Gene expression profiling of somatic and pluripotent cells reveals novel pathways involved in reprogramming

Y.N. Cai[1,2]*, X.H. Dai[1,2]*, Q.H. Zhang[1], R.Hu[1] and Z.M. Dai[1]

[1]School of Information Science and Technology, Sun Yat-sen University,
Guangzhou, Guangdong, China
[2]SYSU-CMU Shunde International Joint Research Institute (JRI), Shunde,
Guangdong, China

Corresponding authors: Y.N. Cai / X.H. Dai
E-mail: caiyanningmitl@sina.com / issdxh@mail.sysu.edu.cn

ABSTRACT. We investigated gene expression in embryonic stem (ES) cells, induced pluripotent stem (iPS) cells, and fibroblasts. Microarray expression data sets obtained from the Gene Expression Omnibus were analyzed using the Partek software. Human genes from ES cells, iPS cells, and fibroblasts were ranked from low to high according to their expression levels. The gene expression mode in iPS cells was much more like the mode in ES cells, and the expression levels of fibroblast genes fluctuated more dramatically than those of iPS and ES cells. The number of genes with significantly different expression was lower in the iPS and ES cells. Several genes with the expression levels that were significantly different between ES and iPS cells were found, including *LEFTY2*, *DLK1*, and *NLRP2*. Four genes belonged to the low expression category in fibroblasts with the high expression category occurring in ES cells, i.e., *HESRG*, *PROM1*, *NTS*, and *LRRN1*. Analyzing the expression of these genes is helpful to elucidate the mechanisms of cell fate regulation and efficiently obtain iPS cells.

Key words: Embryonic stem cell; Induced pluripotent stem cell; Fibroblast; Gene expression

INTRODUCTION

Cellular states are plastic. Pluripotent embryonic stem (ES) cells can differentiate into diverse cell types that make up organisms (Artyomov et al., 2010). There has long been a desire to reverse this process, i.e., reprogram a differentiated cell so that it has pluripotent potential. Recently, this has been achieved by overexpressing specific transcription factors in differentiated cells (Yamanaka, 2007). Induced pluripotent stem (iPS) cells can be obtained from fibroblasts upon ectopic expression of OCT4, SOX2, KLF4, and c-MYC (Takahashi and Yamanaka, 2006). This method does not use embryonic materials and promises the development of patient-specific regenerative medicine. However, reprogramming efficiencies are only 0.0001-29%. Most reports show that successful induction of the pluripotent state is rare even if all required factors are present (Pang et al., 2011; Huang et al., 2011). The regulatory mechanisms that make reprogramming possible and determine its efficiency are poorly understood.

Most cells in an organism have the same DNA. Nevertheless, different cell types express different proteins and carry out different functions. Gene expression intensity is a distinct feature of the cell state (Artyomov et al., 2010). iPS cells hold great promise for the study and therapy of human diseases because they are very like ES cells in their ability to self-renew and give rise to all three germ layers (Sridharan et al., 2009). The similarity between these two cell types is reflected in their gene expression.

In the present study, we investigated the gene expression levels in human ES cells, iPS cells, and fibroblasts. Moreover, we compared the gene expression patterns in ES cells, iPS cells, partially reprogrammed (piPS) cells, and mouse embryonic fibroblasts (MEFs). By investigating gene expression in different cells we sought to shed light on the mechanisms that regulate cell fate.

MATERIAL AND METHODS

To compare the gene expression intensities in different cell types, we selected microarray expression data sets obtained from the Gene Expression Omnibus database with the following ID numbers: GSE9865, GSE12390, GSE26100, GSE14012, and GSE28688 (Lowry et al., 2008; Maherali et al., 2008; Sridharan et al., 2009; Koche et al., 2011; Mah et al., 2011). These data sets were from fibroblasts, iPS cells, and ES cells in humans and mice.

Each expression data set was loaded into the Partek data analysis software and treated with the GC-RMA package, followed by quantile normalization and median polishing. Intensities for each probe set were averaged across replicates and extracted. RefSeq transcripts were assigned intensity scores by averaging the intensities of associated probe sets. Human genes from ES cells, iPS cells, and fibroblasts were ranked from low to high according to their expression levels. We defined the first 1450 genes as having high expression intensities (about 5% of the total genome), and the last 1450 genes as having low expression intensities.
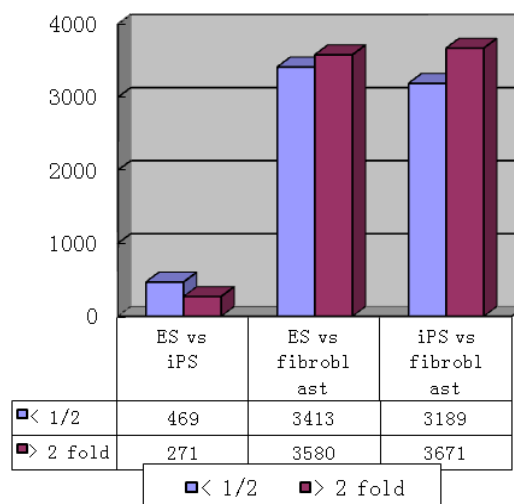
RESULTS

Human gene expression comparisons

Probing the non-infected cells, we obtained accurate expression data sets of 29,034 genes from ES cells, iPS cells, and fibroblasts. Genes were ranked from low to high according to their

expression levels in the iPS cells (Figure 1). We then compared the gene expression intensities in other cell types and determined the similarities and differences. Gene expression intensities showed similar trends in all cell types. However, it was obvious that the gene expression mode in iPS cells was much more like that in ES cells than in fibroblasts. The gene expression levels fluctuated more dramatically in the fibroblasts than in the iPS and ES cells. The similarities and differences of gene expression in the three cell types were further investigated, as shown in Figure 2. We counted the number of genes whose expression changed 2-fold or more in the different cell types. A total of 271, 3580, and 3671 genes showed significant expression changes (larger than 2-fold) between the ES and iPS cells, the ES cells and fibroblasts, and the iPS cells and fibroblasts, accounting for 2.55, 24.09, and 23.63%, respectively.
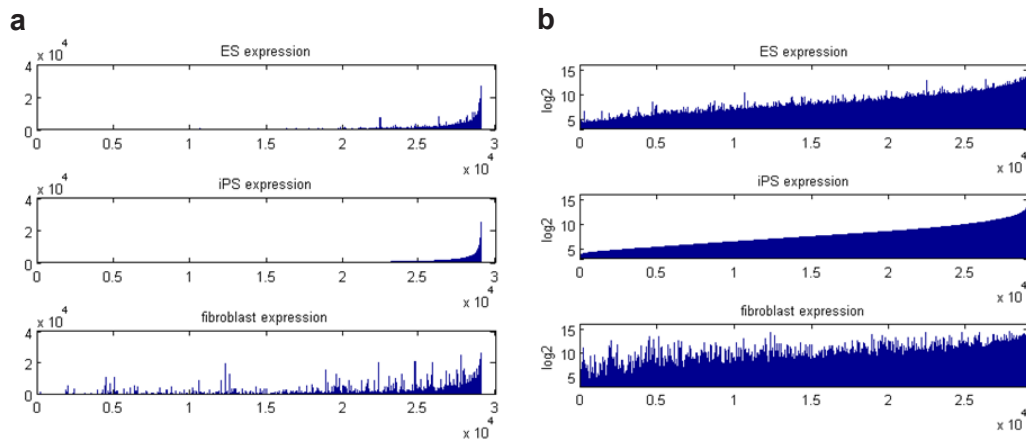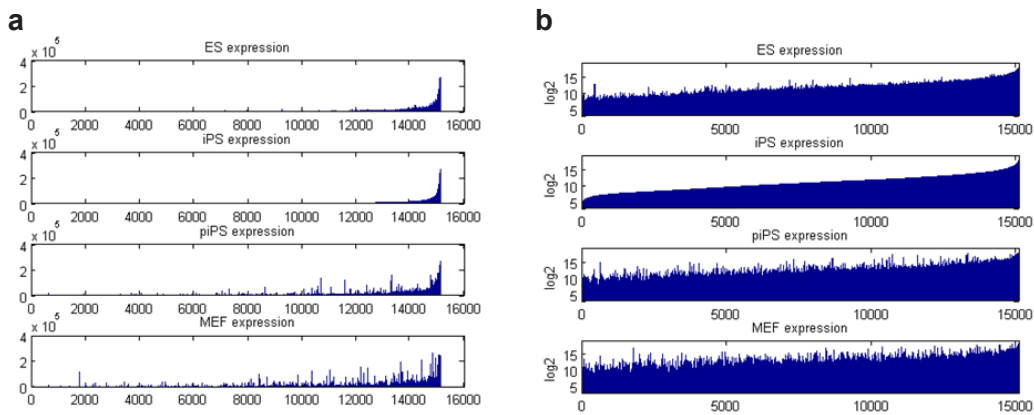

Figure 1. Human gene expression levels. The genes are arranged according to their expression levels in the induced pluripotent stem (iPS) cells using the original data (**a**) and the log2 values of the data (**b**).

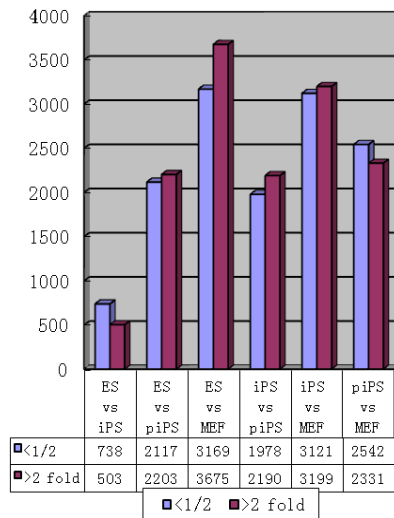|	ES vs iPS	ES vs fibroblast	iPS vs fibroblast
< 1/2	469	3413	3189
> 2 fold	271	3580	3671

Figure 2. Human genes with significant differences in expression levels.

Murine gene expression comparisons

Probing non-infected cells, we obtained accurate expression data sets for 15,085 genes from ES cells, iPS cells, piPS cells, and MEFs. Genes were ranked from low to high according to their expression levels in iPS cells (Figure 3). By comparing expression changes at the three stages, our analysis provided insight into the gene expression patterns in the different cell types. Figure 4 shows the number of genes whose expression changed 2-fold or more in the different murine cell types. A total of 503, 2203, 3675, 2190, 3199, and 2331 genes showed significant expression changes (larger than 2-fold) between the ES and iPS cells, the ES and piPS cells, the ES and MEF cells, the iPS and piPS cells, the iPS and MEF cells, and the piPS and MEF cells, accounting for 8.26, 28.64, 45.37, 27.63, 41.9, and 32.3%, respectively.

a

b

Figure 3. Murine gene expression levels. The genes are arranged according to their expression levels in the induced pluripotent stem (iPS) cells using the original data (**a**) and the log2 values of the data (**b**).

	ES vs iPS	ES vs piPS	ES vs MEF	iPS vs piPS	iPS vs MEF	piPS vs MEF
■<1/2	738	2117	3169	1978	3121	2542
■>2 fold	503	2203	3675	2190	3199	2331

■<1/2   ■>2 fold

Figure 4. Murine genes with significant differences in expression levels.

Majority of genes showed stable expression patterns in various cell states

Most genes belonging to the high or low expression categories overlapped in all cells. The overlap was greatest in the ES and iPS cells. There were 773 genes belonging to the high expression category in all cells. The number of genes highly expressed in ES and iPS cells, ES cells and fibroblasts, and iPS cells and fibroblasts was 1263, 840, and 824, respectively. The largest difference at the high expression level was about 4-fold. Many highly expressed genes were housekeeping genes, which are indispensable to living organisms, such as those that encode the ribosomal proteins. Similarly, there were 1107 genes belonging to the low expression category in all cells. The number of genes slightly expressed in ES and iPS cells, ES cells and fibroblasts, and iPS cells and fibroblasts was 1319, 1176, and 1174, respectively. The largest difference at the low expression level was about 1.5-fold.

Genes with significantly different expression intensities in diverse cell types

Generally, there were no genes belonging to opposing expression categories in the ES and iPS cells. Thus, these two types of cells showed similar expression patterns. Nevertheless, there were genes that had significantly different expression between ES and iPS cells, including *LEFTY2*, *DLK1*, and *NLRP2*. *LEFTY2* belonged to the high expression category in ES cells, but did not belong to the high expression category in iPS cells. *DLK1* and *NLRP2* belonged to the high expression category in the iPS cells, but did not belong to the high expression category in the ES cells. No genes belonged to the high expression category in fibroblasts, or simultaneously to the low expression category in ES and iPS cells. Only a few genes, such as *HOTAIR*, showed a slightly higher expression in fibroblasts compared with ES or iPS cells. A total of four genes, i.e., *HESRG*, *PROM1*, *NTS*, and *LRRN1*, belonged to the low expression category in fibroblasts but they belonged to the high expression category in ES and iPS cells.

DISCUSSION

In the present study, we investigated the gene expression levels in human ES cells, iPS cells, and fibroblasts. Through analyzing gene expression at high and low levels, we obtained different results for gene expression between the three cell types. Moreover, we compared the gene expression patterns in ES cells, iPS cells, piPS cells, and MEFs.

The gene expression mode in iPS cells was much more like the mode in ES cells, and gene expression levels in the fibroblasts fluctuated more dramatically than in the iPS and ES cells. Moreover, the number of genes with significantly different expression was smaller in the iPS and ES cells compared with the other cells types. This indicates that iPS cells might be used to replace ES cells on many occasions. Owing to their ability to self-renew, proliferate, and differentiate, ES cells are widely recognized as a precious source of pluripotent cells (Stojkovic et al., 2004). Nevertheless, the problem of immune rejection following transplantation and the controversy surrounding the use of human embryos inhibit their application (Ghosh et al., 2010). The direct reprogramming of somatic cells to iPS cells was first reported by Takahashi and Yamanaka, who converted adult mouse fibroblasts to iPS cells by ectopic expression of a group of transcription factors. Since then, many groups have shown that both human and mouse somatic cells can be reprogrammed by overexpression of a few transcription factors to what appears to be an embryonic state, i.e.,

iPS cells (Chin et al., 2009; Ghosh et al., 2010). It seems that iPS cells offer a non-controversial alternative source for ES cells. We think that the similarities in gene expression patterns between ES and iPS cells can explain this phenomenon.

Although similar gene expression profiles between ES and iPS cells were found in our research, the number of gene expression differences between the two cellular populations might account for incomplete reprogramming. Therefore, it was important to define the differences between ES and iPS cells. Several genes with significantly different expression between ES and iPS cells were found, including *LEFTY2*, *DLK1*, and *NLRP2*. Genes relevant to tumor development, such as *DLK1* and *NLRP2*, are worthy of attention. *DLK1* plays a significant role in liver cancer tumorigenesis. *NLRP2* is related to many kinds of tumor. They were both highly expressed in iPS cells but showed comparatively low expression in ES cells. This result is consistent with the former studies that showed that the reprogramming process is somewhat similar to tumor development (Halazonetis et al., 2008; Kawamura et al., 2009; Marión et al., 2009; Utikal et al., 2009; Edel et al., 2010; Jasencakova and Groth, 2010; Menendez et al., 2010; Gupta et al., 2011). *LEFTY* expression has been recognized as a stemness marker owing to its enrichment both in undifferentiated ES cells and in blastocysts. *LEFTY1* knockdown in mouse ES cells showed enhanced phosphorylation of *Smad2* and increased differentiation potential (Kim et al., 2014). Thus, the lower expression of *LEFTY1* in iPS cells might result in partial reprogramming.

No genes belonged to the high expression category in fibroblasts, or simultaneously to the low expression category in ES and iPS cells. That is to say, cells with pluripotent ability tend to have higher gene expression levels than differentiated cells. This may reflect the fact that the genes in pluripotent cells are less stable (Sedighi and Sengupta, 2007; Orford and Scadden, 2008; Hanna et al., 2009; Yamanaka, 2009; Fisher and Fisher, 2011; Pasi et al., 2011; Blasco et al., 2011). Four genes, i.e., *HESRG*, *PROM1*, *NTS*, and *LRRN1*, belonged to the low expression category in fibroblasts but the high expression category in ES cells. *HESRG*, also known as *ESRG* (embryonic stem cell-related gene), is expressed specifically in undifferentiated human ES cells (Li et al., 2013). *PROM1*, also called *CD133*, is a pentaspan membrane protein, exhibiting self-renewal and differentiation (Gopisetty et al., 2012). *NTS* is related to the nervous system, and *LRRN1* is relevant to brain formation, and might play a role in signal transduction (Tang et al., 2012). Partially reprogrammed cells, expressing markers of the intermediate reprogramming stage, have failed to transcriptionally activate pluripotency regulators. They can be converted to a pluripotent state with some specific small molecules that affect chromatin modifications or modulate signal transduction (Sridharan et al., 2009). Thus, overexpression of these genes is significant to pluripotency. In other words, their overexpression is specific to embryo development. Analyzing their expression intensities in different types of cells is helpful to the investigation of the function of each cell stage.

CONCLUSION

The majority of genes showed similar expression patterns in all cells, especially in iPS and ES cells. However, some distinct differences among various cell types were noticed. Such differences usually occur when the genes are relevant to specific cell types. Analyzing the expression of these genes is helpful for elucidating the mechanisms of cell fate regulation and efficiently obtaining iPS cells.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Artyomov MN, Meissner A and Chakraborty AK (2010). A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput. Biol.* 6: e1000785.

Blasco MA, Serrano M and Fernandez-Capetillo O (2011). Genomic instability in iPS: time for a break. *EMBO J.* 30: 991-993.

Chin MH, Mason MJ, Xie W, Volinia S, et al. (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 5: 111-123.

Edel MJ, Menchon C, Menendez S, Consiglio A, et al. (2010). Rem2 GTPase maintains survival of human embryonic stem cells as well as enhancing reprogramming by regulating p53 and cyclin D1. *Genes Dev.* 24: 561-573.

Fisher CL and Fisher AG (2011). Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Curr Opin. Genet. Dev.* 21: 140-146.

Ghosh Z, Wilson KD, Wu Y, Hu S, et al. (2010). Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS One* 5: e8975.

Gopisetty G, Xu J, Sampath D, Colman H, et al. (2013). Epigenetic regulation of CD133/PROM1 expression in glioma stem cells by Sp1/myc and promoter methylation. *Oncogene* 32: 3119-3129.

Gupta PB, Fillmore CM, Jiang G, Shapira SD, et al. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146: 633-644.

Halazonetis TD, Gorgoulis VG and Bartek J (2008). An oncogene-induced DNA damage model for cancer development. *Science* 319: 1352-1355.

Hanna J, Saha K, Pando B, van Zon J, et al. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 462: 595-601.

Huang P, He Z, Ji S, Sun H, et al. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* 475: 386-389.

Jasencakova Z and Groth A (2010). Replication stress, a source of epigenetic aberrations in cancer? *Bioessays* 32: 847-855.

Kawamura T, Suzuki J, Wang YV, Menendez S, et al. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* 460: 1140-1144.

Kim DK, Cha Y, Ahn HJ, Kim G, et al. (2014). Lefty1 and lefty2 control the balance between self-renewal and pluripotent differentiation of mouse embryonic stem cells. *Stem Cells Dev.* 23: 457-466.

Koche RP, Smith ZD, Adli M, Gu H, et al. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* 8: 96-105.

Li G, Ren C, Shi J, Huang W, et al. (2013). Identification, expression and subcellular localization of ESRG. *Biochem. Biophys. Res. Commun.* 435: 160-164.

Lowry WE, Richter L, Yachechko R, Pyle AD, et al. (2008). Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* 105: 2883-2888.

Mah N, Wang Y, Liao MC, Prigione A, et al. (2011). Molecular insights into reprogramming-initiation events mediated by the OSKM gene regulatory network. *PLoS One* 6: e24351.

Maherali N, Ahfeldt T, Rigamonti A, Utikal J, et al. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* 3: 340-345.

Marión RM, Strati K, Li H, Murga M, et al. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* 460: 1149-1153.

Menendez S, Camus S and Izpisua Belmonte JC (2010). p53: guardian of reprogramming. *Cell Cycle* 9: 3887-3891.

Orford KW and Scadden DT (2008). Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nat. Rev. Genet.* 9: 115-128.

Pang ZP, Yang N, Vierbuchen T, Ostermeier A, et al. (2011). Induction of human neuronal cells by defined transcription factors. *Nature* 476: 220-223.

Pasi CE, Dereli-Öz A, Negrini S, Friedli M, et al. (2011). Genomic instability in induced stem cells. *Cell Death Differ.* 18: 745-753.

Sedighi M and Sengupta AM (2007). Epigenetic chromatin silencing: bistability and front propagation. *Phys. Biol.* 4: 246-255.

Sridharan R, Tchieu J, Mason MJ, Yachechko R, et al. (2009). Role of the murine reprogramming factors in the induction of pluripotency. *Cell* 136: 364-377.

Stojkovic M, Lako M, Strachan T and Murdoch A (2004). Derivation, growth and applications of human embryonic stem cells. *Reproduction* 128: 259-267.

Takahashi K and Yamanaka S (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126: 663-676.

Tang KH, Ma S, Lee TK, Chan YP, et al. (2012). CD133(+) liver tumor-initiating cells promote tumor angiogenesis, growth, and self-renewal through neurotensin/interleukin-8/CXCL1 signaling. *Hepatology* 55: 807-820.

Utikal J, Polo JM, Stadtfeld M, Maherali N, et al. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* 460: 1145-1148.

Yamanaka S (2007). Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cell* 1: 39-49.

Yamanaka S (2009). Elite and stochastic models for induced pluripotent stem cell generation. *Nature* 460: 49-52.