

# Combined microarray analysis uncovers self-renewal related signaling in mouse embryonic stem cells

Hao Zhu · He Yang · Markus R. Owen

Received: 23 October 2007 / Revised: 9 April 2008 / Accepted: 19 April 2008 / Published online: 8 May 2008  
© The Author(s) 2008

**Abstract** Due to the limited understanding of self-renewal and pluripotency related signaling in stem cells, extracting information from genome-wide expression data is not only important but also challenging. With the combined use of two methods, we analyzed a set of microarray data at 11 time points from three mouse embryonic stem cell lines cultivated with and without leukemia inhibitory factor (LIF) for 14 days. *Albeit* the expression of individual genes in signaling pathways was not noticeably different between cells cultivated with and without LIF, at gene-set level the expression of ERK/MAPK (but not JAK/STAT) and cell cycle related genes was found significantly enriched in cells cultivated with LIF. This indicates that the Ras/Raf/ERK pathway, in addition to JAK/STAT, may also be a key player to carry on external LIF signal into mouse embryonic stem cells to promote self-renewal. When data at the first 7 time points were compared with data at the last 4 time points, the expression of several cell cycle related gene sets was apparently enriched in all three cell lines, indicating the active cell proliferation in the first 2 days. Compared with the slight decay of Oct4/Nanog/Sox2 during the 14 days, the expression of cell differentiation genes such as Gata4/6 underwent a drastic increase, which indicates that the upregulated expression of cell

differentiation genes may better reflect the loss of self-renewal than the down regulated expression of the stemness indicators Oct4, Sox2 and Nanog. Apart from differential expression and gene set enrichment analyses, a clustering algorithm was also used to classify genes into co-expression clusters. The possible regulation of two clusters, whose expression was most changed during cell culture from very low to very high, was explored. The drastic changes of these genes, including Slc39a8 which was a potential indicator of cell differentiation, in contrast the slight changes of self-renewal genes, imply that differentiation may be the default fate of stem cells and self-renewal may rely on a maintenance mechanism. When that mechanism weakens, cell differentiation begins.

**Keywords** Cell cycle · LIF · Microarray · Self-renewal · Embryonic stem cell

## Introduction

Persistent self-renewal is a common feature of both stem cells and cancer cells, although whether they employ common mechanisms to implement self-renewal is not clear. There exist extrinsic and intrinsic mechanisms that work in parallel to promote stem cell self-renewal. In mouse embryonic stem (mES) cells leukemia inhibitory factor (LIF), through the LIF receptor (LIFR) and gp130, activates the JAK/STAT pathway (Smith et al. 1988; Williams et al. 1988; Yoshida et al. 1994; Niwa et al. 1998) (Fig. 4). Afterwards the phosphorylated and dimerized STAT3 is translocated to the nucleus to stimulate self-renewal gene expression. In human embryonic stem (hES) cells, however, LIF appears not to function but ligands of the fibroblast growth factor (FGF) family are required

**Electronic supplementary material** The online version of this article (doi:10.1007/s11693-008-9015-2) contains supplementary material, which is available to authorized users.

H. Zhu (✉) · M. R. Owen  
Center for Mathematical Medicine and Biology, School of  
Mathematical Sciences, University of Nottingham, Nottingham  
NG7 2RD, UK  
e-mail: Hao.Zhu@nottingham.ac.uk

H. Yang  
Bioinformatics Institute of Singapore, Singapore, Singapore

(Thomson et al. 1998; Amit et al. 2000; Daheron et al. 2004). On the other hand, the intrinsic mechanisms, centered at Nanog and Oct4, are more conserved in species and independent of STAT3 (Chambers et al. 2003; Mitsui et al. 2003; Catena et al. 2004; Okumura-Nakanishi et al. 2005). Apart from these extra- and intra-cellular factors, their downstream targets and signaling processes are important. Recent studies reveal that the JAK/STAT pathway may not be the sole player in self-renewal signaling in mES. In the complete absence of LIF, self-renewal is not abolished and undifferentiated mES cell colonies are still obtained (Dani et al. 1998). The possible mechanism is that a co-cultivated, differentiated, and LIF-deficient cell line provides a paracrine factor supporting mES self-renewal. Moreover, intrinsic self-renewal factors like Nanog and Oct4 may play a more fundamental role than extrinsic factors such as LIF and FGF. To reveal how these and other possible genes and pathways, collaborative with or independent of JAK/STAT, contribute to self-renewal of stem cells under different conditions is important for the understanding of stem cell self-renewal mechanisms.

Self-renewal involves promoting cell proliferation and inhibiting cell differentiation. The two may demand coordinated actions of different pathways. A reasonable approach to expanding the understanding of self-renewal is to screen potential proliferation and differentiation related genes in cells under different conditions and at different stages, and furthermore, to identify their function and organization. In this regard, microarray experiments and data analysis are of great help due to the features of high-throughput and genome-wide scope (Sekkaï et al. 2005). We analyzed a set of temporarily dynamic microarray data at 11 time points from 3 mES cell lines for 14 days, 2 cultivated with and 1 without LIF. We found in this dataset that despite the expression of individual genes, including the positive and negative indicators of self-renewal, did not differ significantly in cells cultivated with and without LIF, at signaling module and pathway level the expression of several sets of genes was identified to be much enriched, either particularly in cells cultivated with LIF or predominantly in all cells at the early or late stage of cultivation. These gene sets, involved in Ras/Raf/ERK and cell cycle signaling, suggest that, in addition to the well documented JAK/STAT pathway, Ras/Raf/ERK may be another pathway that transfers external signals to promote mES cell self-renewal (Fig. 4), and that active cell cycle signaling is a key feature of self-renewal. As significant cell cycle activities showed in all 3 cell lines in the first 2 days during which the expression of Nanog/Oct4/Sox2 remained high, these self-renewal factors must play a significant role in promoting cell proliferation independent of LIF. The difference in the expression of cell cycle related gene sets was

more significant at different times than in different cells, and significant down regulation occurred quite accurately after 2 days of culture. Compared with self-renewal and cell cycle genes, the expression of cell differentiation related genes such as Gata4/6 showed a more marked change during the 14 days. This implies that differentiation may be the default cell fate and self-renewal should rely on a maintenance mechanism. As soon as the mechanism becomes even slightly wane, cell differentiation begins.

## Materials and methods

The description of the cell culture and microarray experiment is based on the documentation of the publicly available data; more details can be found at <http://www.scgp.ca:8080/StemBase> (Perez-Iratxeta et al. 2005).

### Cell culture

Three murine embryonic stem cell lines—V6.5, R1 and J1—were cultivated for 14 days in the presence (V6.5 and R1) and absence (J1) of LIF. In V6.5 and R1 LIF was only given at the start of cultivation. High glucose DMEM (Hyclone FCS, non-essential amino acids, Glutamax, Penicillin/Streptomycin, sodium pyruvate, and beta-mercaptoethanol) was equally supplied to the three cell lines. A DR4 cell line cultured also in DMEM (heat inactivated Hyclone FBS, Glutamax, sodium pyruvate, non-essential amino acids, Penicillin/Streptomycin, beta-mercaptoethanol, LIF), a sample of mouse embryonic fibroblasts in triplicate and acting as feeder cells, was used as the control for all the J1, R1 and V6.5 mES samples.

### Microarray experiment

To avoid any potential variability between different microarray platforms, data of all three cell lines at 11 time points (0, 6, 12, 18, 24, 36, 48 h and 4, 7, 9, 14 days) were generated using Affymetrix MOE430. Probe set hybridization values were generated using the MAS5.0 algorithm. Except one sample in duplicate, all other samples were in triplicate. RNA preparation was made with RNeasy. Three data sets (E113 for V6.5, E165 for R1, and E201 for J1) were generated.

### Data selection and preparation

To avoid any variability between different chips, only data of MOE430A were used and analyzed. The probe sets that do not correspond to any known genes were removed. For genes with multiple probe sets, the probe set with the maximal expression value was chosen. The total probe sets

were therefore reduced from 22691 to 11222 for the data analysis. The whole dataset contains 33 samples, 11 for each of the V6.5, R1 and J1 cell lines at the correspondent 11 time points. Sample data are in two forms: original data (The value of each sample is averaged by the sample's all replicates) and log-ratio data (the value of the original data is further treated by comparison with the control using the following equation:

$$\text{value} = \log(\text{original\_sample\_data} / \text{original\_control\_data}) / \log(2).$$

#### Gene set enrichment analysis

A recently reported method (Subramanian et al. 2005) was applied to the 11222 probes in each sample to identify gene sets with an enriched expression in particular samples or at particular time points. 191 gene sets were organized, which were collected from SuperArray (<http://www.superarray.com>), extracted from the signaling pathway databases BioCarta (<http://www.biocarta.com>), GenMapp (<http://www.genmapp.org>) and KEGG (<http://www.genome.ad.jp/kegg/pathway.html>), and curated from published papers. The full list of gene sets is available online (Supplemental File 1). In the gene set enrichment analysis program (Subramanian et al. 2005), the *Permutation number* was set to 1000; *Phenotype permutation* was used; *Enrichment statistic* was weighted; and the *Metric* for ranking genes was *Signal2Noise*. Only gene sets with more than 10 genes were computed in the analysis. As the method compares two data sets each time, the control was not required and the original sample data were used.

#### Clustering

A revised clustering method, which not only classifies genes into clusters but also gives the possible gene regulation by a set of regulators (Segal et al. 2003), was used to classify differentially expressed genes into co-expression clusters and to identify the possible regulation and regulators of each cluster. As the method deals with different data together, the log-ratio sample data all compared with the control were used. If a gene in any of the three cell lines had  $\geq 3$  samples (i.e., at any  $\geq 3$  time points) that had a fold change  $\geq 2.0$  compared with the control sample acquired from the DR4 cell line, it was considered differentially expressed and chosen as a record for clustering. This criterion produced 2135 records from the 11222 records, which is a reasonable size for this method. The ligands, receptors and effectors in known signaling pathways were also selected regardless their fold change. Together a list of 2257 genes was made. Among them, 26 transcriptional factors (Table S2) that showed the most significant change in expression during the 14 days according to their

enrichment score computed by the gene set enrichment analysis were used as the possible regulators to check whether “regulator X regulates cluster Y under condition W” (Segal et al. 2003).

## Results

Differentially expressed genes in known signaling pathways were not noticeably different in cells cultivated with and without LIF

Assuming that LIF would significantly enhance the activity of some pathways, we first examined gene expression in the cells cultivated with and without LIF. Differentially expressed genes were identified according to their fold change compared with the control (see “Clustering” in Materials and methods). 2135 genes were differentially expressed in the two LIF-cultivated cell lines and 1716 genes in the one non-LIF-cultivated cell line (see Materials and methods). Of them, 1607 were shared by both. We then checked differentially expressed genes in known signaling pathways and functional modules, including those stemness specific genes (11 activated and 12 repressed, denoted as Boyer+ and Boyer– in Boyer et al. 2005). Unexpectedly, the expression of genes in these pathways and modules seemed not to be noticeably different (Table 1). Specifically, all the 7 positive indicators (Boyer+) and 3 of 4 negative indicators (Boyer-) of self-renewal were differentially expressed in all cell lines. Also, Nanog, Oct4 and Sox2 did not show a clear, LIF-dependent expression pattern (Fig. 1). These results suggest that LIF might not be the vital or sole factor for self-renewal under this culture condition. As conjectured (Dani et al. 1998), other factors may exist and play a role.

#### LIF stimulated cell cycle activities via ERK/MAPK pathway

Analogous gene expression does not exclude differential signaling in cells. However, to construe differential signaling from the expression of individual genes is difficult and unreliable. Although the expression of individual genes gives only limited and separate information, the expression of a set of correlated genes in a functional module or signaling pathway gives richer and more reliable evidence. This is the basis of gene set enrichment analysis (Subramanian et al. 2005), which demands an inclusive collection and appropriate organization of gene sets. Based on multiple sources, including recently published papers, our gene-set list consists of 191 gene sets which cover genes in all known signaling pathways (Supplemental File 1, see also Materials and methods).

**Table 1** Differentially expressed genes in cell lines cultivated with and without LIF

		Differentially expressed in E113 and E165	Differentially expressed in E201	Shared by E113, E165 and E201
1	Total genes	2135	1716	1607
2	Wnt pathway	41	35	33
3	TGF $\beta$ pathway	25	22	21
4	Notch pathway	3	2	2
5	MAPK pathway	58	49	47
6	JAK pathway	32	21	21
7	Hh pathway	6	3	3
8	Cell cycle	18	18	18
9	Apoptosis	16	16	14
<sup>a</sup>	Boyer+	7	7	7
	Boyer-	4	3	3
2–11	Total pathways genes <sup>a</sup>	170	144	138

<sup>a</sup> Some genes are in multiple pathways. The list of total pathway genes is given in Supplemental File 2

We first divided the data of 33 samples into two groups and, with these gene sets, investigated whether there were differences in signaling in cells cultivated with and without LIF. When E113 and E165 data were compared with E201 data, four gene sets were found to show an enriched expression with significantly small nominal (NOM) *p*-values and false discovery rate (FDR) *q*-values (Table 2). Of them, two were in the ERK/MAPK pathway, one (numbered <1501> and including Creb1, Dusp6, Bad, Hras1, Map3k8, Braf, Raf1, Grb2, Eif4e, Rasa1, Mknk2, Csnk2a1, Csnk2b, Atf1, Nras, Jun, Kras, Rap1a, Mapk3, Shc1, Mapk1, Araf, Mknk1, Map2k1, Map3k1, Elk1) from SuperArray and the other (numbered <1904> and including Map2k2, Myc, Hras1, Eras, Raf1, Nras, Kras, Elk4, Mapk3, Mapk1, Mras, Map2k1, Rras, Elk1) from KEGG (here and henceforth genes underlined had a truly enriched expression and contributed to the enrichment score). This indicates that, in addition to JAK/STAT, Ras/Raf/ERK may be another LIF-stimulated intracellular signaling pathway in mES. This result differs from the previous understanding that ERK antagonizes ES self-renewal (Burdon et al. 2002) and also from the report that ERK phosphorylation is dispensable for the regulation of cyclin D1 and for the progression from G1 to S phase in mES cells (Jirmanova et al. 2002). The other two enriched gene sets were cell cycle related, one (numbered <3201> and including Ccnd1, Ccne2, Myc, Cdk2ap1, Ccne1, Cdk4, E2f3, Cdc25a, Ccnd2, Ccnd3, Cdk2, Cdk2ap2, E2f1, Cdk6, Pim1) directly coming from a recent review (Burdon et al. 2002), confirming the hypothesis that cell cycle control is different in differentiated and mES cells, and the other (numbered <2306> and including Ccne2, Cdk4, E2f3, Sfn, Jun, Kras, Esr1, Aatf, E2f1, Brca2) from SuperArray. The <3201> gene set shows that the LIF induced cell cycle activities, via c-myc and Rb/E2F respectively, were strengthened in E113 and E165. Two overlapping enriched gene sets in each of the two pathways strongly indicate that

strengthened Ras/Raf/ERK signaling and cell cycle activities are an important feature of self-renewal in mES cells cultivated with LIF (Fig. 4).

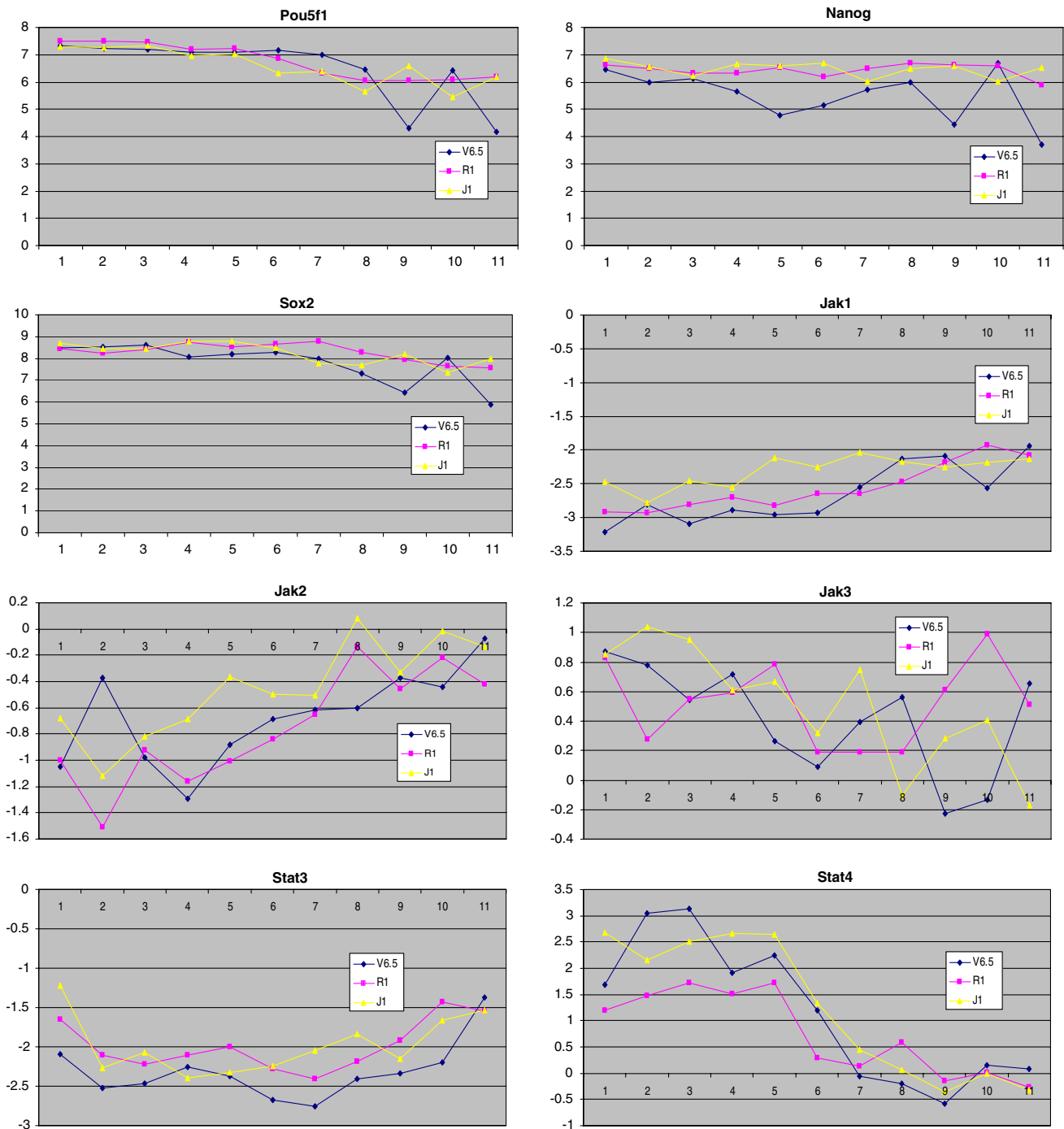
Unexpectedly, the JAK/STAT pathway, as well as other JAK related ones such as JAK/PI3K/Akt, did not appear in the list of enriched gene sets. We checked all JAK and STAT genes and found neither JAK nor STAT3 showed significantly different expression in different cell lines and at different time points (Fig. 1). Instead, JAK and STAT3 expression was even higher both in the J1 cell line and at later time points. STAT4, by contrast, showed an apparently changed expression over time (Fig. 1). These results, not previously reported in detail, revise our understanding of the function of the LIF/JAK/STAT pathway in the maintenance of stem cell self-renewal.

We also inversely compared E201 data with E113 and E165 data, and no gene sets with a significantly low NOM *p*-value and FDR *q*-value (<0.25) were found (Subramanian et al. 2005). This means no gene sets had strengthened expression in J1 cells cultivated without LIF.

#### Gene set expression varies more significantly in cells at different time points

As microarray experiments were taken at 11 time points during the 14 days, we then checked the expression of different gene sets at different time points, regardless of cell lines. We combined the data of all three cell lines (data of single cell line, due to the small sample size, gave unconvincing results) at the first 7 time points (0–48 h) into one group, and the data at the last four time points (4–14 days) into another. Comparing the first group with the second produced enriched gene sets with very low (more convincing) NOM *p*-values and FDR *q*-values (Table 3).

The first gene set (numbered <3205> and including Lefty1, Fgf4, Nodal, Sox2, TdGF1, Utf1, Rex2, Foxd3, Pax9, Pax8, Pax3, Pax7, Pax4, Pax6, Pax1, Pax5, Rex3)



**Fig. 1** The expression of Pou5f1, Nanog, Sox2, JAK1/2/3 and STAT3/4 (against the control) in the 14 days in three cell lines. Only Stat4 showed a clear down regulation

indicates the existence of a self-renewal signaling feedback in mES cells at the early time points (Rao 2004). Also enriched in the first 2 days was the molecular signature of stemness (numbered <3203> and including Pou5f1, Set, Smarcd1, Sox2, Skil, Rest, Zic3, Nanog, Hesx1, Myst3, STAT3) (Boyer et al. 2005), in agreement with the suggested positive self regulation of these genes (Kuroda et al.

2005; Rodda et al. 2005). The two gene sets together illustrate that the three cell lines in the first 2 days maintained their stemness and underwent significant self-renewal.

The remaining six significantly enriched cell cycle related gene sets, all from SuperArray, suggest active cell proliferation in the first 2 days. Set <1201> (including

**Table 2** The enriched gene sets in cells cultivated with LIF

	Gene set	Description	Size	ES <sup>a</sup>	NES <sup>b</sup>	NOM p-val	FDR q-val
1	[1501]	ERK Pathway	26	0.49	1.69	0.013	0.268
2	[1904]	ERK Pathway	14	0.57	1.64	0.014	0.230
3	[3201]	Cell cycle	15	0.62	1.62	0.020	0.183
4	[2306]	Cell cycle	10	0.66	1.59	0.014	0.175
5	[1703]		24	0.41	1.41	0.095	0.607
6	[1406]		33	0.47	1.33	0.130	0.788
7	[1301]		34	0.36	1.31	0.128	0.786
8	[2502]		13	0.50	1.30	0.128	0.722
9	[2003]		22	0.39	1.24	0.167	0.821
10	[1901]		18	0.40	1.17	0.289	0.998

<sup>a</sup> ES: Enrichment score<sup>b</sup> NES: Normalized enrichment score**Table 3** The enriched gene sets in all cell lines at the first 7 time points

	Gene set	Description	Size	ES	NES	NOM p-val	FDR q-val
1	[3205]	Self-renewal feedback genes	17	0.74	1.92	0.000	0.012
2	[3203]	Active targets of Oct, Nanog, Sox2	11	0.84	1.70	0.002	0.103
3	[1201]	Cell cycle control & DNA Damage Repair	18	0.61	1.62	0.019	0.158
4	[2305]	Cell cycle	10	0.69	1.61	0.011	0.122
5	[1303]	DNA Damage Checkpoint/p53 and ATM	23	0.55	1.60	0.009	0.110
6	[1404]	M phase	31	0.60	1.59	0.035	0.106
7	[1402]	S phase and DNA replication	14	0.88	1.57	0.000	0.108
8	[1302]	S phase	12	0.58	1.46	0.098	0.225
9	[2308]		13	0.53	1.40	0.097	0.297
10	[1406]		33	0.49	1.38	0.125	0.315

Trp53, Ccne1, Chek2, Brca2, Pten, Ccnd1, Atm, Brca1, E2f1, Cdk4, Prkdc, Cdk2, Cdc25a, Cdkn1b, Rb1, Mdm2, Cdkn2a, Cdkn1a) is a DNA damage repair gene set; [2305] (including Trp53, Pycard, Pten, Atm, Wt1, Tsc2, Brca1, Trp73, Rb1, Trp63) is a negative cell cycle regulation gene set; [1303] (including Trp53, Mre11a, Bax, Nbn, Hus1, Rad51, Gadd45a, Rad17, Chek1, Ube3a, Atm, Brca1, Rad9, Rad50, Ube1x, Ubc, Nfkb1a, Abl1, Bcl2, Mdm2, Trp63, Apaf1, Timp3) is a DNA damage checkpoint/p53/ATM pathway; [1404] (including Pes1, Ccnb1, Cdc20, Tnfsf5ip1, Brca2, Smc211, Mad211, Ran, Terf1, Ccnb1-rs1, Foxm1, Nek2, Cdk2, Foxm1, Stag1, Shc1, Cdc25b, Rbx1, Ywhae, Nek3, Cdc2a, Smc111, Cdc25a, Prc1, Rad21, Prm1, Prm2, Wee1, Ccna1, Stag2, Cdc16) is a group of M phase genes; [1402] (including Mcm5, PcnA, Mcm2, Mcm3, Mre11a, Mcm6, Mcm4, Rad51, Mcm7, Msh2, Sumo1, Rad17, Rad50, Mki67) is a set of S phase and DNA replication genes; and [1302] (including Ccna2, Cdk7, Ccnh, Cdc6, Ccnc, Cdc25b, Cdc45l, Cdc25a, Cdk8, Mki67, Ccna1, Ccng2) is another S phase related gene set. These gene sets reveal that cell cycle activities, along with the concomitant DNA damage check and repair, are an important feature of stem cell self-renewal.

The opposite comparison, i.e., comparing the data at the last 4 time points with data at the first 7 time points, produced two different enriched gene sets (Table S1). One was protein amino acid phosphorylation related (numbered <1712> and including Raf1, Rps6ka1, Rps6kb1, Mapk1, Prkcc, Pdpk1, Prkci, Map2k2, Mknk1, Prkcb1, Insr, Pik3r1, Akt2, Pik3ca, Map2k1, Araf, Prkcz, Gsk3b, Akt3, Igflr, Akt1), and the other was Ca<sup>2+</sup>/NF-AT signaling related transcriptional factors (numbered <1102> and including Nfkbib, Sp1, Sp3, Fos, Actb, Nfatc2ip, Fosb, Fosl1, Nfkb2, Egr2, Nfkbie, Egr3, Mef2d, Rela, Junb, Fosl2, Cebpb, Mef2a, Jun, Gata4). More studies are needed to interpret why these gene sets were enriched at the late stages of cultivation and how they were linked to a possible reduction in self-renewal.

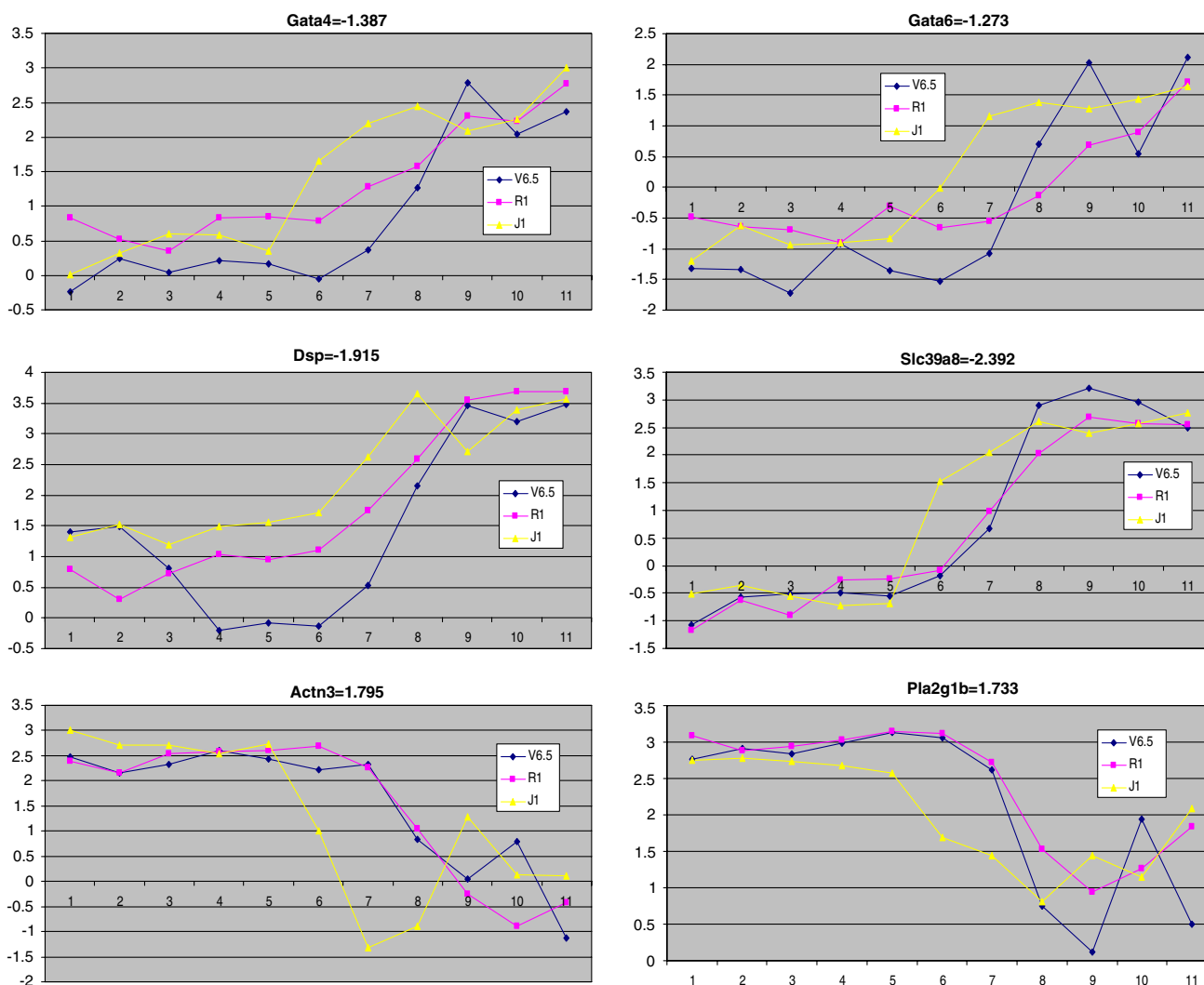
To determine the critical time point for the change in gene expression, comparisons were made with different shuffles of data. When data at the first 8 time points were compared with data at the last 3 time points, similar results were acquired, but with smaller NES and larger (hence less convincing) NOM *p*-values and FDR *q*-values. When data at the first 6 time points were compared with data at the last 5 time points, no significantly changed gene sets were

found. These results indicate that the change of expression of these self-renewal and cell cycle related gene sets occurred at the 7th time point (2 days' cultivation), after which their expression decayed quickly.

Expression profiles of cell differentiation related genes

Two research groups reported that differentiation of mES cells is induced by Gata4/6 (Fujikura et al. 2002; Li et al. 2004), but it is not known when, and with what partners, Gata4/6 induce stem cell differentiation. It is interesting that Oct3/4 over-expression can induce Gata4 transcription, which in turn triggers differentiation (Fujikura et al. 2002; Li et al. 2004). As no knowledge of the Gata4/6 related signaling pathway was available, we simply checked the expression profile of these two genes. Gata4/6 expression was low at the early time points but became significantly

higher at the last four (Fig. 2), with a large negative enrichment score (-1.387 and -1.273, respectively). The increase of expression began earlier in E201 (J1, without LIF) than in E113 and E165 (V6.5 and R1, with LIF), indicating that LIF may play some role in delaying cell differentiation. Nevertheless, we found that Gata4/6 were not the genes with the largest negative enrichment score. The largest and second largest scores, -2.392 and -1.915, came from Slc39a8 and Dsp. Whether the two play a more important role than Gata4/6 in stem cell differentiation and their relationship with Gata4/6 remain unknown. These results confirm that Gata4/6 can indeed reflect stem cell differentiation, but Slc39a8 and Dsp could be alternative, and possibly better, indicators. In addition, the most enriched single genes at the early time points were Actn3 and Pla2g1b, which do not appear in any of our defined gene sets (Fig. 2). Whether they are involved in self-renewal is



**Fig. 2** The expression of Gata4/6, Dsp, Sla39a8, Actn3 and Pla2g1b (against the control) in the 14 days in three cell lines. The number following the gene name is the enrichment score

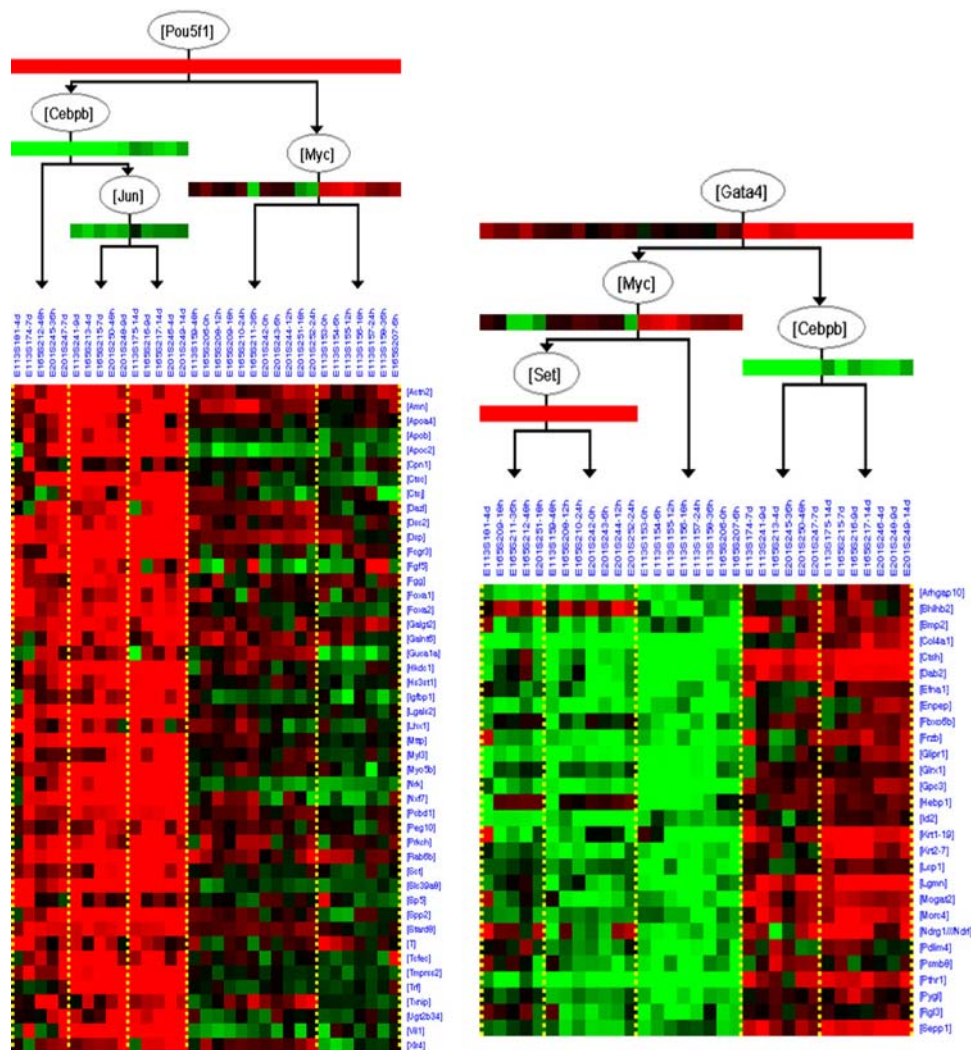
not clear. These genes, with extraordinarily high enrichment scores, provide new and important clues for deciphering signaling in stem cell self-renewal and differentiation.

The possible regulation of genes with the most changed expression

An important question in microarray data analysis is how changed gene expression is regulated, under different conditions and at different times. To answer this question, genes should be clustered according to their expression, and then regulators of each cluster be identified. Judged against the control sample, 2257 genes, either differentially expressed in the three cell lines or important for developmental signaling, were chosen for clustering (see Materials and methods). Using a revised clustering method (Segal et al. 2003), which not only classifies genes into clusters of different expression patterns but also reveals their possible regulation by defined regulators, we found that the expression of a few gene

clusters underwent a drastic change during the 14 days from very low (in green in Fig. 3) to very high (in red in Fig. 3). It is reasonable to postulate that the expression of these gene clusters reflects mES self-renewal and differentiation. To reveal how their drastically altered expression was regulated during the cell culture, we assumed that in the decay of self-renewal and the initiation of differentiation the expression of regulators themselves underwent a significant change, showing either a positive or a negative enrichment, and further that these regulators were transcriptional factors. From the 2257 genes we picked up 26 transcriptional factors (Table S2), which had a high enrichment score computed with the gene set enrichment analysis, and used them as regulators in the clustering program. Apart from a *Max Module Number* of 50, which means the genes are clustered into at most 50 clusters, all other parameters were the default values. The two clusters with the most change in gene expression during the 14 days, with the possible regulation by the chosen regulators (Fig. 3), comprised genes (including *Slc39a8* which was a potential indicator of cell

**Fig. 3** The possible regulation and regulators of the two co-expressed gene clusters with the most significantly changed expression during the 14 days in the three cell lines. Red and green colors indicate high and low expression level. The hierarchy of regulation indicates the possible relationship among regulators. The expression of each regulator is divided into two stages which exert different regulatory impacts on either other regulators or clustered genes





differentiation as aforementioned) that had very low expression at early time points but very high expression at later times. Repeated running of the program gave the similar results, revealing that the low expression at the early time points was probably because of the repression by active self-renewal genes such as Oct4, Nanog, Pcna, Myc and Smar-cad1 (Chambers et al. 2003; Mitsui et al. 2003; Niwa et al. 2000), and that the high expression at the late time points was probably due to both the enhanced expression of differentiation genes (Gata4/6) and the down regulated self-renewal genes (especially Oct4, Jun and Cebpb). In contrast, the typical self-renewal genes, including Nanog, Oct4 and Sox2, only showed slight changes during the 14 days. This implies that differentiation may be the default fate of stem cells and self-renewal relies on a maintenance mechanism. When that mechanism weakens, cell differentiation begins.

## Discussion

### LIF stimulated cell proliferation

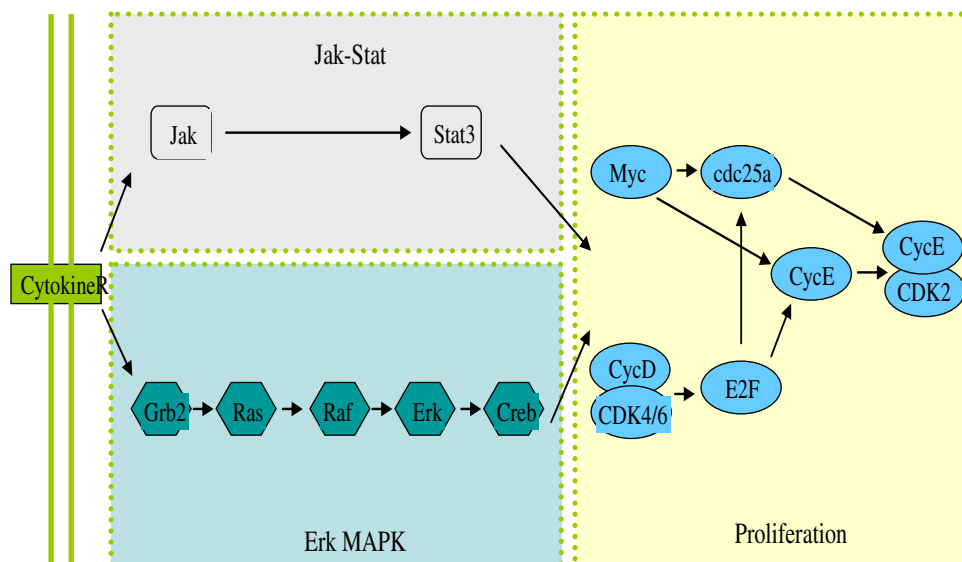
The JAK/STAT pathway has been well documented to play a major role in stem cell self-renewal (Smith et al. 1988; Williams et al. 1988; Yoshida et al. 1994; Niwa et al. 1998). However, the enriched Ras/Raf/ERK, but not JAK/STAT, genes in this dataset, especially the enriched expression of Eras, disclose that another MAPK related pathway may also be a key player to carry on external LIF signal into mES cells to promote self-renewal. As the most enriched gene sets in cells cultivated with LIF are cell cycle related, there exists a possibility that the major impact of LIF on self-renewal could be to promote cell proliferation. However, since cell cycle activity is naturally active in stem cells with high pluripotency, it is difficult to

tell whether active cell cycle activity is required for or just indicates pluripotency. More relevant epistatic analysis upon gene mutation is needed. The assembled picture of JAK/STAT and Ras/Raf/ERK signaling, upon our results, gives a more comprehensive understanding of stem cell self-renewal (Fig. 4). To make the picture more complete, new gene sets need to be identified so as to find links between these genes and those known self-renewal indicators, especially Nanog, Oct4 and Sox2. This would help to better answer the question as to how cell proliferation helps maintain self-renewal and pluripotency.

### Maintenance of self-renewal in the absence of LIF

Compared with the difference in gene expression in cells at different times, the difference in gene expression in cells cultivated with and without LIF is not prominent. It is not clear whether this is due to the high Nanog/Oct4/Sox2 expression in all three cell lines or whether LIF is not essential for self-renewal. The latter is partially supported by recent findings that LIF alone may not play a decisive role in maintaining self-renewal. In addition to extrinsic regulators such as LIF, it is suggested that intrinsic transcriptional determinants, especially Oct4 and Nanog, are required, and possibly more important, for maintaining the undifferentiated state (Chambers and Smith 2004). When overexpressed, Nanog allows ES cells to self-renew without the otherwise required LIF (Chambers et al. 2003). Oct3/4 may also make an independent contribution (Rao 2004). As the precise level of Oct3/4 is reported to govern the three distinct fates of ES cells (Li et al. 2004), to look more intensively into the quantitative relationship between the expression of self-renewal, cell cycle control and cell differentiation genes is important.

**Fig. 4** Two pathways, JAK/STAT and Ras/Raf/ERK, both carry the external LIF signal into mES stem cells to promote cell proliferation and self-renewal



## Indicators of stemness

As in the dataset most self-renewal genes showed only a slight change during the 14 days' culture but cell differentiation indicators had a significantly increased expression during the later stages, we postulate that differentiation may be the default fate for stem cells and self-renewal may rely on a maintenance mechanism, which demands either external signals such as LIF and FGF or internal ones like Nanog, Oct4 and Sox2. When the maintenance mechanism becomes weakened, probably even slightly, cell differentiation begins. Thus the upregulated expression of cell differentiation genes would better reflect the loss of self renewal than the down regulated expression of the core factors Oct4, Sox2 and Nanog.

## On the combined use of different methods

Traditionally, identifying differentially expressed genes and classifying them into different clusters by using subtle algorithms is the major work of microarray data analysis. In this study we show that to carefully use supervised methods and to properly combine different methods (one supervised and one unsupervised here) could produce more interesting and insightful results, which may be valuable for developing and validating hypotheses. Compared with the information indicated by the change of individual genes, the expression of a set of correlated genes in a functional module or a signaling pathway gives richer and more reliable information. In the case of gene set enrichment analysis, we note that an appropriate and comprehensive definition of gene sets can directly lead to the discovery of new pathways or fill in the gaps between known ones (Fig. 4), and in the case of the revised clustering method, results like "regulator X regulates cluster Y under condition W" can be readily confirmed or rejected by experiment. Nevertheless, to decipher molecular signaling from gene expression with whatever methods could be thorny and unreliable, as analogous gene expression does not exclude differential signaling, nor means co-regulation. To validate whether the revealed changes in gene expression indeed indicate changes in molecular signaling, validation steps at protein interaction level, for example, by using specific antibodies, are required to check the concentration of the effectors of these pathways.

## The roles of genes in other signaling pathways

The roles of Notch, Wnt and Hh pathways in promoting mES self-renewal were not found to be significant in the analysis of this dataset with the defined gene sets. The expression of genes in another self-renewal related

pathway, BMP-Smad-Id (Ying et al. 2003; Varga and Wrana 2005), was also not apparently enriched.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Amit M, Carpenter MK, Inokuma MS, Chiu CP, Harris CP, Waknitz MA, Itskovitz-Eldor J, Thomson JA (2000) Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol* 227:271–278
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122:947–956
- Burdon T, Smith A, Savatier P (2002) Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* 12:432–438
- Catena R, Tiveron C, Ronchi A, Porta S, Ferri A, Tatangelo L, Cavallaro M, Favaro R, Ottolenghi S, Reinbold R, Scholer H, Nicolis SK (2004) Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. *J Biol Chem* 279:41846–41857
- Chambers I, Smith A (2004) Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* 23:7150–7160
- Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113:643–655
- Daheron L, Opitz SL, Zaehres H, Lensch WM, Andrews PW, Itskovitz-Eldor J, Daley GQ (2004) LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* 22:770–778
- Dani C, Chambers I, Johnstone S, Robertson M, Ebrahimi B, Saito M, Taga T, Li M, Burdon T, Nichols J, Smith A (1998) Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev Biol* 203:149–162
- Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, Miyazaki J, Niwa H (2002) Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* 16:784–789
- Jirmanova L, Afanassieff M, Gobert-Gosse S, Markossian S, Savatier P (2002) Differential contributions of ERK and PI3-kinase to the regulation of cyclin D1 expression and to the control of the G1/S transition in mouse embryonic stem cells. *Oncogene* 21:5515–5528
- Kuroda T, Tada M, Kubota H, Kimura H, Hatano SY, Suemori H, Nakatsuji N, Tada T (2005) Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* 25:2475–2485
- Li L, Arman E, Ekblom P, Edgar D, Murray P, Lonai P (2004) Distinct GATA6- and laminin-dependent mechanisms regulate endodermal and ectodermal embryonic stem cell fates. *Development* 131:5277–5286
- Mitsui K, Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S (2003) The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and escells. *Cell* 113:631–642
- Niwa H, Burdon T, Chambers I, Smith A (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* 12:2048–2060

- Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24:372–376
- Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F (2005) Oct3/4 and Sox2 regulate Oct3/4 gene in embryonic stem cells. *J Biol Chem* 280:5307–5317
- Perez-Iratxeta C, Palidwor G, Porter CJ, Sanche NA, Huska MR, Suomela BP, Muro EM, Krzyzanowski PM, Hughes E, Campbell PA, Rudnicki MA, Andrade MA (2005) Study of stem cell function using microarray experiments. *FEBS Letters* 579:1795–1801
- Rao M (2004) Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev Biol* 275:269–286
- Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P (2005) Transcriptional regulation of Nanog by Oct4 and Sox2. *J Biol Chem* 280:24731–24737
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedmann N (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34:166–176
- Sekkai D, Sekkai D, Gruel G, Herry M, Moucadel V, Constantinescu SN, Albagli O, Tronik-Le Roux D, Vainchenker W, Bennaceur-Griscelli A (2005) Microarray analysis of LIF/Stat3 transcriptional targets in ES cells. *Stem Cells* 23:1634–1642
- Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 336:688–690
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–15550
- Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282:1145–1147
- Varga A, Wrana J (2005) The disparate role of BMP in stem cell biology. *Oncogene* 24:5713–5721
- Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA, Gough NM (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 336:684–687
- Ying QL, Nichols J, Chambers I, Smith A (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with Stat3. *Cell* 115:281–292
- Yoshida K, Chambers I, Nichols J, Smith A, Saito M, Yasukawa K, Shoyab M, Taga T, Kishimoto T (1994) Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signaling pathways. *Mech Dev* 45:163–171