By extending this analysis to an additional HESC line, H14, we confirmed that the differences in gene expression observed between SSEA3+ and SSEA3− cells were not a unique characteristic of the H7 cell line. Accordingly, cells from maintenance cultures of H14 were sorted for SSEA3 expression (Fig. 2C) and analysed for expression of selected genes by semi-quantitative RT–PCR (Fig. 2D). The results obtained were consistent with those seen in H7 cells. *FLJ10884* and *RPC32* were the most high regulated transcripts in common between the N+ versus N3− and N3+ versus ND analyses of H7 cells and consistently exhibited higher expression in SSEA3+ than in SSEA3− H14 cells (Fig. 2C); neither have been previously associated with the self-renewing pluripotent state. Although the function of *FLJ10884*, which contains LINE retroviral sequences, remains unknown, *RPC32* has been identified as a transcriptional subunit of RNA polymerase III (20). Inspection of the GEO database (21) reveals a restricted expression pattern for *RPC32* in both normal and cancer tissues. Its potential importance was evidenced by a dramatic inhibition in cell growth following the introduction of *RPC32* siRNA into the



	N3+/Nd	N3+/N3-	A3+/Ad	A3+/A3-
SEMA6A	7.94	3.53	2.15	1.53
PTPRZ1	7.97	3.04	4.02	1.22
FLJ10884	41.5	3.83	7.29	1.86
RPC32	40.16	3.82	16.63	1.94
DIAPH2	15.19	2.4	4.15	1.43
GAL	13.99	2.45	27.79	2.1
DNMT3B	12.86	1.81	8.65	1.55
GABRB3	12.18	1.94	9.77	1.73
CYP26A1	10.93	1.55	6.51	0.99
FLJ10713	8.59	2.06	3.01	1.62
ALPL	8.15	1.51	3.74	1.08
MBD2	7.92	1.62	1.94	1.29
LIN28	7.86	1.48	1.71	1.06
TDFG1	39.74	3.11	23.23	2.48
POU5F1	25.37	1.99	13.22	1.83
Nanog	34.29	2.03	22.8	2.2
LECT1	8.24	1.9	4.51	1.3
216405_at	8.59	0.27	3.56	0.25
NTS	28.47	2.36	5.71	2.03
LEFTB	22.22	2.3	5.86	2.16

Figure 2. (**A**) A hierarchical cluster of all genes found to be differentially expressed by a multiclass SAM with minimum FDR. The table adjacent shows the position of the top 20 genes most downregulated between the N3+ and ND cells. Clustering was implemented in Genesis (43). (**B**) RT–PCR analysis for the selection of genes upregulated in SSEA3+ normal H7 HESC. Lanes from left to right: unsorted HESC (U/S); SSEA3+ HESC (3+); SSEA3− HESC (3−) and D, differentiated cells. (**C** and **D**) The results of sorting the independent HESC line H14 for expression of SSEA3 and analysing the expression patterns of selected genes in the SSEA3+ and SSEA3− subsets by RT–PCR. Note that as with the sorted H7 cells (B), these genes show markedly lower expression in the SSEA3− subset. (**E** and **F**) Cultures of NTERA2 EC cells treated 4 days with siRNA directed to β2microglobulin (negative control) and to *RPC32*; note the substantial reduction in growth of cells treated with siRNA to *RCP32*.

pluripotent human EC cell line, NTERA2, which is readily amenable to RNA interference. (Fig. 2E and F). We are currently investigating the mechanisms underlying this observation and the function of *RPC32* in the undifferentiated stem cells. Initial results suggest that *RPC32* is required for cell viability (unpublished data) as opposed to cell proliferation *per se*.

Attention is drawn to a group of 58 genes (Supplementary Material) that are down regulated in N3− cells versus their N3+ counterparts and then up regulated in ND cells. These would not have been revealed by previous studies of unsorted cells and include the key haematopoietic and neural stem cell marker *PROM1* (*prominin 1*) (22). In a similar vein, 14 genes (Supplementary Material), typified by the interstitial collagenase, *MMP1*, were up regulated between N3+ and N3− cells, but were then down regulated in the ND cells. Such genes could be potentially 'miscalled' as stem cell genes if cultures were not first fractionated to remove spontaneously arising differentiated derivatives.

Mechanisms of adaptation in culture

Comparison of the global gene expression profiles of the different subsets of normal and adapted cells (Fig. 3A)

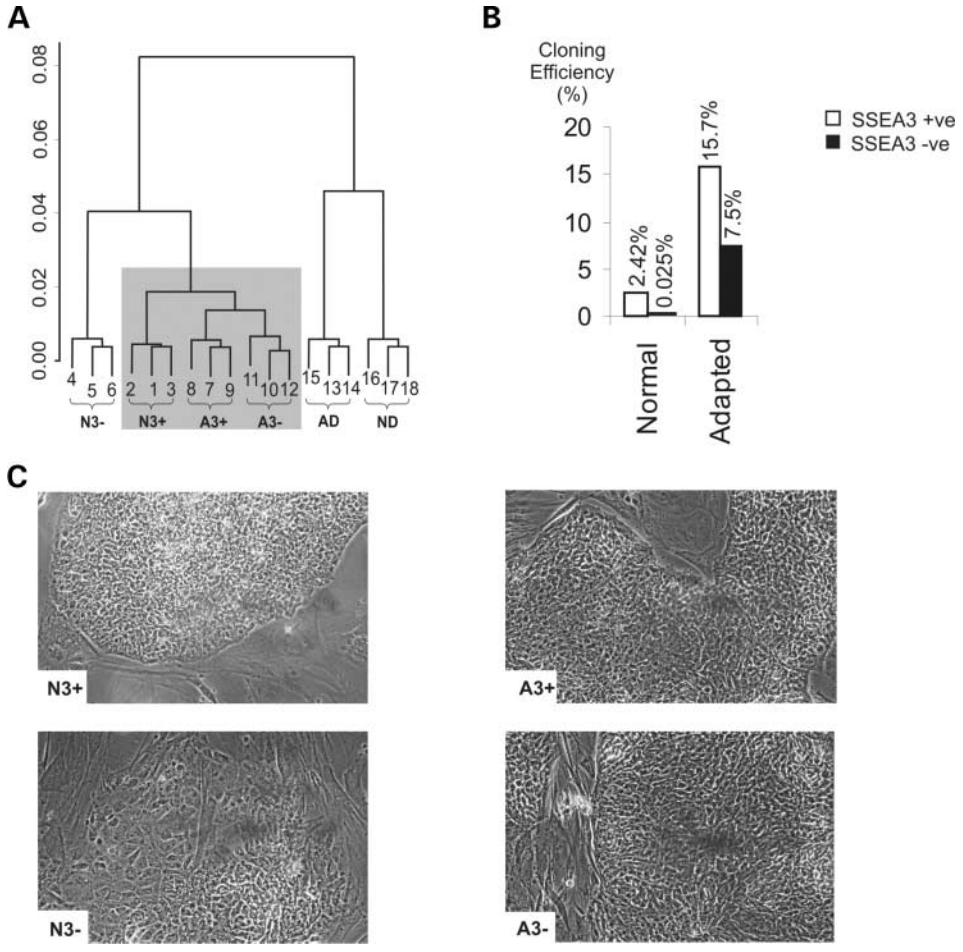


Figure 3. (**A**) A hierarchical cluster of all arrays using the Pearson correlation distance. Note triplicate arrays of the same type cluster together and that the A3− samples cluster together with the N3+ and A3+ samples. (**B**) Plating efficiency of normal and adapted HESCs sorted for expression of SSEA3 and re-plated as single cells; typical cloning efficiencies for unsorted populations of these normal and adapted HESCs were 0.7 and 11%, respectively. (**C**) Photomicrographs of representative colonies in cultures re-established from normal and adapted cells after sorting for expression of SSEA3.

revealed three prominent features. First, the ND and AD samples are, as expected, closely related to each other and most distinct from all other ES cell subsets analysed. Analysis of the genes upregulated in the AD samples confirms that the adapted cells exhibited a full range of differentiation (Supplementary Material). Secondly, the N3+ and A3+ samples segregate together. Thirdly, the N3− and A3− cells do not co-segregate; rather, the A3− samples cluster with the N3+ and A3+ samples, whereas the N3− sample forms a unique group. In principle, the genotypic differences between the normal and adapted cells might have weighted the clustering to some extent. However, if this was a major contributory factor, one would predict that the cells would segregate according to genotype, i.e. the A3+ and A3− cells would co-segregate and be different from the N3+ and N3− which would also co-segregate together. Nevertheless, this is clearly not the case because, despite the acquired karyotypic differences, the A3+ and A3− subsets co-segregate with the N3+ subset in the stem cell territory, whereas the N3− cells, which are genotypically identical to the N3+ cells, segregate separately and distinctly from the N3+ cells. Thus, loss of SSEA3 seems to mark an early stage in the differentiation of the normal cells, whereas this is not evident in the adapted cells. This last, and unexpected, finding suggests that adaptation in culture represents a shift in the balance between self-renewal and differentiation.

Ability to re-initiate stem cell cultures is a stringent test of stem cell status. To examine further the relationship between the SSEA3 expression and the primitive, undifferentiated state of HESC, the clonogenic capacity of SSEA3+ and SSEA3− cells was assessed (Fig. 3B). The N3− subset exhibited a 112-fold diminished cloning efficiency when compared with the N3+ subset. This difference in functionality is striking, given the modest nature of the gene expression differences between these subsets. In contrast, the A3+ cells showed a 6-fold higher cloning efficiency when compared with the N3+ cells, and this was only slightly reduced (2-fold) in the A3− subset. Morphologically, in mass culture, the colonies produced by the N3+, A3+ and A3− cells were mostly typical ES colonies, although with rather more apparent differentiation in the cultures developing from the N3+ cells (Fig. 3C). Taken together, these data suggest that there are two related cellular mechanisms that account for the dominant emergence of adapted cells in HESC

cultures: retention and enhancement of a high clonogenic potential and a reduced capacity for spontaneous differentiation under standard culture conditions revealed by analysis of the SSEA3− subsets.

Ideally, one would wish to examine further these issues in additional 'paired' samples of normal and adapted cells from a variety of independently derived human ES cell lines. Unfortunately, carefully 'curated' samples, where normal versus adapted status has been systematically documented over time in culture, are currently not available. These caveats aside, we have nevertheless compared stem cell status, as judged by clongenicity of SSEA3+ and SSEA3− cells from relatively low and high passages of another HESC line, H1 (Fig. 4). The results exhibit a trend that is broadly consistent with the hypothesis that in HESC cultures, the clonogenic stem cells are predominantly found within the SSEA3+ subpopulation, but that this distinction between SSEA3+ and SSEA3− cells becomes markedly less pronounced with progressive adaptation to culture.

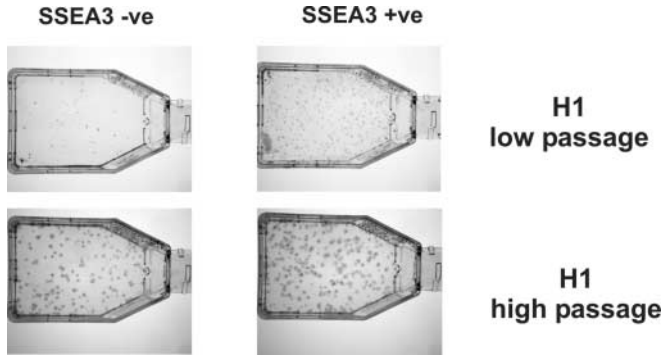


Figure 4. Colony formation from SSEA3+ and SSEA3− cells sorted from cultures of the HESC line H1 at relatively low passage (p39) and high passage (p110) levels. Following sorting, single cells (2000 per flask) were seeded into tissue culture flasks containing inactivated feeder cells: the colonies were fixed and stained with Geimsa after 2 weeks. The mean plating efficiencies were low passage SSEA3− (1.9%), SSEA3+ (13.2%), high passage SSEA3− (6.1%) and SSEA3+ (11.3%).

'Expression karyotyping'

We identified 604 genes that were more highly expressed in the A3+ versus N3+ cells; of these, 67 genes passed an additional 2-fold filter with the highest degree of regulation (7-fold) exhibited by the gene encoding the Notch ligand, *DLK1*. In total, 386 genes were down regulated with 251 genes showing more than a 2-fold decrease. Among these, *MMP1*, for example, also highlighted in our analysis of N3+ and N3− cells, exhibited a 50-fold change in level. (see Supplementary Material for upregulated and downregulated gene lists).

As the adapted cells are aneuploid, we mapped these differentially expressed genes to their chromosomal locations (Fig. 5). Visual inspection and initial statistical analysis reveal that the distribution of downregulated genes is even across all chromosomes, whereas the upregulated genes are more concentrated along the entire length of chromosome 1, on the long arm of chromosome 17 and also on the X-chromosome. The concentration of genes on chromosomes 1 and 17q is consistent with the karyotype of the adapted cells by conventional cytogenetics. However, no obvious abnormalities associated with the X-chromosome were detected (discussed subsequently).

To assess the contribution of chromosome copy number to increased gene expression in the adapted cells, we used 'cumulative expression analysis', in which the expression of each gene along the length of a given chromosome is cumulatively summed (Fig. 5). An increase in gene expression arising from an additional chromosomal copy is predicted to result in a diverging line for the adapted versus normal chromosome. Data from chromosomes 1, 17 and X are shown. The divergence of the traces from normal and adapted cells across the entire length of chromosome 1 is compared with the late divergence seen in chromosome 17 which corresponds to the long-arm region. Increased expression throughout the entire length of the X-chromosome was also seen using this approach.

We next examined whether the magnitude of the increase in gene expression associated with these genomic intervals was simply due to their increased copy number or resulted from some additional level of altered regulation. We therefore re-normalized the microarray data in a chromosome-specific manner using Fastlo (23). Cumulative expression analysis after cyclic loess correction removed the global divergence between adapted and normal cells. These data suggest that the level of overexpression of most chromosome 1 genes was consistent with an additional copy of this chromosome (Fig. 5); similar results were obtained for chromosome 17q (data not shown).

Epigenetic changes associated with adaptation

H7 is a female line with two copies of the X-chromosome, and our gene expression profiling data indicate increased expression along the entire length of X. However, no additional X-chromosomal material was detected by cytogenetics using a whole chromosome paint (Supplementary Material). These facts point to a failure in X-inactivation as a potential mechanism for the increased expression observed in the adapted cells. The array data for the expression of the *XIST* transcript, a non-coding mRNA transcript known to be involved in X-inactivation (Fig. 6) (24), show that the normal cells expressed *XIST*, whereas the adapted cells did not, even after differentiation. We confirmed these results by RT–PCR analysis of *XIST* transcripts in an independent set of sorted cells.

To confirm whether normal H7 cells have an inactive X-chromosome whereas adapted cells have both X-chromosomes active, we stained colonies of normal and adapted cells using (1) an antibody directed against a histone modification associated with the inactive X-chromosome, (2) an antibody directed against the active form of RNA polymerase II and (3) the DNA stain, Topro 3 (Fig. 6). These data show that the normal ES cells contain an inactive X-chromosome that is excluded from the region of the nucleus associated with active RNA polymerase II transcription. The inactive X-chromosome, which can also be appreciated as a Barr body, is located at either the periphery of the nucleus or within the nucleolus. In marked contrast, the adapted ES cells showed no evidence of X-inactivation.

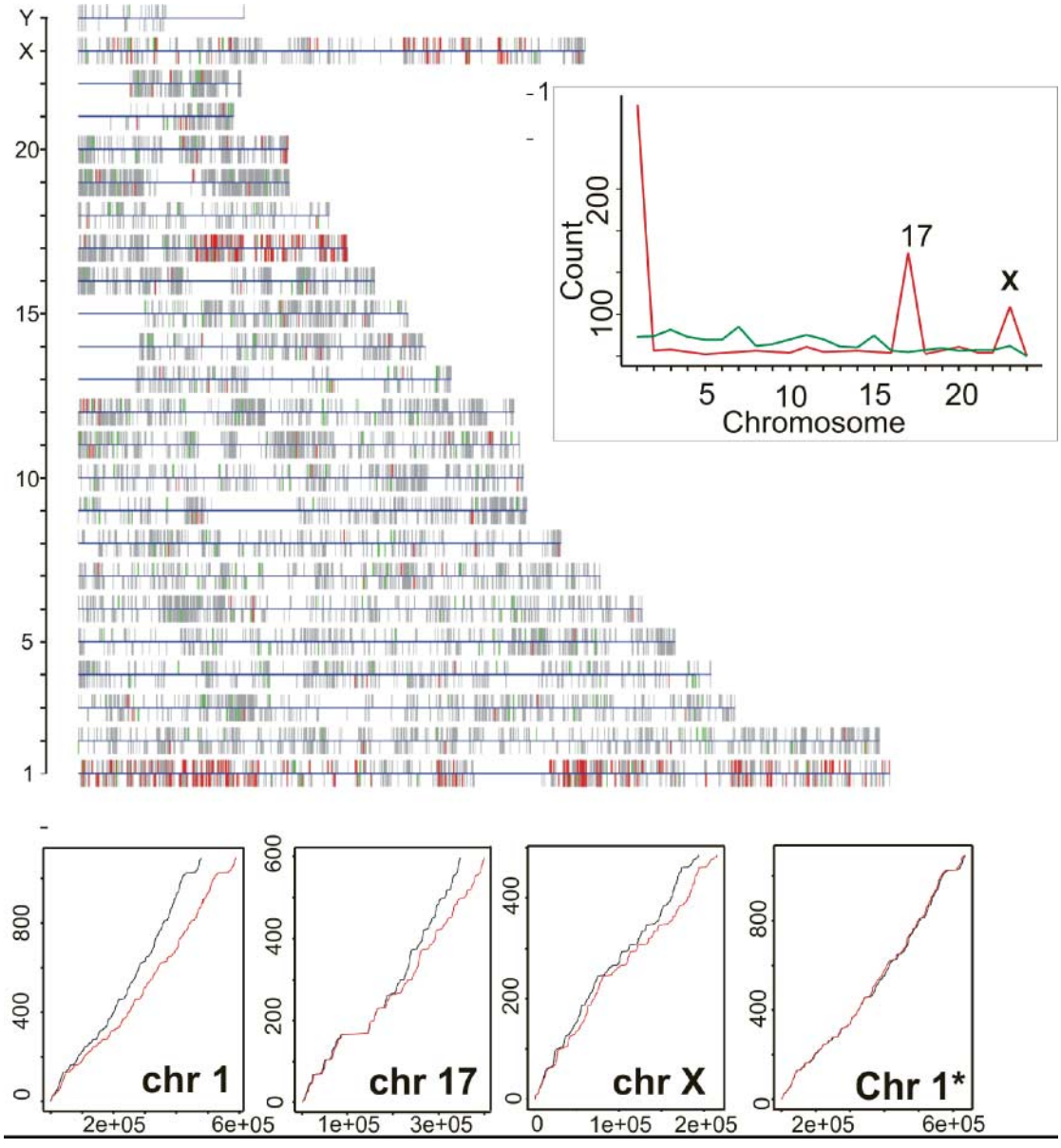


Figure 5. Differentially expressed genes between the N+ and A+ cells are plotted at their chromosomal positions (red: upregulated and green: downregulated); inset: frequencies of up- and downregulated genes for each chromosome. The lower panels show the outcome of cumulative expression analysis when the diverging lines indicate that chromosomes 1, 17 and X have an overall increased transcriptional output. The outcome of chromosome-specific normalization when applied to chromosome 1 indicates that the divergence is a dosage effect (chr 1*).

DISCUSSION

The commitment of ES cells to the alternatives of self-renewal or differentiation involves a complex interaction of multiple genes and regulatory pathways. Although the identity of some of the components of these systems is beginning to emerge (e.g. *Oct4* and *NANOG*), the ways in which they are integrated are poorly understood. Global analysis of gene expression patterns by DNA microarrays is one way to obtain insights into the nature of this integration. In our present study, we have explored the molecular basis for the cellular heterogeneity that characterizes maintenance cultures of HESC, coupling an analysis of the changes in gene expression as HESC loses expression of the surface marker antigen, SSEA3, with an analysis of the changes in the transcriptome of HESC following their adaptation to prolonged culture *in vitro*. Our results highlight the hierarchical nature of commitment in HESC cultures.

Even though the function of glycolipid antigens such as SSEA3 is unclear, and these antigens may be lost without affecting the behaviour of the cells or affecting early development (25), their expression is paradoxically, exquisitely controlled and they provide sensitive markers of cell state (13,14).

Our results show that in the relatively low passage normal HESC, only cells within the SSEA3+ subset had a true

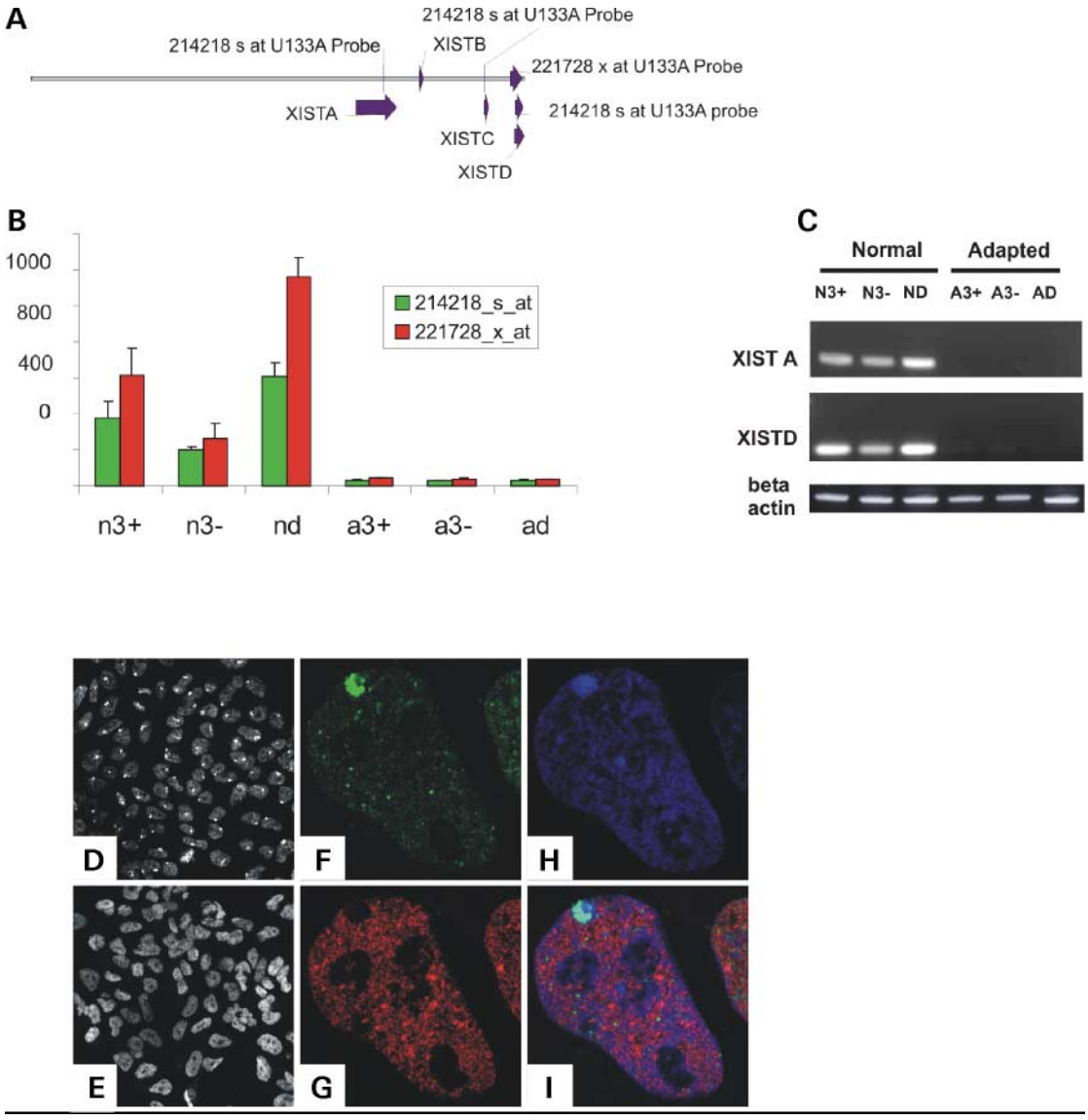


Figure 6. (**A**) Schematic representation of the human *XIST* locus. The four exons are labelled XIST A–D; the relative positions of the U133A probes used in this study are also indicated. (**B**) Expression level of XIST as detected by the Affymetrix U133A probe sets 214218_s_at and 221728_x_at. (**C**) RT–PCR for XISTA and XISTD in normal and adapted HESCs sorted for SSEA3 expression. (**D**–**I**) The presence of an inactive X-chromosome in normal HESC and its absence from adapted HESC. Staining of normal (D) and adapted (E) HESCs with an antibody against histone H3, three-methylated at lysine 27, which labels facultative heterochromatin and highlights inactive X-chromosomes. Note the localization of staining to a single body (the putative inactive X-chromosome) in the normal cells (D), but the diffuse pattern of staining in the adapted cells (E). (F) High power view of a typical normal HESC from (D) showing accumulation of three-methylated lysine 27 histone H3 associated with the inactive X-chromosome. (G) The same cell stained with antibody directed against phosphorylated serine 5 of RNA polymerase II (the active form of this molecule). Note that this active polymerase is excluded from the region surrounding the inactive X-chromosome. (H) DNA staining of the same cell using Topro 3, highlighting the Barr body. (**I**) A merged image of the three previous panels.

clonogenic stem cell capacity and thus the molecular profiling that we have performed on this population represents the most authentic signature of HESC pluripotency yet described. Comparison of this signature with that of the SSEA3− cells, obtained from the same cultures, yielded a significant number of genes that showed upregulation in the SSEA3− cells. Such genes would be incorrectly called as 'stem cell specific' if this prior fractionation of the cultures was not performed. As expected, differential gene expression between these subsets was limited in terms of gene number (124 downregulated and 397 upregulated genes), and maximal fold changes were also less than those observed between N3+ and ND. The limited downregulation (∼2-fold) of genes, such as *Oct4* or *NANOG*, that have been implicated as key regulators of ES cell fate between SSEA3+ and SSEA3− cells is consistent with the view that relatively small changes in the levels of such genes are sufficient to profoundly alter cell fate (16).

Most of the genes identified in Figure 2, notably *SEMA6A*, *FLJ10884*, *GAL*, *DNMT3B*, *GABRB3*, *CYP26A1*, *FLJ10713*, *ALPL*, *LIN28*, *TDGF1*, *POU5F1*, *NANOG* and *LEFT*, were found to be present, or called as enhanced, in the undifferentiated

pluripotent stem cell state in several previous microarrays (1,2,5–7,26), suggesting that they represent a core stem cell genetic signature. Although genes such as *TDGF1*, *POU5F1*, *NANOG* and *LEFTB* have either well characterized or predictable roles in maintaining the undifferentiated stem cell state, the roles of others, such as *GAL* and *GABRB3*, are more obscure. However, further analysis of gene expression within the different HESC subsets allowed us to identify several potentially novel candidate regulators or markers of the pluripotent, clonogenic state defined by SSEA3 expression. Several of these genes, notably *PTPRZ1*, *RPC32*, *DIAPH2 MDB2*, *LECT1* and *NTS*, were not identified in several other array studies using unsorted populations of HESC. A preliminary analysis of one of these, *RPC32*, encoding a subunit of RNA polymerase III, did indeed suggest that it is required for the maintenance of the stem cell state and raises the question of whether a specific subset of RNA transcribed by polymerase III is required to maintain the undifferentiated stem cell state.

Several studies have reported that HESC cultures can become populated with adapted cells, better able to proliferate under prevailing culture conditions (15,27,28), though others have reported that their ES lines do not develop gross karyotypic abnormalities in culture (29). This may be an intrinsic property of the cells *per se* or may relate to the methods of passage which involve physical resection of individual ES colonies as opposed to bulk passage by enzymatic dissociation (28). In any event, our results indicate that absence of karyotypic abnormalities is an insufficient index of normality. The failure of culture-adapted H7 cells to undergo X-inactivation indicates that cells can acquire epigenetic changes that may alter their biological properties.

The X-inactivation status of different HESCs is variable: in a comparison of three female XX lines, the expression of *XIST* was readily detected in two lines (H7 and H13), but not in the third, H9 (30). In another study, undifferentiated H9 cells were shown to possess two active X-chromosomes, whereas random or non-random X-inactivation followed differentiation (31). The reason for the difference between HESC lines is not known, but might depend on the X-inactivation status of the particular ICM cell that gave rise to the individual ES cell lines. Thus, H7 cells may thus have derived from an ICM cell that had already undergone lyonization. As our adapted H7 subline was derived from the initial H7 line, it must have secondarily lost X-inactivation; this is consistent with our observation of the 'mixed' colonies in the normal subline (unpublished data). Alternatively, the differences between the H7 and H9 lines might reflect the outbred nature of HESC. However, somewhat surprisingly, the adapted H7 cells also failed to undergo X-inactivation during differentiation, suggesting that the defect in the X-inactivation mechanism acquired during adaptation is not specific to the undifferentiated cells. In any event, our results indicate that further studies of the epigenetic status of normal ES cells and their differentiated derivatives are warranted.

Understanding the nature of adaptation may lend insights into the mechanisms underlying the self-renewal and differentiation of HESC. It may also be pertinent to oncogenesis and tumour progression. The majority of differences in gene expression between the normal and adapted cells could be attributed to the presence of the additional copies of chromosomes 1 and 17q revealed by cytogenetics. It is notable that in human EC cells, an extra copy of chromosome 17q is almost always noted; additional copies of chromosome 1 and the X-chromosome, sometimes with dysregulation of *XIST* expression, are also frequent (32) (33,34). By normalizing the microarray data on an individual chromosome basis, we were able to show that cells are, for the most part, unable to regulate the level of gene expression from the additional chromosome copies to maintain a normal level of gene expression. However, a few genes on these chromosomes were down regulated despite the increased dosage and a few genes located elsewhere were up regulated despite a normal diploid dosage. There are several possible mechanisms that might be responsible for adaptation. They include (1) a cumulative effect of the marginally increased expression of a large number of genes from the additional genomic intervals (2) an effect due to a relatively marginal increase in only a single or a few genes—this possibility has precedent in the acute dosage dependence seen with genes such as *Oct4* or (3) an effect only of the few more dramatically upregulated genes, i.e. whose increase in expression is significantly more than can be accounted for by the additional genomic copy. In the latter cases, most of the changes in gene expression would be the result of a 'hitch-hiker' effect due to linkage to the key genes located on chromosomes 17q, 1 and X or due to being downstream targets of genes located on those chromosomes.

A comparison of the adapted and normal HESCs did yield a number of candidate genes that may be important for the process of adaptation. Most notable was *Dlk1*, which encodes a Notch ligand, the expression of which was substantially higher in the A3+ when compared with the N3+ cells. The Notch pathway is implicated in several stem cell systems, and our current investigation of the function of the Notch pathway in HESC suggests a role for this signalling system in HESC proliferation and self-renewal (manuscript in preparation). The other notable differentially expressed gene was that encoding matrix metalloproteinase 1, which was down regulated. Several growth factors and molecules bind the extracellular matrix, so that differential expression of an extracellular proteinase could affect the nature of the extracellular matrix and therefore, indirectly, influence any intercellular signalling that mediates community effects and promotes stem cell maintenance.

Our analysis of the relatedness of different subsets of normal and adapted HESCs leads us to propose a hierarchical model to describe HESC self-renewal and differentiation in culture (Fig. 7). This hierarchy has parallels with those described in normal and malignant somatic stem cell systems (35). In our model, N3− cells, represent cells that have exited the stem cell compartment, but retain multilineage differentiation potential. The possibility exists that these cells may, at low probability, have the capacity of re-entering the stem cell compartment, reflected by their low clonogenic capacity. Terminally differentiated cells exist outside the stem and progenitor cell territories. The positioning of adapted HESC within this landscape is particularly informative. The A3+ and AD cells segregate with their normal counterparts. However, the A3− cells fall into the stem cell

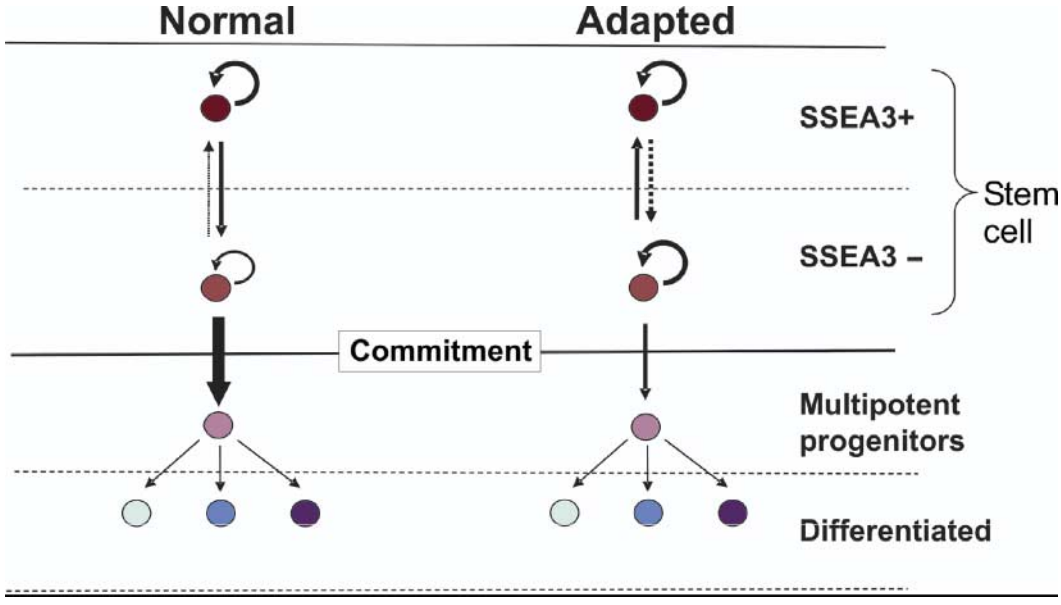


Figure 7. A schematic representation of the relationship between SSEA3+ and SSEA3− cells and their relationships to self-renewal and differentiation in normal and adapted HESCs. In this scheme, loss of SSEA3 expression does not necessarily equate to commitment to differentiate; such an SSEA3− cell can proceed to a commitment step and subsequently differentiate or revert to the SSEA3+ compartment. It is envisaged that the normal and adapted cells differ in the relative probabilities of these alternatives; i.e. the normal cells have a substantially higher probability of continuing to commitment, whereas the adapted cells have a higher probability of reverting to an SSEA3+ state.

territory. This implies that adaptation reflects an alteration in the balance between self-renewal and differentiation. Such a change is an essential feature of cancer (36). There are therefore likely to be strong similarities between the mechanisms of oncogenesis and adaptation, indicating that the study of HESC adaptation may shed light not only on embryonic stem cell biology, but also provide a useful model for assessing the changes that occur during transformation of cells to a tumourigenic phenotype.

MATERIALS AND METHODS

Cell culture

The HESC lines, H7, H1 and H14 (10), were cultured in knockout DMEM with 20% serum replacement (Invitrogen, Carlsbad, CA, USA) and 4 ng/ml basic fibroblast growth factor, on mouse embryonic fibroblast feeders, was inactivated with mitomycin C (12). The adapted aneuploid subline of H7 was described previously (15).

To perform plating efficiency experiments, the HESCs were either harvested as single cell suspensions using 0.05% trypsin and 1 mM EDTA or seeded from single cell suspensions resulting from sorting using the flow cytometer.

Preparation of RNA from SSEA3 subsets of HESC

Live cells were harvested from stock cultures of H7 and sorted by fluorescence activated cell sorting as described previously (9) after staining with monoclonal antibody MC631 which recognizes SSEA3 (37). Myeloma antibody P3X63Ag8 (38) was used as the negative control in the indirect immunofluorescence. The MC631 antibody used was concentrated hybridoma supernatant, pre-titred to ensure maximal binding to target cells. Total cellular RNA was isolated with TRIzol (Invitrogen) and used in the preparation of labelled targets which were hybridized to Affymetrix U133A genechips as described previously (39).

Global gene expression analysis

Intensities were extracted from CEL files using RMA (40), which is implemented in the 'affy' package of the Bioconductor (41) suite written in the R (42) language. For comparisons between two cell types, the data were pre-filtered to remove any genes with an expression level less than 100 over all six arrays. Two-class SAMs (43) was performed using the 'siggenes' package, and those genes with a false discovery rate (FDR)=0 were called differentially expressed. Differentially expressed genes were mapped to their chromosomes using the 'geneplotter' package. Where all arrays from all types were analysed together, a multiclass SAM was performed using 1000 permutation rounds.

Hierarchical clustering and heat map were created in Genesis (44) using the Pearson correlation distance with average linkage. The arrays were clustered using the Pearson correlation distance with complete linkage.

Cumulative expression analysis was implemented by isolating the intensities from one array for one chromosome and rearranging the genes of the positive strand, therefore the order represents the sequence in which the genes appear on the chromosome. The cumulative sum of the resulting vector was then calculated. To normalize the intensities in a

chromosome-specific manner, the data were first extracted using RMA, but omitting any normalization. The data set was then grouped by chromosome membership and each of the sets individually normalized using Fastlo normalization (23).

RT–PCR

Production of cDNA and RT–PCR was carried out as described previously (17). The primer sequences are shown in the Supplementary Material.

Immunofluorescent staining of cells *in situ*

HESCs on cover slips were fixed in 4% paraformaldehyde and immunolabelled (45) using a primary rabbit antibody to trimethyl-histone H3 (lys27) (Upstate, Lake Placid, NY, USA) and a mouse antibody to the phosphoserine 2 on the C-terminal domain of RNA polymerase II (H5) (Covance, Berkeley, CA, USA) (this antibody recognizes the elongating form of the enzyme). Secondary antibodies were donkey anti-mouse IgG tagged with Cy3 (Jackson ImmunoResearch, West Grove, PA, USA) and donkey anti-rabbit IgG tagged with Alexa 488, prepared using a Molecular Probes' kit. DNA was counterstained with Topro 3 (Molecular Probes, Eugene, OR, USA).

RNA interference

Double-stranded, short (21mer) interfering RNA (siRNA) corresponding to ß2-microglobulin (*B2M*) and *RPC32* was designed with the following sense and anti-sense sequences and synthesized by Qiagen (Valencia, CA, USA): *RPC32* (sense) 5′-CCAGUACCACUGAAAACAGdTdT-3′ and (anti-sense) 5′-CUGUUUUCAGUGGUACUGGdTdT-3′; *B2M* (sense) 5′-GAUUCAGGUUUACUCACGUdTdT-3′ and (anti-sense) 5′-ACGUGAGUAAACCUGAAUCdTdT-3′.

DsRNA was delivered to cells as described previously (17).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

T.E. acknowledges the MRC, LRF and Wellcome Trust for funding. P.W.A. was supported by grants from BBSRC, YCR, MRC and JDRF. 'Funding to pay the Open Access publication charges for this article was provided by BBSRC and MRC'.

Conflict of Interest statement. None declared.

REFERENCES

1. Abeyta, M.J., Clark, A.T., Rodriguez, R.T., Bodnar, M.S., Pera, R.A. and Firpo, M.T. (2004) Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum. Mol. Genet.*, **13**, 601–608.
2. Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., Luo, Y., Yang, A.X., Joshi, B.H., Ginis, I., Thies, R.S., Amit, M. *et al.* (2004) Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood*, **103**, 2956–2964.
3. Dvash, T., Mayshar, Y., Darr, H., McElhaney, M., Barker, D., Yanuka, O., Kotkow, K.J., Rubin, L.L., Benvenisty, N. and Eiges, R. (2004) Temporal gene expression during differentiation of human embryonic stem cells and embryoid bodies. *Hum. Reprod.*, **19**, 2875–2883.
4. Rao, R.R. and Stice, S.L. (2004) Gene expression profiling of embryonic stem cells leads to greater understanding of pluripotency and early developmental events. *Biol. Reprod.*, **71**, 1772–1778.
5. Sato, N., Sanjuan, I.M., Heke, M., Uchida, M., Naef, F. and Brivanlou, A.H. (2003) Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev. Biol.*, **260**, 404–413.
6. Brandenberger, R., Khrebtukova, I., Thies, R.S., Miura, T., Jingli, C., Puri, R., Vasicek, T., Lebkowski, J. and Rao, M. (2004) MPSS profiling of human embryonic stem cells. *BMC Dev. Biol.*, **4**, 10.
7. Richards, M., Tan, S.P., Tan, J.H., Chan, W.K. and Bongso, A. (2004) The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells*, **22**, 51–64.
8. Furusawa, T., Ohkoshi, K., Honda, C., Takahashi, S. and Tokunaga, T. (2004) Embryonic stem cells expressing both platelet endothelial cell adhesion molecule-1 and stage-specific embryonic antigen-1 differentiate predominantly into epiblast cells in a chimeric embryo. *Biol. Reprod.*, **70**, 1452–1457.
9. Henderson, J.K., Draper, J.S., Baillie, H.S., Fishel, S., Thomson, J.A., Moore, H. and Andrews, P.W. (2002) Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens. *Stem Cells*, **20**, 329–337.
10. Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S. and Jones, J.M. (1998) Embryonic stem cell lines derived from human blastocysts. *Science*, **282**, 1145–1147.
11. Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A. and Bongso, A. (2000) Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nat. Biotechnol.*, **18**, 399–404.
12. Draper, J.S., Pigott, C., Thomson, J.A. and Andrews, P.W. (2002) Surface antigens of human embryonic stem cells: changes upon differentiation in culture. *J. Anat.*, **200**, 249–258.
13. Fenderson, B.A., Andrews, P.W., Nudelman, E., Clausen, H. and Hakomori, S. (1987) Glycolipid core structure switching from globo- to lacto- and ganglio-series during retinoic acid-induced differentiation of TERA-2-derived human embryonal carcinoma cells. *Dev. Biol.*, **122**, 21–34.
14. Andrews, P.W., Casper, J., Damjanov, I., Duggan-Keen, M., Giwercman, A., Hata, J., von Keitz, A., Looijenga, L.H., Millan, J.L., Oosterhuis, J.W. *et al.* (1996) Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours. *Int. J. Cancer*, **66**, 806–816.
15. Draper, J.S., Smith, K., Gokhale, P., Moore, H.D., Maltby, E., Johnson, J., Meisner, L., Zwaka, T.P., Thomson, J.A. and Andrews, P.W. (2004) Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat. Biotechnol.*, **22**, 53–54.
16. Niwa, H., Miyazaki, J. and Smith, A.G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.*, **24**, 372–376.
17. Matin, M.M., Walsh, J.R., Gokhale, P.J., Draper, J.S., Bahrami, A.R., Morton, I., Moore, H.D. and Andrews, P.W. (2004) Specific knockdown of Oct4 and {beta}2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. *Stem Cells*, **22**, 659–668.
18. Burdon, T., Smith, A. and Savatier, P. (2002) Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol.*, **12**, 432–438.
19. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, **113**, 643–655.
20. Wang, Z. and Roeder, R.G. (1997) Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev.*, **11**, 1315–1326.
21. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
22. Shmelkov, S.V., St Clair, R., Lyden, D. and Rafii, S. (2005) AC133/CD133/prominin-1. *Int. J. Biochem. Cell. Biol.*, **37**, 715–719.

23. Ballman, K.V., Grill, D.E., Oberg, A.L. and Therneau, T.M. (2004) Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics*, **20**, 2778–2786.
24. Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A. and Panning, B. (2002) Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.*, **36**, 233–278.
25. Fenderson, B.A., Ostrander, G.K., Hausken, Z., Radin, N.S. and Hakomori, S. (1992) A ceramide analogue (PDMP) inhibits glycolipid synthesis in fish embryos. *Exp. Cell Res.*, **198**, 362–366.
26. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C. and Melton, D.A. (2002) 'Stemness': transcriptional profiling of embryonic and adult stem cells. *Science*, **298**, 597–600.
27. Inzunza, J., Sahlen, S., Holmberg, K., Stromberg, A.M., Teerijoki, H., Blennow, E., Hovatta, O. and Malmgren, H. (2004) Comparative genomic hybridization and karyotyping of human embryonic stem cells reveals the occurrence of an isodicentric X chromosome after long-term cultivation. *Mol. Hum. Reprod.*, **10**, 461–466.
28. Mitalipova, M.M., Rao, R.R., Hoyer, D.M., Johnson, J.A., Meisner, L.F., Jones, K.L., Dalton, S. and Stice, S.L. (2005) Preserving the genetic integrity of human embryonic stem cells. *Nat. Biotechnol.*, **23**, 19–20.
29. Buzzard, J.J., Gough, N.M., Crook, J.M. and Colman, A. (2004) Karyotype of human ES cells during extended culture. *Nat. Biotechnol.*, **22**, 381–382, author reply 382.
30. Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O. and Thomson, J.A. (2003) Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc. Natl Acad. Sci. USA*, **100**, 13350–13355.
31. Dhara, S.K. and Benvenisty, N. (2004) Gene trap as a tool for genome annotation and analysis of X chromosome inactivation in human embryonic stem cells. *Nucleic Acids Res.*, **32**, 3995–4002.
32. Wang, N., Perkins, K.L., Bronson, D.L. and Fraley, E.E. (1981) Cytogenetic evidence for premeiotic transformation of human testicular cancers. *Cancer Res.*, **41**, 2135–2140.
33. Skotheim, R.I., Monni, O., Mousses, S., Fossa, S.D., Kallioniemi, O.P., Lothe, R.A. and Kallioniemi, A. (2002) New insights into testicular germ cell tumorigenesis from gene expression profiling. *Cancer Res.*, **62**, 2359–2364.
34. Kawakami, T., Okamoto, K., Sugihara, H., Hattori, T., Reeve, A.E., Ogawa, O. and Okada, Y. (2003) The roles of supernumerical X chromosomes and XIST expression in testicular germ cell tumors. *J. Urol.*, **169**, 1546–1552.
35. Warner, J.K., Wang, J.C., Hope, K.J., Jin, L. and Dick, J.E. (2004) Concepts of human leukemic development. *Oncogene*, **23**, 7164–7177.
36. Reya, T., Morrison, S.J., Clarke, M.F. and Weissman, I.L. (2001) Stem cells, cancer, and cancer stem cells. *Nature*, **414**, 105–111.
37. Shevinsky, L.H., Knowles, B.B., Damjanov, I. and Solter, D. (1982) Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells. *Cell*, **30**, 697–705.
38. Kohler, G., and Milstein, C. (1975) Continous cultures of fused cells secreting antibody of predefined specificity. *Nature*, **256**, 495–497.
39. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.
40. Bolstad, B.M., Irizarry, R.A., Astrand, M. and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
41. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
42. Ihaka, R. and Gentleman, R.C. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.
43. Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
44. Sturn, A., Quackenbush, J. and Trajanoski, Z. (2002) Genesis: cluster analysis of microarray data. *Bioinformatics*, **18**, 207–208.
45. Iborra, F.J., Jackson, D.A. and Cook, P.R. (2001) Coupled transcription and translation within nuclei of mammalian cells. *Science*, **293**, 1139–1142.