The Fourth International Conference on Computational Systems Biology (ISB2010)
Suzhou, China, September 9–11, 2010

Potential Linkages Between the Inner and Outer Cellular States of Human Induced Pluripotent Stem Cells

Shigeru Saito[1,2,*] Yasuko Onuma[3,*] Yuzuru Ito[3,*] Hiroaki Tateno[4,*]
Masashi Toyoda[5,*] Hidenori Akutsu[5] Koichiro Nishino[5]
Emi Chikazawa[5] Yoshihiro Fukawatase[5] Yoshitaka Miyagawa[6]
Hajime Okita[6] Nobutaka Kiyokawa[6] Yohichi Shimma[4]
Akihiro Umezawa[5] Jun Hirabayashi[4] Katsuhisa Horimoto[1,7,§]
Makoto Asashima[3,8,§]

1 Computational Biology Research Center, National Institute of Advanced Industrial Science Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
2 INFOCOM CORPORATION, Sumitomo Fudosan Harajuku Building, 2-34-17, Jingumae, Shibuya-ku, Tokyo, 150-0001, Japan
3 Research Center for Stem Cell Engineering, National Institute of Advanced Industrial Science Technology (AIST), Tsukuba Central 4, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8562, Japan.
4 Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science Technology (AIST), Tsukuba Central 2, 1-1-1 Umezono, Ibaraki 305-8568, Japan.
5 Department of Reproductive Biology, National Research Institute for Child Health and Development, 2-10-1 Ookura, Setagaya-ku, Tokyo 157-8535, Japan
6 Department of Developmental Biology and Pathology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan
7 Institute for Systems Biology, Shanghai University, Shangda Road 99, Shanghai 200444, China
8 Department of Life Sciences (Biology), Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
* These authors contributed equally to this work.
§ Correspondence author

Abstract We analyzed both RNA profile to reveal gene expression changes and glycan profile to identify structural changes in glycans between four parental somatic cell (SC) lines and nine hiPSC lines that were originally established. The sum of these information by a combination of standard statistical techniques and a network approach showed significant differences in expression between the iPSCs and SCs. Subsequent network analysis of the gene expression and glycan signatures revealed glycan transfer network associated with known epitopes for differentiation. The present study is the first to uncover the relationships between gene expression patterns and cell surface changes in hiPSCs, and reinforces the importance of the cell surface to identify established iPSCs from SCs.

Keywords Induced pluripotent stem cell; DNA microarray; Lectin microarray; Network Analysis

1 Introduction

Human iPS cells (hiPSCs) attract great attention for the application to drug screening and analysis of the mechanisms of diseases, and even as the next generation materials for regenerative medicine. Genetic reprogramming to a pluripotent state of human somatic cells was first achieved by ectopic expression of four factors (Sox2, Oct4, Klf4 and c-Myc) using retrovirus [1]. Subsequently, this method was applied to various human cells using different combination of defined factors [2, 3]. However, transcription factor-induced acquisition of replication competence and pluripotency raises the question as to how exogenous factors induce changes in the inner and outer cellular states [4-6].

Here, we analyzed both RNA profile to reveal gene expression changes and glycan profile to identify structural changes in glycans between four parental somatic cell (SC) lines and nine hiPSC lines that were originally established. The sum of these information by a combination of standard statistical techniques and a network approach showed significant differences in expression between the iPSCs and SCs. Subsequent network analysis of the gene expression and glycan signatures revealed glycan transfer network associated with known epitopes for differentiation. These results shed new light on the potential linkages between the inner and outer cellular states for acquisition of replication competence and pluripotency in hiPSCs**.**

2 Materials and Methods

2.1 Cell experiments

Cell experiment was performed as described [7, 8]. Somatic cell pellets were harvested by scraping. The hiPSCs were incubated at 37ºC in a solution that contained 1 mg/ml collagenase IV (Invitrogen, Carlsbad, CA), 1 mM $CaCl_2$, 20% KNOCKOUT^{TM} Serum Replacement (KSR), and 10% ACCUMAX (Innovative Cell Technologies, Inc., San Diego, CA). When the edges of the colonies started to dissociate from the bottom of the dish, the collagenase solution was removed and the cells were washed with medium. Colonies were then picked up and collected.

2.2 Gene expression analysis

Changes in mRNA levels were monitored using an Agilent Whole Human Genome Microarray chip (G4112F). This array covers 41,000 well-characterized human genes and transcripts. After background correction using a Normal plus Exponential convolution model, which adjusts the foreground to the background, we used an offset to dampen the variation of the log-ratios for intensities close to zero.

2.3 Network Analysis

We calculated the graph consistency probability [9], which reflects the

consistency of a given network structure with the expression data of the constituent genes monitored in the present study, for each of the 146 network structures constructed using the previous ChIP-on-Chip data of four factors [10] and the gene classification scheme of the Molecular Signatures Database (MSigDB) [11]. The statistical significance level of the thresholds was set at 0.05.

2.4 Glycan analysis

We analyzed cell surface glycans using a lectin microarray [12]. Briefly, 43 lectins were spotted at a concentration of 0.5 mg/ml in triplicate onto epoxysilane-coated glass slides. Cell membrane fractions labeled with Cy3 NHS ester were then incubated with the lectin microarray, and fluorescence images were acquired using an evanescent-field activated fluorescence scanner SC profiler. The fluorescence signal for each spot was quantified with background value correction, and the lectin signals were defined by the average of triplicate spots.

3 Results and Discussion

3.1 Gene expression signature of hiPSCs descended from different parent SCs

To determine the gene expression signatures of hiPSCs, a detailed genome-wide expression analysis was performed to compare iPSCs and their parental SCs from amniotic mesodermal (AM), placental artery endotherial (PAE), uterine endometrium (UtE), and MRC-5 (MRC) cell sources. In total, 51 cell samples of 13 cell lines (39 hiPSCs samples of 9 hiPSC lines [7, 8]) were studied in the present study, so as to statically compare the hiPSCs and parental SCs. Unsupervised hierarchical clustering of the gene expression data across the four hiPSC lines (AM, PAE, UtE, MRC) and their corresponding parental SCs revealed interesting patterns of gene expression heatmap (Fig. 1A). First, the hiPSCs were clearly distinguishable from their respective parental SCs. Second, the four hiPSC lines had gene expression profiles that were linked to those of their parental SCs, while the hiPSCs from different passages clustered more closely with each other than with the hiPSCs from the corresponding parental SCs (Fig. 1A). In support of these findings, a Pearson correlation analysis demonstrated that the gene expression profiles of hiPSCs from different passages were more closely related to each other than to the hiPSCs from the same parental SCs, (Fisher's z-transformation comparison of correlations) (data not shown).

Analyses of the differences in gene expression between the four hiPSC lines and the parental SC lines revealed that 8,287 (out of 16,483) genes in the AM cells, 7,249 genes in the MRC cells, 7,465 genes in the PAE cells, and 6,314 genes in the UtE cells showed significant differences between the hiPSC lines and the corresponding parental lines, as determined using the Student's *t*-test (for a false discovery rate [FDR] < 5% and requiring a≥2.0 -fold change in expression between the cells). In total, 2,502 genes were categorized into a gene expression signature common to the above four gene sets with expression differences (Fig. 1B). In this expression signature, 62% of the genes (1549 genes) were upregulated and 38%

(953 genes) downregulated in the hiPSCs compared to the parental SCs.

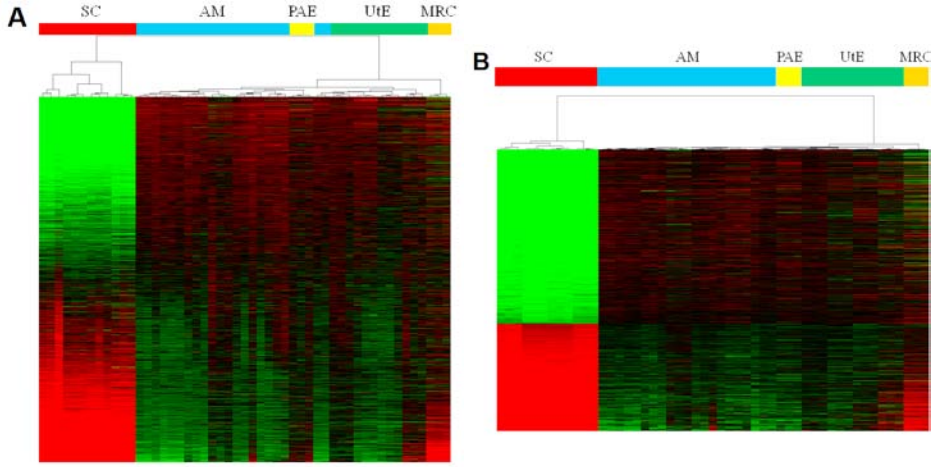


Figure 1 Gene Expression Signature. A: Heat map and hierarchical clustering for all cells and genes. Cell types are indicated by the bars, and the following abbreviations are used for the source of cell types of the human somatic cells (SCs) and induced pluripotent stem cells (hiPSCs): AM, amniotic membrane cell; PAE, placental artery cell; UtE, uterine endometrium cell; MRC, MRC-5 cell. **B**: Heat map of 2,502 genes in the hiPSCs and parental cells. Cell types are indicated by the bars.

3.2 Significant correlation between reprogramming and glycan biosynthesis based on network analysis

To elucidate the nature of the expression signature of the hiPSCs, we incorporated information on gene binding and function into a network analysis approach. To prepare the network analysis, we identified in the expression signature 146 regulatory networks of 313 genes, which were classified with their functions using the gene sets defined previously [11] (see also Methods), among 519 genes that were identified as being bound by four factors in ChIP-on-chip experiments [10]. We then analyzed the 146 reference networks, which were regarded as being directly induced by the four factors (OCT3/4, SOX2, KLF4, c-MYC), to define the network signature of the hiPSC, according to the two following thresholds: 1) the enrichment probability of the genes in the expression signature for each network; and 2) the consistency of the network structure in relation to the gene expression profile [9]. Thus, we defined as the network signature 28 networks of 76 genes that fulfilled these conditions (Fig. 2A).

As expected, the network signature almost completely covered the pathways that were previously implicated in the reprogramming of hiPSC pluripotency (Figs. 2A and B). For example, the relationship between reprogramming for pluripotency and signal transduction was emphasized for the TGF-β [13], Wnt [14], and MAPK

pathways [15]. In addition, pathways related to cell-cell interactions were implicated. Although the molecular mechanisms underlying the cell-cell interactions in the inner cellular states are less understood, several studies have reported the importance of cellular communication through the extracellular matrix with respect to changes in the cellular states, such as those that occur during development and differentiation [16]. Furthermore, relationships to cancer-related pathways were identified, which is consistent with the fact that the four factors induce various cancer cells [17]; this finding may be useful in the prevention of cancer induction by hiPSCs. Although several pathways in the network signature remain to be characterized, the network signature provides clues as to the molecular mechanisms underlying reprogramming for hiPSC pluripotency and self-renewal.

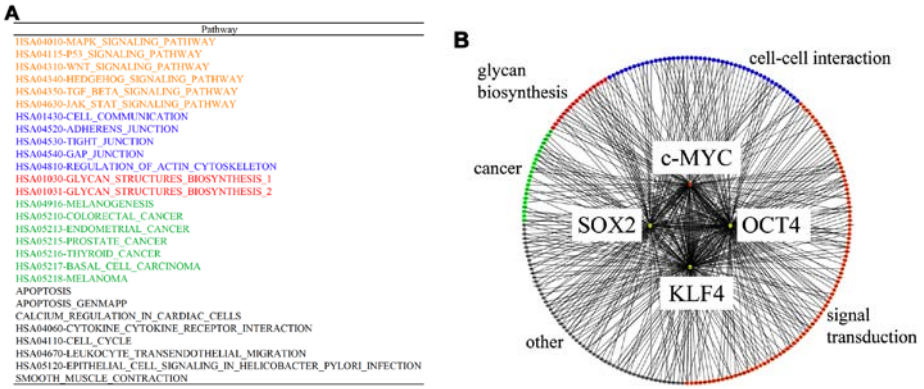


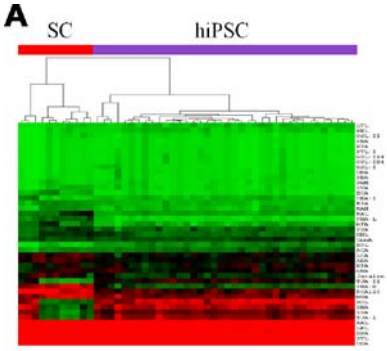
Figure 2 Network signature. **A**: List of network signatures. The pathways with significant probabilities are classified into the following categories: 2^{nd}-7^{th} row, signal transduction; 8^{th}-12^{th}, cell-cell interactions; 13^{th} and 14^{th}, glycan biosynthesis; 15^{th}-21^{st}, cancer; and otherwise, unclassified pathways. **B**: Schematic presentation of networks. The four induced factors are described in the center, and the binding genes, which are grouped according to the classification scheme described in **A**, are connected by thin lines.

Interestingly, two regulatory networks related to the glycome for linkage of the inner and outer cellular states appeared in the network signature (Fig. 2A). In general, glycan biosynthesis is a multi-step process that requires a variety of enzymes, i.e., glycosyltransferases and enzymes involved in cytosolic sugar metabolism, and in many cases, glycan biosynthesis follows a glycan-specific linear pathway. As most glycosyltransferases are regulated at the transcription level, an assessment of the transcriptional profile of glycan biosynthesis genes is warranted. In the two pathways, we found three genes (*ST6GAL1*, *B3GNT3*, and *GCNT2*) related to glycan transfer and two genes (*EXT1* and *HS6ST2*) related to heparan sulfate biosynthesis that were included in the expression signature. These findings are consistent with recent studies that have revealed associations between *N*-glycan and the maintenance of embryonic stem cell (ESC) pluripotency [18] and between heparan sulfate and the reprogramming of ESCs [19]. Therefore, the identified genes in the above two pathways are candidates for the maintenance of the outer

cellular state of iPSCs.

3.3 Unique glycan signatures of hiPSCs distinct from parent SCs

In addition to the expression and network signatures of the inner cell state, we examined the differences in the outer cellular states of the hiPSCs and parental SCs using a lectin array, which detects glycan structures on cell surface proteins based on glycan-lectin interactions [12]. In this analysis, the hiPSCs were clearly distinct from their parental SCs, and the dendrogram of the lectin microarray generated by unsupervised hierarchical clustering showed a clear separation between the hiPSCs and the parental SCs (Fig. 3A). Although the binding relationships between lectins and glycans and the relationships between the changes in glycan structures and the corresponding glycosyltransferases are redundant [20], we summarized the lectin-glycan-glycosyltransferase relationships using KEGG GLYCAN [21] and manual curation of previous papers. We found strong correlations between the gene expression profiles of the glycosyltransferases and the corresponding lectin fluorescence intensities (data not shown). This result indicates that glycosyltransferases are coordinately expressed with reprogramming, which the result that the hiPSCs bear glycan structures that are distinct from those of their parental SCs, reflecting reprogramming of the inner cellular state.



B

Lectin	Reaction	Transferase
AOL	Fuc a1-2 Gal	FUT2
MAL	Neu5Ac a2-3 Gal	ST3GAL1
TJA-II	Fuc a1-2 Gal	FUT2
WFA	GalNAc a1- Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
SNA	Neu5Ac a2-6 Gal	ST6GAL1
BPL	Gal b1-3 GalNAc	B3GALT5, B3GALT1
TJA-I	Neu5Ac a2-6 Gal	ST6GAL1
SSA	Neu5Ac a2-6 Gal	ST6GAL1
ABA	Gal b1-3 GalNAc	B3GALT5
	GalNAc a1- Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
MPA	GalNAc a1- Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
AAL	Fuc a1-2 Gal	FUT2
ACA	Gal b1-3 GalNAc	B3GALT5
Jacalin	Gal b1-3 GalNAc	B3GALT5
	GalNAc a1- Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
VVA	GalNAc a1- Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
STL	GlcNAc b1-3 Gal	B3GNT2
UDA	GlcNAc b1-3 Gal	B3GNT2

Figure 3 Glycan signature. **A**:Heat map and hierarchical clustering of lectins. The hiPSCs and parental SCs are discriminated. **B**: Correspondence between lectin gene expression patterns and glycan signatures. The lectin-glycosyltransferase relationships are described, together with their reactions. The lectins were selected under the condition that the corresponding glycosyltransferases were found in the expression signature.

Based on the Student's *t*-test (FDR <0.01) analysis, 28 of the 43 lectins in the lectin microarray showed significant differences between the hiPSCs and parental SCs (data not shown). We assigned 16 lectins to the glycan signature, which interacted with the 12 glycosyltransferases, which were related to the six patterns of glycan reactions, based on correspondence with the expression signature (Fig. 3B).

3.4 Molecular candidates for linkages between the inner and outer cellular states

Based on the correspondences between the expression and network signatures and between the expression and glycan signatures, we identified a total of 14 glycosyltransferases, owing to the appearance of *ST6GAL1* in both sets of correspondences. These glycosyltransferases are potential candidates for the linkage between the inner and outer cellular states in hiPSCs. Interestingly, these glycosyltransferases may be related to the biosynthesis of a glycolipid that is characteristic of hiPSCs. Indeed, allocation of the above glycosyltransferases to the pathways of "Glycan Biosynthesis and Metabolism" in KEGG GLYCAN revealed that the glycosyltransferases identified in the present study are important in glycolipid biosynthetic pathway. We identified *B3GALT5* in the biosynthetic pathway for the carbohydrate chains of the globo-series glycosphingolipids that carry the well-known SSEA-3 and SSEA -4 epitopes for ESCs and iPSCs, [22 ,23], and although *FUT2* is not directly involved in the synthesis of these glycans, it was found in the neighboring pathway that leads to the type IV H antigen. Furthermore, *B3GALT1* and *GCNT2*, in addition to *B3GALT5* and *FUT2*, were found in the extensive biosynthetic pathway of the carbohydrate chains of the lacto- and neolacto-series glycosphingolipids that carry SSEA-1, which is intensively expressed in ESCs but is absent in cells that have differentiated from ESCs [24]. In addition, members of the *GALNT* family, responsible for the *O*-glycan biosynthetic pathway of sialyl-T antigen, which is the most abundant glycan in several carcinoma cell lines, and *ST6GAL1* were only found in the *N*-glycan biosynthetic pathway, which is involved in the generation of cell-surface carbohydrate determinants and the differentiation antigens HB-6, CDw75, and CD76 [25]. These analyses identify the glycosyltransferases that are directly and indirectly related to known glycan epitopes, thereby indicating the key molecules and the marker epitopes involved in reprogramming.

Acknowledges

We thank to H. Abe, M. Yamazaki - Inoue and M. Machida, H. Sakaguchi for providing expert technical assistance. This work was supported by the project grant titled 'Development of Analysis Technology for Induced Pluripotent Stem (iPS) Cell', from The New Energy and Industrial Technology Development Organization (NEDO).

References

[1] Takahashi K, et al.: Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007, 131:861–872.
[2] Lowry WE, et al.: Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci USA* 2008, 105:2883–2888.
[3] Huangfu D, et al.: Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotechnol* 2008, 26:1269–1275.
[4] Kim J, Chu J, Shen X, Wang J, Orkin SH: An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 2008, 132:1049–1061.
[5] Chen X, et al.: Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008, 133:1106–1117.
[6] Mikkelsen TS, et al.: Dissecting direct reprogramming through integrative genomic

analysis. *Nature* 2008, 454:49–55.
[7] Makino H, et al.: Mesenchymal to embryonic incomplete transition of human cells by chimeric OCT4/3 (POU5F1) with physiological co-activator EWS. *Exp Cell Res* 2009, 288:2727-2740.
[8] Nagata S, et al.: Efficient reprogramming of human and mouse primary extra-embryonic cells to pluripotent stem cells. *Genes Cells* 2009, 14:1395–1404.
[9] Saito S, Aburatani S, Horimoto K: Network evaluation from the consistency of the graph structure with the measured data. *BMC Sys Biol* 2008, 2:84.
[10] Sridharan R, et al.: Role of the murine reprogramming factors in the induction of pluripotency. *Cell* 2009, 136:364–377.
[11] Subramanian A, et al.: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005, 102:15545–15550.
[12] Kuno A, et al.: Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat. Methods* 2005, 2:851-856.
[13] Lin T et al.: A chemical platform for improved induction of human iPSCs. *Nat Methods* 2009, 6:805-808.
[14] Sumi T, Tsuneyoshi N, Nakatsuji N, Suemori H: Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. *Development* 2008, 135:2969–2979.
[15] Eiselleova L, et al.: A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. *Stem Cells* 2009, 27:1847–1857.
[16] Discher DE, Mooney DJ, Zandstra PW: Growth Factors, matrices, and forces combine and control stem cells. *Science* 2009, 26:1673–1677.
[17] Yamanaka S: Elite and stochastic models for induced pluripotent stem cell generation. *Nature* 2009, 460:49-52.
[18] Satomaa T, et al.: The N-glycome of human embryonic stem cells. *BMC Cell Biol* 2009, **10**:42.
[19] Levenstein ME, et al.: Secreted proteoglycans directly mediate human embryonic stem cell-basic fibroblast growth factor 2 interactions critical for proliferation. *Stem Cells* 2008, 26:3099–3107.
[20] Gabius H-J: Glycans: bioactive signals decoded by lectins. *Biochem Soc Trans* 2008, **36**:1491–1496
[21] Hashimoto K, et al.: Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans. *Carbohydr Res* 2009, 12:881–887.
[22] Shevinsky L, Knowles BB, Damjanov I, Solter D: Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells. *Cell* 1982, 30:697–705.
[23] Kannagi R et al.: Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. *EMBO J* 1983, 2:2355–2361.
[24] Gooi HC, et al.: Stage-specific embryonic antigen involves a1→3 fucosylated type 2 blood group chains. *Nature* 1981, 292:156–158.
[25] Bert JE, et al.: The HB-6, CDw75, and CD76 differentiation antigens are unique cell-surface carbohydrate determinants generated by the ß-galactoside a2,6-sialyltransferase. *J Cell Biol* 1992, 116:423–435.