

An integrated systems biology approach identifies positive cofactor 4 as a factor that increases reprogramming efficiency

Junghyun Jo^{1,†}, Sohyun Hwang^{2,†}, Hyung Joon Kim^{3,†}, Soomin Hong¹, Jeoung Eun Lee³, Sung-Geum Lee³, Ahmi Baek³, Heonjong Han², Jin Il Lee⁴, Insuk Lee^{2,*} and Dong Ryul Lee^{1,3,4,*}

¹Department of Biomedical Science, College of Life Science, CHA University, Seoul, Korea, ²Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul, Korea, ³CHA Stem Cell Institute, CHA University, Seoul, Korea and ⁴Fertility Center, CHA Gangnam Medical Center, College of Medicine, CHA University, Seoul, Korea

Received September 09, 2015; Revised November 24, 2015; Accepted December 01, 2015

ABSTRACT

Spermatogonial stem cells (SSCs) can spontaneously dedifferentiate into embryonic stem cell (ESC)-like cells, which are designated as multipotent SSCs (mSSCs), without ectopic expression of reprogramming factors. Interestingly, SSCs express key pluripotency genes such as *Oct4*, *Sox2*, *Klf4* and *Myc*. Therefore, molecular dissection of mSSC reprogramming may provide clues about novel endogenous reprogramming or pluripotency regulatory factors. Our comparative transcriptome analysis of mSSCs and induced pluripotent stem cells (iPSCs) suggests that they have similar pluripotency states but are reprogrammed via different transcriptional pathways. We identified 53 genes as putative pluripotency regulatory factors using an integrated systems biology approach. We demonstrated a selected candidate, *Positive cofactor 4 (Pc4)*, can enhance the efficiency of somatic cell reprogramming by promoting and maintaining transcriptional activity of the key reprogramming factors. These results suggest that *Pc4* has an important role in inducing spontaneous somatic cell reprogramming via up-regulation of key pluripotency genes.

INTRODUCTION

Reprogramming differentiated somatic cells into pluripotent stem cells (PSCs) facilitates the study of disease mechanisms and development of cell-based therapeutics to repair damaged or diseased tissue. A pioneering study in 1962 (1)

discovered that tadpole intestinal cells could be converted into embryonic stem cells (ESCs) via somatic cell nuclear transfer (SCNT). In recent times, SCNT has been demonstrated as a promising way to generate patient-specific human ESCs (2–4). However, successful human SCNT must overcome the extremely low efficiency of blastocyst formation due to reprogramming barriers (3).

Induced pluripotent stem cells (iPSCs) can be generated by introducing only four master regulators—*Oct4* (*Pou5f1*), *Sox2*, *Klf4* and *Myc* (*c-Myc*) (also known as OSKM factors)—into somatic cells (5). This discovery provided a new technology for generating patient-specific PSCs and for understanding the molecular mechanisms of reprogramming processes (6–10). The three distinct phases during somatic cell reprogramming (SCR) are initiation, maturation and stabilization, which are associated with specific genes (11). Ectopic expression of OSKM genes in fibroblasts initiates the first SCR phase by increasing proliferation, changing metabolites, initiating the mesenchymal-to-epithelial transition (MET) and activating DNA repair. The initiation phase correlates with morphological changes because fibroblast cells undergo MET and display epithelial signatures such as *Cdh1* and *Epcam* expression (10). The SCR maturation phase is characterised by major transcriptional changes of the pluripotency-associated genes *Sall4*, *Oct4*, *Nanog*, *Esrrb* and *Sox2* (9,11). Buganim *et al.* (12) claim that the maturation phase starts with *Sox2* activation, which leads to iPSCs. The SCR stabilization phase occurs after cells acquire pluripotency (13). In this last phase, cells can be sustained independently of ectopic OSKM gene expression. The regulatory mechanisms of initiation and maturation phases are unclear, and the efficiency of generating iPSCs from somatic cells is still very low.

*To whom correspondence should be addressed. Tel: +82 2 2123 5559; Fax: +82 2 362 7265; Email: insuklee@yonsei.ac.kr

Correspondence may also be addressed to Dong Ryul Lee. Tel: +82 2 3468 3421; Fax: +82 2 3468 3466; Email: drleedr@cha.ac.kr

†These authors contributed equally to the paper as first authors.

Another method to generate PSCs is via spontaneous conversion of spermatogonial stem cells (SSCs) into ESC-like multipotent SSCs (mSSCs) using a culture-inducing system (14). We previously showed that an intermediate SSCs (iSSCs) phase subsisted during germline stem cell dedifferentiation to PSCs (15). SSCs express key OSKM reprogramming factors at some levels (16), and do not require ectopic expression of any gene for the acquisition of pluripotency during reprogramming to mSSCs. Therefore, we reasoned that additional factors are required to regulate SSC reprogramming.

In this study, we first compared the expression of reprogramming signature genes among somatic cells, iPSCs, SSCs, mSSCs and partially reprogrammed cells, and found that mSSCs and iPSCs appear to have similar pluripotency states based on transcriptional signature, whereas they have different transcriptional pathways for reprogramming. We developed a systems biology approach to prioritise genes for pluripotency regulatory factors by integrating transcriptome and interactome data on the genome-wide functional network. Then, we performed a series of systematic gene prioritisation steps and identified 53 candidates, which included some known reprogramming factors. We experimentally validated one particular candidate, Positive cofactor 4 (*Pc4*), which was expressed in PSCs and yielded a positive RNA interference (RNAi) response in an *Oct4* reporter assay. We demonstrated that *Pc4* enhanced the efficiency of OSKM-mediated reprogramming by promoting the transcriptional activity of key pluripotency factors, and by regulating the expression of many protein- and miRNA-encoding genes involved in reprogramming and suppression of somatic cell-specific genes.

MATERIALS AND METHODS

RNA extraction, RT-PCR and transcriptome profiling

Total RNA was extracted from cultured cells using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. Each RNA sample was quantified by reverse transcriptase-polymerase chain reaction (RT-PCR) as described in Supplementary Methods. Transcriptome profiling for SSC, iSSC and mSSC was performed using Affymetrix Mouse Genome 430 2.0 Array. For hybridization, 10 μ g of total RNA was amplified and labelled using Nugen WT-Ovation One-Direct Amplification system and Nugen FL-Ovation cDNA Biotin Module V2 labelling kits. To examine the effects of *Pc4* overexpression on genome-wide transcriptional regulation in mESCs, we performed sequencing RNAs isolated from mESCs with or without ectopic *Pc4* expression using an Illumina HiSeq2500 instrument. More details about sequencing procedures and data analysis are described in Supplementary Methods. All microarray and RNA sequencing data generated in this study were deposited in Gene Expression Omnibus database (GSE74156).

Comparison of mSSC and iPSC reprogramming by transcriptome analysis

We analysed six microarray data sets consisting of two replicates of three types of germ-lineage stem cells using biocon-

ductor affy (17), limma (18) and *q*-value (19) packages. Affy was used to preprocess microarray data, limma was used to identify significant differentially expressed genes (DEGs) between mSSCs, iSSCs and SSCs, and *q*-values were used to adjust limma *P*-values for multiple testing corrections. We selected 1120 DEGs between mSSCs and iSSCs, and 1427 DEGs between mSSCs and SSCs, with *q*-value < 0.01. Of these, 549 and 685 genes were up-regulated in mSSCs compared to genes in iSSCs and SSCs, respectively.

To compare mSSC reprogramming and iPSC reprogramming, we used two public microarray data sets, which included partially reprogrammed cells (6,7). We preprocessed expression data derived from Polo *et al.* (7) using the oligo (20) package, and those from Mikkelsen *et al.* (6) using the affy (17) package according to their platforms. We calculated mean expression values for 11 cell types: (i) two SSC replicates, (ii) two iSSC replicates, (iii) two mSSC replicates, (iv) two MEF replicates of Polo *et al.*, (v) two MEF Thy1⁻ replicates of Polo *et al.*, (vi) two MEF SSEA⁺ replicates of Mikkelsen *et al.*, (vii) one data set of MEF Oct4-GFP⁺ cells of Polo *et al.*, (viii) five replicates of MEF-derived iPSCs of both Polo *et al.* and Mikkelsen *et al.*, (ix) two B-lymphocyte replicates of Mikkelsen *et al.*, (x) two replicates of B-lymphocytes expressing OSKM and (xi) two B-lymphocyte-derived iPSC replicates of Mikkelsen *et al.*. Then, we used the aroma.light (21) package for normalization.

Clustering analysis was based on 21 mouse genes that show significantly different expression changes in differentiated cells and reprogrammed cells, and that belong to the 40 mouse genes compiled from the literature that indicate three reprogramming phases and regulate stem cell maintenance (11,12) (Supplementary Table S1). The major pluripotency marker *Nanog* was excluded from the gene set for clustering, because the Affymetrix HT Mouse Genome 430A Array *Nanog* probe used by Polo *et al.* has non-unique matched genes, which could potentially provide false expression values.

To compare biological functions involved in mSSC and iPSC reprogramming, we conducted Fisher's exact test for 101 (*P*-value < 3.0e-5) and 107 (*P*-value < 3.0e-7) up-regulated genes in reprogrammed mSSCs and iPSCs, respectively, of Polo *et al.* (7), using gene ontology biological process terms. To improve interpretability, the identified GO terms significantly involved in each reprogramming process were visualised using the REVIGO web server (22), which reduced the redundancy among GO terms.

Integrative prediction of 53 mouse pluripotency regulatory factors

To identify novel mouse pluripotency regulatory factors using an integrated systems biology approach, we first compiled a core pluripotency interactome (core-PI) from three interactome data sets, one for *Nanog* (23) and two for Oct4 (24,25). For *Nanog* interactome mapping, three independent tandem purifications with anti-Flag immunoprecipitation coupled with mass spectrometry analysis were performed (23). We initially selected 57 co-purified proteins confirmed by two or more purification experiments with an average of at least one unique peptide. To construct a high-

confidence Nanog interactome, we assessed co-expression between Nanog and associated proteins using mRNA expression of reprogrammed cells (6) based on Pearson correlation coefficient (PCC). Among 57 co-purified proteins, 34 proteins were significantly co-expressed with Nanog (P -value < 0.01). Similar proteomics analyses have been performed previously for Oct4 (24,25). Using similar criteria as those used for the Nanog interactome, we selected 127 proteins that co-purified with Oct4. The above co-expression analysis validated 60 of them for construction of a high-confidence Oct4 interactome. We combined the Nanog and Oct4 interactomes to reconstruct a mouse Oct4-Nanog interactome of 88 proteins.

Genes functionally associated with the core pluripotency interactome might also be involved in modulating reprogramming. We previously developed HumanNet, a high-accuracy genome-scale functional gene network for humans, using machine learning and Bayesian integration of various types of large-scale genomics data (26). To perform network-based prediction of mouse pluripotency regulatory factors on the human gene network, we identified 87 human orthologs of the mouse core pluripotency interactome genes using Inparanoid 4.1 orthology detection software (27). We used these 87 human genes to guide the prioritisation of 4994 human genes for pluripotency interactome based on sum of edge-weight scores for all connections between each gene and to the 87 core pluripotency interactome genes in HumanNet. We observed an intersection of 244 human genes in the 4994 network-based candidates and 612 human orthologs of 685 up-regulated mouse genes in mSSC, which are considered as more confident candidates due to the multiple supports from orthogonal evidence including expression and interaction data.

Assuming that most pluripotency regulatory factors are transcriptional regulators, the 244 candidate genes were further filtered for transcription factors. We compiled 2464 human transcription factor genes from two published research articles (28,29). Vaquerizas *et al.* reported 1833 human transcription factors by manual curation of sequence-specific DNA-binding factors, their functions, genomic organization and evolutionary conservation (29). Kanamori *et al.* reported 1675 mouse transcription factors based on DNA-binding properties and their regulators (28). Human genes were converted to mouse genes by orthology. Then, we selected 2,464 transcription factors reported in at least one of the articles. This filtration resulted in 53 final candidate genes for analysis.

Pluripotency regulator candidates from Oct4-reporter screens

We further narrowed the set of 53 candidates for pluripotency regulation using both public and in-house experimental data. Genome-wide screens for pluripotency genes have been conducted based on RNAi analysis with the Oct4 reporter system (30–32). We compiled a total of 864 mouse candidate genes for pluripotency regulation identified by three primary RNAi screens: 566 genes from a study of human ESCs (30), 296 genes from a study of mouse ESCs (31) and 148 genes from another study of mouse ESCs (32). We found that 10 of the 53 candidate mouse genes belonged to

the set of 864 potential pluripotency regulators identified by the Oct4-reporter system screens. We independently performed qPCR analysis for the 53 candidates and validated the expression of 46 genes in reprogrammed cells, and finally obtained four mouse genes that passed all steps of the integrative prediction pipeline.

Construction of a transcriptional regulatory network for reprogramming

To construct a transcriptional regulatory network between the two candidate genes, *Pc4* and *Zfp64*, and 21 reprogramming marker genes, we employed the genie3 (33) R package with the same microarray data used for the clustering analysis.

Culture of mouse SSCs, mSSCs, ESCs and iPSCs

The mouse SSCs and mSSCs cultured as previously described (15) and sampled for real-time RT-PCR analysis. The mouse ESC line was derived from a C57BL/6 strain mouse, and maintained on irradiated mouse embryonic fibroblasts as a feeder-cell layer in mESC growth medium, which is described in Supplementary Methods.

Lentiviral vector construction and lentiviral production

The full-length mouse *Pc4* cDNA (MC203765, Origene Technologies, Rockville, MD) and reverse tetracycline-controlled transactivator protein (rtTA; Clontech, Shiga, Japan) were PCR-amplified using the *Pc4* open reading frame (ORF) and a plasmid containing rtTA, and subcloned into the pCR[®]8/GW/TOPO[®] (Invitrogen, Carlsbad, CA) Gateway recombinational cloning entry vector. The *Pc4* ORF and rtTA sequence in pCR[®]8/GW/TOPO[®] (Invitrogen) was transferred to the CSII-EF-RfA-IRES2-Venus lentiviral vector (RIKEN, Ibaraki, Japan) by the Gateway[®] LR clonase[™] II (Invitrogen) reaction. A tetracycline (tet)-inducible lentivirus designated LV-tetO containing mouse *Oct4*, *Sox2*, *Klf4* and *c-Myc* was obtained from Addgene (Cambridge, MA). Lentiviral vectors were produced by transient triple-plasmid transfection into 293FT cells (Invitrogen). For more details, see Supplementary Methods.

Generation of *Pc4*-overexpressing mESC lines

Pc4-overexpressing mESC lines were established by Venus (YFP)-expressing lentiviral transfection. The mESCs were split at a density of 2×10^4 cells onto fresh MEF feeder cells seeded into a 6-well dish (containing mESC growth medium) with virus particles, and 25 μ g/ml polybrene (Sigma Aldrich) was added. After 24 h, the medium was replaced with fresh growth medium. After 4 days later, mESC colonies expressing YFP were picked and replated. Three different *Pc4*-overexpressing mESC lines were established.

RNA interference

Small interfering RNAs (siRNAs) were designed to knock down *Pc4* expression using *Pc4* ORF region-specific sequences that were conjugated to a 3'-UU-overhang. *Pc4*-specific siRNAs were chemically synthesised by Genolution

Pharmaceuticals (Seoul, South Korea). The siRNA duplex sequences were 5'-GGAAAGAUGAGAUUGUCAUU-3' (sense) and 5'-UGACAUAUCUCAUCUUUCCUU-3' (antisense). For efficient *Pc4* knock-down, MEFs were dissociated to single cells by treatment with 0.05% trypsin-EDTA (Hyclone), and 4×10^5 cells were seeded on 5% Matrigel (BD Bioscience, Bedford, MA) in coated 12-well dishes. Then, 100 nM siRNAs were transfected into MEFs using Lipofectamine 2000 (Invitrogen).

Proliferation assay

To analyse phenotypic changes of mESCs resulting from *Pc4* overexpression, a cellular proliferation assay was performed and colonies and cell numbers were counted. Then, the same number of cells were seeded (2×10^3 cells) in 4-well dishes, and mESCs were maintained for four days with mESC growth medium. The number of colonies that stained with alkaline phosphatase (AP; Sigma Aldrich) was determined. The proliferation rate of *Pc4*-OE mESCs was compared with that of control mESCs on every passage.

Generation of iPSCs and comparative study of reprogramming efficiency

For infection, each concentrated lentiviral supernatant for *Oct4*, *Sox2*, *Klf4*, *c-Myc*, rtTA, and with or without *Pc4* was added to the preseeded Oct4-EGFP MEF from B6;CBA-Tg(Pou5f1-EGFP)2Mnn/J mouse cells (Jackson Laboratory, ME) in Dulbecco's modified Eagles medium (DMEM) with high glucose (Hyclone) supplemented with 10% FBS (Gibco), 1% (v/v) non-essential amino acids (Gibco-BRL), 100 U/ml penicillin G, 100 µg/ml streptomycin (Gibco-BRL) and 0.1% β-mercaptoethanol (Gibco-BRL). The cells were plated on a 6-well dish at a density of 1×10^5 cells per well with 25 µg/ml polybrene (Sigma Aldrich). After 1 day, the medium was changed with fresh medium containing 2 µg/ml of doxycycline (Sigma Aldrich) to induce the expression of exogenous genes. We recorded the date as day 0. On day 4, the transfected cells were replated at a density of 2×10^4 cells per well in plates pre-coated with Matrigel (BD), and cultured for 2 more days. On day 6, the medium was changed to iPSC induction medium that consisted of DMEM with high glucose (Hyclone), 20% knockout serum replacement (Gibco-BRL), 1% (v/v) non-essential amino acids (Gibco-BRL), 0.1% β-mercaptoethanol (Gibco-BRL), 100 U/ml penicillin G, 100 µg/ml streptomycin (Gibco-BRL) and 2000 U/ml LIF (ES-GRO). Cells continued growing until GFP-expressing iPSC colonies appeared. The GFP-positive colonies were picked for establishing iPSC lines, stained for the AP cell proliferation assay, and the number of AP-positive colonies were counted to determine the reprogramming efficiency by *Pc4* co-transfection.

Protein extraction and western blot analysis

The total proteins of mESCs were harvested using PRO-PREP (iNtRON Biotechnology) protein extraction solution according to the manufacturer's instructions. Extracted total protein was quantified using a Qubit Fluorometer (Invitrogen) with the Quant-iT protein assay kit (Invitrogen).

Each sample was subjected to western blot analysis as described in Supplementary Methods.

Statistical analysis

All experiments were conducted at least in triplicate, and the results are expressed as the mean ± standard error. Statistical analyses were performed using the one-way ANOVA test and followed by Student's *t*-test if necessary. Statistical significance was considered as $p < 0.05$.

RESULTS

Comparative transcriptome analysis reveals that mSSCs and iPSCs have similar pluripotency states but undergo reprogramming via different transcriptional pathways

To comparatively examine the pluripotency states and reprogramming processes in mSSCs and iPSCs at the molecular level, we compared the transcriptional profiles related to reprogramming/pluripotency among somatic cells, iPSCs, SSCs, mSSCs and partially reprogrammed cells. We utilised publicly available transcriptome data from two previous studies on iPSC reprogramming, which included profiles for partially reprogrammed cells (6,7). Polo *et al.* (7) reported a stepwise progression of reprogramming based on surface markers and the OCT4-GFP knock-in reporter system. After induction of OSKM factors, cell populations converted from Thy1⁺ cells to Thy1⁻ cells and ultimately to SSEA⁺ cells, Oct4-GFP⁺ cells and iPSC cells. Mikkelsen *et al.* (6) generated two different iPSC lines by introducing OSKM factors into two distinct cell types: mouse embryo fibroblast (MEF)-derived iPSCs and mature-B-lymphocyte-derived iPSCs. These authors identified partially reprogrammed cells during cell reprogramming. We performed a comparative analysis using the published expression data (6,7) and the transcriptome data for SSCs, iSSCs and mSSCs generated in the present study. We clustered samples based on the expression of 21 selected genes that exhibit significant variation across samples and belong to the set of 40 reprogramming markers compiled in our previous studies (11,12,14) (Supplementary Table S1).

The clustering analysis based on 21 reprogramming markers revealed a cluster containing mSSCs, iPSCs and Oct4-GFP⁺ cells, whereas SSCs and iSSCs appear to be more similar to partially reprogrammed cells (Figure 1A). SSCs have endogenously higher expression levels of several key pluripotency markers such as *Oct4*, *Epcam*, *Sall4* and *Lin28a* than those of MEFs or B-lymphocytes. In iSSCs, expression levels of other key stem cell marker genes such as *Sox2*, *Foxo15* and *Rex1* (also known as *Zfp42*) increased, whereas somatic cell markers of fibroblast identity such as the mesenchymal master regulator *Snai2* and collagen type V alpha 2 (*Col5a2*) were down-regulated. A pairwise comparison of functional enrichment among up-regulated genes in mSSCs and SSCs (Figure 1B) and in iPSCs and MEFs (Figure 1C) revealed shared gene ontology (GO) biological process terms, such as 'endodermal cell fate specification', 'forebrain development' and 'somatic stem cell maintenance' (Supplementary Table S2).

During reprogramming, *c-Myc* expression dynamics are distinct in iPSCs and mSSCs (Figure 1A). *Myc* expression

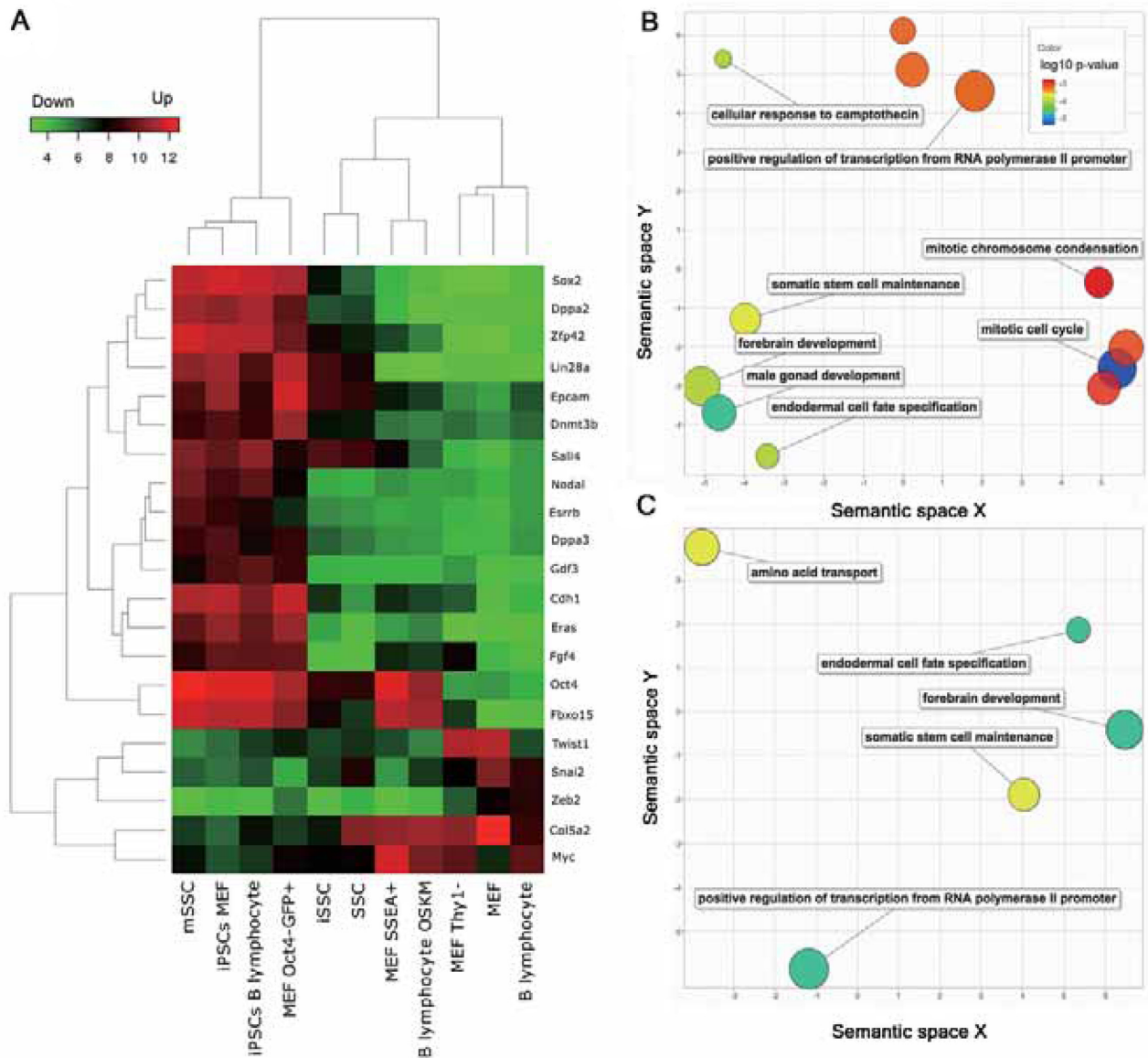


Figure 1. Comparison of reprogramming to mSSCs and iPSCs based on gene expression data analysis. (A) Clustering analysis for 11 cell types during reprogramming: (i) SSC, (ii) iSSC, (iii) mSSC, (iv) MEF, (v) MEF Thy1⁻, (vi) MEF SSEA⁺, (vii) MEF Oct4-GFP⁺, (viii) iPSCs MEFs, (ix) B-lymphocytes, (x) B-lymphocytes OSKM and (xi) iPSCs B-lymphocytes. The results showed that mSSCs were grouped with iPSCs and SSCs and iSSCs were grouped with partially reprogrammed cells. Visualisation of GO terms associated with (B) 101 up-regulated genes in mSSCs compared to SSCs and (C) 107 up-regulated genes in iPSCs compared to MEFs. After performing gene-set enrichment analyses for GO biological process terms using Fisher's exact test, we visualised the enriched GO terms with reduced redundancy using the REVIGO web server.

was higher (q -value = $9.27e-07$) in partially reprogrammed Thy1⁻ cells than in MEF cells, and declined (q -value = $3.22e-07$) in iPSCs. During iPSC reprogramming, *Myc* promotes proliferation and activates a set of pluripotency genes and miRNA (34). However, we did not observe significant changes in *Myc* expression during SSCs reprogramming to mSSCs (q -values are 0.239 and 0.695 for iSSCs and mSSCs, respectively). There are several other pluripotency marker genes that show different expression levels in iSSCs and partially reprogrammed iPSCs. These results suggest that mSSC reprogramming involves a different transcriptional program than that of iPSC reprogramming. We conclude

that mSSC reprogramming generates multipotent stem cells that are similar to iPSCs, but proceeds via different transcriptional pathways than those of iPSC reprogramming.

Integrated systems biology approach to prioritise mouse genes for pluripotency regulatory factors

SSCs express some key pluripotency genes (16) and can be spontaneously reprogrammed to mSSCs without ectopic OSKM gene expression under inducing culture conditions (14). Therefore, we reasoned that the set of up-regulated genes during mSSC reprogramming may contain additional factors that facilitate somatic cell pluripotency reprogram-

ming. We utilised our transcriptome profile data and detected 685 mouse genes that were up-regulated in mSSCs compared to SSCs. One naïve approach to select candidate genes for further experimental testing would be to focus on genes with high fold-changes in expression levels in reprogrammed cells. However, many genes exhibiting high fold-changes in expression levels are effector genes rather than causal reprogramming factors. To maximise our prediction success, we designed a multi-step gene prioritisation pipeline using an integrative approach.

In our prioritisation pipeline, the set of 685 up-regulated mSSC genes was integrated with three types of public data: (i) core pluripotency interactome (core-PI) (23–25), (ii) a probabilistic functional gene network (26) and (iii) transcription factor information (28,29) (see Figure 2A). First, we compiled 88 mouse genes for the key pluripotency interactome by combining three interactome data sets, a Nanog interactome (23) and two Oct4 interactomes, and then applied further filtration for interactions in agreement with co-expression-based expression profiles during iPSC reprogramming (7). To exploit the maximum potential of interactome data, we constructed an expanded gene set by including additional genes functionally associated with the core-PI genes. The candidate gene set expansion by functional links requires a gene network with high accuracy and wide gene coverage. We previously constructed a highly accurate and comprehensive co-functional human network, designated HumanNet (26). To perform network-based prediction of additional pluripotency genes using HumanNet, we identified 87 human orthologs for the mouse core-PI genes using the Inparanoid algorithm (27). When we searched HumanNet with the 87 human core-PI gene orthologs, 68 of them were interconnected in the network, indicating that HumanNet is highly predictive for the key pluripotency interactome in humans (35). Thus, network neighbours of the core-PI genes also are likely to be involved in reprogramming and pluripotency. Consequently, we established an expanded pluripotency interactome (Expanded-PI) of 4994 human genes including 87 core-PI genes and their neighbours in HumanNet.

To prioritise genes for pluripotency regulatory factors, we filtered 612 human orthologs of the 685 genes that were up-regulated in mSSCs for the Expanded-PI genes. We observed that 244 human genes overlap between the two sets of 612 candidates from transcriptome evidence and 4994 candidates from interactome evidence. Given that the majority of pluripotency regulatory factors are transcriptional regulators, we further filtered the 244 genes using data for 2464 human transcription factor genes compiled from two published research studies (28,29). After all filtration steps, we identified 53 mouse genes (and their 53 human orthologs) as the final candidates for novel pluripotency regulatory factors (Supplementary Table S3).

To generate a pathway view for the 53 candidate pluripotency regulatory factors, we constructed a functional network among the 53 genes and 87 core-PI genes based on HumanNet (26) (Figure 2B). The pathway view indicated that two gene sets were highly interconnected (111 of 126 genes were connected), and 10 genes overlapped the two sets (diamond nodes in Figure 2B). We identified three functional modules from the network: (i) DNA repair, (ii) chromatin regulation and (iii) stem cell regulation or development (see Figure 2B). These results indicate that the Oct4-Nanog interactome and the 53 candidate pluripotency regulatory factors are closely associated and contain pathways for DNA repair, chromatin regulation and stem cell regulation or development.

matin regulation and (iii) stem cell regulation or development (see Figure 2B). These results indicate that the Oct4-Nanog interactome and the 53 candidate pluripotency regulatory factors are closely associated and contain pathways for DNA repair, chromatin regulation and stem cell regulation or development.

Assessment and further selection of candidate pluripotency regulatory factors

We performed three different assessments of the final 53 candidate genes using several types of orthogonal information. First, we used quantitative PCR analysis to examine their expression levels in known pluripotent cells such as ESCs, iPSCs and mSSCs. The results indicated that 46 of the 53 candidate genes (86.8%) showed higher expression levels in pluripotent cells than in SSCs (Figure 3A). This suggests that our candidate gene set is highly enriched for genes that are up-regulated in pluripotent cells. Second, we utilised publicly available data on RNAi-responsive genes involved in regulation of self-renewal and pluripotency. We compiled 864 genes that appeared to regulate *Oct4* expression from three genome-wide RNAi screens using the Oct4-reporter assay (30–32). Among the 53 candidate genes, 10 genes showed statistically significant responses to RNAi for regulating *Oct4* activation. Thus, our candidate gene set showed significant enrichment for genes involved in regulating *Oct4* activation (P -value = $1.28e-4$, one-tail Fisher's exact test). Third, we searched the literature for evidence to validate whether the candidate genes were implicated in maintaining pluripotency or promoting reprogramming (Supplementary Table S4). Of the 53 candidates, 24 genes (45.3%) were implicated in pluripotency and reprogramming. Of these 24 genes, the following 19 genes (79.2%) have important roles in ESC maintenance and pluripotency: *Chd1*, *Esrrb*, *Hmgb2*, *Klf5*, *Mtf2* (also known as *Pcl2*), *Myb*, *Mybl2*, *Nanog*, *Nr0b1* (also known as *Dax1*), *Nr5a2*, *Oct4*, *Prdm5*, *Sall1*, *Sox2*, *Tfcp2l1*, *Utf1*, *Zfp462*, *Zic3* and *Zscan10* (also known as *Zfp206*) (see Supplementary Table S4 for more detailed information about these genes). *Nr5a2*, *Utf1*, *Esrrb* and *Nr0b1* also have been reported as factors that accurately mark the rare cells that will become iPSCs (36,37) and as naïve pluripotency markers (38). The remaining five genes appear to regulate rather than promote pluripotency: *Ctbp2*, *Otx2*, *Tcea3*, *Tcf7l1* and *Tgif1* (39–43). The combined results of the three independent validation efforts indicate that our 53 candidate genes are highly enriched for genes involved in promoting and maintaining cellular pluripotency.

One of the candidate genes was selected for further validation by considering multiple lines of evidence. We first sorted the 53 genes by their qPCR expression score in PSCs, and focused on the top 33 genes (orange bars in Figure 3A). Of these 33 genes, six showed significant RNAi responses for regulation of *Oct4* activation: *Sox2*, *Nanog*, *Oct4*, *Pc4*, *Zfp64* and *Zscan10* (red boxes in Figure 3A). Among the remaining 27 candidates, *Esrrb* and *Nr0b1* were previously reported to enhance OSKM-mediated reprogramming (44). Because *Nanog*, *Oct4*, *Sox2* and *Zscan10* were already known to have important roles in pluripotency (5,45), we focused on the other two candidates, *Pc4* and *Zfp64*. We then

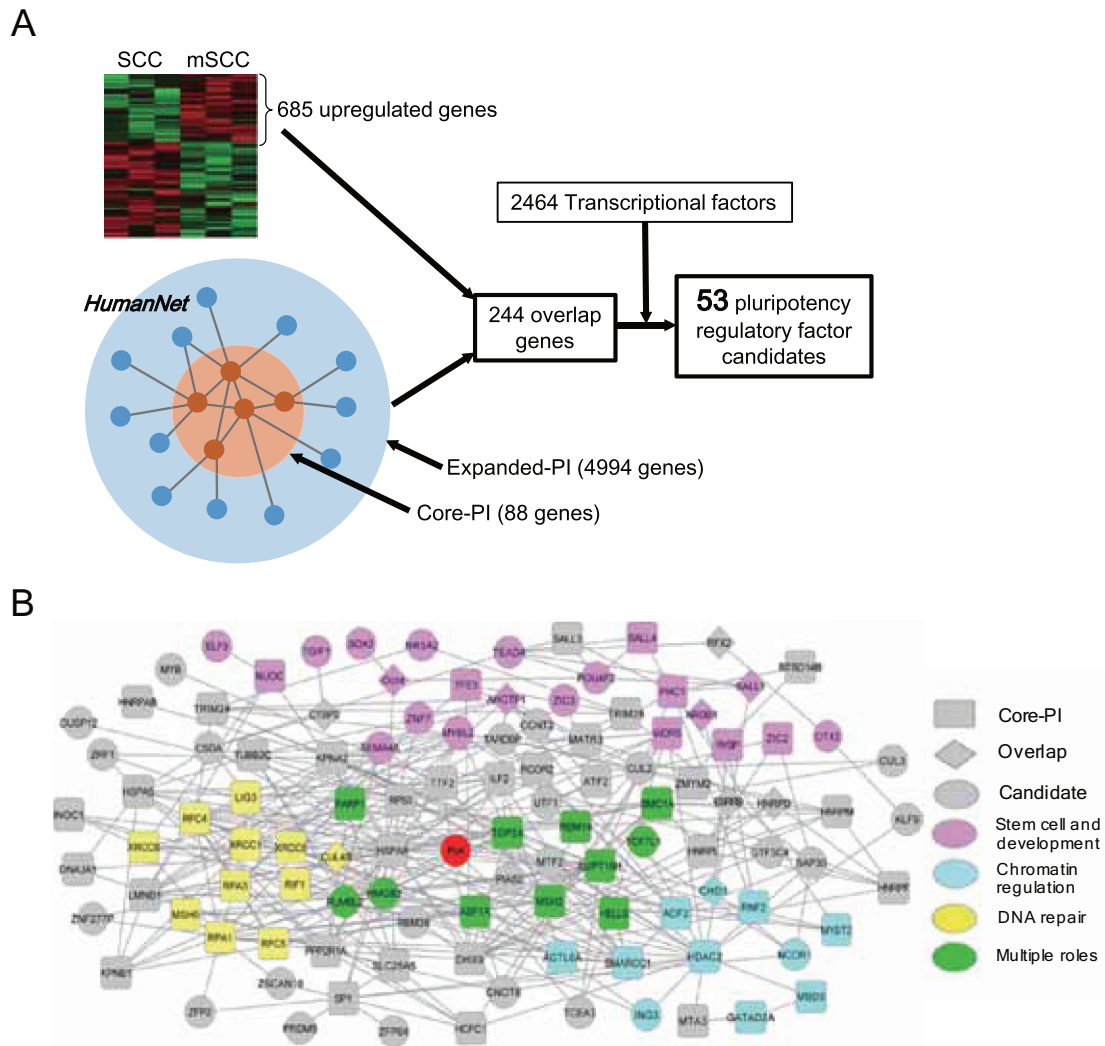


Figure 2. Systematic prediction of mouse pluripotency regulatory factors. (A) Schematic to summarise a systematic prediction pipeline for mouse pluripotency regulatory factors by integrative use of differential expression for reprogrammed cells, expanded pluripotency interactome (expanded-PI) inferred from functional association with the Nanog-Oct4 interactome (core-PI), annotation for transcriptional regulators. (B) Functional gene network for stem cell maintenance and reprogramming. The network consists of 62 core-PI genes, 39 candidate genes, and 10 overlapping genes between them. The node colour represents biological functions: pink, stem cell maintenance and developmental processes; sky blue, chromatin regulation; yellow, DNA repair; purple, multiple processes; grey, genes with no information about biological function. Network edges are based on those of HumanNet.

constructed a gene regulatory network for the two candidate genes and 21 reprogramming marker genes used for the mSSC and iPSC clustering analysis (see Figure 1A). The results suggested that *Pc4* had a more important role than *Zfp64* based on out-degree centrality and hierarchical position in the network (Figure 3B). Notably, qPCR analysis showed that endogenous expression level of *Pc4* mRNA was significantly higher in SSC than MEF, and substantially increased in their reprogrammed cells, mSSC and miPSC (mouse iPSC), respectively (Figure 3C). These results suggest that *Pc4* is a positive regulator for both spontaneous and induced pluripotency, and its high endogenous expression in SSC could initiate spontaneous reprogramming in the culture inducing system. Therefore, we selected *Pc4* as a promising candidate to validate reprogramming competency and dissect molecular mechanisms of reprogramming.

***Pc4* promotes pluripotency marker gene expression and represses somatic cell marker expression**

We performed *Pc4* gain-of-function analysis by establishing *Pc4*-overexpressing (OE) mESC lines. Three different *Pc4*-OE mESC lines that express YFP were established. Mock mESCs were generated by transfecting empty CSII-EF-RfA-IRES2-Venus lentiviral vector (Figure 4A). RT-PCR and real-time RT-PCR were performed to validate *Pc4* expression levels. The three *Pc4*-OE lines showed approximately two-fold higher *Pc4* expression levels than those of control mESCs (Figure 4B and C). Western blot analysis using specific anti-PC4 antibody confirmed enhanced PC4 protein levels in the *Pc4*-OE mESC lines (Figure 4D and E). We tested whether *Pc4* promoted cellular proliferation using a proliferation assay based on the number of colonies for control and *Pc4* OE lines by alkaline phosphatase (AP)

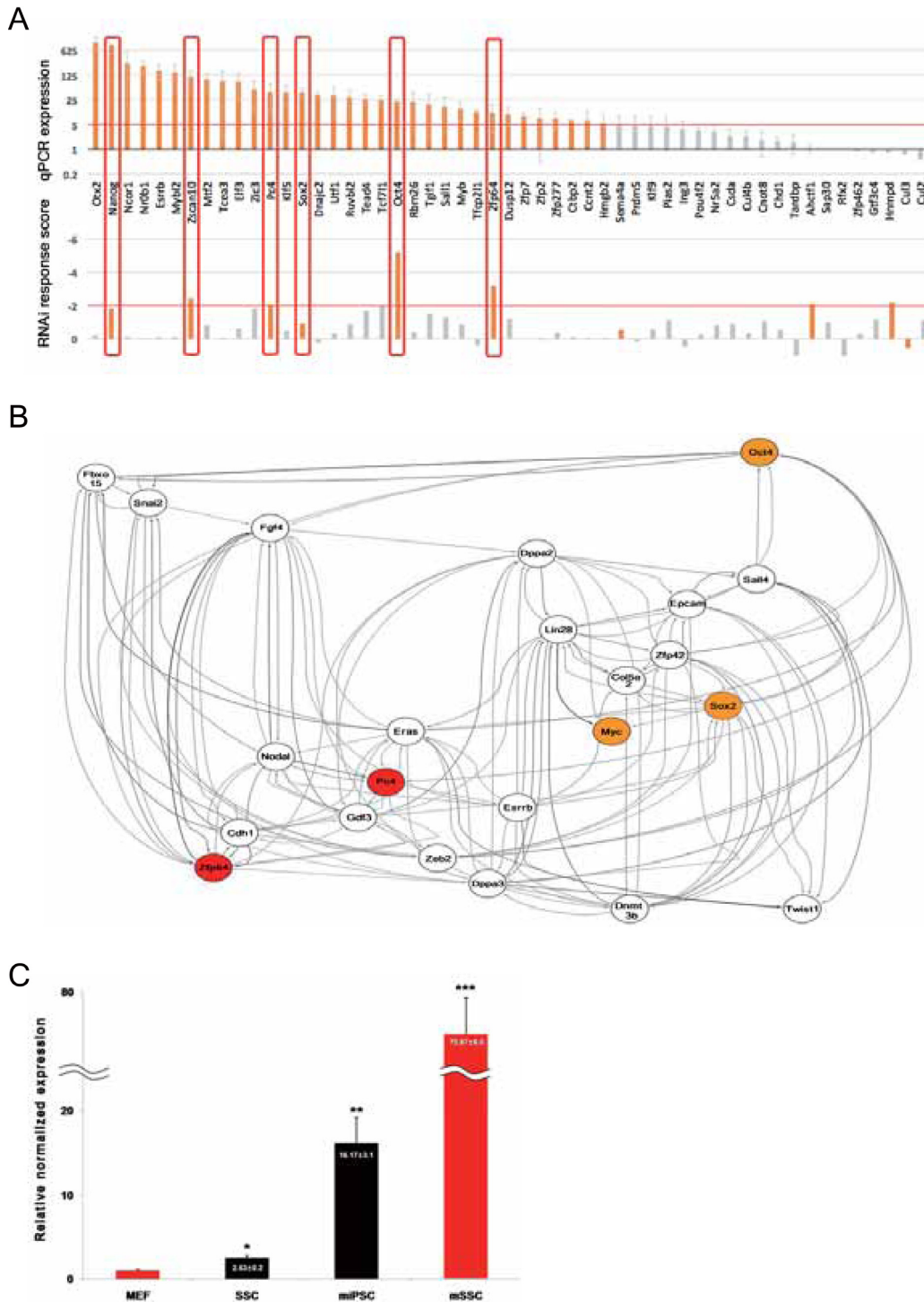


Figure 3. Further selection of candidate pluripotency regulatory factors. **(A)** We selected 864 candidate genes by combining three RNAi response screens. Only Chia *et al.* provided genome-wide RNAi screen results; therefore, we used those results for this plot. Ten candidates were confirmed in three RNAi response analyses. RNAi responsive score is the mean of the Z-score for GFP fluorescence reduction, and its cut-off score for statistical significance is less than -2 . The scores for *Sox2*, *Sema4a* and *Cul3* were not less than -2 , but they have been reported as significantly responsive genes in two other RNAi screens. In qPCR analysis, 46 of the 53 candidate genes (86.8%) were up-regulated in three PSCs—ESCs, iPSCs and mSSCs (upregulation was determined by the average qPCR expression in three PSCs is higher than expression in SSC), but we focused on 33 highly expressed genes (five-fold or higher increase compared to SSC expression levels). Only six overlapping genes between 33 highly expressed genes in PSCs and 10 RNAi-responsive genes were observed: *Nanog*, *Oct4*, *Pc4*, *Sox2*, *Zfp64* and *Zscan10*. **(B)** A gene regulatory network was constructed based on microarray data of 21 reprogramming marker genes and the two selected candidate genes *Pc4* and *Zfp64*. The two candidate genes and the three key reprogramming factor genes, *Oct4*, *Myc* and *Sox2*, are colour-coded as red and orange, respectively. *Oct4* and *Sox2* also are candidate genes, and are coloured by orange because they are key reprogramming factor genes. **(C)** Results of qPCR analysis of *Pc4* in MEF, SSC, miPSC and mSSC. Expression levels of *Pc4* are significantly different among those cells (P -value < 0.01 by ANOVA test for all comparisons, except P -value = 0.015 between MEF and SSC).

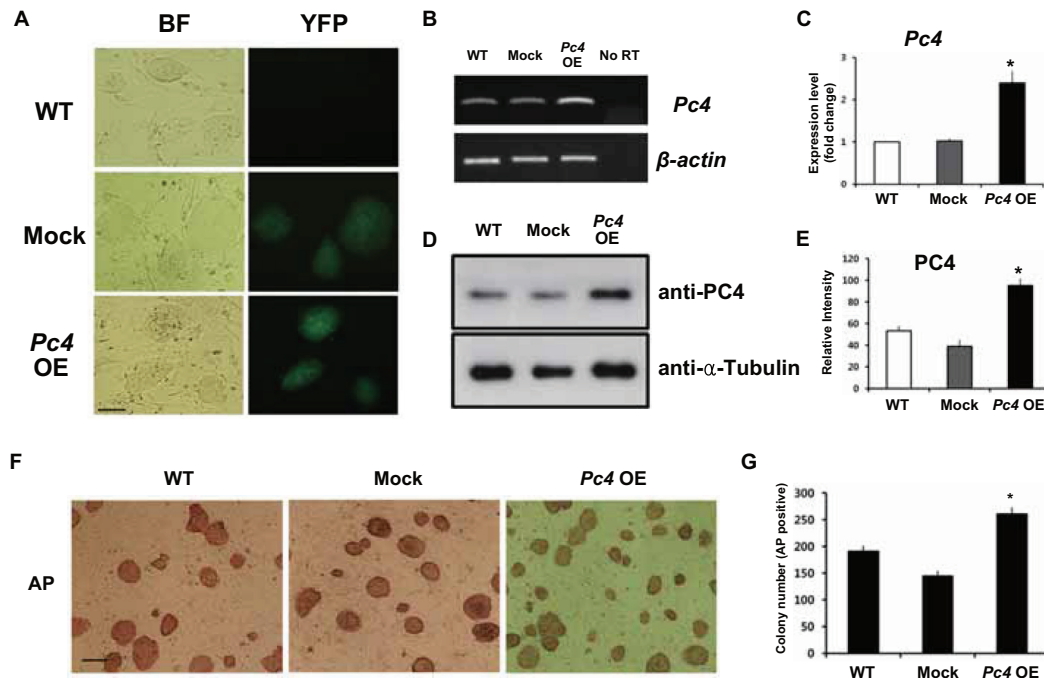


Figure 4. Gain-of-function analysis for *Pc4*. (A) Microscopic images of controls and *Pc4*-overexpressing mESC lines generated by *Pc4*-IRES2-Venus (YFP) lentiviral transfection. Bars = 100 μ m. (B) RT-PCR analysis of *Pc4* expression level and (C) real-time RT-PCR were performed to confirm *Pc4* overexpression ($*P < 0.01$). (D) Increased PC4 protein level was confirmed by western blot analysis. (E) Quantification of western blot bands ($*P < 0.01$). (F) Alkaline phosphatase staining of colonies is shown for undifferentiated wild-type (WT), mock and *Pc4*-overexpressing mESCs. (G) Cellular growth rate measure by colony counting. Scale bars represented 250 μ m ($*P < 0.01$).

staining. We observed a slight but significant increase in the number of colonies in *Pc4*-OE lines compared with those of control mESCs (Figure 4F and G).

We examined the effects of *Pc4* overexpression on genome-wide transcriptional regulation in mESCs by performing RNA sequencing (RNA-seq) analysis in *Pc4*-OE mESC lines. Although mESCs have high expression levels of key reprogramming factors, *Pc4* overexpression further enhanced *Oct4* expression (≈ 1.84 -fold increase). We detected 88 up-regulated genes and 253 down-regulated genes in *Pc4*-OE mESC lines with at least 4-fold differences from those of control samples (Supplementary Table S5). Furthermore, 58 of 88 up-regulated genes and 70 of 253 down-regulated genes encoded miRNAs. Given that 1123 of 24 059 mouse genes encode miRNAs, the differentially expressed genes in *Pc4*-OE mESC lines were highly enriched for miRNA-coding genes (P -values were $1.20e-55$ and $8.969e-35$ for up- and down-regulated genes, respectively; one-tail Fisher's exact test). These results suggest that *Pc4* regulates many miRNAs, the majority of which have not yet been functionally characterised.

An extensive literature search for the regulated genes revealed that many of the up-regulated protein-coding genes in *Pc4*-OE mESCs are involved in the initiation and maturation phases of reprogramming (Supplementary Table S6). *Pc4* stimulates bone morphogenic protein 4 (*Bmp4*) expression and BMP signalling, which has an important role in the initiation stage of mouse iPSC reprogramming (46), and mESCs can be preserved in a naïve state by culturing in medium containing either leukaemia inhibitory factor (LIF) or BMP4 (47,48). *Pc4* also enhances the late

or stabilization phase of reprogramming. *Pc4* overexpression also increased the expression of *Gbx2* (49) and *Lifr* in LIF/STAT3 signalling, which is required for the maturation phase of reprogramming (50), and several genes associated with the late or stabilization phase such as *Lefty2* (9), *Dppa2*, *Dppa4* and *Gdf3* (12). *Pc4* overexpression increased the expression of several known genes such as *Nr0b1* (51), *Klf5* (52), *Foxd3* (53) and *Lin28* (54). *Dppa2*, *Lin28* and *Nr0b1* have been reported as factors that accurately mark the rare cells that will become iPSCs (36,37), and as naïve pluripotency markers (38).

We found that the majority of down-regulated protein-coding genes in *Pc4*-OE mESC lines are involved in the suppression of MEF marker genes such as *Thy1*, *Postn* (12), *Cd44*, *Snail1/2*, *Zeb2* (11), *Nfix*, *Prrx1* and *Tgfb3* (55) (Supplementary Table S7). Down-regulation of several key MEF-specific genes greatly enhanced reprogramming efficiency (55). Suppression of Tgf β (*Tgfb1* and *Tgfb2*) signalling cooperatively induced iPSCs (56) and inhibited the p53/21 pathway (*Cdkn1a* known as *p21*), increased cell division and accelerated iPSC formation (57). The mesenchymal-to-epithelial transition occurs during somatic cell reprogramming to iPSCs, leading to down-regulation of N-cadherin (*Cdh2*) and epithelial-to-mesenchymal transition (EMT) genes (47). Increased miR-145 expression inhibits human ESC self-renewal and represses pluripotency gene expression (58). The polycomb protein SUZ12 together with Oct4, Sox2 and Nanog represses miR-615 expression (59).

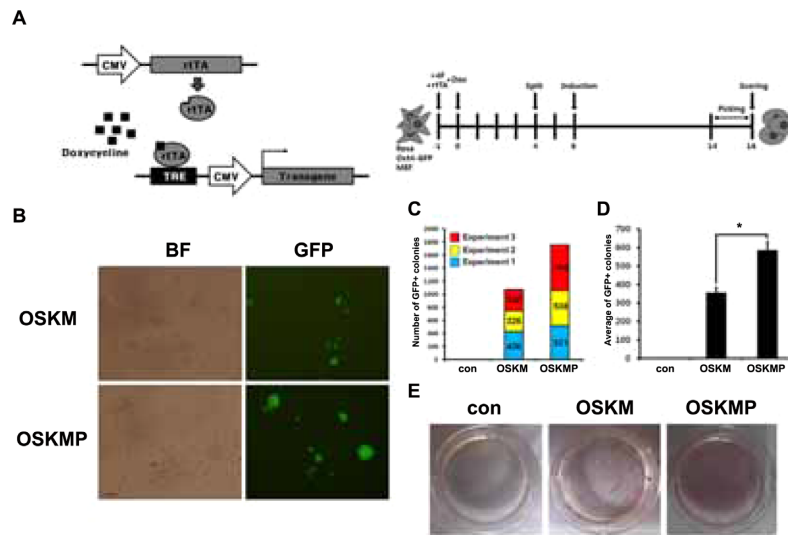


Figure 5. Generation of iPSCs from MEFs with OSKM or OSKMP. (A) Schematic of the inducible lentivirus containing OSKM and rtTA, and experimental outline for generating iPSCs and analysing reprogramming efficiency. (B) Appearance of GFP⁺ colonies in reprogrammed cells. (C) The number of GFP⁺ colonies for reprogrammed cells induced by OSKM or OSKM+Pc4 (OSKMP) in three independent experiments was represented in stacked bar graphs. (D) The average number of GFP⁺ colonies for reprogrammed cells in three replicated experiments indicates that OSKMP has ≈ 1.7 -fold higher reprogramming efficiency than OSKM ($*P < 0.05$). (E) AP staining of colonies reprogrammed by OSKM or OSKMP on day 16.

Pc4 enhances OSKM-mediated reprogramming efficiency

We hypothesised that *Pc4*-overexpression-mediated activation of pluripotency-related gene expression and repression of MEF genes enhanced OSKM-mediated reprogramming efficiency. To test this hypothesis, we compared iPSCs generated by lentiviral transfection of MEF with OSKM or OSKM+*Pc4* (OSKMP) using a transgenic *Oct4*-EGFP reporter. Lentiviral vectors are efficiently integrated into the genome and exogenous gene expression is inducible by administering doxycycline (Figure 5A). Morphological changes were observed in MEF within a few days after doxycycline induction; on day 10, small GFP-positive colonies were observed on the plate (Figure 5B). Although the first GFP-positive colony appeared for both OSKM and OSKMP at the same time point, the colony formed by OSKMP emitted strong GFP fluorescence compared with that of OSKM. The number of GFP-positive colonies was counted on day 16. We observed ≈ 1.7 -fold increase in GFP-positive colonies by OSKMP induction compared with OSKM induction (Figure 5C and D). The results of AP staining also showed that the OSKMP group had much higher proliferation than the OSKM group (Figure 5E).

To confirm the role of *Pc4* in regulating expression of pluripotency-related genes, we performed the reciprocal experiments depleting *Pc4* expression in mESC. After transfection of mESCs with *Pc4* siRNA, we measured expression of *Pc4* by qPCR (Supplementary Figure S1A). Knock-down *Pc4* level was decreased by approximately 80% at 24 h and 48 h. In addition, decrease in PC4 protein level was confirmed by Western blot analysis at 24 h and 48 h (Supplementary Figure S1B). Using qPCR, we found that *Pc4* knock-down in mESCs significantly reduced the expression of pluripotency-related genes, *Oct4* and *Sox2* (Supplementary Figure S1C and S1D). We also tested pluripotency maintenance using AP staining and cellular prolifer-

ation by total cell counting, and observed *Pc4* knock-down in mESCs significantly reduced the number of AP positive colonies and total cell count at 48 h (Supplementary Figure S1E and S1F).

Pc4 promotes and maintains transcriptional activity of key reprogramming factors

To test whether *Pc4* enhances reprogramming efficiency by regulating transcriptional activity of endogenous key reprogramming factors, we profiled expression levels of *Pc4* and the key reprogramming factors *Oct4*, *Sox2*, *Nanog* and *Klf4* during OSKM-mediated reprogramming on day 1, 2, 3, 5 and 7 after doxycycline induction (Figure 6A). We observed a gradual increase in *Pc4* expression, followed by gradual increases (short initial lag time) in key reprogramming factor expression up to day 7. Conversely, siRNA-mediated *Pc4* knock-down during reprogramming reduced expression levels of *Pc4* and other key reprogramming factors by approximately 80% within three days (Figure 6B). These results indicate that the observed enhancement of OSKM-mediated reprogramming efficiency by additional *Pc4* expression results from *Pc4* regulation of the transcriptional activity of endogenous key reprogramming factors.

DISCUSSION

The discovery that iPSCs can be generated by transfection of transcription factors was a pivotal event in the field of stem cell research because it overcame previous ethical and resource limitations. Despite the huge potential of this technology, the process still suffers from extremely low efficiency and genomic instability. There have been considerable efforts in recent years to improve reprogramming efficiency, including molecular dissection of reprogramming processes (6,7) and screening for additional factors to facilitate the

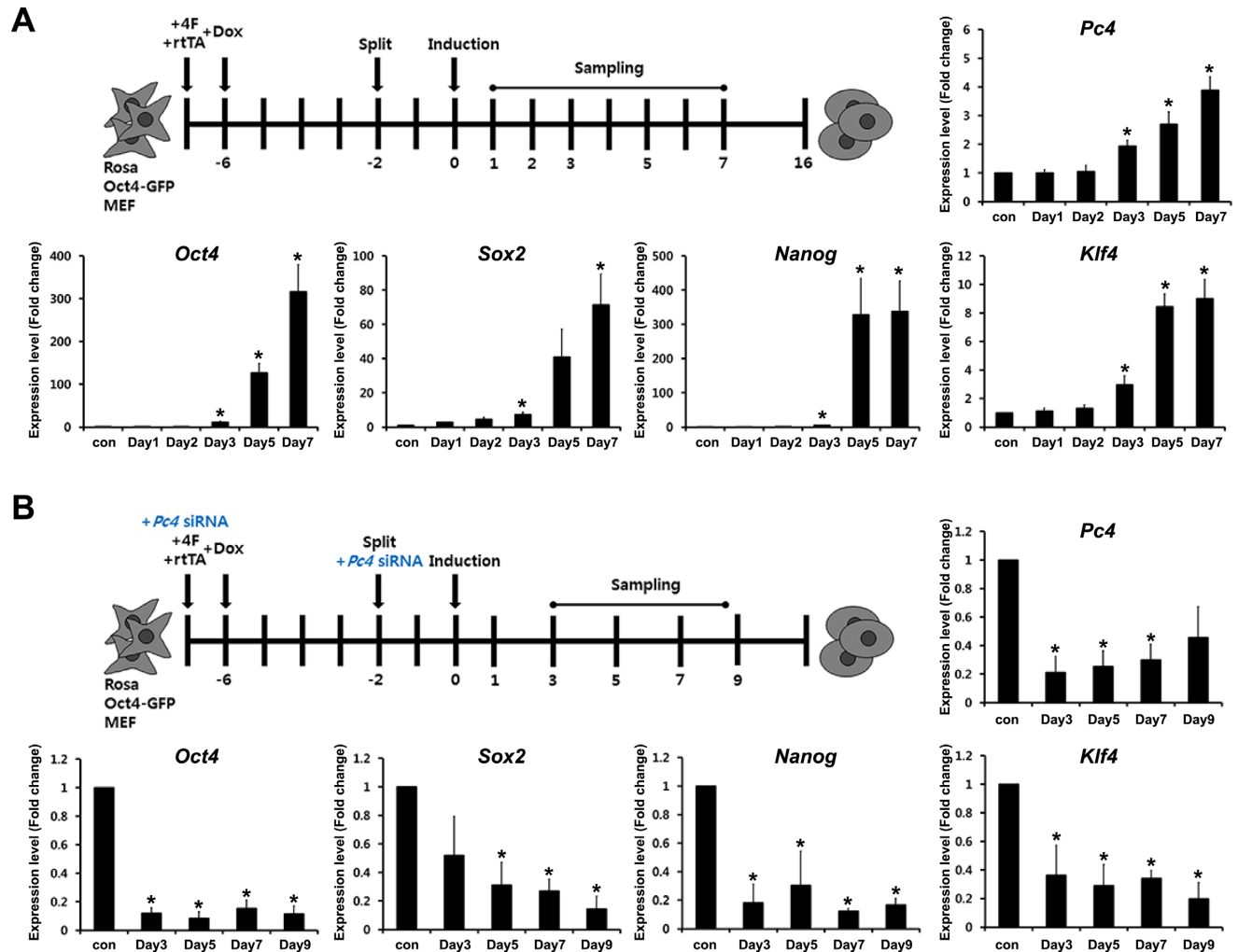


Figure 6. Time-course analysis of PC4-overexpression effects on key reprogramming factors during reprogramming into iPSCs. Experimental outline of doxycycline-mediated reprogramming to iPSCs, and time points for transcriptome profiling. Changes in expression of endogenous key reprogramming factors mediated by *Pc4* overexpression with (A) OSKM, and (B) OSKM + *Pc4* knock-down (* $P < 0.05$).

process (44,60). In this study, we performed a comparative transcriptome analysis to identify differences in transcriptional programs of spontaneous SSC conversion into mSSCs and iPSC technology. We did not observe *c-Myc* expression dynamics during mSSC conversion, whereas *c-Myc* expression increased in partially reprogrammed cells and then decreased in iPSCs (Figure 1A). It has been suggested that *c-Myc* enhances iPSC reprogramming efficiency by promoting cell proliferation, but it simultaneously increases tumorigenicity risk (61). Therefore, in-depth comparative transcriptional analysis of the two reprogramming systems could provide novel insights for the development of safe cellular reprogramming protocols.

We utilised multiple types of orthogonal information related to reprogramming and pluripotency maintenance to prioritise the initial candidates, such as interactome of core pluripotency factors, co-functional links between genes and transcription factor information. This integrated systems biology approach identified 53 candidate pluripotency regulatory factors that were significantly enriched for genes that are relevant to reprogramming/pluripotency in the lit-

erature, expressed in PSCs, and involved in the regulation of *Oct4* activation. These results indicate that our approach is robust and efficient for the prediction of pluripotency regulatory factors.

We used a multi-step filtering and evaluation process, and ultimately selected the *Pc4* candidate gene for further analysis and validation. We demonstrated that *Pc4* enhances the efficiency of OSKM-mediated MEF reprogramming by promoting the transcriptional activity of key reprogramming factors. We also observed that *Pc4* regulates the expression of many genes involved in reprogramming/pluripotency and somatic cell marker genes. *Pc4* is known as a transcriptional co-activator that facilitates transcriptional machinery activity. PC4 physically interacts with the activators and transcription machinery of RNAP II, and is a component of the transcriptional pre-initiation complex (PIC) on melted promoter DNA. These results suggest that PC4 promotes transcriptional activity of key reprogramming factors by facilitating transcriptional machinery activity during the initiation and maturation stages of reprogramming. Stem cells must have high

pluripotency to achieve a robust cell resource for clinical applications. For somatic cell reprogramming into a pluripotent state, fully reprogrammed iPSCs have to be generated in large quantities for research and clinical uses. *Pc4* may be an important factor for generating fully reprogrammed iPSCs by regulating pluripotency-related gene expression. *Pc4* could support a high-quality pluripotent state in PSCs in *in vitro* cultures. We also found that the expression level of *Pc4* was significantly higher in SSC compared to MEF, and substantially increased in reprogrammed cells, more in spontaneously reprogrammed mSSC than induced pluripotent cells (see Figure 3C). These observations suggest that *Pc4* is a positive regulator for both spontaneous and induced pluripotency, and its relatively high endogenous level in SSC could initiate spontaneous reprogramming into mSSC in the culture inducing system.

A similar prediction strategy may be applied to identify transcriptional regulatory factors involved in state transitions of other cell types, such as transdifferentiation (also known as direct conversion) (62). If cell-type specific transcriptome profiles and some prior information (e.g. known genes for cell-type specific differentiation) are available, we may be able to integrate the information with an established genome-wide co-functional or regulatory network to identify additional candidates. Such approaches were not previously feasible due to the lack of multiple and orthogonal omics data for cellular differentiation. However, high-throughput molecular profiling and genome engineering technology will generate an unprecedented volume of data for many cellular lineages in the coming years. At that time, integrated systems biology approaches to predict master transcriptional regulators for lineage conversion would be a pragmatic strategy for the *in vitro* production of therapeutic cells from PSCs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Author contributions: D.L. and I.L. conceived the project. J.J. and H.K. performed experiments, and S.H. performed bioinformatics analysis. J.L. and H.H. assisted bioinformatics analysis, and S.H., S.L. and A.B. assisted experimental analysis. I.L. supervised bioinformatics analysis, and D.L. supervised experimental analysis. J.J., S.H. and I.L. wrote the manuscript. J.J., S.H., I.L. and D.L. edited the manuscript.

FUNDING

This research was partly supported by grants from the Bio & Medical Technology Development Program [2012M3A9C6049723] funded by the Ministry of Science, ICT and Future Planning, and Basic Science Research through the National Research Foundation of Korea [2009-0093821], the Stem Cell Research Program [2006–2004127 to D.L.], and grants from the National Research Foundation of Korea [2012M3A9B4028641, 2012M3A9C7050151 to I.L.]. Funding for open access charge: National Research Foundation of Korea.

Conflict of interest statement. None declared.

REFERENCES

- Gurdon, J.B. (1962) The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.*, **10**, 622–640.
- Chung, Y.G., Eum, J.H., Lee, J.E., Shim, S.H., Sepilian, V., Hong, S.W., Lee, Y., Treff, N.R., Choi, Y.H., Kimbrel, E.A. *et al.* (2014) Human somatic cell nuclear transfer using adult cells. *Cell Stem Cell*, **14**, 777–780.
- Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.S., Sritanadomchai, H. *et al.* (2013) Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell*, **153**, 1228–1238.
- Yamada, M., Johannesson, B., Sagi, I., Burnett, L.C., Kort, D.H., Prosser, R.W., Paull, D., Nestor, M.W., Freeby, M., Greenberg, E. *et al.* (2014) Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells. *Nature*, **510**, 533–536.
- Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663–676.
- Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S. and Meissner, A. (2008) Dissecting direct reprogramming through integrative genomic analysis. *Nature*, **454**, 49–55.
- Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J. *et al.* (2012) A molecular roadmap of reprogramming somatic cells into iPSCs. *Cell*, **151**, 1617–1632.
- Tonge, P.D., Corso, A.J., Monetti, C., Hussein, S.M., Puri, M.C., Michael, I.P., Li, M., Lee, D.S., Mar, J.C., Cloonan, N. *et al.* (2014) Divergent reprogramming routes lead to alternative stem-cell states. *Nature*, **516**, 192–197.
- Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C.L., Treka, D. and Wrana, J.L. (2012) A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell*, **11**, 769–782.
- Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q. *et al.* (2010) A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell*, **7**, 51–63.
- David, L. and Polo, J.M. (2014) Phases of reprogramming. *Stem Cell Res.*, **12**, 754–761.
- Buganim, Y., Faddah, D.A. and Jaenisch, R. (2013) Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.*, **14**, 427–439.
- Stadtfeld, M. and Hochedlinger, K. (2010) Induced pluripotency: history, mechanisms, and applications. *Genes Dev.*, **24**, 2239–2263.
- Kanatsu-Shinohara, M., Inoue, K., Lee, J., Yoshimoto, M., Ogonuki, N., Miki, H., Baba, S., Kato, T., Kazuki, Y., Toyokuni, S. *et al.* (2004) Generation of pluripotent stem cells from neonatal mouse testis. *Cell*, **119**, 1001–1012.
- Kim, H.J., Lee, H.J., Lim, J.J., Kwak, K.H., Kim, J.S., Kim, J.H., Han, Y.M., Kim, K.S. and Lee, D.R. (2010) Identification of an intermediate state as spermatogonial stem cells reprogram to multipotent cells. *Mol. Cells*, **29**, 519–526.
- Kanatsu-Shinohara, M., Lee, J., Inoue, K., Ogonuki, N., Miki, H., Toyokuni, S., Ikawa, M., Nakamura, T., Ogura, A. and Shinohara, T. (2008) Pluripotency of a single spermatogonial stem cell in mice. *Biol. Reprod.*, **78**, 681–687.
- Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.
- GK, S. (2005) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, NY, pp. 397–420.
- Storey, J.D. (2002) A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**, 479–498.
- Carvalho, B.S. and Irizarry, R.A. (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, **26**, 2363–2367.
- Bengtsson, H., Ray, A., Spellman, P. and Speed, T.P. (2009) A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. *Bioinformatics*, **25**, 861–867.

22. Supek,F., Bosnjak,M., Skunca,N. and Smuc,T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, **6**, e21800.
23. Wang,J., Rao,S., Chu,J., Shen,X., Levasseur,D.N., Theunissen,T.W. and Orkin,S.H. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature*, **444**, 364–368.
24. Pardo,M., Lang,B., Yu,L., Prosser,H., Bradley,A., Babu,M.M. and Choudhary,J. (2010) An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell*, **6**, 382–395.
25. van den Berg,D.L., Snoek,T., Mullin,N.P., Yates,A., Bezstarosti,K., Demmers,J., Chambers,I. and Poot,R.A. (2010) An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell*, **6**, 369–381.
26. Lee,I., Blom,U.M., Wang,P.I., Shim,J.E. and Marcotte,E.M. (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.*, **21**, 1109–1121.
27. O'Brien,K.P., Rimm,M. and Sonnhammer,E.L. (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.*, **33**, D476–D480.
28. Kanamori,M., Konno,H., Osato,N., Kawai,J., Hayashizaki,Y. and Suzuki,H. (2004) A genome-wide and nonredundant mouse transcription factor database. *Biochem. Biophys. Res. Commun.*, **322**, 787–793.
29. Vaquerizas,J.M., Kummerfeld,S.K., Teichmann,S.A. and Luscombe,N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
30. Chia,N.Y., Chan,Y.S., Feng,B., Lu,X., Orlov,Y.L., Moreau,D., Kumar,P., Yang,L., Jiang,J., Lau,M.S. *et al.* (2010) A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*, **468**, 316–320.
31. Ding,L., Paszkowski-Rogacz,M., Nitzsche,A., Slabicki,M.M., Heninger,A.K., de Vries,I., Kittler,R., Junqueira,M., Shevchenko,A., Schulz,H. *et al.* (2009) A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell*, **4**, 403–415.
32. Hu,G., Kim,J., Xu,Q., Leng,Y., Orkin,S.H. and Elledge,S.J. (2009) A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.*, **23**, 837–848.
33. Huynh-Thu,V.A., Irrthum,A., Wehenkel,L. and Geurts,P. (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, **5**, e12776.
34. Wernig,M., Meissner,A., Cassady,J.P. and Jaenisch,R. (2008) c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell*, **2**, 10–12.
35. Hwang,S., Rhee,S.Y., Marcotte,E.M. and Lee,I. (2011) Systematic prediction of gene function in Arabidopsis thaliana using a probabilistic functional gene network. *Nat. Protoc.*, **6**, 1429–1442.
36. Buganim,Y., Faddah,D.A., Cheng,A.W., Itskovich,E., Markoulaki,S., Ganz,K., Klemm,S.L., van Oudenaarden,A. and Jaenisch,R. (2012) Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, **150**, 1209–1222.
37. Welling,M. and Geijsen,N. (2013) Uncovering the true identity of naive pluripotent stem cells. *Trends Cell Biol.*, **23**, 442–448.
38. Nichols,J. and Smith,A. (2009) Naive and primed pluripotent states. *Cell Stem Cell*, **4**, 487–492.
39. Acampora,D., Di Giovannantonio,L.G. and Simeone,A. (2013) Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development*, **140**, 43–55.
40. Yi,F., Pereira,L. and Merrill,B.J. (2008) Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells*, **26**, 1951–1960.
41. Yan,L., Womack,B., Wotton,D., Guo,Y., Shyr,Y., Dave,U., Li,C., Hiebert,S., Brandt,S. and Hamid,R. (2013) Tgflf1 regulates quiescence and self-renewal of hematopoietic stem cells. *Mol. Cell Biol.*, **33**, 4824–4833.
42. Park,K.S., Cha,Y., Kim,C.H., Ahn,H.J., Kim,D., Ko,S., Kim,K.H., Chang,M.Y., Ko,J.H., Noh,Y.S. *et al.* (2013) Transcription elongation factor Tcea3 regulates the pluripotent differentiation potential of mouse embryonic stem cells via the Lefty1-Nodal-Smad2 pathway. *Stem Cells*, **31**, 282–292.
43. Tarleton,H.P. and Lemischka,I.R. (2010) Delayed differentiation in embryonic stem cells and mesodermal progenitors in the absence of CtBP2. *Mech. Dev.*, **127**, 107–119.
44. Fishedick,G., Klein,D.C., Wu,G., Esch,D., Hoing,S., Han,D.W., Reinhardt,P., Hergarten,K., Tapia,N., Scholer,H.R. *et al.* (2012) Zfp296 is a novel, pluripotent-specific reprogramming factor. *PLoS One*, **7**, e34645.
45. Wang,Z.X., Kueh,J.L., Teh,C.H., Rossbach,M., Lim,L., Li,P., Wong,K.Y., Lufkin,T., Robson,P. and Stanton,L.W. (2007) Zfp206 is a transcription factor that controls pluripotency of embryonic stem cells. *Stem Cells*, **25**, 2173–2182.
46. Samavarchi-Tehrani,P., Golipour,A., David,L., Sung,H.K., Beyer,T.A., Datti,A., Woltjen,K., Nagy,A. and Wrana,J.L. (2010) Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell*, **7**, 64–77.
47. Pieters,T. and van Roy,F. (2014) Role of cell-cell adhesion complexes in embryonic stem cell biology. *J. Cell Sci.*, **127**, 2603–2613.
48. Ying,Q.L., Stavridis,M., Griffiths,D., Li,M. and Smith,A. (2003) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.*, **21**, 183–186.
49. Tai,C.I. and Ying,Q.L. (2013) Gbx2, a LIF/Stat3 target, promotes reprogramming to and retention of the pluripotent ground state. *J. Cell Sci.*, **126**, 1093–1098.
50. Tang,Y. and Tian,X.C. (2013) JAK-STAT3 and somatic cell reprogramming. *JAKSTAT*, **2**, e24935.
51. Zhang,J., Liu,G., Ruan,Y., Wang,J., Zhao,K., Wan,Y., Liu,B., Zheng,H., Peng,T., Wu,W. *et al.* (2014) Dax1 and Nanog act in parallel to stabilize mouse embryonic stem cells and induced pluripotency. *Nat. Commun.*, **5**, 5042.
52. Ema,M., Mori,D., Niwa,H., Hasegawa,Y., Yamanaka,Y., Hitoshi,S., Mimura,J., Kawabe,Y., Hosoya,T., Morita,M. *et al.* (2008) Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. *Cell Stem Cell*, **3**, 555–567.
53. Liu,Y. and Labosky,P.A. (2008) Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3. *Stem Cells*, **26**, 2475–2484.
54. Viswanathan,S.R., Daley,G.Q. and Gregory,R.I. (2008) Selective blockade of microRNA processing by Lin28. *Science*, **320**, 97–100.
55. Yang,C.S., Lopez,C.G. and Rana,T.M. (2011) Discovery of nonsteroidal anti-inflammatory drug and anticancer drug enhancing reprogramming and induced pluripotent stem cell generation. *Stem Cells*, **29**, 1528–1536.
56. Maherali,N. and Hochedlinger,K. (2009) Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr. Biol.*, **19**, 1718–1723.
57. Hanna,J., Saha,K., Pando,B., van Zon,J., Lengner,C.J., Creighton,M.P., van Oudenaarden,A. and Jaenisch,R. (2009) Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*, **462**, 595–601.
58. Xu,N., Papagiannakopoulos,T., Pan,G., Thomson,J.A. and Kosik,K.S. (2009) MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell*, **137**, 647–658.
59. Anokye-Danso,F., Snitow,M. and Morrissey,E.E. (2012) How microRNAs facilitate reprogramming to pluripotency. *J. Cell Sci.*, **125**, 4179–4187.
60. Singhal,N., Graumann,J., Wu,G., Arauzo-Bravo,M.J., Han,D.W., Greber,B., Gentile,L., Mann,M. and Scholer,H.R. (2010) Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. *Cell*, **141**, 943–955.
61. Nakagawa,M., Koyanagi,M., Tanabe,K., Takahashi,K., Ichisaka,T., Aoi,T., Okita,K., Mochiduki,Y., Takizawa,N. and Yamanaka,S. (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.*, **26**, 101–106.
62. Vierbuchen,T. and Wernig,M. (2011) Direct lineage conversions: unnatural but useful? *Nat. Biotechnol.*, **29**, 892–907.