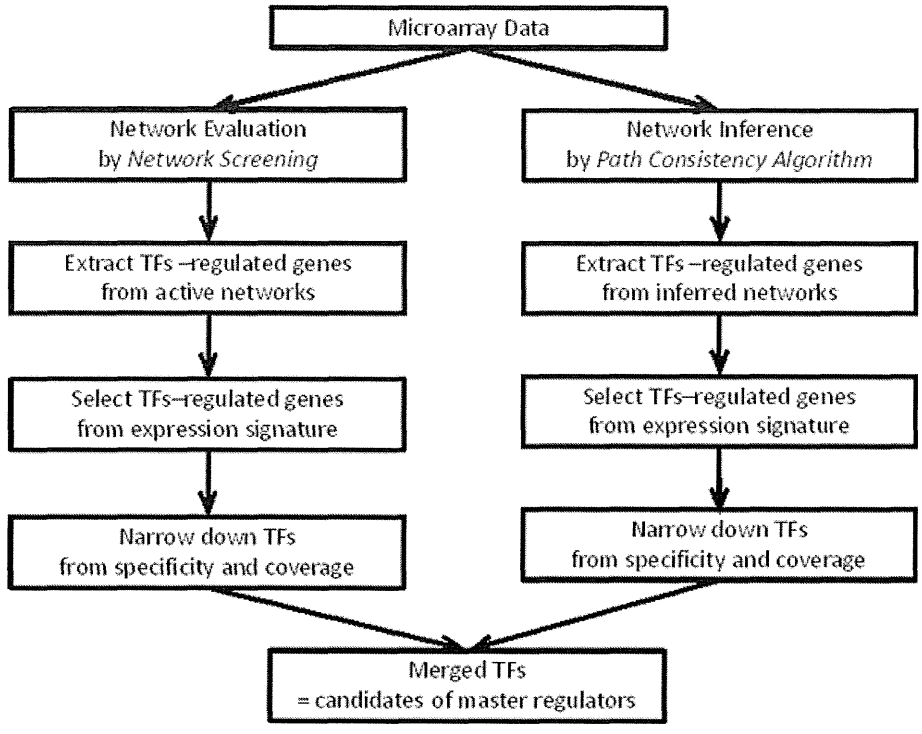

Figure 1 Workflow of the MR identification procedure.

gene relationships emerged iteratively for different periods in GK and WKY rats.

Among the TF-gene relationships selected above, the TFs were narrowed down in two ways. First, the TF-gene relationships were selected by the specificity, which means that the TFs emerge only in GK, but not in WKY. As a result, we found a total of 21 TFs, as shown in Table 1. Second, the TF-gene relationships were selected by the coverage, which means how many genes the TFs regulate, among the genes in the expression signature. The TFs thus selected were sorted according to the coverage, and then the MR candidates were further selected by a statistical test (see the Methods) for each period in GK and WKY listed in Table 2. As seen in the table, most of the TFs emerged in both GK and WKY, in terms of the coverage selection. We finally found 3 TFs (EGR1, NRF1, and TCFAP2A) among the genes by the initial selection in Table 2.

MR candidates inferred by the modified path consistency algorithm

We first inferred six networks of all genes on the microarray for each of the three periods in GK and WKY rats, by the modified path consistency algorithm [10,11], and then the TF-gene relationships were extracted from each network. After the extraction, only the relationships that included the genes with a significant difference between GK and WKY rats were further selected for the 6 sets of relationships.

Using the same procedure as that described in the preceding subsection, the TFs were narrowed down. First, we chose the relationships in terms of the gene-emergence specificity. As a result, 108 TFs were identified as the MR candidates in Table 3. The number of candidates seems to be large, even in comparison with the candidate number, 27 TFs, in the previous case of the brain tumor [3]. While one network was considered to identify the candidates in the previous paper, three networks for the three periods in GK rats were surveyed to select the candidates in the present study. Thus, the number of TFs extracted from one network, 36 TFs on average, is similar to that in the previous study. Second, the TF-gene relationships were selected by the coverage. We chose the TF-gene relationships by a statistical test (see the details in the Methods) for each period in GK and WKY, as shown in Table 4. In contrast to the

Table 1 TFs identified by network screening in terms of specificity.

Ar, Bcl6, Brca1, Etv4, Fus, Gli1, Hes1, Hnf1b, Hnrnpk, Klf10, Klf4, Lyl1, Mef2c, Nfia, Nr2f1, Nrl, Pax6, Sp2, Sp4, Tcfap2b, Wt1

All of the gene names are cited from the Rat Genome Database http://rgd.mcw.edu/ in all of the tables, the figures, and the text.

Table 2 TFs identified by network screening in terms of coverage.

4w				8w_12w				16w_20w			
GK		WKY		GK		WKY		GK		WKY	
TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes
SP1	10	**SP1**	19	**SP1**	39	**SP1**	18	**SP1**	12	**SP1**	5
		SP3	8	**SP3**	11	HNF4A	6	**SP3**	3	**FOXO3**	3
		TP53	4	**TP53**	11	**FOXO3**	4				
				EGR1	6						
				NRF1	6						
				TCFAP2A	5						

TFs found in both GK and WKY are indicated by bold letters.

coverage selection in network screening, only a few TFs emerged in both GK and WKY. Indeed, among the 44 TFs in Table 4, only two TFs (Tbpl1 and Cbfb) emerged in both GK and WKY. Finally, we found 42 TFs as MR candidates.

MR selection by comparison of the TF sets detected by the two methods

We obtained the final MR candidates by selecting the overlapped TFs detected by the two methods in terms of two criteria (Tables 1, 2, 3, 4), as shown in Table 5. Indeed, 21 TFs detected by network screening in terms of specificity overlapped with only 4 TFs (Etv4, Nr2f1, Sp2, and Tcfap2b) and 2 TFs (Fus and Sp2) by the modified path consistency algorithm by two criteria, respectively. In contrast, 3 TFs detected by network screening in terms of coverage showed no overlapped TFs by the path consistency algorithm by two criteria. This difference might reflect the restriction of the known TF-gene relationships in network screening.

As a result, we merged the MR candidates identified by the two methods, and 5 TFs were finally identified as the candidates of MRs for diabetes progression in GK rats. Note that Sp2 emerged in both the 4 TFs and 2 TFs. The 5 final MR candidates with their regulated genes, in total 54 genes, are listed in Table 6.

Discussion

In this study, we have identified the candidates of master regulators based on our previous study [5], by using an improved method for their identification [8]. The MR candidates were extracted from the active networks of many genes characterized by biological pathways, as the feasible gene candidates for experimental verification. From the methodological aspect, the method was improved by considering the coverage of TFs in a statistical manner, in addition to the specificity that was considered in the previous method. Although the experiments are beyond the scope of the present study, we consider experimental verification studies of the present candidates as our future research topic. Our study clearly illustrated a rational way to narrow down the genes of MR candidates, and is fundamentally different from metaphysical presentations, such as biological pathways or large network forms.

Our study intended to identify the MR candidates, which are those genes with large impacts on phenotype changes, in a biological sense [3]. Here, we logically identified MR candidates by the specificity of the TF appearance and the coverage of the regulated genes to the gene expression signature in the networks of GK and WKY rats. Apart from a biological sense, we further investigated the meaning of "master" from the viewpoint of the network structure. To do this, we revealed the hierarchical structures of the 8w-12w and 16w-20w networks by network screening, using a vertex sort algorithm [12], and allocated the present 5 TFs into the hierarchical structures (Figure 2A). As seen in the figures, all 5 TFs were allocated into the highest level. Indeed, Nr2f1 in the 8w-12w network and Tcfap2b in the 16w-20w network were definitely allocated into the highest level of the hierarchical structures. In addition, the remaining TFs were allocated into the levels including the highest and middle levels, but not into the lowest level. Furthermore, we investigated the hierarchical structure by another method, the BFS-level algorithm

Table 3 TFs identified by network inference in terms of specificity.

Alx1, Arnt, Cebpg, Ddit3, Dlx5, Dmrt2, Dnmt1, Dr1, Ebf1, Elf5, Elk3, Elk4, Erg, Etv4, Etv5, Fev, Fosl1, Foxe1, Foxg1, Foxo3, Foxp4, Gabpb1l, Gfi1, Gtf2a1, Gtf2b, Gtf2e1, Gzf1, Hcfc1, Hey1, Hhex, Hoxb3, Hoxb7, Ilf3, Irx2, Kcnip4, Klf1, Klf15, Klf3, Klf5, Klf7, Ldb2, LOC680117, Mafk, Meis2, Mnat1, Msx1, Msx2, Mybl2, Myc, Myocd, Myod1, Mzf1, Neurod2, Nfix, Nfx1, Nkx6-1, Notch1, Nr1h4, Nr2f1, Nr4a1, Nr5a1, Pax8, Pbx2, Phox2a, Pitx1, Pitx3, Pou2f3, Pou3f1, Ppard, Pparg, Ppargc1a, Rbl1, RGD1566107, Rreb1, Runx1, Shh, Six5, Six6, Skp2, Sox10, Sox11, Sp1, Sp2, Spdef, Srebf1, Ss18l1, Stat5a, Stat5b, Taf2, Tbx18, Tbx2, Tcf12, Tcfap2b, Tead1, Tfdp2, Tfec, Tmf1, Tp53bp1, Twist1, Vdr, Zbtb5, Zfhx3, Zfp191, Zfp238, Zfp423, Zfp444, Zhx1, Zic1

Table 4 TFs identified by network inference in terms of coverage.

4w				8w_12w				16w_20w			
GK		WKY		GK		WKY		GK		WKY	
TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes	TF	No. of regulated genes
Arntl	31	Max	10	Lhx5	24	Ywhae	18	Fus	10	Foxq1	32
Lhx2	22	Otx2	10	Etv1	23	Pfdn5	13	Smad5	10	Hoxa1	16
Sp2	18	Daxx	9	Ctnnb1	8	Atf1	11	Nfx1	9	Rbl2	16
Gabpa	13	Sim1	9	Rpa3	8	Cdk9	11	Hsf1	8	Zic2	12
Xpa	4	Tcf21	8	Zfp105	8	Hmgb2	11	Tlx3	8	Rorc	8
Foxs1	3	Gata5	7	Foxo3	7	Sfpq	9	Tp53	8	Tcfap4	6
		Tcfap2c	7	Hoxc5	6	Zfp281	9	Foxs1	7	Pttg1	5
		Meis3	5	Litaf	6	Cdk7	8	LOC679869	7	Ncoa3	4
		Rorc	5	Nr2f2	6	Ets2	8	**Cbfb**	6	Ccnh	3
		Snapc1	5	Foxo1	5	Hoxa1	8	Ctcf	6	Hif1a	3
		Zic2	5	Msx1	5	Nfe2l2	8	Glis2	6	Junb	3
		Meis1	4	Myocd	5	Nfil3	8	Irf7	6	Kcnip1	3
		Pou2af1	4	Pbx1	5	Six4	8	Nfkbib	6	Mtf1	3
		Srf	4	**Tbpl1**	5	Cux2	7	Nr1i2	6	Zfp148	3
		Stox2	4	Vdr	5	Mafg	7	Hdac1	5		
		Tcfcp2l1	4	Hltf	4	Nfkbia	7	Rfx5	5		
		Gtf2h2	3	Htt	4	Pgr	7	Tle1	5		
		Zfx	3	LOC680117	4	Ppp1r13b	7	Xpa	5		
				Mbd1	4	**Tbpl1**	7				
				Parp1	4	**Cbfb**	6				
				Rreb1	4	Ezh2	6				
				Smarcc1	4	Hbp1	6				
						Junb	6				
						Taf13	6				
						Tef	6				

TFs found in both GK and WKY are indicated by bold letters.

[13]. As shown in Figure 2B, the positions of the MR candidates are similar to those in Figure 2A. Indeed, previous hierarchical analyses of the regulatory networks by the BSF method in *Escherichia coli* and *Saccharomyces cerevisiae* suggested that the MRs were in the middle of the hierarchy [13]. In general, the vertex sort algorithm reports a linear ordering of nodes that contains all feasible solutions, while the BSF-level algorithm reports just a single solution, as shown in Figures 2A and 2B. Subsequently, unlike the BFS-level algorithm, the ordering in the vertex sort algorithm permits nodes to span an entire interval of possible positions with any feasible ordering. Despite this difference in the computational algorithms, the 5 TFs showed the common property as MRs. At any rate, although the verification experiments remain to be performed for the justification of the MRs in a biological sense, the 5 TFs may be regarded as the plausible MR candidates from the viewpoint of network structure.

Table 5 Summary of TFs identified by the two methods, in terms of specificity and coverage.

		path consistency algorithm	
		specificity (108)	coverage (42)
network screening	specificity (21)	4	2
	coverage (3)	0	0

A preliminary survey revealed that all 5 of the TFs have no reported causal relationship to diabetes. The 5 TFs are sequence-specific DNA-binding proteins, and they function as both transcriptional activators and repressors of large numbers of genes that are closely related to the cell cycle and tumorigenesis. Notably, the relationships of ETS translocation variant 4 (Etv4) and transcription factor AP-2 beta (Tcfap2b) to adipogenesis, which is strongly related to diabetes, have been reported, together with their association with the other pathways [14,15]. Nuclear Receptor subfamily 2, group F, member 1 (NR2F1) is a member of the steroid hormone receptor family, and has been shown to interact with estrogen receptor alpha (ESR1) [16]. There is a gender difference in the incidence of type 2 diabetes,

Table 6 Candidates of MRs and their regulated genes for diabetes progression in GK rat.

TF	Regulated genes						No. of genes	
Etv4	Mcm10	**Erbb2**	**Mmp7**	**Nid1**	Plau	Ptgs2	6	
Fus	Mcpt8l2	Mcpt9	**Paics**	**Ppat**	Ugt1a1	Ugt1a2	12	54
	Ugt1a3	Ugt1a5	Ugt1a6	Ugt1a7c	Ugt1a8	Ugt1a9		
Nr2f1	Alox5	**Cpt1b**	Cyp11b2	**Tf**	Ugt1a3	Ugt1a5	6	
Sp2	**Capns1**	**Irs2**	LOC685183	LOC685226	LOC685291	LOC685759	24	
	LOC688519	LOC688603	LOC689083	LOC689312	LOC689338	LOC689690		
	LOC689999	LOC690179	LOC690328	LOC690379	LOC690577	LOC691712		
	LOC691735	LOC691754	Papss2	Vom2r45	Vom2r46	Vom2r47		
Tcfap2b	Aqp1	Egfr	Krt14	Ptgds	Sod2	Tgm1	6	

The genes in bold characters are included in known TF-gene relationships detected by network screening.

which is largely due to the role of the sex hormone estrogen. The Sp family proteins, containing the conserved DNA-binding domain, are localized primarily within subnuclear foci associated with the nuclear matrix. Recent unpublished data from our lab have shown that another Sp family member, Sp1, has a major impact on the insulin signaling pathway. The Sp2 transcription factor interacts with E2F1, which mediates both cell proliferation and p53-dependent/independent apoptosis [17]. The recently discovered close relationships between diabetes and tumors in terms of these TFs are quite likely to play a crucial role in the control of diabetes. RNA-binding protein (FUS) is able to bind DNA, RNA and protein [18]. The interactions between the FUS recognition sites and Tcfap2, GCF, and Sp1 were identified recently. Thus, although direct evidence was not found in the previous knowledge, the 5 TFs are expected to be MR candidates, in consideration of the circumstantial evidence of their relationships to diseases, the hierarchical analysis of the 5 TFs, and the successful discovery of new MRs in brain tumor, by the previous version of the procedure. Actually, our current information in terms of important diabetes-related genes includes mostly functional proteins, located at the lowest level of our hierarchical structure, while the MR is deeply hidden and therefore must be revealed by systems biology methods. Thus, in addition to analyses of their regulated genes, some experimental verification of the MR candidates may be desirable to further examine

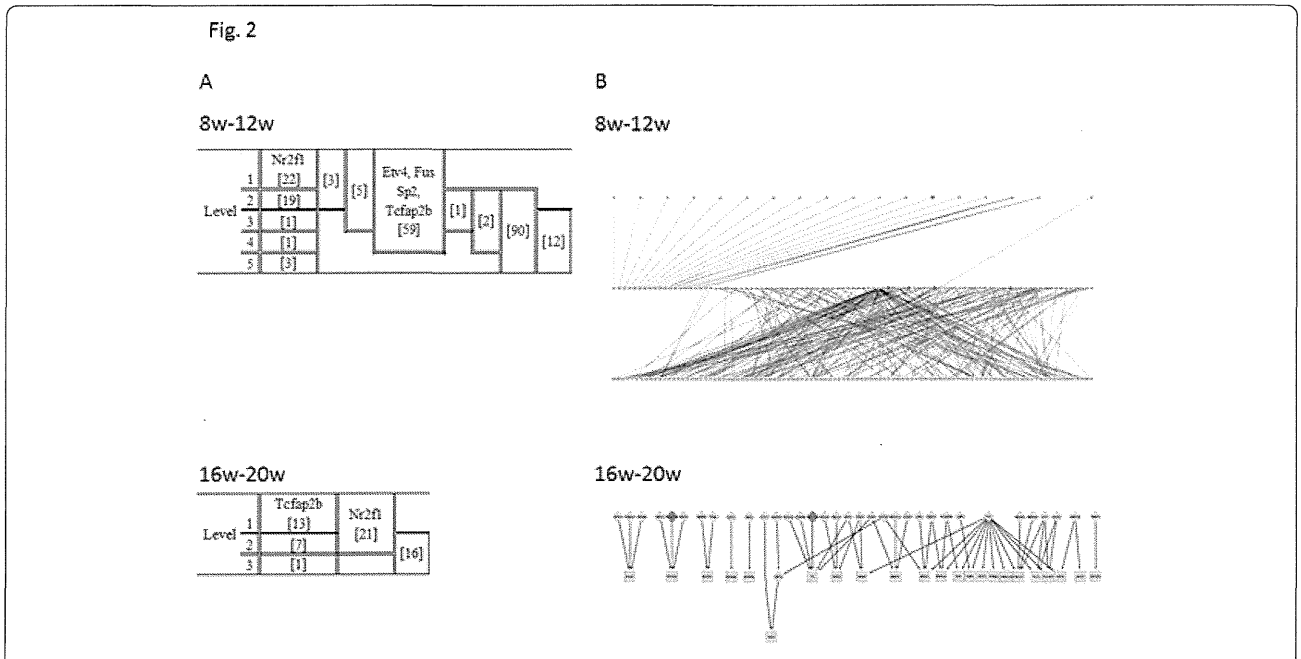

Figure 2 Hierarchical structures of networks for 8w-12w and 16w-20w by two previous methods. The 5 TFs are indicated at the levels in hierarchical structures obtained by the vertex-sort algorithm (A) [12] and those by the BFS method (B) [13], and the numbers of TFs in each level are indicated in parentheses in (A), and by red circles in (B). In (B), the TFs and the regulated genes are indicated by diamonds and rectangles, respectively.

their plausibility as MR candidates for diabetes progression.

Conclusions

In this work, using our new method, we identified the MR candidates for diabetes progression, 5 TFs and their regulated genes, in GK rats. This number of candidates is very small, and thus the results can be used as a basis for biological experiments for verification. Furthermore, the recent availability of the next-gen sequencer may provide another way to confirm the effectiveness of our method, and to test its performance further with other datasets. Indeed, RNA-seq and ChIP-seq are useful for more accurate measurements of gene expression, and yield detailed information about the regulated genes. Thus, the combined use of the two approaches may compensate for the pitfalls inherent in each method, and will provide important clues about the transcriptional networks that regulate transitions into physiological or pathological cellular states.

Methods

Network screening

The candidates of the active regulatory networks were detected by network screening [5-7]. Here, we briefly summarize the network screening in the present study, as follows.

First, the regulatory network sets were generated in the same manner as in the previous study [5], as follows. The mouse binary relationships compiled in the TRANSFAC database [19] were used. Based on the correspondence between the mouse and rat gene ids, 3,015 binary relationships of 1,507 genes between 503 TFs and 1,123 regulated genes were achieved. Based on those binary relationships, transcriptional networks were constructed according to the functional gene sets previously defined in the Molecular Signatures Database (MSigDB) [20]. In each gene set, the regulated genes in the binary relationships were searched, and if at least one gene was found in the gene set, then the corresponding binary relationships were regarded as a regulatory network characterized by the gene set. In present study, the reference network comprised 1,760 regulatory networks characterized by biological functions that are composed of 1,195 genes. The numbers of TFs and regulated genes were 335 and 860, respectively.

Then, we calculated the graph consistency probability (GCP) [6], which expressed the consistency of a given network structure with the monitored expression data of the constituent genes in this study. The consistency of a directed acyclic graph (DAG), $G(V_i, E_j)$, where V_i is a vertex (i = 1, 2, ..., n_v) and E_j is an edge (j = 1, 2, ..., n_e) in the graph, and the joint density function $f(X_i)$, corresponding to V_i for the graph G with the measured data, is quantitatively expressed by the logarithm of the likelihood based on the Gaussian graphical model (*GN*: Gaussian Network), i.e.,

$$\begin{aligned} l(G_0) &= \ln \prod_{i=1}^{n_v} f\left(X_i | pa\{X_i\}\right) \\ &= -\frac{1}{2}\sum_{i=1}^{n_v}\sum_{j=1}^{n_i}\left\{\frac{1}{\sigma_i^2}\sum_{k=1}^{m}\left(x_{ik}-\sum_{j=1}^{n_i}\beta_{ij}x_{kj}\right)^2 + \ln\left(2\pi\sigma_i^2\right)\right\}, \end{aligned} \quad (1)$$

where $pa\{X_i\}$ is the set of variables corresponding to the parents of V_i in the graph, x_{ik} is the measured value of X_i, at the k-th point, and n_i is the number of variables corresponding to the parents of V_i. Since the likelihood depends on the graph size, we designed a simple procedure to transform the likelihood to the probability for the expression of the graph consistency with the data [6]. First, we generated N_r networks under the condition that the networks shared the same numbers of nodes and edges as those of the given networks. Then we defined GCP, as follows,

$$GCP = \frac{N_s}{N_r}, \quad (2)$$

where N_s is the number of networks with larger log-likelihoods than the log-likelihood of the tested network. In the present study, N_r was set to 2,000, and the *GCP* significance of the given network was set at 0.05.

Path consistency algorithm

The path consistency (PC) algorithm [9] is an algorithm to infer a causal graph composed of two parts: the undirected graph inference by a partial correlation coefficient and the following directed graph construction by the orientation rule. The present method partially exploits the first part of the PC algorithm for the inference of the network structures. A simple example of the PC algorithm is illustrated in Figure 3.

We assume that five variables, X_1, X_2, X_3, X_4, X_5, have the following five relationships: i) $X_1 \perp\!\!\!\perp X_2$,

ii) $X_2 \perp\!\!\!\perp (X_1, X_4)$,

iii) $X_3 \perp\!\!\!\perp X_4 | (X_1, X_2)$,

iv) $X_4 \perp\!\!\!\perp (X_2, X_3) | X_1$, and

v) $X_5 \perp\!\!\!\perp (X_1, X_2) | (X_3, X_4)$,

where the symbol, $\perp\!\!\!\perp$, in the above relationships, means the independence between variables. The PC algorithm reconstructs the above relationships as follows.

1) Prepare a complete graph, C, between the five variables.

2) Test the correlation between two variables by calculating the zeroth-order of the partial correlation coefficient (Pearson's correlation coefficient). From the test,

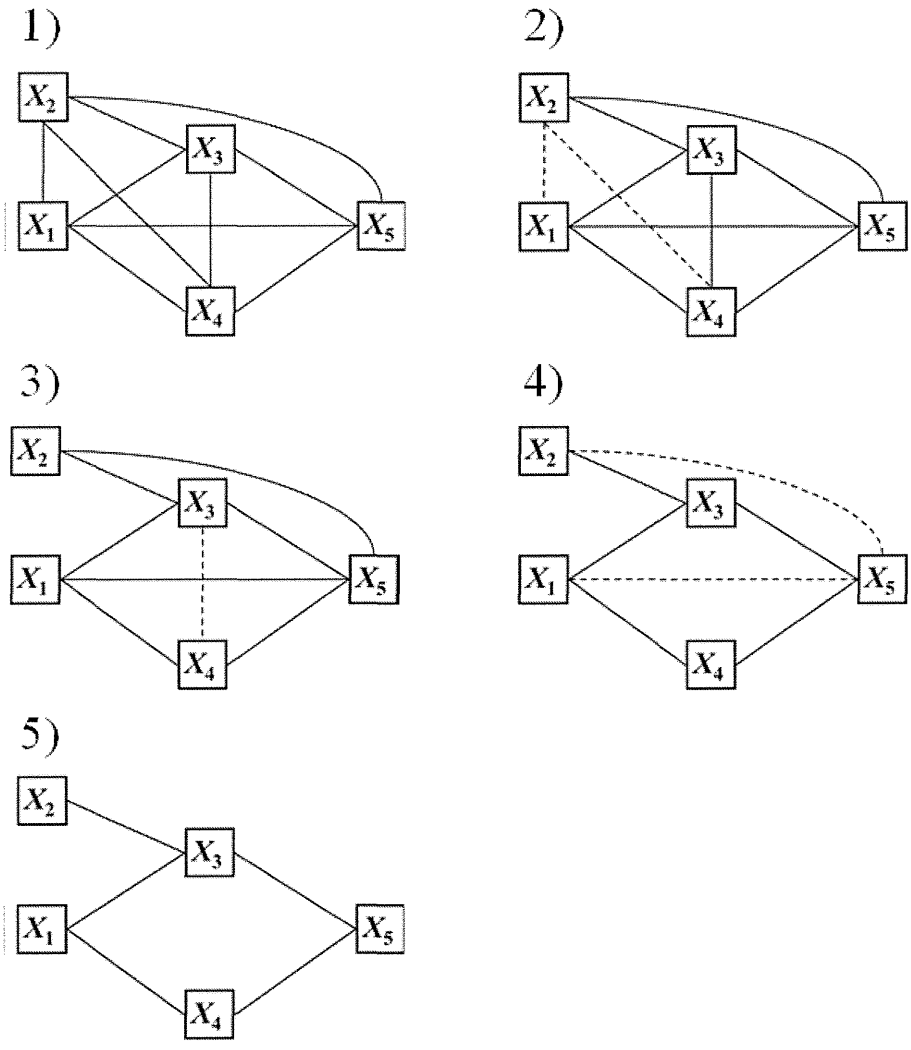

Figure 3 Example of the path consistency algorithm.

two variable pairs, (X_1, X_2) and (X_2, X_4), are excluded (dashed lines in Figure 2), due to the relationships, i) and ii).

3) Test the correlation between three variables by calculating the first-order of the partial correlation coefficient of the variable pairs, given one variable. Then, one variable pair, (X_3, X_4), is further excluded from the updated graph by 2), due to iii) and iv).

4) Test the correlation between four variables by calculating the second-order of the partial correlation coefficient of the variable pairs, given two variables. Then, two variable pairs, (X_1, X_5) and (X_2, X_5), are excluded, due to iv).

5) We could not find any edges adjacent to the three edges in the updated C. Thus, the algorithm naturally stops. As seen in the final graph, the five relationships emerged completely.

In general, the (m-2)-th order of the partial correlation coefficient is calculated between two variables, given (m-2) variables; i.e., $r_{ij,\ rest}$, between X_i and X_j, given the '*rest*' of the variables, $\{X_k\}$ for k = 1, 2, ..., m, and $k \neq i, j$, and after calculating the (m-2)-th order of the partial correlation coefficient, the algorithm naturally stops. However, the algorithm does not usually request the (m-2)-th order of the correlation coefficient for the natural stop. This is because after excluding the variables, the adjacent variables are often not found, even in the calculation of the lower orders of partial correlation coefficients.

Modification of the path consistency algorithm for microarray data analysis

In the actual expression profile data, many genes frequently show profiles with similar patterns. This makes

the numerical calculation of correlation coefficients difficult, due to the multi-colinearity between the variables. The original PC algorithm accidentally stops, if only one correlation between a pair of variables shows a violation of the numerical calculation. However, in a biological sense, the gene pairs that cause the accidental stop can be interpreted as a case of their high association with each other, in terms of gene expression. Thus, we modified the original PC algorithm to prevent it from accidentally stopping with the highly associated gene pairs, as follows [10,11]. If the calculation of any order of the partial correlation coefficient between the variables is violated, then the corresponding pair of variables is regarded as being dependent. For example, if the first-order correlation coefficient, $r_{ij,\ k}$, cannot be calculated numerically, due to the multi-colinearity between X_i and X_j, then the edge X_i-X_j is kept without the statistical test. The other parts remain unchanged in the modified algorithm. Note that the above modification ensures that the algorithm will naturally stop for the data including a high correlation.

As seen in the original algorithm, the output is not unique, depending on the calculation order of pairs [9]. A permutation test for the calculation order is a convenient way to partly resolve this issue. In this study, the estimation without permutation was empirically adopted as the first approximation, based on the successful estimations of the relationships in our previous studies [10,11]. In addition, one of the most remarkable features of the PC algorithm is that the algorithm removes the pseudo-correlations between the variables (genes) by considering the higher-order partial correlations. If we have the measurement data for a complex network, then we frequently face the more serious issue of the pseudo-correlation, rather than the correlation level. The merit of the PC algorithm may be its ability to identify real relationships between TFs and their regulated genes.

Definition of MR candidates by network screening and network inference

We first referred to two sets of networks obtained by the network screening [5-7] and the network inference [10,11]. From each network set, the binary relationships between the TFs and their regulated genes were extracted, only if the regulated genes were included in the expression signature, which is the ensemble of genes with significant differences in gene expression, as statistically estimated by the false discovery rate (FDR) test for multiple comparisons (FDR < 0.05) [21]. In the extraction of TFs and their regulated genes, the TF was also cited from the TRANSFAC database [19], but the expression degree of the TF was not considered, due to the small expression changes even under different conditions. Only the regulated genes that were estimated to directly bind TFs were extracted. The numbers of genes in the three gene expression signatures of the three periods (period of 4w, period of 8w and 12w, and period of 16w and 20w) were 1,582, 2,719, and 2,777, respectively.

Then, we defined the MR candidates from the binary relationships by two criteria. One was the specificity of the TF, which was the same criterion as in the previous method [8], and the other was the coverage of the TF, which was newly introduced in the present MR candidate identification. Here, the specificity simply means that the TF emerged only in the GK networks, but not in the WKY networks. To select the TFs in terms of the specificity, we selected the TFs that emerged in the three periods in GK, but not in WKY, as the MR candidates. Note that in the selection of the TFs, we only selected those that were estimated to regulate the genes including the expression signature, to consider the enrichment of the regulated genes in the signature. The coverage means how many genes each TF regulates. To select the TFs in terms of the coverage, we first counted the genes regulated by each TF for each period in GK and WKY, and then also considered the enrichment of their regulated genes in the expression signature, by sorting the numbers of regulated genes for each case. To consider the coverage in a rational way, we used the Smirnov-Grubbs outlier test [22] for the numbers of regulated genes, by setting a threshold ($p < 0.05$). Thus, the TFs with the larger number of regulated genes that fulfilled the threshold are selected in a statistical manner. Finally, the two sets of MR candidates that were selected in terms of the specificity and the coverage were compared, to define the final MR candidates.

Data analyzed in this study

We analyzed the gene expression data measured in GK and WKY rats [23], which were cited from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/projects/geo/) database (GSE 13271). The data were composed of 31,099 probes that were measured by using Affymetrix Microarray Suite 5.0 (Affymetrix), and were further reduced into 14,506 genes, for 5 samples of male spontaneously diabetic GK rats and WKY controls at each of 5 time points (4, 8, 12, 16, and 20 weeks of age). In this analysis, the 5 periods were classified into three periods: period of 4w, period of 8w and 12w, and period of 16w and 20w.

Acknowledgements

This work was supported by a grant, "Joint Seminar 2011 in NSFC-JSPS Scientific Cooperation Program". This work was also partly supported by a project grant, entitled "Development of Analysis Technology for Induced Pluripotent Stem (iPS) Cell" from NEDO of Japan; Major State Basic Research Development Program of China (973 Program) under No. 2011CB504003; NSFC under Nos. 61134013, 81070657, 31100949, 61072149 and 91029301;

the Chief Scientist Program of SIBS of CAS under Grant No. 2009CSP002; the Knowledge Innovation Program of SIBS of CAS with Grant Nos. 2011KIP203 and KSCX2-EW-R-01; Shanghai NSF under Grant No. 11ZR1443100; and the SA-SIBS Scholarship Program.
This article has been published as part of *BMC Systems Biology* Volume 6 Supplement 1, 2012: Selected articles from The 5th IEEE International Conference on Systems Biology (ISB 2011). The full contents of the supplement are available online at http://www.biomedcentral.com/bmcsystbiol/supplements/6/S1.

Author details
[1]School of Life Sciences, University of Science and Technology of China, Hefei 230026, China. [2]Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China. [3]Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan. [4]INFOCOM Corporation, Tokyo 150-0001, Japan. [5]Key Laboratory of Human Functional Genomics of Jiangsu Province, Nanjing Medical University, Nanjing 210029, China. [6]National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.

Authors' contributions
HZ, LC and KH conceived the research. SS, GP, YS and ZPL performed the study. JW, YW and XH provided valuable suggestions and improvements. HZ, LC and KH supervised the project. HZ, ZPL, SS and KH drafted a version of the manuscript. All authors wrote and approved of the manuscript.

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

Published: 16 July 2012

References

1. Margolin AA, *et al*: **Reverse engineering cellular networks.** *Nature Protocols* 2006, **1**:662-671.
2. Mani KM, *et al*: **A systems biology approach to prediction of oncogenes and perturbation targets in B cell lymphomas.** *Mol Syst Biol* 2008, **4**:169-178.
3. Carro MS, *et al*: **The transcriptional network for mesenchymal transformation of brain tumours.** *Nature* 2010, **463**:318-325.
4. Chen L, Wang RS, Zhang XS: *Biomolecular Networks: Methods and Applications in Systems Biology* Wiley; 2009.
5. Zhou H, *et al*: **Network Screening of Goto-Kakizaki Rat Liver Microarray Data during Diabetic Progression.** *BMC Syst Biol* 2011, **5(Suppl 1)**:S16.
6. Saito S, *et al*: **Network evaluation from the consistency of the graph structure with the measured data.** *BMC Syst Biol* 2008, **2**:84.
7. Saito S, *et al*: **Potential linkages between the inner and outer cellular states of human induced pluripotent stem cells.** *BMC Syst Biol* 2011, **5(Suppl 1)**:S17.
8. Saito S, *et al*: **Identification of Master Regulator Candidates in Conjunction with Network Screening and Inference.** *Int J Data Mining and Bioinformatics* .
9. Spirtes P, Glymour C, Scheines R: *Causation, Prediction, and Search (Springer Lecture Notes in Statistics, 2nd edition, revised)* MIT Press, Cambridge; 2001.
10. Saito S, Horimoto K: **Co-Expressed Gene Assessment Based on the Path Consistency Algorithm: Operon Detention in Escherichia coli.** *Proceedings of IEEE International Conference on Systems, Man and Cybernetics* 2009, 4280-4286.
11. Saito S, *et al*: **Discovery of Chemical Compound Groups with Common Structures by a Network Analysis Approach.** *J Chem Inf Model* 2011, **51**:61-68.
12. Jothi R, *et al*: **Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture.** *Mol Syst Biol* 2009, **5**:294.
13. Yu H, Gerstein M: **Genomic analysis of the hierarchical structure of regulatory networks.** *Proc Natl Acad Sci USA* 2006, **103**:14724-14731.
14. Park KW, *et al*: **The small molecule phenamil is a modulator of adipocyte differentiation and PPARγ expression.** *J Lipid Res* 2010, **51**:2775-2784.
15. Tao Y, *et al*: **The transcription factor AP-2beta causes cell enlargement and insulin resistance in 3T3-L1 adipocytes.** *Endocrinology* 2006, **147**:1685-1696.
16. Brown KK, *et al*: **NR2F1 deletion in a patient with a de novo paracentric inversion, inv(5)(q15q33.2), and syndromic deafness.** *Am J Med Gen Part A* 2009, **149A**:931-938.
17. Letourneur M, *et al*: **Sp2 regulates interferon-gamma-mediated socs1 gene expression.** *Mol Immunol* 2009, **46**:2151-2160.
18. Kwiatkowski TJ Jr, *et al*: **Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis.** *Science* 2009, **323**:1205-1208.
19. Wingender E: **TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation.** *Brief Bioinformatics* 2008, **9**:326-332.
20. 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.
21. Benjamini Y, Yekutieli D: **The control of the false discovery rate in multiple testing under dependency.** *Ann Statistics* 2001, **29**:1165-1188.
22. Grubbs FE: **Sample criteria for testing outlying observations.** *Ann Math Statistics* 1950, **21**:27-58.
23. Almon RP, DuBois DC, Lai W, Xue B, Nie J, Jusko WJ: **Gene expression analysis of hepatic roles in cause and development of diabetes in Goto-Kakizaki rats.** *J Endocrinol* 2009, **200**:331-346.

doi:10.1186/1752-0509-6-S1-S2
Cite this article as: Piao *et al.*: A computational procedure for identifying master regulator candidates: a case study on diabetes progression in Goto-Kakizaki rats. *BMC Systems Biology* 2012 **6**(Suppl 1):S2.

BIOLOGY OF REPRODUCTION (2012) **86**(6):182, 1–11
Published online before print 14 March 2012.
DOI 10.1095/biolreprod.111.096792

Tracing the Conversion Process from Primordial Germ Cells to Pluripotent Stem Cells in Mice[1]

Go Nagamatsu,[2,3,6] Takeo Kosaka,[3,5] Shigeru Saito,[7,8] Keiyo Takubo,[4] Hideo Akiyama,[9] Tetsuo Sudo,[9] Katsuhisa Horimoto,[7,10] Mototsugu Oya,[5] and Toshio Suda[4]

[4]*Department of Cell Differentiation, The Sakaguchi Laboratory, School of Medicine, Keio University, Tokyo, Japan*
[5]*Department of Urology, School of Medicine, Keio University, Tokyo, Japan*
[6]*Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, Kawaguchi, Japan*
[7]*Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan*
[8]*Chem & Bio Informatics Department, INFOCOM Corporation, Tokyo, Japan*
[9]*Toray New Frontiers Research Laboratories, Kanagawa, Japan*
[10]*Institute of Systems Biology, Shanghai University, Shanghai, China*

ABSTRACT

To understand mechanisms underlying acquisition of pluripotency, it is critical to identify cells that can be converted to pluripotent stem cells. For this purpose, we focused on unipotent primordial germ cells (PGCs), which can be reprogrammed into pluripotent embryonic germ (EG) cells under defined conditions. Treatment of PGCs with combinations of signaling inhibitors, including inhibitors of MAP2K (MEK), GSK3B (GSK-3beta), and TGFB (TGFbeta) type 1 receptors, induced cells to enter a pluripotent state at a high frequency (12.1%) by Day 10 of culture. When we employed fluorescence-activated cell sorting to monitor conversion of candidate cells to a pluripotent state, we observed a cell cycle shift to S phase, indicating enrichment of pluripotent cells, during the early phase of EG formation. Transcriptome analysis revealed that PGCs retained expression of some pluripotent stem cell-associated genes, such as *Pou5f1* and *Sox2*, during EG cell formation. On the other hand, PGCs lost their germ lineage characteristics and acquired expression of pluripotent stem cell markers, such as *Klf4* and *Eras*. The overall gene expression profiles revealed by this system provide novel insight into how pluripotency is acquired in germ-committed cells.

acquisition of pluripotency, primordial germ cell, purification, tracing

INTRODUCTION

Somatic cells can be reprogrammed into pluripotent cells in vitro through introduction of four defined factors: *Pou5f1*, *Sox2*, *Klf4*, and *Myc* [1, 2]. Currently, it is possible to produce induced pluripotent stem (iPS) cells with characteristics identical to those of embryonic stem (ES) cells. This procedure is a major breakthrough in stem cell biology and has also been expected to be used for clinical applications in regenerative medicine. Several recent studies also report improved induction methods, including integration-free induction of reprogramming factors [3–9]. However, little is known about how cells alter their characteristics to those of pluripotent cells [10]. Although an understanding of reprogramming mechanisms is essential to manipulate iPS cells and iPS cell-derived differentiated cells, it is experimentally challenging to analyze changes occurring during the process due to a low frequency of conversion and the long culture periods required.

In vivo, germ cells only acquire totipotency after fertilization [11]. Germ cell development involves a sequence of reprogramming events, including drastic epigenetic changes [12]. Germ cells are specified from the epiblast at mouse Embryonic Day 7 (E7). At that time, primordial germ cells (PGCs) suppress the somatic program, including repression of genes of the *Hox* clusters; reactivate *Sox2*, which may function in the recapture of pluripotency potential; and commence genome-wide epigenetic reprogramming, including histone modification and DNA demethylation. In addition, an additional reprogramming process occurring in the embryonic gonad involves both erasure of imprinting and X chromosome reactivation [11]. Intriguingly, these germ cell specification processes are also related to somatic cell reprogramming [13]. Thus, an investigation of germ cell specification may provide useful clues to our understanding of somatic cell reprogramming.

Germ cells can be converted to pluripotent cells under appropriate culture conditions, as seen in the development of embryonic germ (EG) cells from PGCs and multipotent germline stem cells from spermatogonia [14, 15], both of which exemplify acquisition of pluripotency from unipotent germ lineages. The EG cell formation is more efficient and has a shorter culture period than somatic cell reprogramming, and it does not require exogenous genetic manipulations. Therefore, EG cell formation provides a good model to analyze mechanisms by which committed cells acquire pluripotency.

[1]Supported by PRESTO of the Japan Science and Technology Agency and Scientific Research (C) from MEXT (G.N.). This study was also supported in part by a grant from the Project for Realization of Regenerative Medicine, and support for the core institutes for iPS cell research was provided by MEXT, a Grant-in-Aid for the Global Century COE program from MEXT.
[2]Correspondence: Go Nagamatsu, Department of Cell Differentiation, The Sakaguchi Laboratory, School of Medicine, Keio University, 35 Shinano-machi, Shinjuku, Tokyo 160-8582, Japan.
E-mail: gonag@z2.keio.jp
[3]These authors contributed equally to this work.

Received: 7 October 2011.
First decision: 14 November 2011.
Accepted: 8 March 2012.

eISSN: 1529-7268 http://www.biolreprod.org
ISSN: 0006-3363

In this study, we developed an efficient culture method to produce EG cells from E11.5 PGCs using inhibitors of MAP2K (MEK), GSK3B (GSK-3β), and TGFB (TGF-β) type 1 receptor. Furthermore, we employed fluorescence-activated cell sorting (FACS) to purify pluripotent candidate cells during the culture period. Combining both techniques, we were able to enrich pluripotent candidate cells from PGCs at different time points and follow how germ cells acquire pluripotency. We observed significant changes in cell cycle status and expression patterns of the whole transcription by microarray.

MATERIALS AND METHODS

Mice

Nanog-GFP-IRES-puro transgenic mice (RBRC02290) were provided by RIKEN BRC, which participates in the National Bio-Resource Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. ICR and BALB/c nude mice were purchased from Japan SLC. Animal care was performed in accordance with guidelines established by Keio University (Tokyo, Japan) for animal use and recombinant DNA experiments. The PGCs were prepared by crossing *Nanog-GFP* with ICR mice.

Isolation and Culture of PGCs

Gonads from *Nanog-GFP-IRES-puro* transgenic embryos at E11.5 and E8.5 were dissociated to form a single-cell suspension by incubation with 0.05% trypsin and 0.02% ethylenediamine tetraacetic acid for 10 min. Suspensions were sorted by green fluorescent protein (GFP) fluorescence using a BD FACS AriaII cell sorter (BD Biosciences). Sorted PGCs were cultured on Sl/Sl^4-m220 or STO feeder cells with Knockout DMEM (Invitrogen) supplemented with 15% KnockOut Serum Replacement (KSR; Invitrogen), 2 mM glutamine, 1 mM nonessential amino acids, 2-mercapto-ethanol, leukemia inhibitory factor (LIF), and basic fibroblast growth factor (bFGF). All compounds were added from Day 1 until Day 7 of culture, except during the process of seeding on STO cells. When PGCs were seeded on STO cells, chemical compounds were added at the beginning of the culture. At Day 7, bFGF was removed by changing to fresh medium lacking bFGF. In the case of seeding on Sl/Sl^4-m220 cells, cells were collected and reseeded onto STO feeder cells on Culture Day 3.

Feeder Cell Preparation

Sl/Sl^4-m220 cells were treated with 5 μg/ml mitomycin C for 1 h and plated at a density of 4×10^5 cells per well in 24-well plates 1 day before use. STO cells were treated with 12 μg/ml mitomycin C for 2 h and plated at a density of 1×10^6 cells per 55 cm^2.

Chemical Compounds

The chemical compounds purchased for this study were trichostatin A (TSA; Sigma), valproic acid (VPA; Sigma), BIX (Alexis Biochemicals), 5-azacytidine (5AZA; Sigma), A83-01 (Sigma), and dorsomorphin (Sigma). Drs. S. Nishiyama, H. Okano, and W. Akamatsu (Keio University) kindly donated the inhibitors PD173074, PD325901, and CHIR99021, which were prepared by organic synthesis. The concentrations of compounds were TSA (1 ng/ml, 5 ng/ml), VPA (0.2 mM, 1.0 mM), BIX (0.1 μM, 0.5 μM), 5AZA (0.2 μM, 1 μM), A83-01 (25 nM, 250 nM), dorsomorphin (0.2 μM, 2 μM), PD173074 (10 nM, 100 nM), PD325901 (0.1 μM, 1 μM), and CHIR99021 (0.3 μM, 3 μM). For the combined use of PD325901, CHIR99021, and A83-01 (2i + A83), the concentrations were 1 μM PD325901, 3 μM CHIR99021 and 250 nM A83-01.

Antibodies and Flow Cytometry

Cultured PGCs were harvested in 0.05% trypsin. After washing, cells were incubated with anti-FcγR (FCER1G) antibody (2.4G2; eBioscience) at 4°C for 30 min. Then, cells were simultaneously incubated with APC-conjugated anti-SSEA-1 mAb (MC-480) (BioLegend) for 30 min at 4°C. Antibodies were used at 0.2 μg per 1×10^6 cells. After washing, samples were analyzed and sorted using the FACSAria II cell sorter (BD Biosciences). For cell cycle analysis, stained samples were fixed by 4% paraformaldehyde (PFA) and treated with 0.005% saponin (Sigma), 0.25 mg/ml RNase A, and 50 μg/ml propidium iodide (Molecular Probes) for 20 min at 37°C. Then, a BD FACSCalibur flow cytometer (BD Biosciences) was used to analyze the cells.

Microarray Data Analysis

A significant number of PGCs, pluripotent candidate cells at Days 3 and 6 of culture, and EG cells were collected by FACS purification. Expression profiles were analyzed using the 3D-Gene Mouse Oligo chip 24K (Toray Industries). Fluorescence intensities were detected using the Scan-Array Life Scanner (Perkin-Elmer), and PMT levels were adjusted to achieve 0.1%–0.5% pixel saturation. Each TIFF image was analyzed with GenePix Pro version 6.0 software (Molecular Devices). We made biological replicates and used average value. The data were filtered to remove low-confidence measurements and were globally normalized per array, such that the median of the signal intensity was adjusted to 50 after normalization (accession number GSE37261). Similarities of whole-gene expression profiles in four arrays were measured by Pearson correlation coefficients. Hierarchical sample clustering was performed by UPGMA method with Pearson correlation distance, and the whole genes were manually sorted based on their maximum expression values; the expression values were displayed as normalized values (i.e., the log2 gene expression value divided by the median). Furthermore, we designed the following procedure to perform the gene set enrichment analysis for the present case without any replicates. First, a gene list that sorted in the descending order based on fold-change between the PGCs, Day 3 (2i + A83) cells, Day 6 (2i + A83) cells, and EG (2i + A83) cells was calculated. Secondly, preranked gene set enrichment analysis was applied for the gene list to the manually curated gene sets for pluripotent and germ cell markers with default parameters (http://www.broadinstitute.org/gsea/index.jsp).

Alkaline Phosphatase Staining

The ES cells were fixed in 4% PFA/PBS for 10 min at 4°C, washed twice with PBS, and then stained using SCIP/NBT liquid substrate (B-1911; Sigma) for 30 min at 37°C.

Teratoma Formation

To produce teratomas, 1.0×10^6 cells were suspended in BD Matrigel (BD Biosciences) and injected into nude mice. Three to four weeks later, tumors were fixed with 4% PFA in PBS, embedded in paraffin, sectioned, and stained with hematoxylin and eosin.

Bisulfite Sequencing

Bisulfite reactions were performed with the EpiTect Bisulfite kit (Qiagen) according to the manufacturer's instructions. Primers used for PCR were described previously [16]. The PCR products were cloned into pGEM-T-easy (Promega) and sequenced by conventional means.

Time-Lapse Bioimaging

Cells were harvested on Culture Day 3, reseeded on STO feeder cells, and cultured overnight. Cells were analyzed using an LCV110 incubator microscope system (Olympus) and incubated at 37°C in 5% CO_2 during experiments. MetaMorph software (Universal Imaging) was used for image analysis.

RESULTS

Highly Efficient Induction of Pluripotent Stem Cells from PGCs

The PGCs were collected from E11.5 mouse embryos and cultured on S1/S1^4-m220 feeder cells in the presence of bFGF and LIF. After 10 days of culture, *Nanog-GFP*-positive colonies were counted [17]. Primordial germ cells undergo changes in characteristics such as proliferative capacity during various developmental stages. When they enter the embryonic gonad, PGCs undergo cell cycle arrest, making it difficult to generate EG cells at that time point. Here, because PGCs were obtained at E11.5 immediately prior to cessation of proliferation (Supplemental Fig. S1; all supplemental data are available online at www.biolreprod.org), the rate of conversion of PGCs to EG cells remained low, making it difficult to conduct a prospective analysis. Therefore, we aimed to develop a more efficient induction method.

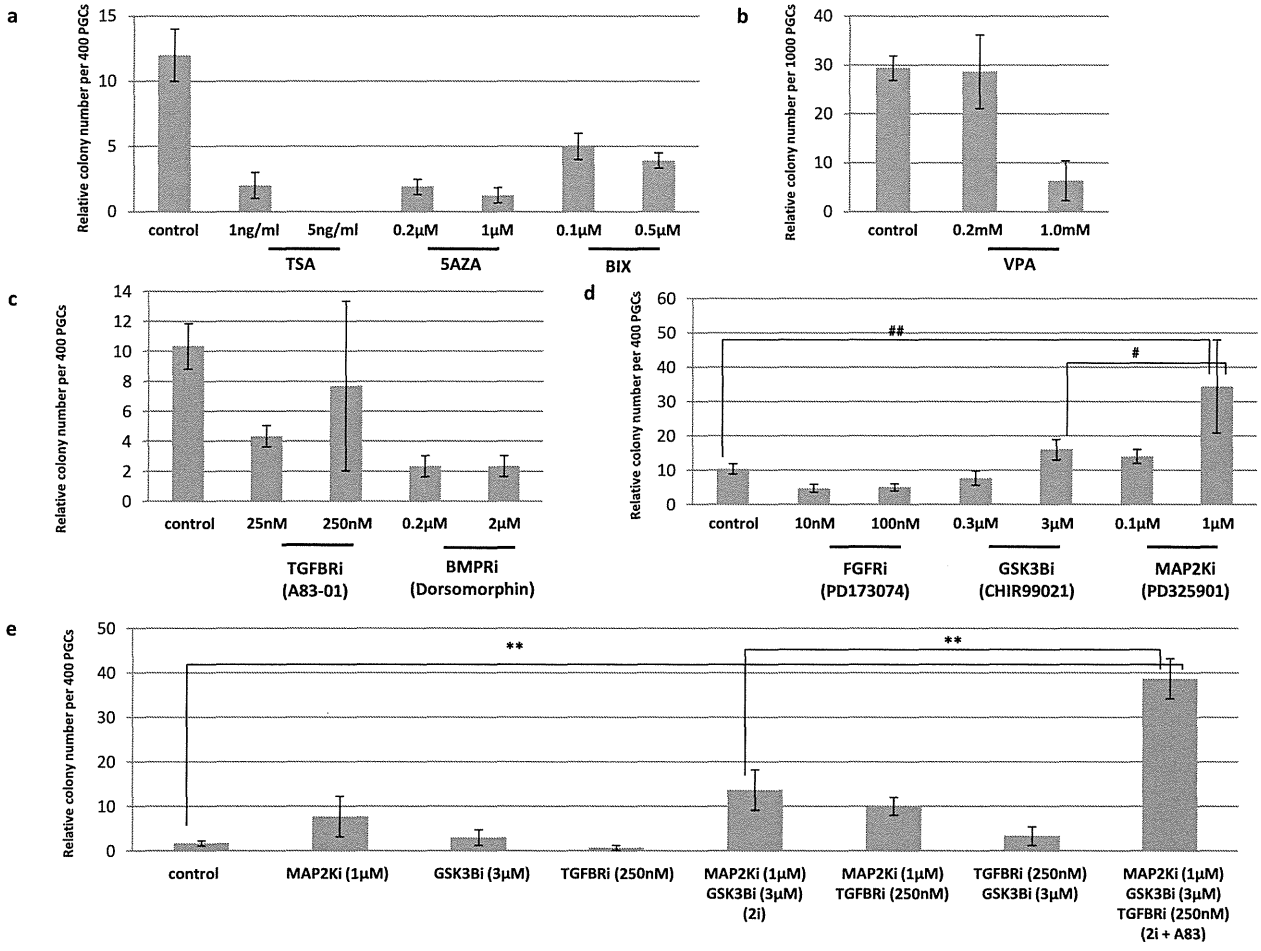

FIG. 1. Establishment of highly efficient culture conditions for EG cell formation. Chemical compounds promoting epigenetic modification (TSA, 5AZA, BIX-01294, and VPA) were screened for capacity to enhance EG cell formation (**a** and **b**). Each chemical was added at the indicated concentrations 1 day after cells were seeded. *Nanog-GFP*-positive colonies were counted on Day 10. **c** and **d**) Analysis of signaling inhibitors. $\#P = 0.0136$, $\#\#P = 0.0014$. Effects of the TGFB type 1R (A83-01) inhibitor or BMP type 1R inhibitor (dorsomorphin) are shown in **c**, whereas effects of the FGFR inhibitor (PD173074), the GSK3B inhibitor (CHIR99021), or the MAP2K inhibitor (PD325901) on EG cell formation are shown in **d**. The effects of different combinations of these inhibitors are shown in **e**. Each chemical except for those shown in **e**, was added at the concentrations indicated 1 day after the culture was seeded. For the chemicals in **e**, the concentration of each inhibitor was: MAP2K inhibitor (3 μM), the GSK3B inhibitor (1 μM), and the TGF-β type 1R inhibitor (25 mM). *Nanog-GFP*-positive colonies were counted on Day 10. $**P < 0.001$. All of the culture conditions contained LIF. Error bars indicate SD.

We first asked whether epigenetic modifications altered reprogramming efficiency. To do so, we undertook experiments employing histone deacetylase inhibitors TSA and VPA, the DNA demethylator 5AZA, and BIX-01294, a diazepinylquinazolinamine derivative with a specific inhibitory effect on G9a as an H3K9-specific methyl transferase, all of which reportedly increase reprogramming efficiency in nuclear transfer and/or induction of iPS cells [9, 18–20]. Each compound was applied to PGCs 1 day after seeding because we did not observe changes in efficiency when chemicals were applied immediately after culture initiation (data not shown). Although several colonies per hundred PGCs were observed in untreated control cultures, all of these small-molecule compounds decreased colony growth dose dependently following application, indicating an inhibitory effect (Fig. 1, a and b). These epigenetic alterations are not critical for EG cell formation from PGCs.

Induction of iPS cells from rat liver progenitors and human fibroblasts is reportedly enhanced by treatment of cells with small-molecule inhibitors, including inhibitors of MAP2K, GSK3B, and TGFB type 1R [21]. Embryonic stem cells can also be maintained without feeder cells, LIF, or serum in the presence of a combination of signaling inhibitors [22]. Such inhibitors include the MAP2K inhibitor PD325901, the GSK3B inhibitor CHIR99021, and the FGF receptor (FGFR) inhibitor PD173074. Significantly, combining two inhibitors (2i) PD325901 (for MAP2K) and CHIR99021 (for GSK3B) is reportedly particularly potent and allows cultured ES cells to remain in a self-renewal ground state [23]. Thus, we investigated the role of these signaling pathways during EG cell formation. The TGFB signaling and BMP signaling exert their effects via SMAD signaling. The addition of either a TGFB type 1R inhibitor (A83-01) or dorsomorphin, which inhibits BMP type 1R, had no effect (Fig. 1c). On the other hand, we found that PGCs cultured with the MAP2K inhibitor or the GSK3B inhibitor showed increased colony numbers, whereas addition of the FGFR inhibitor PD173074 had an inhibitory effect (Fig. 1d). The inhibitory effect of the FGFR

inhibitor for EG cell formation supports the idea that FGF signaling positively regulates EG cell formation [24]. Treatment of the MAP2K inhibitor and the GSK3B inhibitor showed additive effect (Fig. 1e). Furthermore, combining the MAP2K inhibitor and the GSK3B inhibitor (2i) with the TGFB type 1R inhibitor A83-01 produced a significant increase in EG colony formation (Fig. 1e). Although 2i was originally used in a serum-free condition [22], we refer to the combination of MAP2K inhibitor and the GSK3B inhibitor in the culture containing KSR as 2i in this paper.

Because *Nanog-GFP*-positive cells in colonies generated in the presence of bFGF + 2i + A83 proliferated for more than 15 passages and exhibited a morphology similar to that seen in EG cells generated using bFGF alone, we further analyzed their characteristics. First, we examined expression patterns of nine genes specific for germ cells or pluripotent cells in PGCs, EG cells, and ES cells. The EG cells induced by bFGF + 2i + A83 showed gene expression profiles similar to those of EG cells cultured in the presence of bFGF alone and of ES cells, but distinct from either of the parental PGCs (Fig. 2, a–i). Microarray analysis showed highly correlated gene expression patterns between EG cells generated in the presence of bFGF + 2i + A83 compared with those without 2i + A83 (Supplemental Fig. S2). To determine the differentiation capacity of these cells, we transplanted 1×10^6 EG cells subcutaneously into nude mice and observed teratomas containing all three germ layers in EG cells generated by bFGF or by bFGF + 2i + A83 (Fig. 2j). To confirm that these EG cells originated from PGCs, we evaluated methylation patterns of imprinted loci. When differentially methylated regions of *Igf2r* and *Peg1* were analyzed by bisulfite sequence, both types of EG cells exhibited erased methylation at imprinted loci compared with mouse embryonic fibroblasts (MEFs) and TT2 ES cells (Fig. 2k). These data indicate that pluripotent cells established from E11.5 PGCs and cultured with bFGF + 2i + A83 are identical in whole-transcription and in vitro differentiation capacities.

Establishment of the Purification Method for Pluripotent Candidate Cells from PGCs

Enhanced efficiency of EG cell formation observed in the presence of bFGF + 2i + A83 led us to investigate the process of pluripotency acquisition. The PGCs were cultured with 2i + A83, bFGF, bFGF + 2i + A83, or without treatment and then monitored on Days 3 and 6 for SSEA-1 and *Nanog-GFP* double positivity (Fig. 3a). Whereas SSEA-1-positive or *Nanog-GFP*–positive cells were detected in all culture conditions at Day 3, the double-positive (DP) population was detected only in bFGF or bFGF + 2i + A83 cultures on Day 6 (Fig. 3a).

To examine whether the DP population on Days 3 and 6 included potentially pluripotent stem cells likely to form by Day 10, the DP population and the others derived in the presence of bFGF were sorted by FACS on Days 3 and 6, plated at 400 cells per well on feeder layers, and cultured. By Day 10, only cells from the DP population formed EG cell colonies, whereas cells from the non-DP population did not (Fig. 3b). Because we replated the cells, the colony formation by nascent PGC/EG cells may be included in addition to the secondary colonies. Therefore, colony formation efficiency was calculated as relative numbers. The relative colony formation efficiencies from DP populations cultured with bFGF + 2i + A83 and sorted at Days 3 and 6 of culture were 18.8% and 43.6%, respectively, whereas DP PGCs cultured with bFGF alone showed efficiencies of only 1.6% and 1.8% on Days 3 and 6, respectively. These findings indicate that culturing in the presence of bFGF + 2i + A83 enhances efficiency of EG cell formation, and combined with FACS sorting facilitates enrichment of candidate EG cells.

We also used time-lapse imaging to examine whether colonies were derived from a single cell. The DP cells were sorted on Day 3 of the culture, and then time-lapse analyses were performed for more than 70 h. This technique revealed that a single EG cell colony was not always derived from a single cell, but from the fusion of smaller colonies derived from single cells (Fig. 4, a and b, and Supplemental Movies S1 and S2). To estimate the frequency of fusion, we compared numbers of colonies formed from a single cell with those derived from bulk culture. Using single-cell deposition from sorted cells, single cells on Day 6 were cultured in the presence of bFGF + 2i + A83. These cells gave rise to EG colonies with an efficiency of 51.3%, slightly higher than that seen in bulk culture of 400 cells (43.5%; Fig. 4b). Based on these data, the frequency of fusion was calculated to be roughly 15.2%. Thus, approximately 1 colony per 6.6 colonies was generated by fusion in the bulk culture.

Cell Cycle Shifts to S Phase During Acquisition of Pluripotency

Pluripotent cells, such as ES and iPS cells, show a greater percentage of cells in S phase compared with somatic and germ stem cells [25]. Therefore, we asked whether cells in our culture system showed cell cycle progression patterns reminiscent of stem cells during pluripotency acquisition. The PGCs from an E11.5 embryo showed distinct peaks at the G_1 and G_2/M phases (G_1: 27.7%, S: 31.6%, G_2/M: 40.0%; Fig. 5b). By Day 3 of culture, almost half of those cells (49.5%) exhibited fragmented DNA (Fig. 5), indicative of cell death. Interestingly, 2i + A83 treatment in the presence of bFGF suppressed cell death to 29.0% at that time point. On Day 6, the dead cell population was drastically reduced with or without 2i + A83 treatment, and the cell cycle pattern was similar to that of established EG cells grown in either bFGF or bFGF + 2i + A83 culture conditions. These data indicate that cell death is induced at early phases of EG cell formation and that a positive effect of bFGF + 2i + A83 on cell survival may underlie the observed increase in frequency of EG cell formation.

Purified Pluripotent Candidate Cells Undergo Changes in Gene Expression

To identify pathways responsible for pluripotency acquisition, pluripotent candidate populations cultured with 2i + A83 were sorted on Days 3 and 6. The expression levels of genes were analyzed by microarray. The coefficients of correlation between all periods of culture cells were calculated and summarized in Figure 6a. From E11.5 PGCs to EG cells, the coefficients of correlation were monotonically decreased. Therefore, it is indicated that the change in gene expression was a gradual process. Principal component analysis also showed gradual changes of cell state (Fig. 6b).

To further resolve these temporal changes, we made unsupervised hierarchical clustering (Fig. 6c). The cells were divided to two clusters between Days 3 and 6 of the culture. From Day 3 to Day 6 of the culture, dead cells were decreased, and the cell cycle progression was observed (Fig. 5). Clustering analysis indicated that the transition also occurred on a transcriptional level. Interestingly, the up-regulation of *Myc*, which regulated cell survival and proliferation, was observed on Day 6 of the culture (Supplemental Table S1). Furthermore, the clustering data revealed the specific clustered genes at each

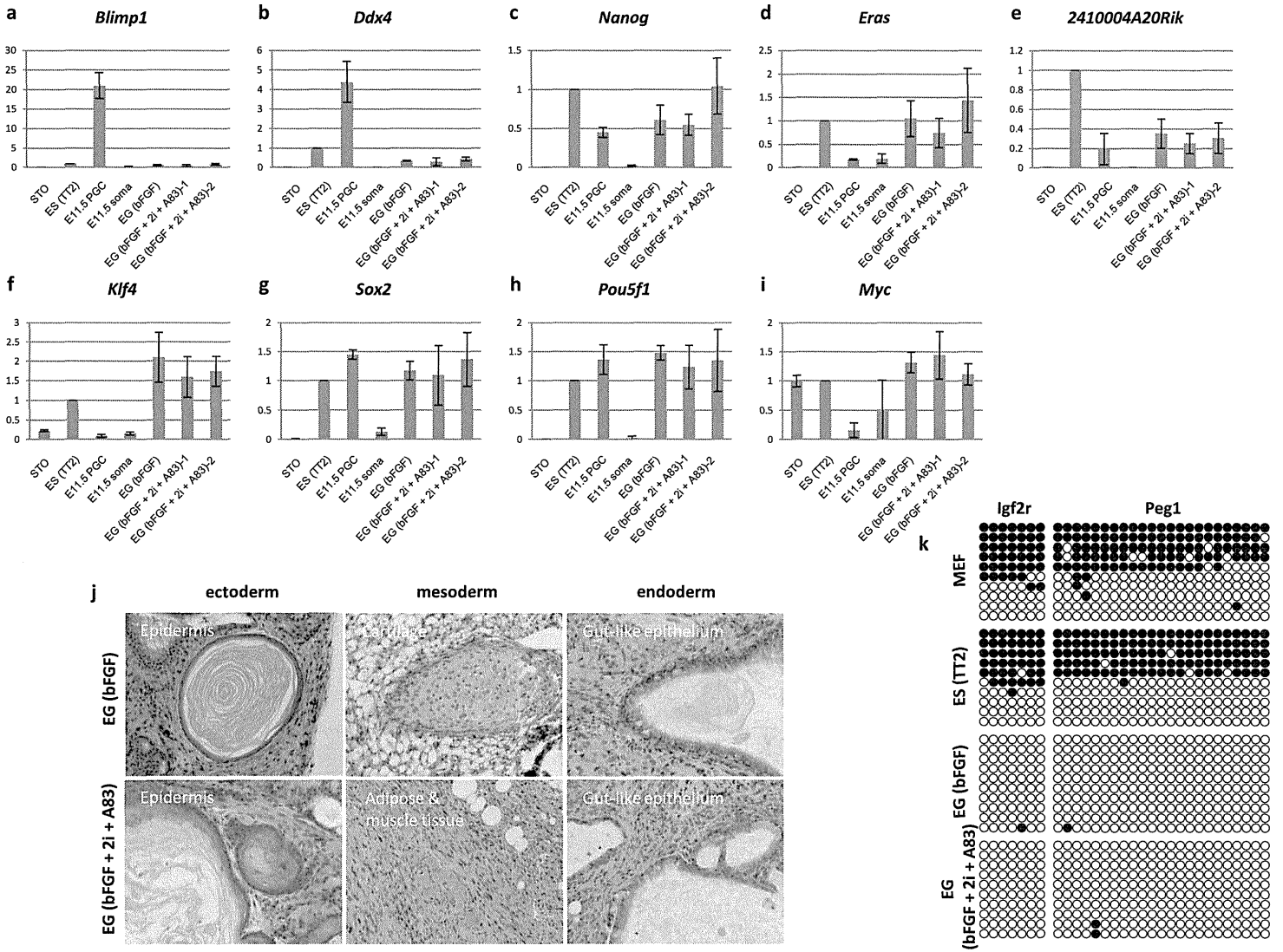

FIG. 2. Characteristics of 2i + A83-treated EG cells. Gene expression patterns were studied in the following cells. EG (bFGF + 2i + A83)-1 and –2: induced EG cells supported by bFGF + 2i + A83; EG (bFGF): induced EG cells supported by FGF alone; ES (TT2), E11.5 PGC: parental PGC; E11.5 soma: somatic cells; and STO cells. Germ cell markers: *Blimp1* (**a**) and *Ddx4* (**b**). Pluripotent cell markers: *Nanog* (**c**), *Eras* (**d**), and *2410004A20Rik* (ECAT1; **e**). Known reprogramming factors: *Klf4* (**f**), *Sox2* (**g**), *Pou5f1* (**h**), and *Myc* (**i**). The data showed relative gene expression levels. Error bars indicate SD. **j**) In vivo differentiation capacity of bFGF + 2i + A83-treated EG cells or bFGF-treated EG cells. Teratoma formation from EG cells generated by each culture showed equivalent differentiation capacity. Original magnification ×10. **k**) Bisulfite sequencing of EG cells at the differentially methylated region (DMR) of the imprinted genes *Igf2r* and *Peg1*. White circles indicate unmethylated CpG dinucleotides, whereas black circles indicate methylated CpG dinucleotides. The methylation pattern in EG cells (bFGF + Ti; bFGF) was different from that of MEFs and ES cells (TT2).

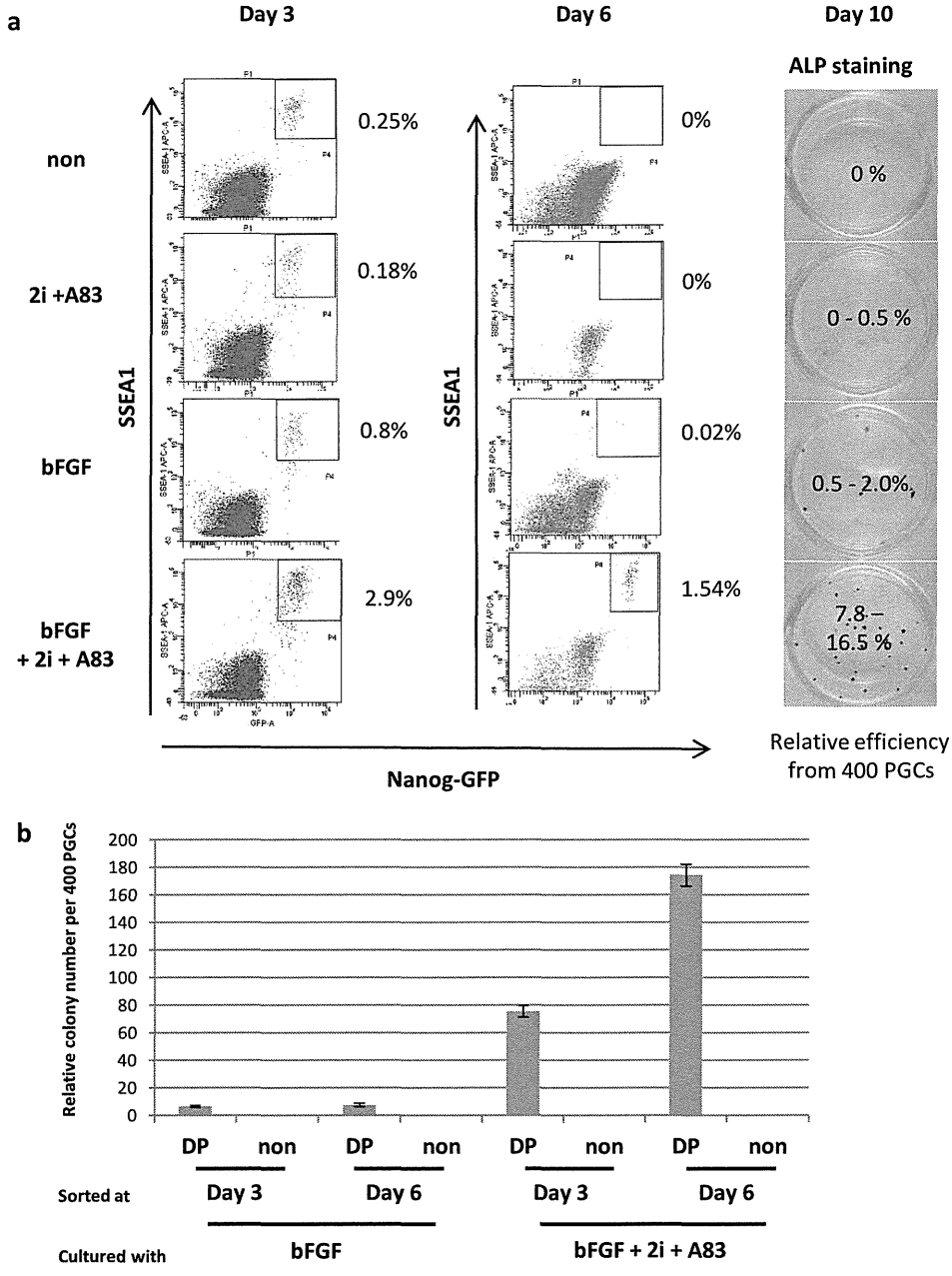

FIG. 3. Tracing PGC conversion to EG cells. **a**) *Nanog-GFP*-positive and SSEA1-positive cells were detected by FACS during EG cell formation from PGCs. Double-positive (DP) cells were analyzed on Culture Days 3 and 6 under each condition. Data representing the average percentage of DP cells from five independent experiments are displayed. At right, EG colonies on Culture Day 10 were visualized by alkaline phosphatase (ALP) staining, and colony formation efficiencies are shown. **b**) Numbers of *Nanog-GFP*-positive colonies from sorted DP and non-DP cells on Days 3 and 6. All of the culture conditions contained LIF. Error bars indicate SD.

time point. Before culture, E11.5 PGCs expressed germ cell-specific genes, such as *Dnd1* and *Ddx4*. On Day 3, the early phase of culture, some ES cell-associated genes, such as *Klf4* and *Eras*, started to become up-regulated (pluripotent cell marker; Supplemental Table S2). At that time, the expression of the germ cell markers *Dnd1* and *Ddx4* started to decrease (germ cell marker; Supplemental Table S2). On Day 6, a further decrease was observed in the expression of germ cell-characteristic genes.

To find cell fate transition from germ lineage to pluripotency on the whole-transcription level, we did enrichment plot analysis (Fig. 6, d and e). Analyzed makers were listed in Supplemental Table S2. According to the culture periods, germ cell markers decreased ($P = 0.01$) and pluripotency markers increased ($P = 0.005$) in whole transcriptome. These data suggest the cell fate transition from the transcriptome level. Recently, it has been reported that mesenchymal-to-epithelial transition is a characteristic event in the early phase of iPS cell

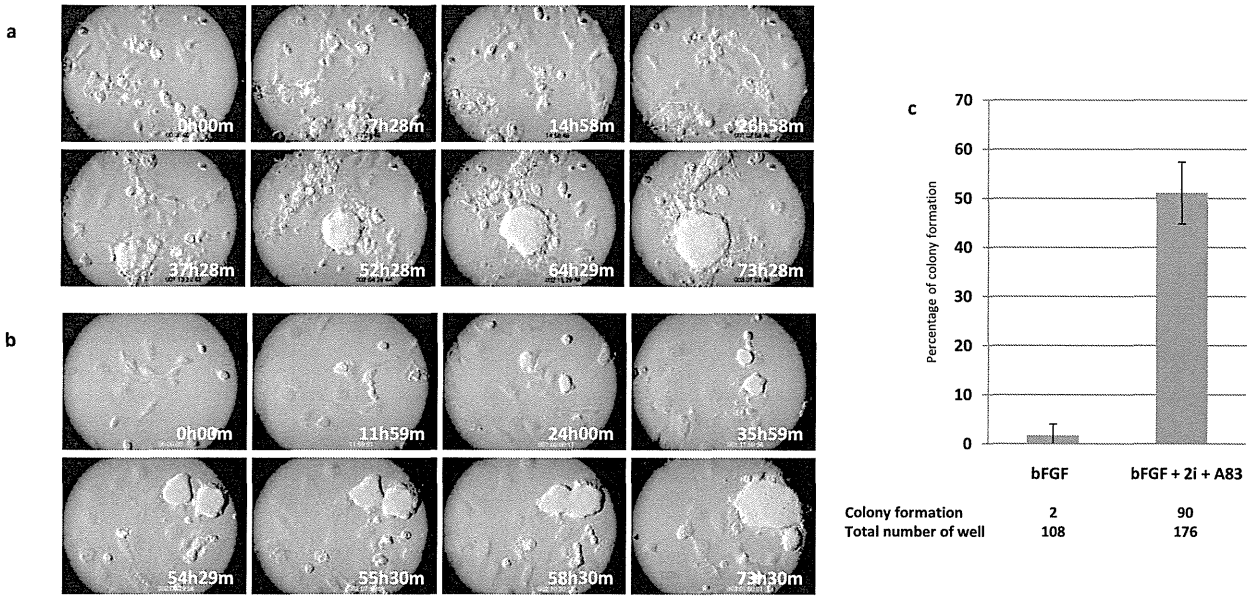

FIG. 4. Time-lapse analysis of EG cell formation. **a** and **b**) A single cell sorted from PGCs cultured with bFGF and 2i + A83 on Day 6 was traced by time-lapse imaging analysis. **a**) Formation of a single cell-derived colony (Supplemental Movie S1). **b**) Formation of a colony by cell fusion (Supplemental Movie S2). Original magnification ×20. **c**) Single DP cells supported by bFGF or bFGF + 2i + A83 were sorted into individual wells of 96-well plates on Day 6. The conversion efficiency is shown as the percentage of colony formation. All of the culture conditions contained LIF. Error bars indicate SD.

formation from fibroblasts [26]. However, our microarray data did not reveal a marked increase in epithelial gene expression (epithelial cell marker and mesenchymal cell marker; Supplemental Table S2). This might be due to the fact that PGCs avoid epithelial-to-mesenchymal transition during specification [27].

DISCUSSION

Low conversion efficiency and long culture periods have hampered the investigation of iPS cell formation from somatic cells. To overcome these experimental difficulties, we focused on the process of conversion of PGCs to pluripotent stem cells. In this study, we established more effective conditions for induction by screening different chemical compounds and purifying pluripotent candidate cells during EG cell formation using FACS. This system facilitated the analysis of cell cycle state and dynamic changes in gene expression.

Establishment of Highly Efficient Induction of Pluripotent Stem Cells

We analyzed various factors known to increase efficiency of nuclear transfer and/or iPS cell induction. Contrary to our expectations, epigenetic modifiers did not alter EG cell generation from PGCs. Although all of the chromatin-remodeling compounds tested are known to induce a relaxed chromatin state, PGCs reportedly already have a relatively loose chromatin structure compared with somatic cells [28], which may explain why chromatin modifiers were ineffective. Although TSA can reportedly replace bFGF signaling during EG cell formation [29], in our hands TSA did not have a positive effect on the presence of bFGF. However, MEK inhibitors did have a positive effect on PGC conversion, suggesting that the MAPK pathway functions in the induction of growth arrest and cell death during conversion. GSK3B inhibition is also known to induce the growth of many cell types, especially certain stem cell populations [30, 31], whereas aberrant activation of WNT/β-catenin signaling delays cell cycle progression of PGCs in vivo [32]. TGFB signaling in PGCs suppresses growth but supports cell migration [33]. In this study, inhibition of TGFB signaling also induced cell growth, but only in the presence of MAP2K and GSK3B inhibitors, indicating that TGFB has independent signaling functions in PGCs. Cell cycle progression is a potential risk for genetic mutation, and germ cells in particular should avoid mutation for the creation of the next generation. It could be possible that forced cell cycle progression induces cell death in PGCs. In this context, the inhibitors for cell cycle progression may promote EG cell formation. Actually, we found the inhibitors suppressed cell death at Day 3 of the culture (Fig. 5).

Establishment of the Purification Method for Pluripotent Candidate Cells from PGCs

It was previously reported that bFGF is required for the first 24 h of culture during EG cell formation [24]. The major pathways requiring bFGF are the MAPK and PIK3 signaling pathways, and activation of AKT, a serine/threonine kinase, enhances EG cell formation downstream of PIK3 [34]. Recently, it was reported that EG cells can be generated using LIF and 2i [35]. That study employed PGCs from embryos at E8.5, at which time point the efficiency of EG cell formation is higher [36]. Although our culture medium contained KSR, which may contain TGFB/activin activity, the enhancement effect of 2i + A83 was also seen in E8.5 PGCs (Supplemental Fig. S3). By contrast, culture conditions using LIF and 2i without bFGF did not generate EG cells from E11.5 PGCs (Supplemental Fig. S4). These data indicate that different signals are required for derivation of EG cells from PGCs at different developmental stages. Actually, from E8.5 to E11.5,

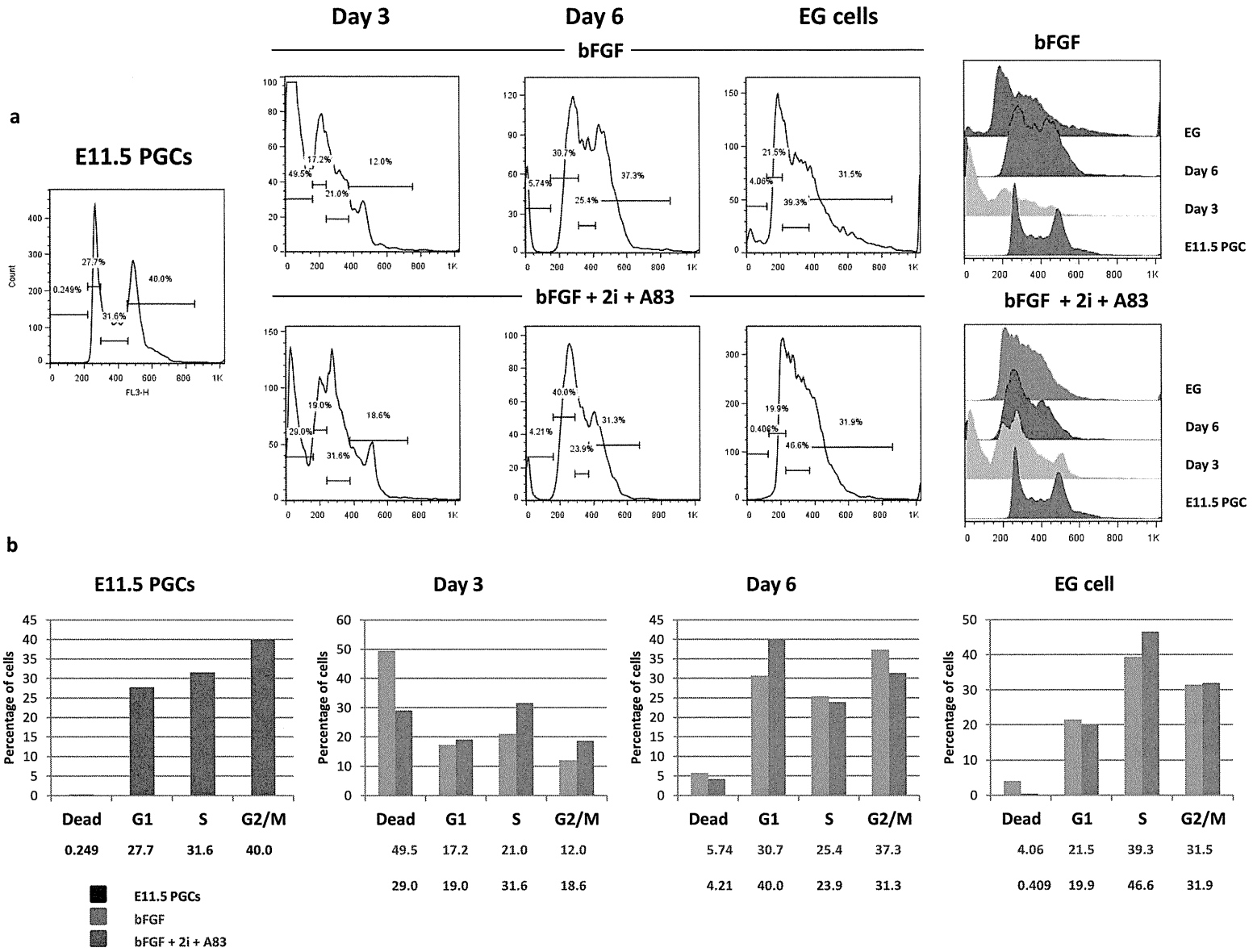

FIG. 5. Cell cycle analysis of DP cells on Days 3 and 6 of culture in the presence of bFGF or bFGF + 2i + A83 during conversion of E11.5 PGCs to EG cells. The cell cycles of E11.5 PGCs and EG cells are also shown. Histograms show DNA content as estimated by propidium iodide staining at indicated time points and culture conditions (**a**). Based on the gates indicated in the histograms above, percentages of cells at each stage of the cell cycle are shown (**b**). All of the culture conditions contained LIF.

a

CCM	E11.5PGCs	Day 3 (2i + A83)	Day 6 (2i + A83)	EG (2i + A83)
E11.5PGCs	1.00000	0.96534	0.94383	0.93126
Day 3 (2i + A83)	0.96534	1.00000	0.97616	0.95945
Day 6 (2i + A83)	0.94383	0.97616	1.00000	0.96814
EG	0.93126	0.95945	0.96814	1.00000

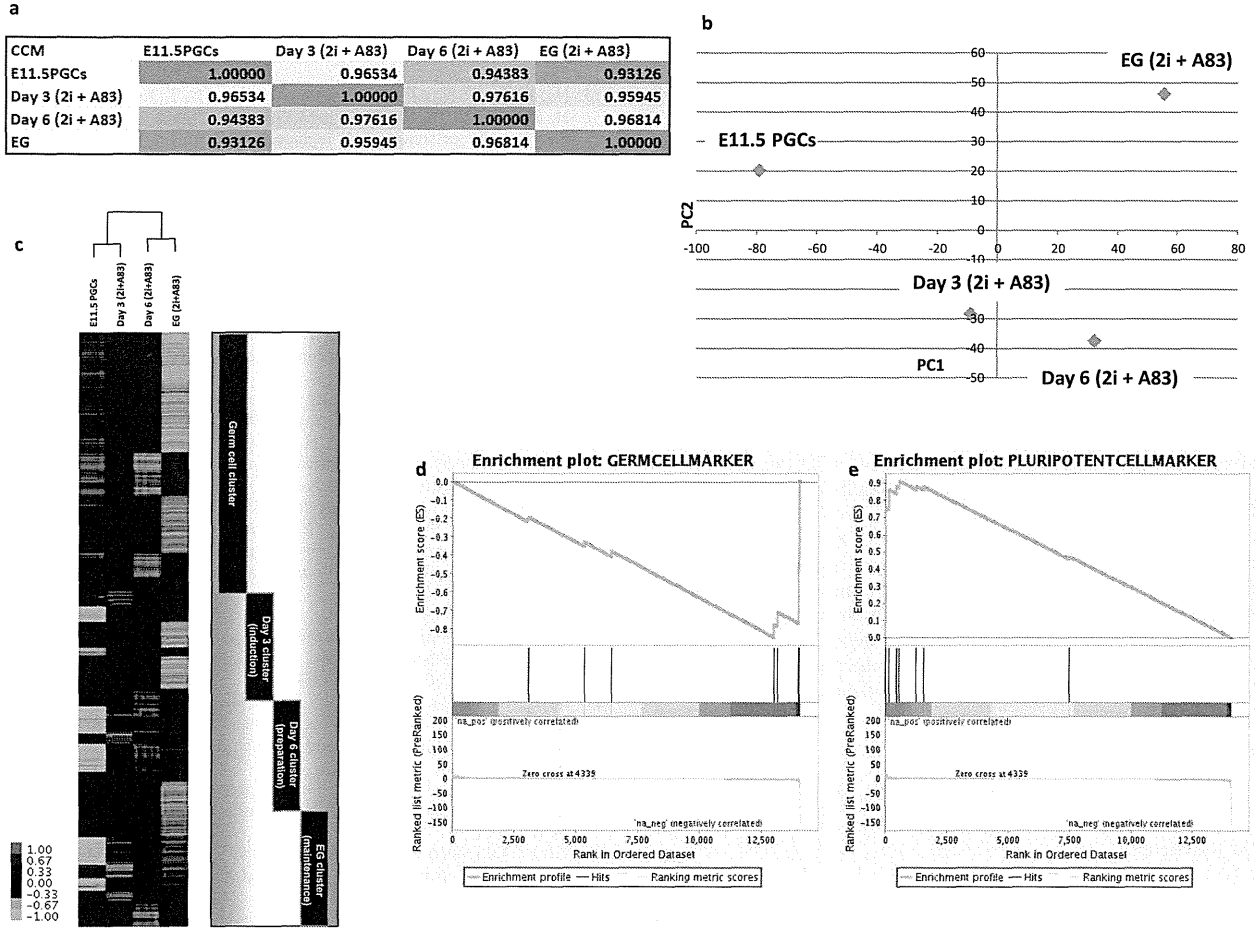

FIG. 6. Microarray analysis during EG cell formation. **a**) The coefficients of correlation between all periods of culture cells. **b**) Principal component analysis during EG cell formation. **c**) Cluster analysis of each time point. **d**) Enrichment plot analysis of germ cell markers ($P = 0.01$). **e**) Enrichment plot analysis of pluripotent cell markers ($P = 0.005$). All of the culture conditions contained LIF.

numerous events occur in PGCs, such as G_2 arrest, repression of RNA polymerase II, and environmental changes.

Pluripotent Candidate Cells Progress S Phase of Cell Cycle and Change Gene Expression Patterns During Acquisition of Pluripotency

PGCs cultured in the presence of bFGF + 2i + A83 showed highly efficient (12.1%) conversion to EG cells. When PGCs were sorted based on *Nanog* and SSEA-1 expression at Days 3 and 6 of culture, only DP cells generated pluripotent EG cells. Therefore, the 2i + A83 culture method established here clearly enhanced the efficiency of EG cell formation, and FACS facilitated purification of candidate pluripotent cells. Although technologies enabling gene expression analysis from single cells are under development, it is necessary to stringently purify candidate cells in order to accurately trace the acquisition of pluripotency.

Interestingly, cell cycle analyses indicated that many PGCs die after induction of pluripotency. The effect of 2i + A83 treatment suppressed cell death rather than the cell cycle progression (Fig. 5). The relationship between cell death and acquisition of pluripotency also indicates the presence of a subpopulation of PGCs. Primordial germ cells can reportedly be subdivided into two populations based on expression levels of the cell-surface antigen α6-integrin: cells with negative or low α6-integrin levels are likely to form EG cells, whereas high α6-integrin expressors tend to be apoptotic [37]. The fact that 2i + A83 treatment does not promote colony production from E11.5 PGCs in the absence of bFGF (Supplemental Fig. S4) suggests that treatment acts to enhance candidate EG cell survival rather than trigger acquisition of pluripotency.

Using the 2i + A83 culture and purification method, we performed gene expression analyses during the acquisition of pluripotency (Fig. 6). It has been reported that transcriptome characteristics of PGCs are different from those of pluripotent cells [38]. Our microarray data uncovered the sequential changes in transcription during the conversion of PGCs to pluripotent cells. During the acquisition of pluripotency, the loss of the original characteristics of germ cells along with the enhancement of pluripotency-associated gene expression was noted. In other words, there were at least induction, preparation, and maintenance phases during EG cell formation. This phase process is comparable to a previous microarray analysis of second iPS cell generation [26]. These results

indicate the existence of a common mechanism underlying the acquisition of pluripotency.

Considering important genes involved in the generation of EG cells from PGCs, we focused on Day 6 of the culture and gradually up-regulated genes. At Day 6 of the culture, some pluripotency markers, such as *Nanog*, *Klf4* and *Zfp42*, were expressed at their highest level during the cultures. Interestingly, at that time the *Meis* family of transcription factors also reached its highest expression (*Meis1* and *Meis2* in Supplemental Table S1). *Meis1* is known to work to maintain hematopoietic stem cells [39]. On the other hand, in the gradually up-regulated genes we found some family members of reprogramming factors, such as *Klf9* and *Sox11*. It is interesting to examine the effect of *Meis* family genes *Klf9* and *Sox11* during the acquisition of puripotency.

Human PGCs also convert to EG cells. Human EG cells are cultured in the presence of LIF, bFGF, and forskolin [40]. Although human EG cells do not form teratoma when injected into mice, they form embryoid body and differentiate into all three germ layers [41, 42]. In mice, EG cells are very similar to ES cells. However, in humans EG cells showed some differences from ES cells. Human EG cells are expressed SSEA-1, whereas human ES cells are not [42]. Although human EG cells are dome-shaped like mouse ES colonies, human ES cells are flat-shaped, which resembles mouse epiblast stem cells [40]. To understand the differences between EG cells and ES cells among species, our microarray data and culture system would provide useful information.

ACKNOWLEDGMENT

We thank Drs. H. Okano, S. Nishiyama, and W. Akamatsu (Keio University) for providing the inhibitors PD173074, PD325901, and CHIR99021, and Drs. J.-W. Cheong and A. Ishizu for critical reading of the manuscript.

REFERENCES

1. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006; 126: 663–676.
2. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007; 131:861–872.
3. Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW, Cowan C, Hochedlinger K, Daley GQ. Disease-specific induced pluripotent stem cells. Cell 2008; 134:877–886.
4. Dimos JT, Rodolfa KT, Niakan KK, Weisenthal LM, Mitsumoto H, Chung W, Croft GF, Saphier G, Leibel R, Goland R, Wichterle H, Henderson CE, et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science 2008; 321:1218–1221.
5. Ebert AD, Yu J, Rose FF Jr., Mattis VB, Lorson CL, Thomson JA, Svendsen CN. Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature 2009; 457:277–280.
6. Lee G, Papapetrou EP, Kim H, Chambers SM, Tomishima MJ, Fasano CA, Ganat YM, Menon J, Shimizu F, Viale A, Tabar V, Sadelain M, et al. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. Nature 2009; 461:402–406.
7. Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. Generation of mouse induced pluripotent stem cells without viral vectors. Science 2008; 322:949–953.
8. Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hamalainen R, Cowling R, Wang W, Liu P, Gertsenstein M, Kaji K, Sung HK, et al. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature 2009; 458:766–770.
9. Zhou H, Wu S, Joo JY, Zhu S, Han DW, Lin T, Trauger S, Bien G, Yao S, Zhu Y, Siuzdak G, Scholer HR, et al. Generation of induced pluripotent stem cells using recombinant proteins. Cell Stem Cell 2009; 4:381–384.
10. Yamanaka S. Elite and stochastic models for induced pluripotent stem cell generation. Nature 2009; 460:49–52.
11. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet 2008; 9: 129–140.
12. Saitou M. Germ cell specification in mice. Curr Opin Genet Dev 2009; 19: 386–395.
13. Nagamatsu G, Kosaka T, Kawasumi M, Kinoshita T, Takubo K, Akiyama H, Sudo T, Kobayashi T, Oya M, Suda T. A germ cell specific gene, Prmt5 works as somatic cell reprogramming. J Biol Chem 2011; 286: 10641–10648.
14. Matsui Y, Zsebo K, Hogan BL. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell 1992; 70: 841–847.
15. Kanatsu-Shinohara M, Inoue K, Lee J, Yoshimoto M, Ogonuki N, Miki H, Baba S, Kato T, Kazuki Y, Toyokuni S, Toyoshima M, Niwa O, et al. Generation of pluripotent stem cells from neonatal mouse testis. Cell 2004; 119:1001–1012.
16. Lucifero D, Mertineit C, Clarke HJ, Bestor TH, Trasler JM. Methylation dynamics of imprinted genes in mouse germ cells. Genomics 2002; 79: 530–538.
17. Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. Nature 2007; 448:313–317.
18. Kishigami S, Mizutani E, Ohta H, Hikichi T, Thuan NV, Wakayama S, Bui HT, Wakayama T. Significant improvement of mouse cloning technique by treatment with trichostatin A after somatic nuclear transfer. Biochem Biophys Res Commun 2006; 340:183–189.
19. Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen AE, Melton DA. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol 2008; 26:795–797.
20. Blelloch R, Wang Z, Meissner A, Pollard S, Smith A, Jaenisch R. Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. Stem Cells 2006; 24:2007–2013.
21. Li W, Wei W, Zhu S, Zhu J, Shi Y, Lin T, Hao E, Hayek A, Deng H, Ding S. Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. Cell Stem Cell 2009; 4:16–19.
22. Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. The ground state of embryonic stem cell self-renewal. Nature 2008; 453:519–523.
23. Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A. Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol 2008; 6:e253.
24. Durcova-Hills G, Adams IR, Barton SC, Surani MA, McLaren A. The role of exogenous fibroblast growth factor-2 on the reprogramming of primordial germ cells into pluripotent stem cells. Stem Cells 2006; 24: 1441–1449.
25. Fujii-Yamamoto H, Kim JM, Arai K, Masai H. Cell cycle and developmental regulations of replication factors in mouse embryonic stem cells. J Biol Chem 2005; 280:12976–12987.
26. Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell 2010; 7:64–77.
27. Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev 2008; 22: 1617–1635.
28. Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development 2007; 134:2627–2638.
29. Durcova-Hills G, Tang F, Doody G, Tooze R, Surani MA. Reprogramming primordial germ cells into pluripotent stem cells. PLoS ONE 2008; 3:e3531.
30. Zechner D, Fujita Y, Hulsken J, Muller T, Walther I, Taketo MM, Crenshaw EB, 3rd, Birchmeier W, Birchmeier C. beta-Catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system.Dev Biol 2003; 258:406–418.
31. Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med 2004; 10:55–63.
32. Kimura T, Nakamura T, Murayama K, Umehara H, Yamano N, Watanabe S, Taketo MM, Nakano T. The stabilization of beta-catenin leads to impaired primordial germ cell development via aberrant cell cycle progression. Dev Biol 2006; 300:545–553.
33. Godin I, Wylie CC. TGF beta 1 inhibits proliferation and has a

chemotropic effect on mouse primordial germ cells in culture. Development 1991; 113:1451–1457.
34. Kimura T, Tomooka M, Yamano N, Murayama K, Matoba S, Umehara H, Kanai Y, Nakano T. AKT signaling promotes derivation of embryonic germ cells from primordial germ cells. Development 2008; 135:869–879.
35. Leitch HG, Blair K, Mansfield W, Ayetey H, Humphreys P, Nichols J, Surani MA, Smith A. Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. Development 2010; 2010:2279–2287.
36. Labosky PA, Barlow DP, Hogan BL. Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development 1994; 120:3197–3204.
37. Matsui Y, Tokitake Y. Primordial germ cells contain subpopulations that have greater ability to develop into pluripotential stem cells. Dev Growth Differ 2009; 51:657–667.
38. Mise N, Fuchikami T, Sugimoto M, Kobayakawa S, Ike F, Ogawa T, Tada T, Kanaya S, Noce T, Abe K. Differences and similarities in the developmental status of embryo-derived stem cells and primordial germ cells revealed by global expression profiling. Genes Cells 2008; 13: 863–877.
39. Hisa T, Spence SE, Rachel RA, Fujita M, Nakamura T, Ward JM, Devor-Henneman DE, Saiki Y, Kutsuna H, Tessarollo L, Jenkins NA, Copeland NG. Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. EMBO J 2004; 23:450–459.
40. Shamblott MJ, Axelman J, Wang S, Bugg EM, Littlefield JW, Donovan PJ, Blumenthal PD, Huggins GR, Gearhart JD. Derivation of pluripotent stem cells from cultured human primordial germ cells. Proc Natl Acad Sci U S A 1998; 95:13726–13731.
41. Shamblott MJ, Axelman J, Littlefield JW, Blumenthal PD, Huggins GR, Cui Y, Cheng L, Gearhart JD. Human embryonic germ cell derivatives express a broad range of developmentally distinct markers and proliferate extensively in vitro. Proc Natl Acad Sci U S A 2001; 98:113–118.
42. Turnpenny L, Brickwood S, Spalluto CM, Piper K, Cameron IT, Wilson DI, Hanley NA. Derivation of human embryonic germ cells: an alternative source of pluripotent stem cells. Stem Cells 2003; 21:598–609.

⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 43, pp. 36273–36282, October 19, 2012

Optimal Ratio of Transcription Factors for Somatic Cell Reprogramming*[S]

Received for publication, May 11, 2012, and in revised form, August 27, 2012 Published, JBC Papers in Press, September 6, 2012, DOI 10.1074/jbc.M112.380683

Go Nagamatsu[‡§1,2], Shigeru Saito[¶‖1], Takeo Kosaka[**], Keiyo Takubo[‡], Taisuke Kinoshita[‡], Mototsugu Oya[**], Katsuhisa Horimoto[¶‡‡], and Toshio Suda[‡]

From the [‡]Department of Cell Differentiation, The Sakaguchi Laboratory, and the [**]Department of Urology, School of Medicine, Keio University, Tokyo 160-8582, Japan, the [¶]Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan, the [§]Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan, the [‖]Chem & Bio Informatics Department, INFOCOM Corporation, Tokyo 150-0001, Japan, and the [‡‡]Institute of Systems Biology, Shanghai University, Shanghai 200444, China

Background: The somatic cell reprogramming factors do not always induce pluripotency.
Results: The optimal ratio of the reprogramming factors is *Oct3/4*-high, *Sox2*-low, *Klf4*-high, and c-*Myc*-high.
Conclusion: Among the various reprogramming transcription factor combinations, high *Oct3/4* and low *Sox2* produced the most efficient results.
Significance: The overall gene expression profiles between the high and low efficiency conditions provide novel insights for somatic cell reprogramming.

Somatic cell reprogramming is achieved by four reprogramming transcription factors (RTFs), *Oct3/4*, *Sox2*, *Klf4*, and c-*Myc*. However, in addition to the induction of pluripotent cells, these RTFs also generate pseudo-pluripotent cells, which do not show *Nanog* promoter activity. Therefore, it should be possible to fine-tune the RTFs to produce only fully pluripotent cells. For this study, a tagging system was developed to sort induced pluripotent stem (iPS) cells according to the expression levels of each of the four RTFs. Using this system, the most effective ratio (*Oct3/4*-high, *Sox2*-low, *Klf4*-high, c-*Myc*-high) of the RTFs was 88 times more efficient at producing iPS cells than the worst effective ratio (*Oct3/4*-low, *Sox2*-high, *Klf4*-low, c-*Myc*-low). Among the various RTF combinations, *Oct3/4*-high and *Sox2*-low produced the most efficient results. To investigate the molecular basis, microarray analysis was performed on iPS cells generated under high (*Oct3/4*-high and *Sox2*-low) and low (*Oct3/4*-low and *Sox2*-high) efficiency reprogramming conditions. Pathway analysis revealed that the G protein-coupled receptor (GPCR) pathway was up-regulated significantly under the high efficiency condition and treatment with the chemokine, C-C motif ligand 2, a member of the GPCR family, enhanced somatic cell reprogramming 12.3 times. Furthermore, data from the analysis of the signature gene expression profiles of mouse embryonic fibroblasts at 2 days after RTF infection revealed that the genetic modifier, *Whsc1l1* (*variant 1*), also improved the efficiency of somatic cell reprogramming. Finally, comparison of the overall gene expression profiles between the high and low efficiency conditions will provide novel insights into mechanisms underlying somatic cell reprogramming.

In 2006, Yamanaka and colleagues (1) showed that somatic cells in mice could be reprogrammed to the pluripotent state in the presence of four reprogramming transcription factors (RTFs),[3] *Oct3/4*, *Sox2*, *Klf4*, and c-*Myc*. The following year, induced pluripotent stem (iPS) cell technology was established in human cells (2), and since then, the number of potential applications for iPS technology in regenerative medicine has been growing rapidly. The technology, using transplanted iPS cells, has been used successfully in mouse and rat models of sickle cell anemia and Parkinson disease (3, 4). However, there are several problems that need resolving before iPS cells can be used safely in clinical applications. For example, although iPS cells can differentiate into any cell type in the body, it is necessary to exclude any undifferentiated cells before iPS cell-derived cells are transplanted, as the presence of undifferentiated cells may lead to tumor formation (5).

To use iPS cells in clinical applications, it is important to understand the mechanism that induces pluripotency. It is clear that the process of iPS cell generation involves certain steps (6). These can be broadly summarized as follows. Upon introduction of the four RTFs, fibroblasts down-regulate THY-1 expression; next, the genes used as markers for pluripotency are activated, including alkaline phosphatase and stage-

* This work was supported by PRESTO of the Japan Science and Technology Agency and Scientific Research (C), a grant from the Project for Realization of Regenerative Medicine, support for the core institutes for iPS cell research was provided by MEXT, a grant-in-aid for the Global century COE program from MEXT to Keio University, and the Keio University Medical Science Fund.
⌘ *Author's Choice*—Final version full access.
[S] This article contains supplemental Figs. S1–S10, Tables S1–S3, and Movies S1 and S2.
[1] Both authors contributed equally to this work.
[2] To whom correspondence should be addressed: Dept. of Cell Differentiation, The Sakaguchi Laboratory, School of Medicine, Keio University, 35 Shinano-machi, Shinjuku, Tokyo 160-8582, Japan. Tel./Fax: 81-3-5363-3475; E-mail: gonag@z2.keio.jp.

[3] The abbreviations used are: RTF, reprogramming transcription factor; iPS, induced pluripotent stem; MEF, mouse embryonic fibroblast; ES, embryonic stem.