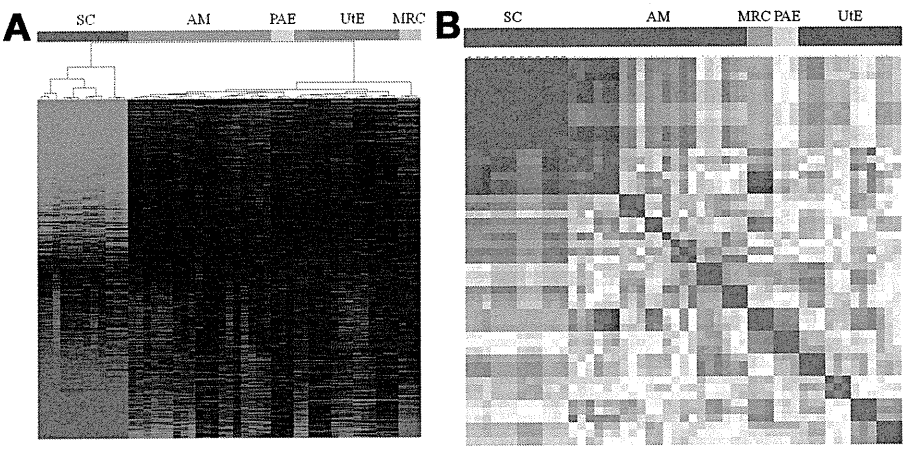


Figure 1 Classification of hiPSCs and SCs from gene expression data (*A*) Heat map and hierarchical clustering for all cells and genes. Cell types are indicated by colored bars, and the following abbreviations are used for the sources of the cell types of the human somatic cells (SCs) and induced pluripotent stem cells (hiPSCs): AM, amniotic membrane cell; PAE, placental artery cell; UtE, uterine endometrium cell; MRC, MRC-5 cell. Details are available in additional file 1. (*B*) Heat map of the correlation coefficient matrix for all of the cells. See also additional file 3.

hiPSCs from different passages were clustered more closely with each other, rather than with those of 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 the hiPSCs from different passages were more closely related to each other than to the hiPSCs from the same parental SCs (Fig. 1B and see also additional file 4: Correlation coefficient matrix for all cells). Furthermore, the above relationship between the hiPSCs and the parental SCs was verified by estimating the classification accuracy by leave-one-out cross-validation (LOOCV) on the nearest-neighbor classifier, based on Pearson's correlation distance (see additional file 5: Cross validation of cell classification).

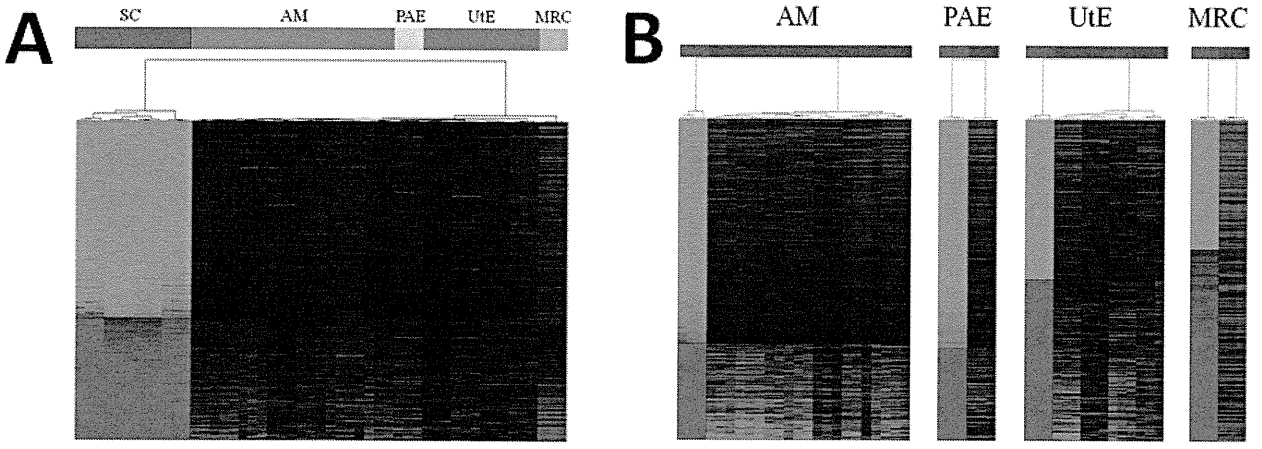
Gene expression signature for hiPSCs descended from different parent SCs

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) (Fig. 2A). In total, 2,502 genes were categorized into a gene expression signature common to the above four gene sets with expression differences (Fig. 2B and see also additional files 6: Number matrix for common genes, and 7: List of 2,502 genes in the expression signature, together with the fold-changes in expression levels and FDR values).

In this expression signature, 62% of the genes (1,549 genes) were upregulated and 38% (953 genes) were downregulated in the hiPSCs, as compared to the parental SCs (Fig. 2B and see also additional file 7). From the 953 genes in the gene signature that were expressed at lower levels, gene ontology analyses revealed 60 terms with significant probability (family-wise error rate [FWER] <0.05), whereas the 1,549 genes that were expressed at higher levels were characterized by 89 terms (see additional file 8: List of enriched GO terms with significant probabilities (FWER < 0.05)). In total, 149 terms were found, and the GO analysis was determined to be inadequate for defining the biological functions of the expression signature in hiPSCs. The 149 terms were summarized into 33 terms as a macroscopic view; these terms shared 9 terms between the higher and lower expression levels (Fig. 2C).

Network signature of hiPSCs by network screening

To elucidate the nature of the expression signature of the hiPSCs, we incorporated information on gene binding and function into a network analysis approach, named network screening [21] (see additional file 9: Schematic representation of the network screening used to obtain the network signature, and Methods). To prepare the network analysis, we identified 146 regulatory networks of 313 genes in the expression signature, which were classified with their functions using the gene sets defined previously [24] (see additional file 10: Reference networks and constituent genes, and Methods), among 519 genes that were identified as being bound by the four factors in ChIP-on-chip experiments [20]. 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



C

iPSC>SC		iPSC<SC	
GO TERM	FWER	GO TERM	FWER
multicellular organismal development	2.18E-23	**anatomical structure morphogenesis**	2.03E-08
cell differentiation	4.34E-18	**response to external stimulus**	8.06E-08
regulation of biological process	1.41E-16	**multicellular organismal development**	2.32E-07
cell communication	2.57E-14	response to stress	2.58E-07
cell-cell signaling	8.03E-12	**regulation of biological process**	5.70E-07
anatomical structure morphogenesis	7.75E-10	**cell proliferation**	7.84E-07
signal transduction	8.13E-08	**signal transduction**	3.03E-06
behavior	7.74E-07	cell death	3.15E-06
transport	6.69E-05	death	3.64E-06
cell proliferation	1.52E-04	**cell communication**	3.71E-05
response to external stimulus	4.09E-04	protein modification process	7.08E-04
embryonic development	1.91E-03	cell growth	6.74E-03
ion transport	3.78E-03	protein transport	2.20E-02
reproduction	6.27E-03	**cell differentiation**	3.96E-02
response to endogenous stimulus	1.90E-02	**biological_process**	4.07E-02
growth	2.67E-02	response to abiotic stimulus	4.08E-02
biological_process	3.11E-02		

Figure 2 Gene expression signature (*A*) Heat map of 2,502 genes in the hiPSCs and parental cells. Cell types are indicated by colored bars. (*B*) Heat maps for the MRC, AM, UtE, and PAE hiPSCs and their corresponding parental cells. The hiPSCs and the parental SCs are discriminated by colored bars (red, SCs; purple, hiPSCs). (*C*) GO terms in the Biological Process, with the significance probabilities for the genes with higher and lower levels of expression. GO terms were summarized into 33 terms, as a macroscopic view, and the overlapped terms in the higher and lower expression classes are indicated by bold characters. See also additional file 6.

network signature of the hiPSCs, according to the following two thresholds (Fig. 3): 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 [21]. Thus, as the network signature, we defined 28 networks of 76 genes that fulfilled these conditions (Fig. 3A and see also additional file 11: Details of the network signature).

As expected, the network signature almost completely covered the pathways that were previously implicated in the reprogramming of hiPSC pluripotency (Figs. 3A and B). For example, the relationship between reprogramming for pluripotency and signal transduction was emphasized for the TGF-β [25], Wnt [26], and MAPK pathways [27]. 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 highlighted 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 [28]. Furthermore, relationships to cancer-related pathways were identified, consistent with the fact that the four factors induce various cancer cells [29]; this finding may be useful in the prevention of cancer induction by hiPSCs. Although several pathways in the network signature remain to be characterized, it provides clues as to the molecular mechanisms underlying the reprogramming for hiPSC pluripotency and self-renewal, in contrast to the information obtained from a characterization based simply on GO terms.

Networks with significant correlation between reprogramming and glycan biosynthesis

Interestingly, two regulatory networks related to the glycome, for linkage of the inner and outer cellular states,

A

Pathway
HSA04010-MAPK_SIGNALING_PATHWAY
HSA04115-P53_SIGNALING_PATHWAY
HSA04310-WNT_SIGNALING_PATHWAY
HSA04340-HEDGEHOG_SIGNALING_PATHWAY
HSA04350-TGF_BETA_SIGNALING_PATHWAY
HSA04630-JAK_STAT_SIGNALING_PATHWAY
HSA01430-CELL_COMMUNICATION
HSA04520-ADHERENS_JUNCTION
HSA04530-TIGHT_JUNCTION
HSA04540-GAP_JUNCTION
HSA04810-REGULATION_OF_ACTIN_CYTOSKELETON
HSA01030-GLYCAN_STRUCTURES_BIOSYNTHESIS_1
HSA01031-GLYCAN_STRUCTURES_BIOSYNTHESIS_2
HSA04916-MELANOGENESIS
HSA05210-COLORECTAL_CANCER
HSA05213-ENDOMETRIAL_CANCER
HSA05215-PROSTATE_CANCER
HSA05216-THYROID_CANCER
HSA05217-BASAL_CELL_CARCINOMA
HSA05218-MELANOMA
APOPTOSIS
APOPTOSIS_GENMAPP
CALCIUM_REGULATION_IN_CARDIAC_CELLS
HSA04060-CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION
HSA04110-CELL_CYCLE
HSA04670-LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION
HSA05120-EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTION
SMOOTH_MUSCLE_CONTRACTION

B

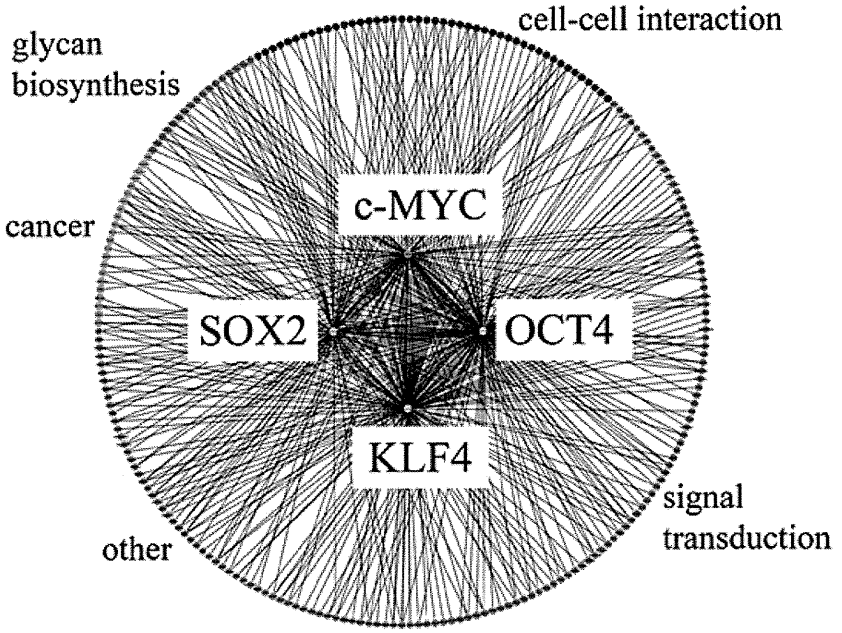


Figure 3 Network signature (*A*) List of network signatures. The pathways with significant probabilities are classified into the following categories: orange, pathways related to signal transduction; blue, pathways related to cell-cell interactions; red, pathways related to glycan biosynthesis; green, pathways related to cancer; and black, unclassified pathways. (*B*) Schematic presentation of networks. The four induced factors are described in the center, and the binding genes, which are colored according to the classification scheme described in (*A*), are connected by thin lines.

HSA01030-GLYCAN_STRUCTURES_BIOSYNTHESIS_1	**ALG3**, **ALG8**, B4GALT3, **DPAGT1**, **EXT1**, GALNT7, HS3ST3B1, **HS6ST2**, **MAN1A2**, MGAT1, **OGT**, **ST6GAL1***,STT3B
HSA01031-GLYCAN_STRUCTURES_BIOSYNTHESIS_2	A4GALT, **B3GNT3***, B3GNT5, B4GALT3, **GCNT2***, ST8SIA1,UGCG

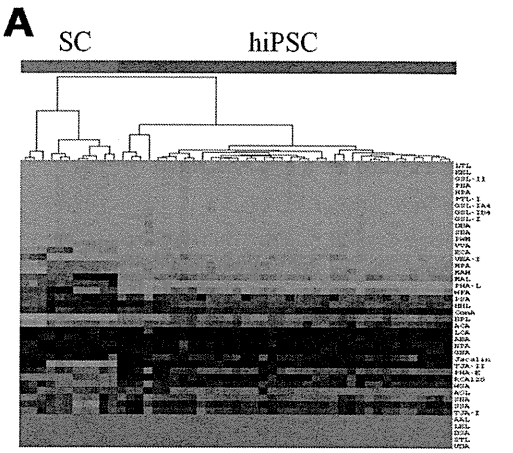
Figure 4 Genes involved in two glycome biosynthesis pathways The genes found in the expression signature are indicated by bold characters. Three genes related to glycan transfer are indicated by asterisks.

appeared in the network signature (Fig. 3A). 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. Most glycosyltransferases are regulated at the transcriptional level, thus warranting an assessment of the transcriptional profile of the glycan biosynthesis genes. 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 (Fig. 4). These findings are consistent with recent studies that revealed the association between *N*-glycans and the maintenance of embryonic stem cell (ESC) pluripotency [5] and that between heparan sulfate and the reprogramming of ESCs [30]. Therefore, the genes identified in the above two pathways are candidates for the maintenance of the outer cellular state of iPSCs.

Glycan signature unique to hiPSCs

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 the parental SCs using a lectin array, which detects glycan structures on cell surface proteins, based on glycan-lectin interactions [31]. 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. 5A). Although the binding relationships between lectins and glycans and the relationships between the changes in glycan structures and the corresponding glycosyltransferases are redundant [32], we summarized the lectin-glycan-glycosyltransferase relationships using KEGG GLYCAN [33] and by manual curation of previous reports. We found strong correlations between the gene expression profiles of the glycosyltransferases and the corresponding lectin fluorescence intensities (see additional file 12: Lectin-glycan-glycosyltransferase relationships and correlations of lectin array intensities with glycosyltransferase expression patterns). This result indicates that the glycosyltransferases are coordinately expressed with the reprogramming, with the result that the hiPSCs bear glycan structures that are distinct from those of their parental SCs, reflecting the reprogramming of the inner cellular state.

Based on the Student's *t*-test (FDR <0.05) analysis, 28 of the 43 lectins in the lectin microarray showed significant differences between the hiPSCs and the parental SCs (see also additional file 12). For the glycan signature, we assigned 16 lectins, which interacted with the 12 glycosyltransferases that were related to the six patterns of glycan reactions, based on the correspondence with the expression signature (Fig. 5B).



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 5 Glycan signature (*A*) Heat map and hierarchical clustering of lectins. The hiPSCs and parental SCs are depicted by colored bars (red, SCs; purple, hiPSCs). (*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.

Candidates of possible 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, since ST6GAL1 appeared 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 (see additional file 13: Knowledge-based relationships between glycosyltransferases and their biosynthetic pathways). Indeed, the allocation of the above glycosyltransferases to the pathways of "Glycan Biosynthesis and Metabolism" in KEGG GLYCAN (Table 1 and see also additional file 14: Locations of the glycosyltransferases detected in the present study in the pathways of "Glycan Biosynthesis and Metabolism") revealed that the glycosyltransferases identified in the present study are important in the glycolipid biosynthetic pathway. We identified B3GALT5 in the biosynthetic pathway for the carbohydrate chains of the globo-series of glycosphingolipids bearing the well-known SSEA-3 and SSEA-4 epitopes for ESCs and iPSCs [34,35], 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 [36]. In addition, the 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 [37]. These analyses identified the glycosyltransferases that are directly and indirectly related to known glycan epitopes, thereby indicating the key molecules and the marker epitopes involved in reprogramming.

Further remarks on the present study

We analyzed more than 50 hiPSCs that were originally established from parental SCs, and the correspondence between each hiPSC and its parental SC was strictly controlled, which supports the present results based on a comparison with a clear genetic relationship. To further clarify the molecular mechanisms of the pluripotency, embryonic stem cells (ESCs) should be analyzed, following the context of the present study. Indeed, the pluripotency of hiPSCs has been extensively evaluated with reference to that of human ESCs, by various comparisons [38-43]. At present, we have prepared more than 100 hiPSCs with higher passages, and their comparisons with ESCs will be reported in the near future.

As for the experimental measurements, two types of data, gene expression and glycan structure, were analyzed by using microarrays and lectin arrays in the present study. To comprehensively understand the features of hiPSCs, more experimental data should be utilized, such as DNA-methylation and mi-RNA data. In particular, the recent availability of the next-gen sequencer will produce RNA-seq and ChIP-seq data with more

Table 1 Relationships between glycosyltransferase expression, network, and glycan signature

Glycosyltransferase	Functions	Glycan structure
ST6GAL1	N-, O-Glycan and glycolipid biosynthesis	Siaa2,6Galb1,4GlcNAc-R
B3GNT3	O-Glycan biosynthesis	core1 extension
GCNT2	N-, O-Glycan and glycolipid biosynthesis	I antigen Siaa2,3Galb1,3GalNAca1-
ST3GAL1	O-Glycan biosynthesis	Ser/Thr
FUT2	N-, O-Glycan and glycolipid biosynthesis	H antigen
GALNT6	O-Glycan biosynthesis	GalNAca1-Ser/Thr
GALNT8	O-Glycan biosynthesis	GalNAca1-Ser/Thr
GALNT10	O-Glycan biosynthesis	GalNAca1-Ser/Thr
GALNT12	O-Glycan biosynthesis	GalNAca1-Ser/Thr
GALNT14	O-Glycan biosynthesis	GalNAca1-Ser/Thr
GALNTL2	Unknown	
B3GALT5	N-, O-Glycan and glycolipid biosynthesis	Galb1,3GlcNAc-R, SSEA-3
B3GALT1	N-, O-Glycan and glycolipid biosynthesis N- and O-Glycan, keratan sulfate	Galb1,3GlcNAc-R
B3GNT2	biosynthesis	polylactosamine

The fourteen glycosyltransferases with identified correspondences between expression, network, and glycan signatures were allocated to biosynthetic pathways, using the KEGG GLYCAN database with modifications. The names of the pathways are listed. See also additional file 12 for the detailed pathways of notable glycosyltransferases, according to the KEGG GLYCAN database.

accurate measurements of gene expression and concrete information about the regulated genes. In addition, vast amounts of protein interaction data are accumulating. A comprehensive analysis integrating the various data from more hiPSCs will be reported in the near future.

Conclusions

The present study is the first to reveal the relationships between gene expression patterns and cell surface changes in hiPSCs, and it reinforces the importance of the cell surface to identify established iPSCs from SCs. In addition, given the variability of iPSCs, which is related to the characteristics of the parental SCs, a glycosyltransferase expression assay should be established that allows more precise definition of hiPSCs and facilitates their standardization, which are important steps towards eventual therapeutic applications of hiPSCs.

Methods

Cell experiments

Somatic cell pellets were harvested by scraping. The hiPSCs were incubated at 37°C, in a solution containing 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.

MRC-5 and amniotic mesodermal (AM) cells were maintained in POWEREDBY 10 medium (MED Shiratori Co., Ltd., Tokyo, Japan). The human placental artery endothelial (PAE) cells were harvested from human placenta. To isolate the arterial endothelium, we used the explant culture method, in which the cells were outgrown from pieces of the placenta's arterial vessels. Briefly, arterial vessels were separated from arteries in the chorionic plate, and chopped into approximately 5-mm^3 pieces. The pieces were washed in endothelial basal medium-2 (EBM-2; Cambrex, Walkersville, MD) and cultured in EGM-2MV medium (Cambrex), which consisted of EBM-2, 5% fetal bovine serum (FBS), and the supplemental growth factors VEGF, bFGF, EGF, and IGF. The arterial vessels attached to the substrata of the culture dishes (BD Falcon; Becton Dickinson, San Jose, CA). Cells migrated out from the surface of the tissues after about 20 days of incubation at 37°C in 5% CO_2. The cells were harvested in PBS containing 0.1% trypsin and 0.25 mM EDTA, and were re-seeded at a density of 3×10^5 cells in a 10-cm dish. Confluent monolayers of cells were subcultured. The culture medium was replaced every 3-4 days. Human uterine endometrium (UtE) was harvested from a patient with endometriosis. The endometrium was sterilized in PBS and cut into small pieces with dissection scissors. These pieces were placed in Dulbecco's Modified Eagle's Medium (Sigma Chemical Co. St. Louis, MO), supplemented with 10% FBS and an antibiotic-antimycotic (100×) solution (Invitrogen), and incubated for 10-14 days at 37°C in a humidified 5% CO_2 atmosphere. Subconfluent adherent cells were harvested in PBS containing 0.06% trypsin and 0.005% EDTA, and were subcultured. The culture medium was replaced every 4 days. This study was approved by the Ethical Committee of the National Institute for Child Health and Development. The purpose of this study was explained thoroughly to the patients, who gave their written informed consent.

hiPSCs were cultivated on irradiated MEFs in iPSellon medium (Cardio, Osaka, Japan), supplemented with 10 ng/ml recombinant human bFGF (Wako Pure Chemicals, Osaka, Japan). hiPSCs were established from MRC-5 and AM cells, as previously described [21,22]. In addition, hiPSCs were established from PAE and UtE cells in the present study. Briefly, 1×10^5 cells were infected overnight with pooled viral supernatants, obtained by the transfection of HEK293FT cells (TransIT-293 reagent; Mirus, Madison, WI) with the retroviral vector pMXs, which encodes the cDNAs for OCT3/4, SOX2, KLF4, and c-MYC, together with the packaging plasmids pCLGagPol and pHCMV-VEV-G (a gift from T. Kiyono, National Cancer Center Research Institute, Tokyo, Japan). Four days after infection, the cells were split, plated on irradiated MEFs in 100-mm dishes, and maintained in iPSellon medium until colonies formed.

The immunocytochemical analysis was performed as described previously [22,23]. Human cells were fixed with 4% paraformaldehyde in PBS for 10 min at 4 °C. After washing with 0.1% Triton X-100 in PBS (PBST), the cells were prehybridized in blocking buffer for 1–12 h at 4 °C, and then incubated for 6–12 h at 4°C with the following primary antibodies: anti-SSEA4 (1 : 300 dilution; Chemicon, Temecula, CA), anti-TRA-1–60 (1 : 300; Chemicon), anti-Oct4 (1 : 50; Santa Cruz Biotechnology, Santa Cruz, CA), anti-Nanog (1 : 300; ReproCELL, Tokyo, Japan), and anti-Sox2 (1 : 300; Chemicon). The cells were then incubated with anti-rabbit IgG, anti-mouse IgG or anti-mouse IgM conjugated with Alexa Fluor 488 or Alexa Fluor 546 (1: 500; Molecular Probes, Eugene, OR) in blocking buffer for 1 h at room temperature. The cells were counterstained with DAPI, and then mounted using a SlowFade light antifade kit (Molecular Probes).

Teratoma formation was performed as described previously [22,23]. The 1:1 mixtures of the AM-hiPSC suspension and Basement Membrane Matrix (BD Biosciences, San Jose, CA) were implanted subcutaneously, at 1.0×10^7 cells / site, into immunodeficient,

non-obese diabetic (NOD)/severe combined immunodeficiency (SCID) mice (CREA, Tokyo, Japan). Teratomas were surgically dissected out 6–10 weeks after implantation, and were fixed with 4% paraformaldehyde in PBS and embedded in paraffin. Sections of 10-μm thickness were stained with hematoxylin-eosin.

Gene expression analysis

Total RNA samples were extracted using ISOGEN (NipponGene). The global gene expression patterns and changes in mRNA levels were monitored using Agilent Whole Human Genome Microarray chips (G4112F) with one-color (Cyanine 3) dye. This microarray chip covers 41,000 well-characterized human genes and transcripts. The raw microarray data were submitted to the GEO (Gene Expression Omnibus) microarray data archive (http://www.ncbi.nlm.nih.gov/geo/) at the NCBI (accession number: GSE 20750). 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.

Among the 41,000 probes, 16,483 representative probes corresponding to MAQC unique genes were used for the following analyses [44]. Global array clustering was performed by the complete linkage method with Euclidean distance, and was visualized using the Java TreeView 1.1.0 software; the gene expression values are displayed as normalized log ratios. Cell line similarities were measured using Pearson correlation coefficients. To further validate whether the global gene expression is different in each origin cell, we evaluated the classification accuracy by leave-one-out cross-validation (LOOCV) on the nearest-neighbor classifier, based on Pearson's correlation distance. To obtain the expression signatures, we performed a differential analysis for each origin cell: differences between the two arbitrary datasets were evaluated by the Student's t-test for the expression of each gene. Thereafter, the false discovery rate (FDR) was estimated using the Benjamini–Hochberg procedure. Differentially expressed genes were selected if they satisfied both FDR <0.05 and a 2.0-fold change in the average values for the cell lines being compared. The gene ontology analysis was performed using the GO Term Finder Perl script [45] (http://go.princeton.edu/cgi-bin/GOTermFinder), with EBI human GO annotations and generic GO slim annotations (http://www.geneontology.org/).

Network screening

Network screening was performed as described previously [21]. This analysis is based on the procedure for estimating the consistency of a network structure (directed acyclic graph) with the measured data for the constituent variables in the graph. The joint density function for a given network (reference network) was recursively factorized into conditional density functions, according to the parent-child relationship in the graph. The conditional functions were quantified into log-likelihoods, using linear regression for the measured data, with the assumption that the data followed a normal distribution. The probability of the log-likelihood for the network structure (graph consistency probability; GCP) was then estimated from the distribution of log-likelihoods for 2,000 networks, generated under the condition that the networks shared the same numbers of nodes and edges as those of the given network. The significance probability of the given network was set at 0.05 in this analysis.

In the present study, the GCP was estimated for the ensemble of reference networks, to extract the candidate activated networks in the hiPSCs, in a process termed 'network screening'. The reference networks were constructed using the ChIP-on-Chip data and the classification scheme for gene function. The genes bound by four factors were cited from a previous report [20], and were divided into sub-networks according to the functional gene sets previously defined in the Molecular Signatures Database (MSigDB) [24]. The sub-networks that included at least one gene of the expression signature were then selected. The set of selected sub-networks was used as the reference network for network screening.

Glycan analysis

We analyzed cell surface glycans with a lectin microarray [31]. The 43 lectins were dissolved at a concentration of 0.5 mg/ml in spotting solution (Matsunami Glass, Osaka, Japan), and were spotted onto epoxysilane-coated glass slides (Nexterion Slide E Epoxysilane-coated Substrate 25 × 75.6 × 1 mm; Schott, Mainz, Germany) attached to a silicone rubber sheet, using a non-contact microarray printing robot (MicroSys 4000; Genomic Solutions, Ann Arbor, MI). The lectins were spotted in triplicate, with a spot diameter of 500 μm. The glass slides were incubated at 25°C for 3 h, to allow lectin immobilization. The lectin-immobilized glass slides were then washed with probing buffer (25 mM Tris-HCl [pH 7.5], 140 mM NaCl, 2.7 mM KCl, 1 mM CaCl2, 1 mM MnCl2, 1% [v/v] Triton X-100), and incubated with the blocking reagent N102 (NOF, Tokyo, Japan) at 20°C for 1 h. Finally, the lectin-immobilized glass slides were flooded with TBS containing 0.1% NaN_3 and stored at 4°C. The cell membrane faction was prepared using the CelLytic MEM Protein Extraction Kit (Sigma-Aldrich, Tokyo, Japan), and the protein concentration was determined using the MicroBCA Protein Assay Reagent kit (Thermo Fisher Scientific, Waltham,

MA). After dilution in PBST (10 mM PBS [pH 7.4], 140 mM NaCl, 2.7 mM KCl, 1% Triton X-100), the cell membrane fraction was labeled with Cy3 NHS ester (GE Healthcare Ltd., Buckinghamshire, England). After dilution in probing buffer to the desired concentration, the Cy3-labeled cell membrane fraction was applied to the lectin microarray and incubated at 20°C overnight. After washing with the probing buffer, fluorescence images were acquired using an evanescent-field activated fluorescence scanner (SC-Profiler; GP BioScience, Kanagawa, Japan). The fluorescence signal of each spot was quantified using the Array Pro Analyzer ver. 4.5 software (Media Cybernetics, Bethesda, MD), and the background value was subtracted. The values shown for the lectin signals represent the average of triplicate spots.

Additional files

There are 14 additional files in the present analysis. For convenience, we provide an overview of the additional files. Additional files 1, 2, 3, 4, 5 are related to the cell classification in Figure 1: the details of the cell lines and their experimental establishment are described in files 1 and 2, and the details of the analyses of the expression data are described in files 3-5. Additional figures 6-8 are related to the gene expression signature in Figure 2: the details of the analyzed data are described in files 6 and 7, and the results obtained by a standard analysis are described in file 8. Additional files 9, 10, 11 are related to the network signature in Figure 3: the methodological aspects of the network screening are described in files 9 and 10, and in file 11, the detailed results are presented. Additional files 12, 13, 14 are related to the glycan signature in Figure 5: all of the information for interpreting the analyzed results is presented in the three files.

Additional material

Additional file 1: Cell lines and numbers of passages analyzed in the present study. The following abbreviations are used for the human somatic cell (SC) and induced pluripotent stem cell (hiPSC) sources: AM, amniotic membrane; PAE, placental artery endothelial; UtE, uterine endometrium; and MRC, MRC-5 cell line. The AM and MRC cell lines were named previously [22,23]. The number of passages for each cell line is indicated by the letter 'p' followed by an Arabic number.

Additional file 2: Generation of iPSCs from human PAE cells. (*A*) PAE cells from the arterial endothelium of a human placenta (a), and generation of hiPSCs through epigenetic reprogramming by retrovirus infection-mediated expression of OCT4, SOX2, KLF4, and c-MYC (b). (*B*) Expression patterns of the pluripotent cell markers, TRA-1-60, SSEA-4, NANOG, OCT3/4, and SOX2. The cell nuclei were stained with DAPI. (*C*) Hematoxylin-eosin staining of sections of teratomas generated by PAE-hiPSC implantation. The histological examination revealed that the tumors contain neural tissues (a: ectoderm), cartilage (b: mesoderm), and a gut-like epithelial tissue (c: endoderm).

Additional file 3: Clustering for all cells by another method. Another clustering was performed by the WARD method, instead of the complete linkage method of Figure 1, with Euclidean distance, and was visualized using the Java TreeView 1.1.0 software. The gene expression values are displayed as normalized log ratios. The abbreviations used are the same as those listed in Figure 1 and additional file 1.

Additional file 4: Correlation coefficient matrix for all cells. Pearson's correlation coefficients between 51 cells for the expression profiles of all genes were calculated. The abbreviations used are the same as those listed in Figure 1 and additional file 1.

Additional file 5: Cross-validation of cell classification. The classification accuracy was evaluated by leave-one-out cross-validation (LOOCV) on the nearest-neighbor classifier, based on the Pearson's correlation distance.

Additional file 6: Number matrix for common genes. The numbers of genes that were different between the iPSCs and SCs are listed on the diagonal of the matrix, and those that were shared between the four gene sets that showed expression differences between the iPSCs are listed above the diagonal. The abbreviations used are the same as those listed in Figure 1.

Additional file 7: List of 2,502 genes in the expression signature, together with the fold-changes in expression levels and FDR values. The fold-change values are listed for the minimum values among the four sets of comparisons between iPSCs and SCs (+, iPSCs>SCs; -, iPSCs<SCs), and the FDR values shown are the maximum values among these sets.

Additional file 8: List of enriched GO terms with significant probabilities (FWER < 0.05).

Additional file 9: Schematic representation of the procedure used to obtain the network signature. The procedure for obtaining the network signature from the expression signature is shown schematically. The detailed procedure is as follows: 1) We first prepare the information for the gene sets to which the transcriptional factors bind, as deduced from the ChIP-on-chip experiments [20]; 2) Next, we prepare the information for the gene sets that were classified using knowledge of biological functions [24]; 3) The large gene sets in step 1 are divided into smaller subsets, according to the classification scheme of the gene sets in step 2; 4) If at least one gene in the expression signature is included in each gene subset in step 3, then the subset is regarded as a reference network; 5) In each reference network, the enrichment probability of the genes in the expression signature is tested with a significance probability of 0.05. Thus, we narrow down the network signature from the reference networks, in terms of gene numbers; 6) The significant reference networks identified in step 5 are further tested by calculating the graph consistency probability, which assesses the consistency between the network structure and the expression data for the constituent genes [24]. In this step, we further refine the network signature, in terms of both the network structure and the extent of gene expression; 7) Finally, we define the network signature, using the reference networks that passed the tests in steps 5 and 6.

Additional file 10: Reference networks and constituent genes.

Additional file 11: Details of the network signature. The characters in the above list are colored, according to the classification of biological function shown in Figure 2A.

Additional file 12: Lectin-glycan-glycosyltransferase relationships and correlations of lectin array intensities with glycosyltransferase expression patterns. Lectins with FDR<0.05 are colored red. The glycosyltransferases in the expression signature are indicated by a circle in the column "Expression signature" in "Gene expression". The Pearson's correlation coefficients between the lectin signal intensities and the expression profiles of the corresponding glycosyltransferases are listed, together with the significance probabilities. The original lectin array data can be obtained by request to HT or JH.

Additional file 13: Knowledge-based relationships between glycosyltransferases and their biosynthetic pathways.

Additional file 14: Locations of the glycosyltransferases detected in the present study in the pathways of "Glycan Biosynthesis and Metabolism". The glycosyltransferases listed in Table 1 were allocated to the pathways in "1.7 Glycan Biosynthesis and Metabolism" of the KEGG GLYCAN program (http://www.genome.jp/kegg/pathway.html#glycan). The glycosyltransferases and epitopes related to differentiation are indicated by red-colored boxes and red lines, respectively, in each pathway (see the text for details).

Acknowledgements
We thank H. Abe, M. Yamazaki-Inoue, M. Machida, and H. Sakaguchi for providing expert technical assistance. This work was supported by the project grant entitled 'Development of Analysis Technology for Induced Pluripotent Stem (iPS) Cell', from The New Energy and Industrial Technology Development Organization (NEDO).
This article has been published as part of *BMC Systems Biology* Volume 5 Supplement 1, 2011: Selected articles from the 4th International Conference on Computational Systems Biology (ISB 2010). The full contents of the supplement are available online at http://www.biomedcentral.com/1752-0509/5?issue=S1.

Author details
[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.

Authors' contributions
S.S. (computational analysis and manuscript preparation), Y.O. (cell experiments and DNA microarray), Y.I. (DNA microarray and manuscript preparation), H.T. (lectin microarray and manuscript preparation), M.T. (cell experiments and manuscript preparation), H.A. (cell experiments), K.N. (cell experiments), E.C. (cell experiments), Y.F. (cell experiments), Y.M. (vector construction), H.O. (vector construction), N.K. (vector construction), Y.S. (lectin microarray), A.U. (supervision of cell experiments), J.H. (lectin microarray), K.H. (computational analysis and manuscript preparation), and M.A. (project leader and coordination).

Competing interests
The authors declare that they have no competing interests.

Published: 20 June 2011

References

1. Takahashi K, *et al*: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell* 2007, **131**:861-872.
2. Muramatsu T, *et al*: **Carbohydrate antigens expressed on stem cells and early embryonic cells.** *Glycocon J* 2004, **21**:41-45.
3. Schopperle WM, *et al*: **The Tra-1-60 and Tra-1-81 human pluripotent stem cell markers are expressed on podocaluxin in embryonal carcinoma.** *Stem Cells* 2006, **25**:723-730.
4. Natunen S, *et al*: **The binding specificity of the marker antibodies Tra-1-60 and Tra-1-81 reveals a novel pluripotency associated type 1 lactosamine epitope.** *Glycobiology* 2010.
5. Satomaa T, *et al*: **The N-glycome of human embryonic stem cells.** *BMC Cell Biol* 2009, **2**:10.
6. Toyoda M, *et al*: **Lectin microarray analysis of pluripotent and multipotent stem cells.** *Genes to Cells* 2010, **16**:1-11.
7. Chen T, *et al*: **E-cadherin-Mediated Cell-Cell Contact is Critical for Induced Pluripotent Stem Cell Generation.** *Stem Cells* 2010, **28**:1315-25.
8. Tchieu J, *et al*: **Female human iPSCs retain an inactive X chromosome.** *Cell Stem Cell* 2010, **7**:329-42.
9. Chin MH, *et al*: **Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures.** *Cell Stem Cell* 2009, **5**:111-123.
10. Guenther MG, *et al*: **Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells.** *Cell Stem Cell* 2010, **7**:249-257.
11. Newman AM, *et al*: **Lab-specific gene expression signatures in pluripotent stem cells.** *Cell Stem Cell* 2010, **7**:258-262.
12. Chin MH, *et al*: **Molecular analyses of human induced pluripotent stem cells and embryonic stem cells.** *Cell Stem Cell* 2010, **7**:263-9.
13. Lowry WE, *et al*: **Generation of human induced pluripotent stem cells from dermal fibroblasts.** *Proc Natl Acad Sci USA* 2008, **105**:2883-2888.
14. Maherali N, *et al*: **A high-efficiency system for the generation and study of human induced pluripotent stem cells.** *Cell Stem Cell* 2008, **3**:340-345.
15. Yu J, *et al*: **Human induced pluripotent stem cells free of vector and transgene sequences.** *Science* 2009, **324**:797-801.
16. Boyer LA, *et al*: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122**:947-56.
17. Loh YH, *et al*: **The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.** *Nat Genet* 2006, **38**:431-40.
18. Chen X, *et al*: **Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.** *Cell* 2008, **133**:1106-17.
19. Kim J, *et al*: **An extended transcriptional network for pluripotency of embryonic stem cells.** *Cell* 2008, **132**:1049-61.
20. Sridharan R, *et al*: **Role of the murine reprogramming factors in the induction of pluripotency.** *Cell* 2009, **136**:364-77.
21. 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.
22. 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.
23. Nagata S, *et al*: **Efficient reprogramming of human and mouse primary extra-embryonic cells to pluripotent stem cells.** *Genes Cells* 2009, **14**:1395-1404.
24. 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.
25. Lin T, *et al*: **A chemical platform for improved induction of human iPSCs.** *Nat Methods* 2009, **6**:805-808.
26. 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.
27. 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.
28. Discher DE, Mooney DJ, Zandstra PW: **Growth Factors, matrices, and forces combine and control stem cells.** *Science* 2009, **26**:1673-1677.
29. Yamanaka S: **Elite and stochastic models for induced pluripotent stem cell generation.** *Nature* 2009, **460**:49-52.
30. Sasaki N, *et al*: **Heparan sulfate regulates self-renewal and pluripotency of embryonic stem cells.** *J Biol Chem* 2008, **283**:3594-606.
31. Kuno A, *et al*: **Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling.** *Nat. Methods* 2005, **2**:851-856.
32. Gabius H-J: **Glycans: bioactive signals decoded by lectins.** *Biochem Soc Trans* 2008, **36**:1491-1496.
33. Hashimoto K, *et al*: **Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans.** *Carbohydr Res* 2009, **12**:881-887.
34. 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.
35. 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.
36. Gooi HC, *et al*: **Stage-specific embryonic antigen involves a1⊠3 fucosylated type 2 blood group chains.** *Nature* 1981, **292**:156-158.
37. Bert JE, *et al*: **The HB-6, CDw75, and CD76 differentiation antigens are unique cell-surface carbohydrate determinants generated by the ß-galactoside α2,6-sialyltransferase.** *J Cell Biol* 1992, **116**:423-435.
38. Chin MH, *et al*: **Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures.** *Cell Stem Cell* 2009, **5**:111-123.
39. Marchetto MC, *et al*: **Transcriptional signature and memory retention of human-induced pluripotent stem cells.** *PLoS One* 2009, **4(9)**:e7076.

40. Chin MH, Pellegrini M, Plath K, Lowry WE: **Molecular Analyses of Human Induced Pluripo-tent Stem Cells and Embryonic Stem Cells.** *Cell Stem Cell* 2010, 7:263-269.
41. Newman AM, Cooper JB: **Lab-Specific Gene Expression Signatures in Pluripotent Stem Cells.** *Cell Stem Cell* 2010, 7:258-262.
42. Guenther MG, *et al*: **Chromatin Structure and Gene Expression Programs of Human Embryonic and Induced Pluripotent Stem Cells.** *Cell Stem Cell* 2010, 7:249-257.
43. Ghosh Z, *et al*: **Persistent Donor Cell Gene Expression among Human Induced Pluripotent Stem Cells Contributes to Differences with Human Embryonic Stem Cells.** *PLoS One* 2010, 5(2):e8975.
44. MAQC Consortium: **The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.** *Nat Biotechnol* 2006, 24:1151-1161.
45. Boyle EI, *et al*: **GO::TermFinder–open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes.** *Bioinformatics* 2004, 20:3710-3715.

doi:10.1186/1752-0509-5-S1-S17
Cite this article as: Saito *et al.*: Possible linkages between the inner and outer cellular states of human induced pluripotent stem cells. *BMC Systems Biology* 2011 **5**(Suppl 1):S17.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 23, pp. 20345–20353, June 10, 2011

Glycome Diagnosis of Human Induced Pluripotent Stem Cells Using Lectin Microarray*[S]

Received for publication, February 15, 2011, and in revised form, March 28, 2011 Published, JBC Papers in Press, April 6, 2011, DOI 10.1074/jbc.M111.231274

Hiroaki Tateno[‡], Masashi Toyota[§], Shigeru Saito[¶‖], Yasuko Onuma[**], Yuzuru Ito[**], Keiko Hiemori[‡], Mihoko Fukumura[‡], Asako Matsushima[‡], Mio Nakanishi[‡‡], Kiyoshi Ohnuma[‡‡], Hidenori Akutsu[§], Akihiro Umezawa[§], Katsuhisa Horimoto[¶], Jun Hirabayashi[‡1], and Makoto Asashima[‡‡]**

From the [‡]Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, the [§]Department of Reproductive Biology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan, the [¶]Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan, [‖]Infocom Corporation, Sumitomo Fudosan Harajuku Building, 2-34-17, Jingumae, Shibuya-ku, Tokyo 150-0001, Japan, the [**]Research Center for Stem Cell Engineering, National Institute of Advanced Industrial Science and Technology, Tsukuba Central 4, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8562, Japan, and the [‡‡]Department of Life Sciences (Biology), Graduate School of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

Induced pluripotent stem cells (iPSCs) can now be produced from various somatic cell (SC) lines by ectopic expression of the four transcription factors. Although the procedure has been demonstrated to induce global change in gene and microRNA expressions and even epigenetic modification, it remains largely unknown how this transcription factor-induced reprogramming affects the total glycan repertoire expressed on the cells. Here we performed a comprehensive glycan analysis using 114 types of human iPSCs generated from five different SCs and compared their glycomes with those of human embryonic stem cells (ESCs; nine cell types) using a high density lectin microarray. In unsupervised cluster analysis of the results obtained by lectin microarray, both undifferentiated iPSCs and ESCs were clustered as one large group. However, they were clearly separated from the group of differentiated SCs, whereas all of the four SCs had apparently distinct glycome profiles from one another, demonstrating that SCs with originally distinct glycan profiles have acquired those similar to ESCs upon induction of pluripotency. Thirty-eight lectins discriminating between SCs and iPSCs/ESCs were statistically selected, and characteristic features of the pluripotent state were then obtained at the level of the cellular glycome. The expression profiles of relevant glycosyltransferase genes agreed well with the results obtained by lectin microarray. Among the 38 lectins, rBC2LCN was found to detect only undifferentiated iPSCs/ESCs and not differentiated SCs. Hence, the high density lectin microarray has proved to be valid for not only comprehensive analysis of glycans but also diagnosis of stem cells under the concept of the cellular glycome.

Increasing attention has been paid to iPSCs[2] and ESCs in their pluripotency and medical applications (1, 2). However, establishment of a robust evaluation system of their properties, including differentiation propensity and risk of possible contamination of xenoantigens and even potential of tumorigenesis, has been hampered by the lack of comprehensive methodology directly applicable to target stem cells, although this is an emerging issue essential for the safe use of iPSCs in regenerative medicine. From many aspects, cell surface glycans are considered to be ideal targets for analyzing or identifying the phenotype of each cell in a direct manner by the following reasons (3, 4). (*a*) Glycans are located at the outermost cell surface. (*b*) The total repertoire of cell surface glycans varies at every level of biological organization (*i.e.* species, tissues, cell types, and molecules). (*c*) Global alterations of the cellular glycome also occur during development, cellular activation, differentiation, malignant transformation, and inflammation. The cell surface glycans are therefore referred to as the "cell signature" that closely reflects cellular backgrounds and conditions, probably because they are actually functioning as cell-to-cell mediators in extensive biological phenomena. This fundamental nature of glycans should be understood with the fact that they are not encoded directly in the genome but are generated by a complex system of a number of glycosidases and glycosyltransferases, whose expressions and activities are significantly affected by both intracellular and extracellular environmental changes. Indeed, cell surface molecules, such as stage-specific embryonic antigens (SSEA1 and -3/4) (5) and tumor rejection antigens (Tra-1-60 and Tra-1-81) (6–8) are glycobiomarkers widely used to evaluate pluripotency. Notably, however, these "representative" glycomarkers have been identified following rather fortuitous development of their specific antibodies, because most carbohydrate structures are poorly antigenic between mammals. In this context, a systematic search is necessary to draw a whole picture of the stem cell glycome and harness its effect on stem cell biology (8, 9). For instance, the growth and directed differentiation of stem cells to specific progeny lineages in cell culture remain problematic. Understanding how stem cells

* This work was supported by the New Energy and Industrial Technology Development Organization in Japan.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S5 and Figs. S1–S4.

[1] To whom correspondence should be addressed: AIST, Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan. Tel.: 81-29-861-3124; Fax: 81-29-861-3125; E-mail: jun-hirabayashi@aist.go.jp.

[2] The abbreviations used are: iPSC, induced pluripotent stem cell; AM, amniotic mesodermal; ESC, embryonic stem cell; MEF, mouse embryonic fibroblast; PAE, placental artery endothelial; SC, somatic cell; UtE, uterine endometrium; rMOA, recombinant MOA; FWER, familywise error rate.

communicate with one another and feeder cells through cell surface glycans may lead to rational design of specific culture systems. However, the glycome is a quite difficult target to predict solely based on any genomic data base because the biosynthetic process of the glycan moieties of glycoproteins is not template-driven and is subject to multiple sequential and competitive enzymatic pathways. In this sense, a rapid and sensitive system enabling direct monitoring of cell surface glycans is essential.

Several methods have been developed for glycan analysis based on physicochemical principles, such as liquid chromatography and mass spectrometry (10–12). Lectin microarray is an alternative technology for structural glycomics, where a panel of lectins with various glycan-binding specificities is printed on a microarray, providing a versatile platform for rapid and high throughput analysis of glycan structures without liberation of glycans (13, 14). Lectins are a class of decoder molecules of cell surface glycans distributed throughout organisms, which mediate various functions through specific glycan recognition. Analytical protocols using lectin microarray have been developed for various sample types: free oligosaccharides (14, 15), tissue sections (16), cell membrane hydrophobic fractions (17, 18), and even whole cells (19, 20). This technology has just begun to be applied to a wide variety of biological researches, including virus profiling (21) and cell profiling (17, 20), and development of cancer glycobiomarkers (22–24). For cell profiling, less than 100 ng of proteins in hydrophobic fractions are sufficient for each analysis (25, 26). Data processing and normalization procedures were optimized to ensure the proper interpretation of the data (25, 26). More recently, we have demonstrated that lectin microarray is also applicable to stem cells (27, 28), although we have yet to reach a clear conclusion as to how the cellular glycome changes upon induction of pluripotency. Moreover, practical applicability of this technology to the quality control of stem cells has not been attained.

Here, we developed an advanced platform of high density lectin microarray with the increased number of probe lectins (96 lectins) to expand the glycome coverage for more precise comparison of various stem cell glycomes. A systematic survey of the cellular glycome was then performed toward 135 cell types in total, including iPSCs (114 cell types) and ESCs (nine cell types). Through this comprehensive analysis, we obtained strong evidence that all of the four SCs with originally distinct glycan profiles have acquired those similar to ESCs upon induction of pluripotency. We also found structural features common to iPSCs and ESCs, which corresponded well to the results of gene expression analysis of glycosyltransferases. Finally, we demonstrate the applicability of lectin microarray in the stem cell diagnosis of multiple factors, including discrimination between undifferentiated and differentiated cells as well as detection of the contamination of the xenoantigen, αGal epitope.

EXPERIMENTAL PROCEDURES

Cells—Endometrium (UtE) (29), placental artery endothelium (PAE) (30), and amnion (AM) (31, 32) were independently established. UtE, AM, and MRC5 cells were maintained in the POWEREDBY10 medium (MED SHIROTORI CO., Ltd.). PAE cells were cultured in EGM-2MV BulletKit (Lonza) containing 5% FBS. Human iPSCs from UtE, PAE, and AM cells were generated according to the procedures described by Yamanaka and colleagues (1) with slight modification (27, 28, 33). The iPSCs derived from MRC5, UtE, PAE, and AM cells were maintained in iPSellon medium (Cardio Inc.) supplemented with 10 ng/ml recombinant human basic fibroblast growth factor (Wako Pure Chemical Industries, Ltd.) on irradiated MEF feeder cells. hiPS201B7 and hiPS253G1 cells were maintained in DMEM/F-12 medium (Invitrogen) supplemented with 20% KSR (Invitrogen), 0.1 mM 2-mercaptoethanol (Sigma-Aldrich), minimum essential medium non-essential amino acids (Invitrogen), and 5–10 ng/ml recombinant human basic FGF (Wako) on mitomycin C-treated mouse embryo fibroblast feeder cells. ESCs were generated and maintained as described previously (34).

Lectins—Lectins from natural sources (58 lectins) were purchased from J-OIL MILLS, Vector Laboratories, EY Laboratories, and Seikagaku Corp. (see the lectin list in supplemental Table S2). Recombinant lectins were prepared as follows. Briefly, genes of carbohydrate recognition domains were cloned into pET27b (Stratagene) and were overexpressed in the *Escherichia coli* BL21-CodonPlus (DE3)-RIL strain under the control of isopropyl-β-D-thiogalactopyranoside (Fermentas Hanover) at appropriate temperatures. All recombinant lectins were purified by affinity chromatography using appropriate sugar-immobilized Sepharose 4B-CL (GE Healthcare) based on the glycan binding specificity of each lectin. They were then dialyzed against diluted PBS (final concentration 2.5 mM phosphate buffer containing 0.015 M NaCl). The protein concentration was determined by a BCA protein assay (Bio-Rad). Lectins were freeze-dried and stored at 4 °C until use. The purity was checked by SDS-PAGE and gel filtration chromatography on Shodex PROTEIN KW-802.5 (Shodex). The glycan binding activity and specificity were analyzed by hemagglutinating activity using 4% rabbit erythrocytes, frontal affinity chromatography (35), and glycoconjugate microarray (36).

Lectin Microarray Production—The lectin microarray was produced as described previously with minor modifications (14, 15). Fifty-eight natural lectins and 38 recombinant lectins (see supplemental Table S2 for a lectin list) were dissolved at a concentration of 0.5 mg/ml in a spotting solution (Matsunami Glass), and spotted onto epoxysilane-coated glass slides (Schott) in triplicate using a non-contact microarray printing robot (MicroSys4000, Genomic Solutions). The glass slides were then incubated at 25 °C overnight to allow lectin immobilization. The lectin-immobilized glass slides were then washed with probing buffer (25 mM Tris-HCl, pH 7.5, 140 mM NaCl (TBS) containing 2.7 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, and 1% Triton X-100) and incubated with blocking reagent N102 (NOF Co.) at 20 °C for 1 h. Finally, the lectin-immobilized glass slides were washed with TBS containing 0.02% NaN_3 and stored at 4 °C until use. The spot quality and reproducibility of the produced microarrays were checked before use, using a Cy3-labeled test probe containing 250 μg/ml asialofetuin (Sigma-Aldrich), 25 ng/ml Siaα2–3Galβ1–4GlcNAc-BSA (Dextra), 10 ng/ml Fucα1–2Galβ1–3GlcNAcβ1–3Galβ1–4Glc-BSA (Dextra), 10 ng/ml βGlcNAc-BSA (Dextra), 10 ng/ml GalNAcα1–3(Fucα1–2)Gal-BSA (Dextra), 10 ng/ml Galα1–3Galβ1-

4GlcNAc-BSA (Dextra), 10 ng/ml Manα1–3(Manα1–6)Man-BSA (Dextra), 10 ng/ml αFuc-BSA (Dextra), 10 ng/ml αGalNAc-BSA (Dextra), and 10 ng/ml Siaα2–6Galβ1–4Glc-BSA (Dextra) dissolved in probing buffer.

Lectin Microarray Analysis—Hydrophobic fractions were prepared using CelLytic minimum essential medium protein extraction (Sigma-Aldrich) in accordance with the manufacturer's procedures (25, 27). After protein quantification using a BCA assay (Thermo Fisher Scientific), hydrophobic fractions were fluorescently labeled with Cy3 monoreactive dye (GE Healthcare), and excess Cy3 was removed with Sephadex G-25 desalting columns (GE Healthcare). After adjusting the protein concentration to 2 μg/ml with PBST (10 mM PBS, pH 7.4, 140 mM NaCl, 2.7 mM KCl, 1% Triton X-100), the hydrophobic fraction was labeled with Cy3 NHS ester (GE Healthcare). After dilution with probing buffer at 0.5 μg/ml, the Cy3-labeled hydrophobic fraction was applied to the lectin microarray and incubated at 20 °C overnight. After washing with probing buffer, fluorescence images were acquired using an evanescent field-activated fluorescence scanner (GlycoStation™ reader 1200; GP BioSciences). The fluorescence signal of each spot was quantified using Array Pro Analyzer version 4.5 (Media Cybernetics, Bethesda, MD), and the background value was subtracted. The background value was obtained from the area without lectin immobilization. The lectin signals of triplicate spots were averaged and normalized to the mean value of 96 lectins immobilized on the array. An inhibition assay was performed by incubating Cy3-labeled cell membrane fractions of MEF(#1) or MRC5-iPS#25(P22)(#13) with a lectin microarray either in the absence or presence of 100 μg/ml of Galα1–3Galβ1–4GlcNAc-PAA (catalog no. 01-079, Glycotech) or a negative control PAA (catalog no. 01-000, Glycotech).

Gene Expression Analysis—Total RNA was extracted from each sample by using ISOGEN (NipponGene). The global gene expression patterns were monitored using Agilent whole human genome microarray chips (G4112F) with one-color (cyanine 3) dye. This microarray covers 41,000 well characterized human genes and transcripts. Of the 41,000 probes, 16,483 representative probes corresponding to the microarray quality control unique genes were used for the following analyses (37).

Statistics—Unsupervised clustering was performed by employing the average linkage method using Cluster 3.0 software. The heat map with clustering was acquired using Java Treeview. Differences between the two arbitrary data sets were evaluated by Student's t test to each lectin signal using SPSS Statistics 19 (SPSS). Significantly different lectin signals or the glycosyltransferase expression were selected if they satisfied a familywise error rate (FWER) by the Bonferroni method of <0.001.

Immunocytochemistry—Immunocytochemical analysis was performed as described previously (29, 33, 38). Human iPSCs were fixed with 4% paraformaldehyde in PBS for 10 min at 4 °C. After washing with 0.1% Triton X-100 in PBS (PBST), the cells were prehybridized in blocking buffer for 1–12 h at 4 °C and then incubated for 6–12 h at 4 °C with the following primary antibodies: anti-SSEA4 (1:300 dilution; Chemicon), anti-TRA-1-60 (1:300; Chemicon), anti-Oct4 (1:50; Santa Cruz Biotechnology, Inc.), anti-Nanog (1:300; ReproCELL), and anti-Sox2 (1:300; Chemicon). The cells were then incubated with anti-rabbit IgG, anti-mouse IgG, or anti-mouse IgM conjugated with Alexa Fluor 488 or Alexa Fluor 546 (1:500; Molecular Probes) in blocking buffer for 1 h at room temperature. The cells were counterstained with DAPI and then mounted using the SlowFade light antifade kit (Molecular Probes).

Teratomas—Teratoma formation was performed as described previously (1, 2). The 1:1 mixtures of the human iPSC suspension and basement membrane matrix (BD Biosciences) were implanted subcutaneously at 1.0×10^7 cells/site into immunodeficient, non-obese diabetic/severe combined immunodeficiency mice. Teratomas were surgically dissected out 8–12 weeks after implantation and were fixed with 4% paraformaldehyde in PBS and embedded in paraffin. Sections of 10-μm thickness were stained with hematoxylin-eosin.

Glycoconjugate Microarray Analysis—Glycoconjugate microarray production and analysis were performed as described previously (36). Briefly, glycoproteins and glycoside-polyacrylamide conjugates were dissolved in the Matsunami spotting solution at a final concentration of 0.5 and 0.1 mg/ml, respectively. After filtration, they were spotted on the Schott epoxy-coated glass slide using the Microsys non-contact microarray printing robot.

Cy3-labeled lectins dissolved in the probing solution (10 or 1 μg/ml) were applied to each chamber of the glycoconjugate microarray (100 μl/well) and were incubated at 20 °C overnight. After washing the chambers with the probing solution, fluorescent images were immediately acquired using an evanescent field-activated fluorescence scanner, the GlycoStation™ Reader 1200, under Cy3 mode. Data were analyzed with the Array Pro analyzer version 4.5 (Media Cybernetics, Inc.). The net intensity value for each spot was determined by signal intensity minus background value. The lectin signals of triplicate spots were averaged and normalized to the highest signal intensity among 98 glycoconjugates immobilized on the array.

RESULTS

Development of High Density Lectin Microarray—In order to increase glycome coverage and the selection range of lectins suitable for stem cell evaluation, we first increased the number of immobilized lectins from 43 to 96, which is the largest number of immobilized lectins reported (39). For this purpose, lectins with defined structures were first categorized into lectin families with different protein scaffolds. We then selected lectins from various lectin families, intending to cover a wider range of glycan binding specificities. Especially, we increased lectins specific to terminal modifications, such as Sia and Fuc, which often change dramatically depending on cell properties. For production of recombinant lectins, the *E. coli* expression system was chosen to avoid glycosylation of the produced lectins, which might cause nonspecific binding to lectin-like molecules in the objective samples. The recombinant lectins thus produced were purified by affinity chromatography using the most appropriate sugar-immobilized Sepharose. The glycan-binding specificities of 96 lectins used in this study were analyzed by both glycoconjugate microarray (supplemental Fig. S1 and Table S1; also see "Experimental Procedures") (36) and,

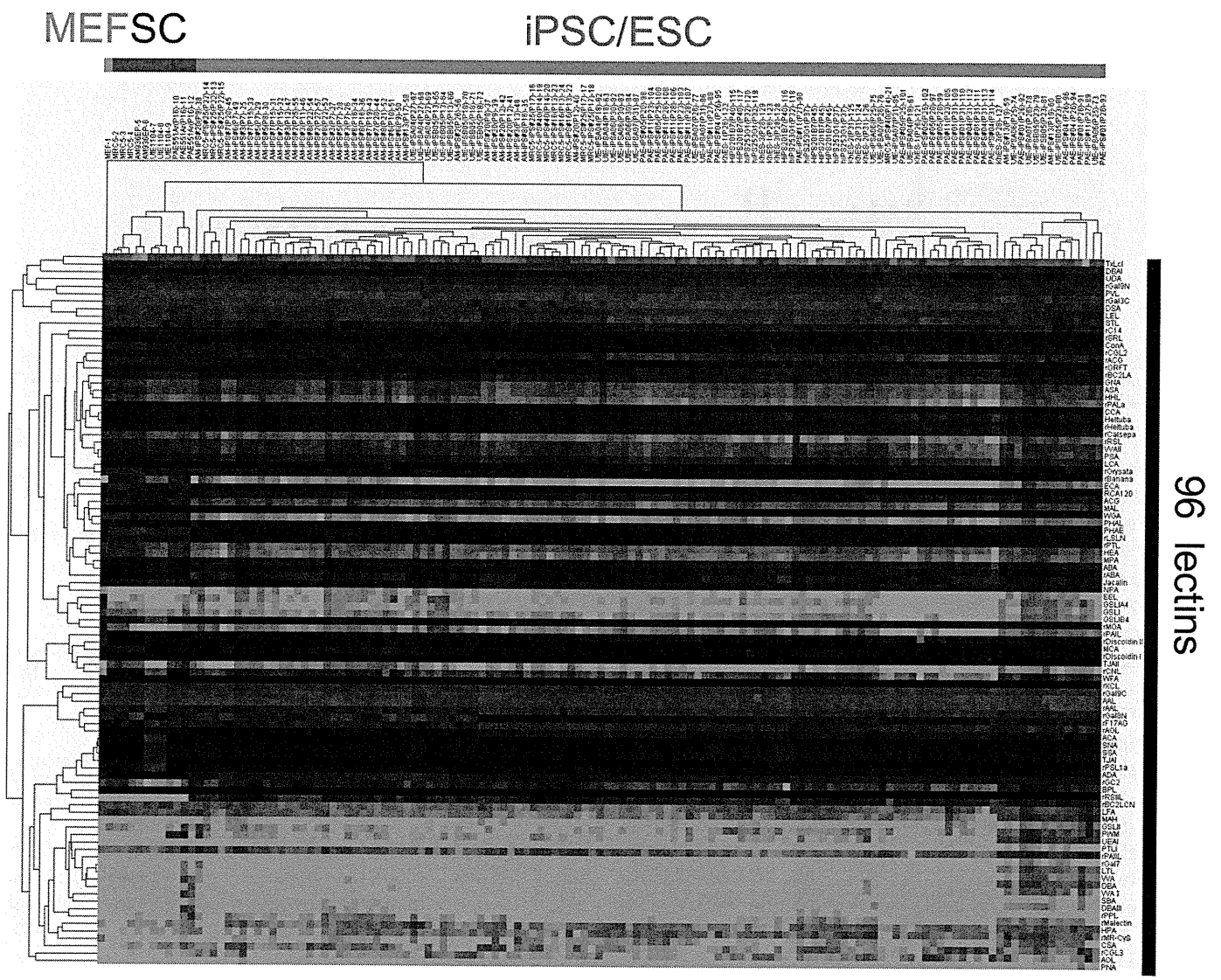


FIGURE 1. **Unsupervised cluster analysis.** Lectin microarray data of iPSCs (n = 123), their parental SCs (n = 11), ESCs (9), and MEF (n = 1) were mean-normalized and log-transformed and then analyzed by Cluster 3.0. The zero value of the lectin signal was converted to 1. *Yellow*, positive; *blue*, negative. Clustering method was average linkage. The heat map with clustering was acquired using Java Treeview.

more quantitatively, frontal affinity chromatography (see the Lectin Frontier Database Web page) (35, 40). Their basic specificities evaluated by the above two analytical methods are briefly summarized in supplemental Table S2. The 96 lectins were spotted onto epoxy-activated glass slides by a non-contact spotter (supplemental Fig. S2), and their quality was extensively assessed using a Cy3-labeled test probe (25). Lot-to-lot variance (coefficients of variation) of the developed high density lectin microarray was confirmed to be low (0.14) after mean normalization (25).

Transcription Factor-induced Reprogramming Leads to a Global Reversion Down to the Pluripotent State at a Cellular Glycome Level as Well—Using the developed lectin microarray, we have analyzed 135 cell samples in total, including 114 iPSCs, 11 SCs, and nine ESCs, all from human origins, as well as one mouse embryonic fibroblast (MEF). Human iPSCs were generated from four different SC lines: MRC5, AM, UtE, and PAE (supplemental Table S3) (28). We have also analyzed human iPSCs generated from human dermal fibroblasts with four (201B7) (1) and three transcription factors (253G1) (41) and three cell lines of human ESCs (42). All iPSCs used in this study were morphologically similar to ESCs, and their pluripotency was confirmed by staining with the established undifferentiation markers (SSEA4, Tra1–60, Oct4, Nanog, and Sox2) and DNA microarray (28).

Cell membrane hydrophobic fractions were prepared, and the extracted glycoproteins were then labeled with Cy3-*N*-hydroxysuccinimide ester and analyzed by lectin microarray (25). We have analyzed cell membrane fractions because they can be stored in a freezer until use and are easy to handle, allowing comprehensive analysis of a large number of samples (25, 26).

After being mean-normalized, the obtained data were first analyzed by unsupervised hierarchical clustering (Fig. 1). As a result, differentiated SCs and undifferentiated iPSCs/ESCs were clearly separated into two large clusters, whereas the four SCs were further separated according to their origins. This indicates that SCs (MRC5, AM, UtE, and PAE) with different glycan profiles have acquired profiles quite similar to one

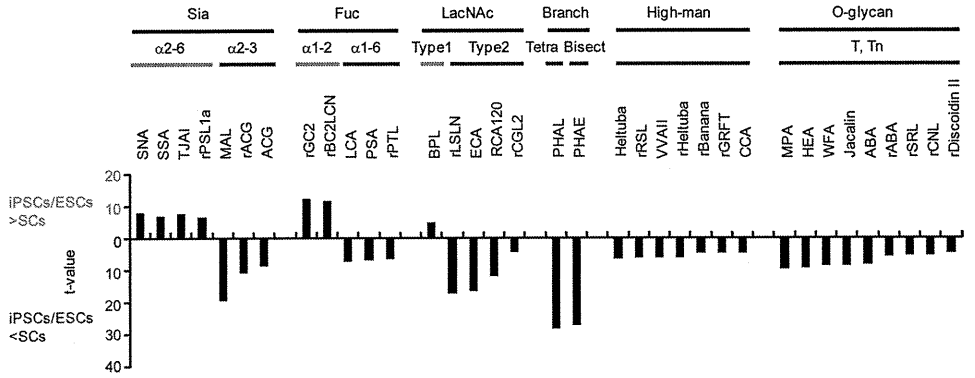


FIGURE 2. **Alterations of the lectin signals upon induction of pluripotency.** Lectin microarray data were mean-normalized and analyzed by Student's *t* test. Lectins with significantly different signals (FWER < 0.001) between undifferentiated iPSCs/ESCs ($n = 123$) and differentiated SCs ($n = 11$) were categorized into six groups based on the glycan binding specificities of lectins. Data are shown with *t* values. Also see supplemental Table S4.

another and even to ESCs upon induction of pluripotency. Thus, transcription factor-induced reprogramming was found to lead to a global reversion down to the pluripotent state at a cellular glycome level as well (27, 28).

Characteristic Features of Glycome Alteration upon Induction of Pluripotency—We then examined in more detail how glycan structures altered during the induction of pluripotency. The mean-normalized data were processed by Student's *t* test to select significant probe lectins discriminating between SCs and iPSCs/ESCs (supplemental Table S4). As a result, 38 lectins were selected with FWER of <0.001. Among them, nine gave higher signals in iPSCs than SCs, whereas 29 exhibited lower signals. Among the 38 lectins, 35 lectins were then categorized into six groups based on their glycan binding specificities, from which glycan alterations having occurred upon induction of pluripotency were estimated (Fig. 2), whereas the three lectins with broader specificities (wheat germ agglutinin (WGA), a Sia binder; rRSIIL and aleuria aurantia lectin (AAL), broad Fuc binders) were not included in this categorization. Here, the lectins with higher signals in iPSCs/ESCs than SCs are shown *below gray lines*, whereas lower signals are shown *below black lines*. The characteristic features of glycan structures of undifferentiated iPSCs/ESCs relative to differentiated SCs are summarized as follows. 1) The signals of α2–6Sia-binding lectins (SNA, SSA, TJAI, and rPSL1a) were increased, whereas those of α2–3Sia-binding lectins (MAL, rACG, and ACG) were decreased correspondingly (28). This agrees well with the previous report that α2–6-sialylated glycan expression is higher in undifferentiated (human ESCs) than differentiated cells (embryoid body) (43). 2) In terms of fucosylation, the signals of α1–2Fuc-specific lectins (rGC2 and rBC2LCN) were increased, whereas those of α1–6Fuc-specific lectins (LCA, PSA, and rPTL) were decreased. This is consistent with the recent report that human ESCs are stained with anti-Globo H (Fucα1–2Galβ1–3GalNAcβ1–3Galα1–4Galβ1–4Glc) and anti-H type 1 (Fucα1–2Galβ1–3GlcNAc), whose antigens contain α1–2Fuc (9). 3) The signals of type 1 LacNAc (Galβ1–3GlcNAc)-binding lectins (BPL) were increased, whereas those of type 2 LacNAc (Galβ1–4GlcNAc)-binding lectins (rLSLN, ECA, RCA120, and rCGL2) were decreased. This agrees well with the recent finding that type 1 LacNAc is the glycan epitope recognized by the well known pluripotency markers Tra-1-60 and Tra-1-81 (7). 4) The

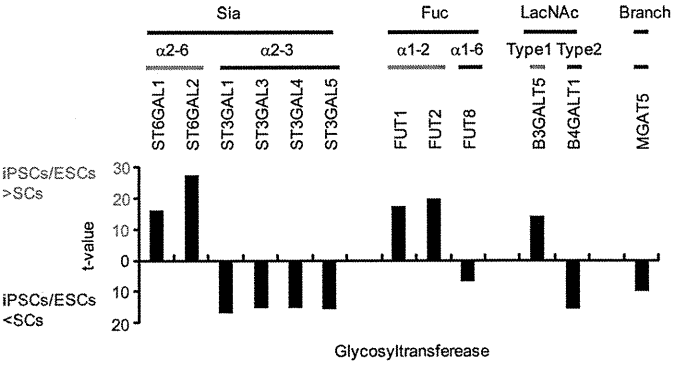


FIGURE 3. **Alterations in the expression of glycosyltransferases upon induction of pluripotency.** Glycosyltransferases related to the lectin signals in Fig. 2 are shown with *t* values. All data are shown in supplemental Table S5.

lectin signals specific to bisecting GlcNAc (PHAE), tetra-antennary *N*-glycans (PHAL), high mannose type *N*-glycans (Heltuba, rRSL, VVAII, rHeltuba, Heltuba, rBanana, rGRFT, and CCA), and *O*-glycans (MPA, HEA, WFA, Jacalin, ABA, rABA, rSRL, rCNL, and rDiscoidin II) were decreased.

The expression profiles of glycosyltransferases synthesizing glycans agreed well with the results obtained by lectin microarray (Fig. 3 and supplemental Table S5); the expression of α2–6-sialyltransferases (ST6GAL1 and -2) (28), α1–2-fucosyltransferases (FUT1 and -2), the major glycosyltransferase involved in the synthesis of type 1 LacNAc (B3GALT5) (28), and MGAT5, a glycosyltransferase involved in the synthesis of tetra-antennary *N*-glycans, was increased, whereas that of α2–3-sialyltransferases (ST3GAL1, -3, -4, and -5), α1–6-fucosyltransferase (FUT8), and the major glycosyltransferase related to the synthesis of type 2 LacNAc (B4GalT1) was decreased correspondingly in iPSCs/ESCs relative to SCs. Based on the results obtained by lectin and DNA microarrays, it is conceivable that the expression of α2–6-sialylation, α1–2-fucosylation, and type 1 LacNAc is increased, whereas that of α2–3-sialylation and tetra-antennary *N*-glycans is decreased upon induction of pluripotency (Fig. 4).

Selection of the Best Lectin Probe to Discriminate Pluripotency—We then addressed the challenge to develop a lectin-based procedure to discriminate between differentiated SCs and undifferentiated iPSCs/ESCs, which could be utilized to monitor the state of differentiation. As described, rGC2, rBC2LCN, SNA, TJAI, SSA, rPSL1a, rRSIIL, BPL, and AAL gave

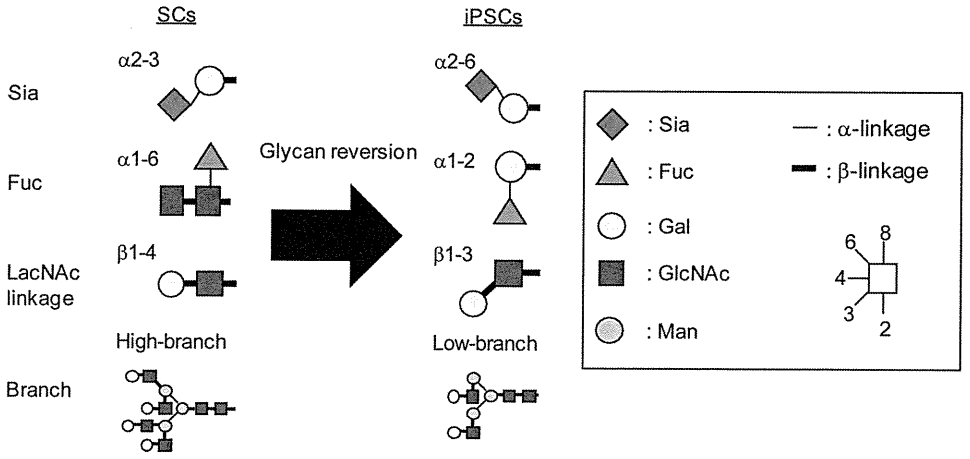


FIGURE 4. **Schematic representation of the putative glycan alterations upon induction of pluripotency.**

TABLE 1

Selection of the best lectin probe to evaluate pluripotency among 96 lectins

Lectins with significantly higher signals (FWER < 0.001) for iPSCs/ESCs than SCs are shown. Also see supplemental Fig. S4.

Lectin	SC (mean, $n = 11$)	SC (S.D.)	iPSC/ESC (mean, $n = 123$)	iPSC/ESC (S.D.)	FWER (Bonferroni)
rGC2	125	31	304	47	1E−21
rBC2LCN	1	1	23	6	2E−19
SNA	69	43	124	18	4E−11
TJAI	96	58	170	28	9E−10
SSA	97	55	156	22	9E−09
rPSL1a	61	32	107	21	2E−07
rRSIIL	78	23	109	20	3E−04
BPL	9	5	20	8	7E−04
AAL	518	77	677	114	1E−03

significantly higher signals in iPSCs/ESCs than SCs with FWER < 0.001 (Table 1). Among them, rBC2LCN showed the best performance as a probe to detect only undifferentiated iPSCs/ESCs but never reacted with differentiated SCs and MEF, whereas other lectins also reacted with SCs (Table 1). Namely, although rGC2 showed a better score in terms of FWER (1×10^{-21}) than rBC2LCN (2×10^{-19}), the former reacted strongly with MEF (Fig. 5). Similarly, SNA (4×10^{-11}), a representative α2–6Sia-binding lectin, showed significant cross-reactivity with a part of SCs derived from PAE in addition to MEF (Fig. 5).

Monitoring the Contamination of the Xenoantigen, αGal Epitope—From a practical viewpoint, monitoring possible contamination by xenotransplantation antigens in iPSCs/ESCs is essential for their safe use in regenerative medicine. A recombinant MOA (rMOA) recognizes the xenotransplantation antigen Galα1–3Galβ1–4GlcNAc (44) present in most cells from New World monkeys and non-primate mammals, including mice, but not in humans. Indeed, rMOA strongly bound to MEFs but not to any human SCs (Fig. 6). Therefore, rMOA signals should not be detected in human iPSCs. However, triplicate samples of the two cell lines MRC5-iPS#25(P22)(#13–15) and UtE-iPSB05(P13)(#64–66) exhibited significant signals on rMOA. In order to validate whether the binding of rMOA is mediated by a carbohydrate recognition domain of rMOA, we then performed inhibition assay. As shown in Fig. 6*B*, the binding of rMOA to cell membrane fractions of MEF and MRC5-iPS#25(P22,#13) were abolished in the presence of 100 μg/ml Galα1–3Galβ1–4GlcNAc-PAA (Fig. 6*B*), but no inhibitory effect was observed for 100 μg/ml of a negative control (PAA without sugar moiety), indicating that the binding is due to specific interactions via the rMOA carbohydrate recognition domain. As expected, no inhibitory effect of Galα1-3Galβ1–4GlcNAc-PAA on a Fuc-binding lectin, rAAL, was observed. These data unambiguously reflect contamination by the xenoantigen αGal epitope, in the above two cell lines, which were most probably contaminated with MEF.

DISCUSSION

Using the developed high density lectin microarray, we performed a systematic analysis of cell surface glycans of a large set of human iPSCs (114 cell types) and ESCs (nine cell types). As a result, a basis for a rational stem cell evaluation system was established, which can reveal both the state of undifferentiation and inclusion of αGal epitope (a representative xenoantigen). Such a comprehensive glycome analysis targeting iPSCs and ESCs has never been carried out so far. There are at least three key advantages in using a lectin microarray. 1) An overall glycan profile of each cell type is readily obtained using a relatively small number of cells (~1×10^3), and thus, the method is widely applicable to stem cells. 2) The proposed evaluation system includes selection of the best probe by a statistical strategy among a number of lectins, which are immobilized on the array. As candidate probes, carbohydrate-binding antibodies developed so far could also be included. 3) Various properties of stem cells can be assessed simultaneously (*i.e.* with "one-chip" technology). Using the same strategy described in this study, lectin-based evaluation methods targeting tumorigenesis and the differentiation propensity of stem cells could also be developed.

Based on the lectin signals and the expression profiles of glycosyltransferases, we concluded that the expression of α2–6-sialylation, α1–2-fucosylation, and type 1 LacNAc increases, whereas that of α2–3-sialylation and tetra-antennary *N*-glycans decreases correspondingly upon the induction of pluripotency. These changes are consistent with the recent reports that relevant glycans (*i.e.* the expression of Globo H and H type 1 with an α1–2Fuc and α2–6Sia in human ESCs) are higher than that in differentiated embryoid body (9, 43). Interestingly, the increased expression of α1–2Fuc and type 1 LacNAc, which are synthesized by the action of FUT1/2 and B3GalT5, respectively, is closely related to the synthesis of the well known pluripo-

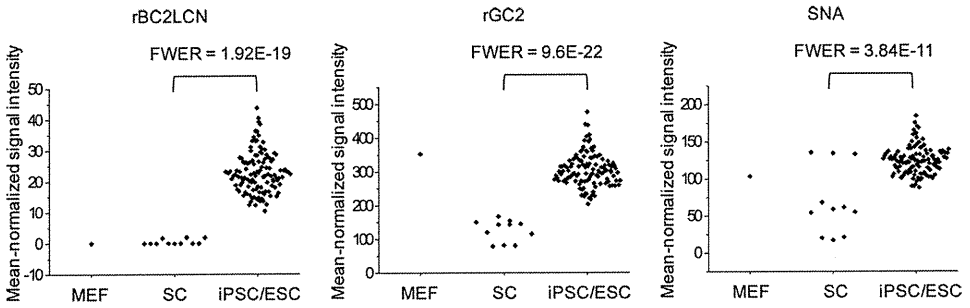


FIGURE 5. **Selection of the best lectin probe to discriminate pluripotency.** The mean-normalized signal intensities of rGC2, rBC2LCN, and SNA to MEF (n = 1), SCs (n = 11), and iPSCs/ESCs (n = 123) are shown.

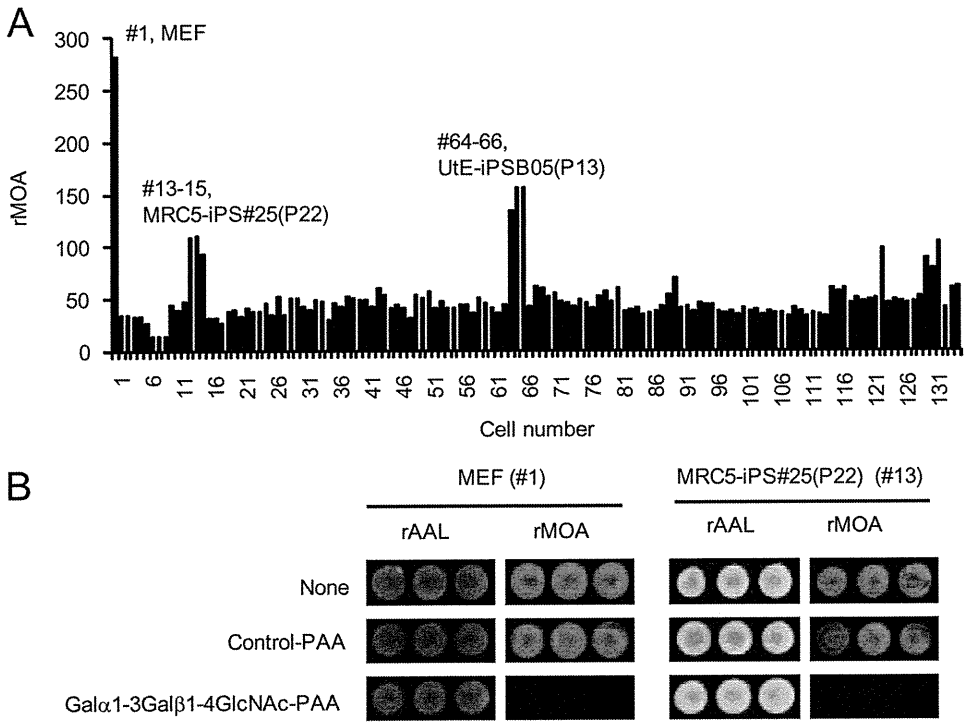


FIGURE 6. **Monitoring the contamination of the xenoantigen αGal epitope using the rMOA lectin.** *A*, mean-normalized lectin microarray data are represented by a *bar graph*. *Numbers* correspond to cell types described in supplemental Table S3. *B*, inhibition assay. Cy3-labaled cell membrane fractions of MEF(#1) or MRC5-iPS#25(P22)(#13) were incubated with lectin microarray either in the absence (*None*) or presence of 100 μg/ml Galα1–3Galβ1–4GlcNAc-PAA or negative control PAA without sugar moiety. Data shown were obtained at gain 110 for MEF and gain 120 for MRC5-iPS#25(P22).

tency markers, SSEA3/4 and Tra-1-60/81 (for a scheme, see supplemental Fig. S4).

In this study, rBC2LCN was selected as the best lectin probe to evaluate pluripotency among the 96 lectins. BC2LCN is a TNF-like lectin molecule identified from a Gram-negative bacterium *Burkholderia cenocepacia* (45). Glycoconjugate microarray analysis revealed that rBC2LCN binds specifically to Fucα1–2Galβ1–3GlcNAc (GalNAc)-containing glycans, such as H type 1 (Fucα1–2Galβ1–3GlcNAc), H type 3 (Fucα1–2Galβ1–3GalNAc), and Lewis b (Fucα1–2Galβ1–3(Fucα1-4)GlcNAc), which include the two structural characteristics related to the pluripotency (α1–2Fuc and type 1 LacNAc) as described above (supplemental Fig. S3). This observation is consistent with the previous report (45) in which Sulak *et al.* studied the glycan-binding specificity of BC2LCN in detail using glycan microarray and titration microcalorimetry. They also demonstrated that this lectin also binds to Globo H (Fucα1–2Galβ1–3GalNAcβ1–3Galα1–4Galβ1–4Glc), which was recently proposed as a glycosphingolipid type pluripotency marker (supplemental Fig. S4) (9, 45). These results explain the mechanism of how this lectin could be used as the probe to discriminate pluripotency. rBC2LCN could be used to probe glycoproteins and possibly all glycoconjugates carrying Fucα1–2Galβ1–3GlcNAc (GalNAc), whereas anti-SSEA3 and anti-SSEA4 specifically target glycosphingolipids. From a practical viewpoint, rBC2LCN is cost-effective because it can be produced in large amounts by the conventional *E. coli* expression system (84 mg/liter). Thus, this lectin could be a versatile probe to evaluate pluripotency.

In contrast, Globo H has also been reported to be overexpressed in epithelial cell tumors (46). Furthermore, α2–6Sia up-regulated in iPSCs/ESCs has been reported to be overexpressed in many types of human cancers, and its high expression positively correlates with tumor metastasis and poor prognosis (47). Thus, the glycan alterations upon induction of pluripotency observed in this study are apparently similar to

those occurring during malignant transformation, as was implied recently (9). Although the reason for this similarity remains to be elucidated, the characteristic glycan changes should be related to the ability of eternal cell proliferation and maintenance, properties common to both cancer cells and pluripotent stem cells.

Glycans are located at the outermost cell surface, where various events take place on the basis of cell-to-cell recognition and interactions. Endogenous lectins, major counterpart molecules of glycans, should play crucial roles in the events (*e.g.* by regulating several signaling pathways). In this context, interactions occurring between cell surface glycans and endogenous lectins are considered to be essential for the maintenance of pluripotency, self-renewal, and differentiation of iPSCs/ESCs (48). Indeed, heparan sulfate proteoglycans were reported to regulate self-renewal and pluripotency of embryonic stem cells (49). Moreover, reduced sulfation on heparan sulfate and chondroitin sulfate were demonstrated to direct neural differentiation of mouse ESCs and human iPSCs (50). Recently, synthetic substrates recognizing cell surface glycans were reported to facilitate the long term culture of pluripotent stem cells (48). Thus, global analysis of the cellular glycomes of iPSCs and ESCs performed in this study will be necessary to provide the basis to explore the functions and applications of the stem cell glycobiology. They includes rational design of the effective substrates and culture conditions to support the long term propagation of ESCs and iPSCs (48). Of course, the results obtained in this study could also be readily applied to staining (specification of the place the event occurs), enrichment (*e.g.* lectin-aided capturing of necessary cells), and targeting of specific cells (*e.g.* elimination of unwanted undifferentiated cells). In this regard, stem cell glycoengineering with the aid of a lectin microarray is a key issue in realization of regenerative medicine in the near future.

Acknowledgments—We thank Yoshiko Kubo and Jinko Murakami for help in lectin microarray production, Dr. Atsushi Kuno and Dr. Yoko Itakura for advice, and Dr. Jun Iwaki for providing galectin expression vectors.

REFERENCES

1. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) *Cell* **131,** 861–872
2. Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007) *Science* **318,** 1917–1920
3. Gagneux, P., and Varki, A. (1999) *Glycobiology* **9,** 747–755
4. Varki, A. (1993) *Glycobiology* **3,** 97–130
5. Muramatsu, T., and Muramatsu, H. (2004) *Glycoconj. J.* **21,** 41–45
6. Schopperle, W. M., and DeWolf, W. C. (2007) *Stem Cells* **25,** 723–730
7. Natunen, S., Satomaa, T., Pitkanen, V., Salo, H., Mikkola, M., Natunen, J., Otonkoski, T., and Valmu, L. (2011) *Glycobiology*, in press
8. Lanctot, P. M., Gage, F. H., and Varki, A. P. (2007) *Curr. Opin. Chem. Biol.* **11,** 373–380
9. Liang, Y. J., Kuo, H. H., Lin, C. H., Chen, Y. Y., Yang, B. C., Cheng, Y. Y., Yu, A. L., Khoo, K. H., and Yu, J. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 22564–22569
10. Hirabayashi, J. (2008) *J. Biochem.* **144,** 139–147
11. Wearne, K. A., Winter, H. C., and Goldstein, I. J. (2008) *Glycoconj. J.* **25,** 121–136
12. Wearne, K. A., Winter, H. C., O'Shea, K., and Goldstein, I. J. (2006) *Glycobiology* **16,** 981–990
13. Pilobello, K. T., Krishnamoorthy, L., Slawek, D., and Mahal, L. K. (2005) *Chembiochem* **6,** 985–989
14. Kuno, A., Uchiyama, N., Koseki-Kuno, S., Ebe, Y., Takashima, S., Yamada, M., and Hirabayashi, J. (2005) *Nat. Methods* **2,** 851–856
15. Uchiyama, N., Kuno, A., Tateno, H., Kubo, Y., Mizuno, M., Noguchi, M., and Hirabayashi, J. (2008) *Proteomics* **8,** 3042–3050
16. Matsuda, A., Kuno, A., Ishida, H., Kawamoto, T., Shoda, J., and Hirabayashi, J. (2008) *Biochem. Biophys. Res. Commun.* **370,** 259–263
17. Ebe, Y., Kuno, A., Uchiyama, N., Koseki-Kuno, S., Yamada, M., Sato, T., Narimatsu, H., and Hirabayashi, J. (2006) *J. Biochem.* **139,** 323–327
18. Pilobello, K. T., Slawek, D. E., and Mahal, L. K. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 11534–11539
19. Hsu, K. L., Pilobello, K. T., and Mahal, L. K. (2006) *Nat. Chem. Biol.* **2,** 153–157
20. Tateno, H., Uchiyama, N., Kuno, A., Togayachi, A., Sato, T., Narimatsu, H., and Hirabayashi, J. (2007) *Glycobiology* **17,** 1138–1146
21. Krishnamoorthy, L., Bess, J. W., Jr., Preston, A. B., Nagashima, K., and Mahal, L. K. (2009) *Nat. Chem. Biol.* **5,** 244–250
22. Kuno, A., Kato, Y., Matsuda, A., Kaneko, M. K., Ito, H., Amano, K., Chiba, Y., Narimatsu, H., and Hirabayashi, J. (2009) *Mol. Cell Proteomics* **8,** 99–108
23. Narimatsu, H., Sawaki, H., Kuno, A., Kaji, H., Ito, H., and Ikehara, Y. (2010) *FEBS J.* **277,** 95–105
24. Matsuda, A., Kuno, A., Kawamoto, T., Matsuzaki, H., Irimura, T., Ikehara, Y., Zen, Y., Nakanuma, Y., Yamamoto, M., Ohkohchi, N., Shoda, J., Hirabayashi, J., and Narimatsu, H. (2010) *Hepatology* **52,** 174–182
25. Tateno, H., Kuno, A., Itakura, Y., and Hirabayashi, J. (2010) *Methods Enzymol.* **478,** 181–195
26. Kuno, A., Itakura, Y., Toyoda, M., Takahashi, Y., Yamada, M., Umezawa, A., and Hirabayashi, J. (2008) *J. Proteomics Bioinform.* **1,** 68–72
27. Toyoda, M., Yamazaki-Inoue, M., Itakura, Y., Kuno, A., Ogawa, T., Yamada, M., Akutsu, H., Takahashi, Y., Kanzaki, S., Narimatsu, H., Hirabayashi, J., and Umezawa, A. (2011) *Genes Cells* **16,** 1–11
28. Saito, S., Onuma, Y., Ito, Y., Tateno, H., Toyoda, M., Akutsu, H., Nishino, K., Chikazawa, E., Fukawatase, Y., Miyagawa, Y., Okita, H., Kiyokawa, N., Shimma, Y., Umezawa, A., Hirabayashi, J., Horimoto, K., and Asashima, M. (2010) *The Fourth International Conference on Computational Systems Biology (ISB2010), Suzhou, China, September 9–11, 2010,* pp. 381–388
29. Nishino, K., Toyoda, M., Yamazaki-Inoue, M., Makino, H., Fukawatase, Y., Chikazawa, E., Takahashi, Y., Miyagawa, Y., Okita, H., Kiyokawa, N., Akutsu, H., and Umezawa, A. (2010) *PLoS One* **5,** e13017
30. Cui, C. H., Miyoshi, S., Tsuji, H., Makino, H., Kanzaki, S., Kami, D., Terai, M., Suzuki, H., and Umezawa, A. (2011) *Hum. Mol. Genet* **20,** 235–244
31. Tsuji, H., Miyoshi, S., Ikegami, Y., Hida, N., Asada, H., Togashi, I., Suzuki, J., Satake, M., Nakamizo, H., Tanaka, M., Mori, T., Segawa, K., Nishiyama, N., Inoue, J., Makino, H., Miyado, K., Ogawa, S., Yoshimura, Y., and Umezawa, A. (2010) *Circ. Res.* **106,** 1613–1623
32. Kawamichi, Y., Cui, C. H., Toyoda, M., Makino, H., Horie, A., Takahashi, Y., Matsumoto, K., Saito, H., Ohta, H., Saito, K., and Umezawa, A. (2010) *J. Cell Physiol.* **223,** 695–702
33. Makino, H., Toyoda, M., Matsumoto, K., Saito, H., Nishino, K., Fukawatase, Y., Machida, M., Akutsu, H., Uyama, T., Miyagawa, Y., Okita, H., Kiyokawa, N., Fujino, T., Ishikawa, Y., Nakamura, T., and Umezawa, A. (2009) *Exp. Cell Res.* **315,** 2727–2740
34. Suemori, H., Yasuchika, K., Hasegawa, K., Fujioka, T., Tsuneyoshi, N., and Nakatsuji, N. (2006) *Biochem. Biophys. Res. Commun.* **345,** 926–932
35. Tateno, H., Nakamura-Tsuruta, S., and Hirabayashi, J. (2007) *Nat. Protoc.* **2,** 2529–2537
36. Tateno, H., Mori, A., Uchiyama, N., Yabe, R., Iwaki, J., Shikanai, T., Angata, T., Narimatsu, H., and Hirabayashi, J. (2008) *Glycobiology* **18,** 789–798
37. Shi, L., Reid, L. H., Jones, W. D., Shippy, R., Warrington, J. A., Baker, S. C., Collins, P. J., de Longueville, F., Kawasaki, E. S., Lee, K. Y., Luo, Y., Sun, Y. A., Willey, J. C., Setterquist, R. A., Fischer, G. M., Tong, W., Dragan, Y. P., Dix, D. J., Frueh, F. W., Goodsaid, F. M., Herman, D., Jensen, R. V., Johnson, C. D., Lobenhofer, E. K., Puri, R. K., Schrf, U., Thierry-Mieg, J.,

Wang, C., Wilson, M., Wolber, P. K., Zhang, L., Amur, S., Bao, W., Barbacioru, C. C., Lucas, A. B., Bertholet, V., Boysen, C., Bromley, B., Brown, D., Brunner, A., Canales, R., Cao, X. M., Cebula, T. A., Chen, J. J., Cheng, J., Chu, T. M., Chudin, E., Corson, J., Corton, J. C., Croner, L. J., Davies, C., Davison, T. S., Delenstarr, G., Deng, X., Dorris, D., Eklund, A. C., Fan, X. H., Fang, H., Fulmer-Smentek, S., Fuscoe, J. C., Gallagher, K., Ge, W., Guo, L., Guo, X., Hager, J., Haje, P. K., Han, J., Han, T., Harbottle, H. C., Harris, S. C., Hatchwell, E., Hauser, C. A., Hester, S., Hong, H., Hurban, P., Jackson, S. A., Ji, H., Knight, C. R., Kuo, W. P., LeClerc, J. E., Levy, S., Li, Q. Z., Liu, C., Liu, Y., Lombardi, M. J., Ma, Y., Magnuson, S. R., Maqsodi, B., McDaniel, T., Mei, N., Myklebost, O., Ning, B., Novoradovskaya, N., Orr, M. S., Osborn, T. W., Papallo, A., Patterson, T. A., Perkins, R. G., Peters, E. H., Peterson, R., Philips, K. L., Pine, P. S., Pusztai, L., Qian, F., Ren, H., Rosen, M., Rosenzweig, B. A., Samaha, R. R., Schena, M., Schroth, G. P., Shchegrova, S., Smith, D. D., Staedtler, F., Su, Z., Sun, H., Szallasi, Z., Tezak, Z., Thierry-Mieg, D., Thompson, K. L., Tikhonova, I., Turpaz, Y., Vallanat, B., Van, C., Walker, S. J., Wang, S. J., Wang, Y., Wolfinger, R., Wong, A., Wu, J., Xiao, C., Xie, Q., Xu, J., Yang, W., Zhang, L., Zhong, S., Zong, Y., and Slikker, W., Jr. (2006) *Nat. Biotechnol.* **24,** 1151–1161

38. Nagata, S., Toyoda, M., Yamaguchi, S., Hirano, K., Makino, H., Nishino, K., Miyagawa, Y., Okita, H., Kiyokawa, N., Nakagawa, M., Yamanaka, S., Akutsu, H., Umezawa, A., and Tada, T. (2009) *Genes Cells* **14,** 1395–1404
39. Gupta, G., Surolia, A., and Sampathkumar, S. G. (2010) *OMICS* **14,** 419–436
40. Nakamura-Tsuruta, S., Kominami, J., Kamei, M., Koyama, Y., Suzuki, T., Isemura, M., and Hirabayashi, J. (2006) *J. Biochem.* **140,** 285–291
41. Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008) *Nat. Biotechnol.* **26,** 101–106
42. Miyazaki, T., Futaki, S., Hasegawa, K., Kawasaki, M., Sanzen, N., Hayashi, M., Kawase, E., Sekiguchi, K., Nakatsuji, N., and Suemori, H. (2008) *Biochem. Biophys. Res. Commun.* **375,** 27–32
43. Satomaa, T., Heiskanen, A., Mikkola, M., Olsson, C., Blomqvist, M., Tiittanen, M., Jaatinen, T., Aitio, O., Olonen, A., Helin, J., Hiltunen, J., Natunen, J., Tuuri, T., Otonkoski, T., Saarinen, J., and Laine, J. (2009) *BMC Cell Biol.* **10,** 42
44. Grahn, E., Askarieh, G., Holmner, A., Tateno, H., Winter, H. C., Goldstein, I. J., and Krengel, U. (2007) *J. Mol. Biol.* **369,** 710–721
45. Sulák, O., Cioci, G., Delia, M., Lahmann, M., Varrot, A., Imberty, A., and Wimmerová, M. (2010) *Structure* **18,** 59–72
46. Bremer, E. G., Levery, S. B., Sonnino, S., Ghidoni, R., Canevari, S., Kannagi, R., and Hakomori, S. (1984) *J. Biol. Chem.* **259,** 14773–14777
47. Zhuo, Y., and Bellis, S. L. (2011) *J. Biol. Chem.* **286,** 5935–5941
48. Klim, J. R., Li, L., Wrighton, P. J., Piekarczyk, M. S., and Kiessling, L. L. (2010) *Nat. Methods* **7,** 989–994
49. Sasaki, N., Okishio, K., Ui-Tei, K., Saigo, K., Kinoshita-Toyoda, A., Toyoda, H., Nishimura, T., Suda, Y., Hayasaka, M., Hanaoka, K., Hitoshi, S., Ikenaka, K., and Nishihara, S. (2008) *J. Biol. Chem.* **283,** 3594–3606
50. Sasaki, N., Hirano, T., Kobayashi, K., Toyoda, M., Miyakawa, Y., Okita, H., Kiyokawa, N., Akutsu, H., Umezawa, A., and Nishihara, S. (2010) *Biochem. Biophys. Res. Commun.* **401,** 480–486

CHEMICAL REVIEWS

REVIEW

pubs.acs.org/CR

Biomaterials for the Feeder-Free Culture of Human Embryonic Stem Cells and Induced Pluripotent Stem Cells

Akon Higuchi,*[,†,‡,§] Qing-Dong Ling,[§,∥] Yi-An Ko,[∥] Yung Chang,[⊥] and Akihiro Umezawa[‡]

[†]Department of Chemical and Materials Engineering, National Central University, Jhongli, Taoyuan, 32001 Taiwan

[‡]Department of Reproduction, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan

[§]Cathay Medical Research Institute, Cathay General Hospital, No. 32, Ln 160, Jian-Cheng Road, Xi-Zhi Dist., New Taipei City, Taiwan

[∥]Institute of Systems Biology and Bioinformatics, National Central University, No. 300, Jhongda RD., Jhongli, Taoyuan, 32001 Taiwan

[⊥]Department of Chemical Engineering, R&D Center for Membrane Technology, Chung Yuan Christian University, 200 Chung-Bei Rd., Jhongli, Taoyuan 320, Taiwan

CONTENTS

1. INTRODUCTION

Human embryonic stem cells (hESCs) are derived from the inner cell mass of 3- to 5-day-old blastocysts.[1−5] hESCs are characterized by a high nucleus-to-cytoplasm ratio, prominent nucleoli, and distinct colony morphology.[6] Recently, pluripotent stem cells that are similar to ESCs were derived from an adult somatic cell by the "forced" expression of certain pluripotency genes,[7−11] such as Oct3/4, Sox2, c-Myc, and klf-4, or their proteins[12] and microRNAs.[13] These cells are known as induced pluripotent stem cells (iPSCs). iPSCs are believed to be similar to ESCs in many respects, including the expression of certain stem cell genes and proteins, chromatin methylation patterns, doubling time, embryoid body formation, teratoma formation, viable chimera formation, potency, and differentiability. However, the full extent of their similarities to ESCs is still under investigation.[7,10]

hESCs and human iPSCs have significant potential in therapeutic applications for many diseases because they have the specific ability to differentiate into all types of somatic cells.[14] For example, hESCs and human iPSCs that have been differentiated into nerve cells that secrete dopamine or β cells that secrete insulin can be transplanted for the treatment of Parkinson's disease[15−17] and diabetes,[18−20] respectively. The pluripotent nature of these cells could permit the development of a wide range of potential stem cell-based regenerative therapies and possible drug discovery platforms.[14]

However, the tentative clinical potential of hESCs and human iPSCs is restricted by the use of mouse embryonic fibroblasts (MEFs) as a feeder layer. While the addition of the leukemia inhibitory factor (LIF) to the culture medium can allow mouse ESCs to proliferate and remain undifferentiated in the absence of a feeder layer of MEFs, this method is not effective for hESCs.[1,2] The addition of LIF to the culture medium is insufficient to maintain the pluripotency and self-renewal of hESCs in a feeder layer-free culture.[6] The possibility of xenogenic contamination during culture restricts the clinical use of transplanted hESCs and human iPSCs. Furthermore, the process of culturing hESCs and human iPSCs using feeder layers is elaborate and costly, limiting the large-scale culture of those cells. The variability of MEFs between laboratories and across batches also affects the characteristics and differentiation abilities of hESCs and human iPSCs. The development of feeder-free cultures using synthetic polymers or biomacromolecules as stem cell culture materials will offer more reproducible culture conditions and lower the cost of production without introducing xenogenic contaminants. These improvements will increase the potential clinical applications of differentiated hESCs and human iPSCs.[6]

Several factors in the microenvironment and niches of stem cells influence their fate: (1) several soluble factors, such as growth factors or cytokines, nutrients, and bioactive molecules; (2) cell−cell interactions; (3) cell−biomacromolecule (or biomaterial) interactions; (4) and physical factors, such as the rigidity of the environment (Figure 1). Mimicking the stem cell microenvironments and niches

Received: October 27, 2010
Published: February 23, 2011

   dx.doi.org/10.1021/cr1003612 | Chem. Rev. 2011, 111, 3021–3035