Transcription factors interfering with dedifferentiation induce cell type-specific transcriptional profiles

Takafusa Hikichi[a,b], Ryo Matoba[c], Takashi Ikeda[a,b], Akira Watanabe[b], Takuya Yamamoto[b], Satoko Yoshitake[a], Miwa Tamura-Nakano[a], Takayuki Kimura[a], Masayoshi Kamon[d], Mari Shimura[a], Koichi Kawakami[e], Akihiko Okuda[d,f], Hitoshi Okochi[a], Takafumi Inoue[g,1], Atsushi Suzuki[h,i,1], and Shinji Masui[a,b,i,2]

[a]Research Institute, National Center for Global Health and Medicine, Tokyo 162-8655, Japan; [b]Center for iPS Cell Research and Application, Kyoto University, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan; [c]DNA Chip Research Inc., Suehirocho, Tsurumi-ku, Yokohama 230-0045, Japan; [d]Research Center for Genomic Medicine, Saitama Medical University, Saitama 350-1241, Japan; [e]Division of Molecular and Developmental Biology, National Institute of Genetics, Department of Genetics, Graduate University for Advanced Studies (SOKENDAI), Mishima, Shizuoka 411-8540, Japan; [f]CREST (Core Research for Evolutional Science and Technology), Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan; [g]Department of Life Science and Medical Bioscience, Waseda University, Shinjuku-ku, Tokyo 162-8480, Japan; [h]Division of Organogenesis and Regeneration, Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka 812-8582, Japan; and [i]PRESTO (Precursory Research for Embryonic Science and Technology), Japan Science and Technology Agency, Saitama 332-0012, Japan

Edited* by Shinya Yamanaka, Kyoto University, Kyoto, Japan, and approved March 5, 2013 (received for review November 27, 2012)

Transcription factors (TFs) are able to regulate differentiation-related processes, including dedifferentiation and direct conversion, through the regulation of cell type-specific transcriptional profiles. However, the functional interactions between the TFs regulating different transcriptional profiles are not well understood. Here, we show that the TFs capable of inducing cell type-specific transcriptional profiles prevent the dedifferentiation induced by TFs for pluripotency. Of the large number of TFs expressed in a neural-lineage cell line, we identified a subset of TFs that, when overexpressed, strongly interfered with the dedifferentiation triggered by the procedure to generate induced pluripotent stem cells. This interference occurred through a maintenance mechanism of the cell type-specific transcriptional profile. Strikingly, the maintenance activity of the interfering TF set was strong enough to induce the cell line-specific transcriptional profile when overexpressed in a heterologous cell type. In addition, the TFs that interfered with dedifferentiation in hepatic-lineage cells involved TFs with known induction activity for hepatic-lineage cells. Our results suggest that dedifferentiation suppresses a cell type-specific transcriptional profile, which is primarily maintained by a small subset of TFs capable of inducing direct conversion. We anticipate that this functional correlation might be applicable in various cell types and might facilitate the identification of TFs with induction activity in efforts to understand differentiation.

ES cells | polycomb | bivalent | neural progenitors | hepatoblasts

The phenotype of a cell is determined by the specificity of its expressed genes, which is primarily regulated by transcription factors (TFs) (1). Prominent examples include the production of induced pluripotent stem cells (iPSCs) from differentiated cell types through the overexpression of pluripotency-associated TFs (2) and the induction of various somatic cell types via the overexpression of the respective cell type-specific TFs in heterologous cells (direct conversion) (3). However, such induction is generally inefficient, and many of the mechanisms remain unclear (4).

Global analyses using several diverse cell types have revealed that core TFs, which directly regulate a large number of genes for cell type specificity, reciprocally regulate one another to maintain a stable, cell type-specific transcriptional profile (5). This regulatory mode may confer mutual exclusivity between different cell type-specific transcriptional profiles, which may explain the inefficiency of induction (1). However, the functional interactions between the TFs regulating different cell type-specific transcriptional profiles remain mostly unknown to date.

Although the dedifferentiation of cells to produce iPSCs is an inefficient process, this induction results in the loss of the cell type-specific transcriptional profile of the differentiated cells (6–8). In this study, we hypothesized that the genes required for iPSC induction may be repressed, in part, by TFs that maintain the respective cell type-specific transcriptional profile. Supporting evidence in B lymphocytes has indicated the repression of iPSC induction by paired box gene 5 (Pax5, a core TF) (8). Therefore, we speculated that the overexpression of a core TF or its activator might augment the maintenance of the cell type-specific transcriptional profile, thus inhibiting alterations in the expression of genes that are required for iPSC induction.

Results and Discussion

To address functional interactions of TFs regulating different cell type-specific transcriptional profiles, we used a neural-lineage cell line as a model, NSEB5-2C, established through the in vitro differentiation of mouse embryonic stem cells (ESCs) (9). The NSEB5-2C line has three hallmark characteristics: (*i*) fibroblast growth factor-2/epidermal growth factor (FGF2/EGF)-dependent proliferation; (*ii*) the capability of differentiating into TuJ1 (β3 Tubulin antibody)-positive cells and GFAP-positive cells; and (*iii*) the expression of marker genes for neural progenitors (Fig. S1*A*; see below). To circumvent the redundancy problem, we used a gain-of-function assay in which a single TF, which is expressed endogenously in NSEB5-2C cells, was overexpressed during iPSC induction. We expected that the TFs at the top of the transcriptional hierarchy would strongly maintain the transcriptional profile and might therefore interfere with dedifferentiation, which would ultimately reduce the iPSC colony number (Fig. 1*A*). To detect genes expressed in higher abundance in NSEB5-2C cells, we compared the global expression between the NSEB5-2C line and mixed RNAs extracted from various organs. The TFs that were more highly expressed in the NSEB5-2C cells were selected using the Gene Ontology (GO) terms "DNA binding" and "transcription activity," yielding 158 TFs (Fig. S1*B* and Dataset S1). We constructed retroviruses with these TFs and adjusted the titer by G418 resistance in NSEB5-2C cells, showing that ∼95% of the cells expressed a transgene, as assayed using enhanced green fluorescent protein (*EGFP*) retrovirus (Fig. S1*C*). At this titer, we coinfected a single TF retrovirus together with retroviruses

Author contributions: S.M. designed research; T.H., T. Ikeda, A.W., S.Y., M.T.-N., T.K., M.S., H.O., T. Inoue, and S.M. performed research; M.K., K.K., A.O., A.S., and S.M. contributed new reagents/analytic tools; T.H., R.M., A.W., T.Y., and S.M. analyzed data; and T.H. and S.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo [accession nos. GSE29875 (microarray) and GSE41252 (ChIP-seq)].

[1]T. Inoue and A.S. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: smasui@cira.kyoto-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1220200110/-/DCSupplemental.

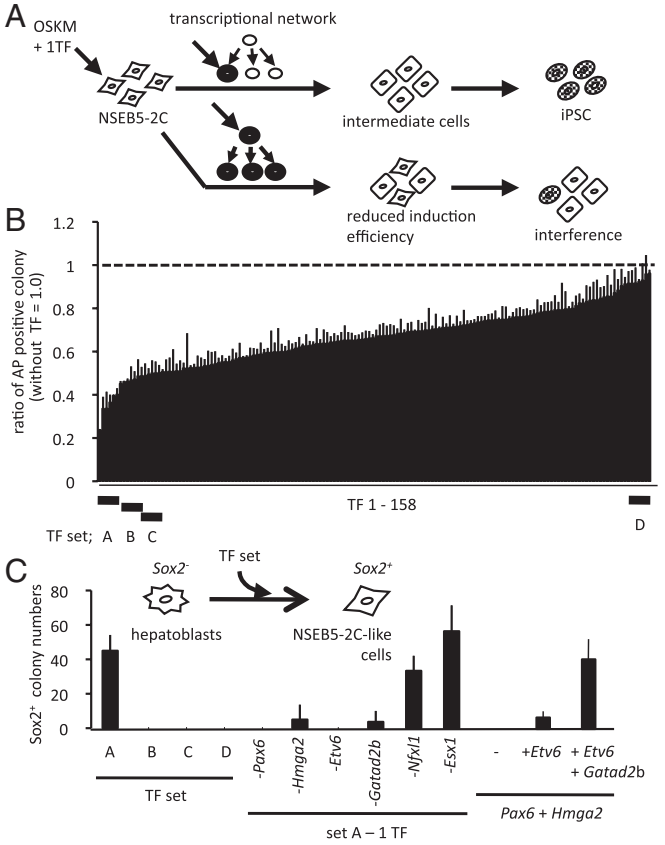

Fig. 1. TFs interfering with iPSC induction in NSEB5-2C induce NSEB5-2C–like cells. (*A*) Principle of the interference assay in NSEB5-2C cells. (*B*) The ratio of AP-positive colonies induced from NSEB5-2C ($n = 3$). OSKM (without an added TF) = 1. Error bars indicate SEM. The detailed results are shown in Dataset S1. (*C Upper*) A schematic drawing of the induction experiment. (*C Lower*) Hygromycin-resistant colonies obtained by coinfection with the indicated TF sets (Dataset S1).

containing octamer-binding transcription factor 3/4 (*Oct3/4*), SRY-box containing gene 2 (*Sox2*), Kruppel-like factor 4 (gut) (*Klf4*), and myelocytomatosis oncogene (*Myc*) (OSKM). After culturing for 11 d, the *Oct3/4*-positive colonies were selected using a blasticidin-resistance gene inserted at the *Oct3/4* locus [also known as *Pou5f1* (POU domain, class 5, transcription factor 1)] and considered to be presumptive iPSC colonies (Fig. S1 *D* and *E*). The number of iPSC colonies in the control without coinfection with a TF virus was set at 1.0. The number of iPSC colonies relative to the control showed substantial variation in "interference" by the different TFs (Fig. 1*B* and Dataset S1). However, the interfering TFs in the NSEB5-2C cells did not interfere with iPSC induction in another cell type, mouse embryonic fibroblasts (MEFs) (Fig. S1*F*), suggesting the specific function of these TFs in the NSEB5-2C line.

To date, various TFs have been reported to have induction activity for cell type-specific transcriptional profiles [e.g., iN (induced neuronal cells) and iHep (induced hepatocyte-like cells)] (3). Thus, we speculated that, because these interfering TFs might substantially contribute to the NSEB-2C–specific transcriptional profile, they may be able to induce NSEB5-2C–like cells. To test this possibility, we initially used MEFs, which are amenable to the induction of heterologous transcriptional profiles (3). We coinfected retroviruses of TF set A, containing the six strongest interfering TFs [paired box gene 6 (*Pax6*); high mobility group AT-hook 2 (*Hmga2*); ets variant gene 6 (*Etv6*) (TEL oncogene); GATA zinc finger domain containing 2B (*Gatad2b*); nuclear transcription factor, X-box binding-like 1 (*Nfxl1*); and extraembryonic, spermatogenesis, homeobox 1 (*Esx1*)], in MEFs carrying the hygromycin-resistance gene at the *Sox2* locus (MEFSH), which is highly expressed in NSEB5-2C cells (Fig. S2*A* and see below). After culturing for 4 wk, we obtained hygromycin-resistant cells (MEFSH-6) at an efficiency >0.01% (Fig. S2 *B* and *C*) (a comparable efficiency with induction of similar cell type; ref. 10) that showed a FGF2/EGF dependency for proliferation and could differentiate into TuJ1-positive cells, as observed in the NSEB5-2C line (Fig. S2 *D* and *E*). In contrast, the induction of hygromycin-resistant colonies was not observed when we used TF sets B, C, or D, which contained weak interfering TFs or noninterfering TFs (i.e., they did not contain the six strongest interfering TFs) (Fig. S2*B*). Although these results suggested that the six TFs might have induced the NSEB5-2C–specific transcriptional profile, there was a possibility that the MEFSH line contained NSEB5-2C–like cells, the proliferation of which was activated by the transgenes. To rule out this possibility, we used hepatoblasts that were derived from a fetal liver primary culture and expressed various hepatic lineage-specific genes, such as *Alb* (Fig. S2*F*). The six TFs were introduced into these cells using a retrovirus with a doxycycline (Dox)-regulated expression system (Fig. S2*G*). After 4 wk (Fig. S2*H*), at an efficiency >0.01% (Fig. 1*C*), we found cells that were very similar to the NSEB5-2C cells in morphology (Fig. 2*A*) and in the ubiquitous expression of nestin (Fig. 2*A* and Fig. S2*I*). As observed in the induction of the MEFSH line, the NSEB5-2C–like cells were not induced by the introduction of the weak interfering TFs or noninterfering TFs (Fig. 1*C*). The induction efficiency was reduced when *Pax6*, *Hmga2*, or *Etv6* (or *Gatad2b*) was removed from the six TFs (Fig. 1*C*), suggesting that multiple interfering TFs synergistically acted to induce the NSEB5-2C–like cells. According to the interference scores, the three TFs necessary for the induction were significantly enriched versus other TFs

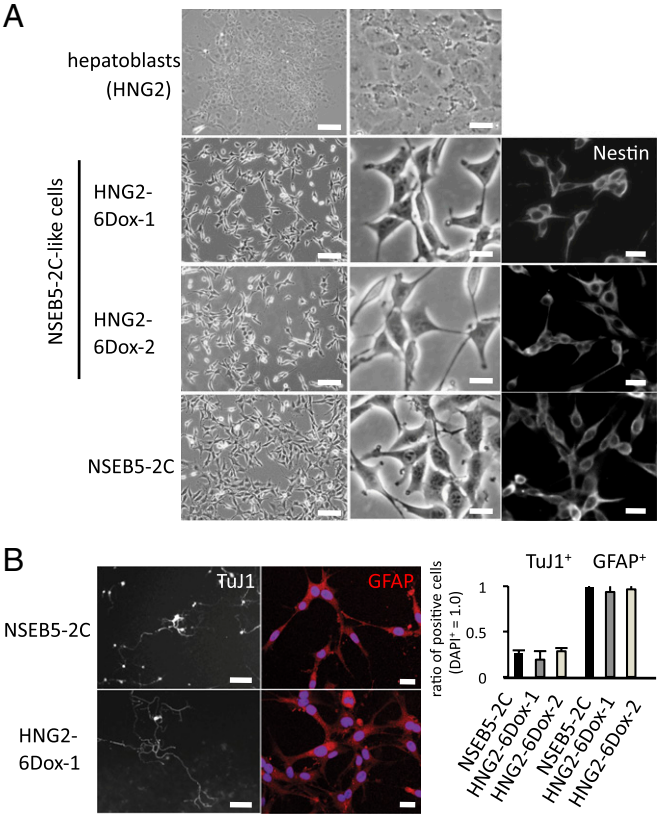

Fig. 2. NSEB5-2C–like cells have similar characteristics to that of NSEB5-2C. (*A Left*) Morphologies of NSEB5-2C-like cells. (*Middle*) Magnified view of the morphology. (*Right*) Immunostaining with a nestin antibody. (Scale bars: 200 μm in *Left*; 100 μm in *Middle* and *Right*). (*B Left*) Immunostaining of differentiated cells with indicated antibody (blue, DAPI). (Scale bars: 100 μm for TuJ1; 20 μm for GFAP.) (*Right*) Ratio of immunostaining-positive cells ($n = 3$). Number of DAPI-positive cells = 1. Error bars indicate SD.

($P < 10^{-6}$, Student t test). After the NSEB5-2C–like cells were obtained, they were characterized in the absence of Dox. Two representative clones of the induced cells (HNG2-6Dox-1 and -2) rapidly proliferated in the presence of FGF2/EGF, whereas they stopped proliferating in the absence of FGF2/EGF (Fig. S2*J*), and a population of the cells became TuJ1-positive and GFAP-positive with an efficiency comparable to that of the NSEB5-2C cells (Fig. 2*B*). We then compared the global transcriptional profiles by microarray analysis and unsupervised hierarchical clustering. We found that the transcriptional profile of four HNG2-6Dox clones clustered with that of the NSEB5-2C line (Fig. 3*A*). A principal component analysis (PCA) also showed that the HNG2-6Dox clones plotted closer to the NSEB5-2C cells than to hepatoblasts (Fig. S2*K*). To gain insight into the category of the gene set induced by the interfering TFs, we selected the genes that were up-regulated in the HNG2-6Dox clones compared with HNG2 and analyzed the enrichment of GO terms. We found an enrichment of the terms for developmental processes, including neural differentiation (Fig. 3*B*), suggesting that the induced genes were involved in the cell type specificity of NSEB5-2C cells. In fact, we found that many marker genes for neural progenitors, including *Sox2*, and endogenous genes for interfering TFs were up-regulated (Fig. 3*C* and Fig. S2*L*), whereas many hepatoblast-marker genes, including such core TFs for hepatic-lineage cells as forkhead box A2 (*Foxa2*) and hepatic nuclear factor 4, alpha (*Hnf4a*) (11, 12), were down-regulated (Fig. 3*C* and Fig. S2*L*). These results indicated that the six TFs induced an NSEB5-2C–specific transcriptional profile that was accompanied by three hallmark characteristics of NSEB5-2C–FGF2/EGF-dependent proliferation, the capability of differentiating into TuJ1- or GFAP-positive cells and the expression of marker genes for neural progenitors. These results indicated that the TFs interfering with iPSC induction in NSEB5-2C cells tended to possess the induction activity of the NSEB5-2C–specific transcriptional profile.

We next addressed the mechanism of interference in the NSEB5-2C line. First, we excluded the possibility that interference was merely an artifact of overexpression, such as a proliferation defect, which reduces the iPSC induction efficiency (13). We did not observe a difference in the proliferation rate of six NSEB5-2C lines overexpressing each of the six strongest interfering TFs (hereafter, at the same titer used in the interference assay) and five NSEB5-2C lines overexpressing each of the five noninterfering TFs (Fig. S3*A*). In addition, there was no significant difference in the correlation coefficients of the global expression between the NSEB5-2C lines overexpressing the interfering TFs and those overexpressing the noninterfering TFs (Fig. S3*B*), suggesting that the overexpression of interfering TFs did not induce a large change in the cellular state, such as differentiation, but acted neutrally in these cells. Next, to gain functional insight into the interference score and TF function, we performed a GO analysis using relatively interfering and noninterfering groups (each 60 TFs). All of the terms enriched in the interfering group were related to morphogenesis and development, whereas most of the terms enriched in the noninterfering group were involved in transcriptional processes (Fig. S3*C*). This finding suggested that the interfering TFs may tend to be involved in the regulation of cell type specificity. Although the knockdown of the interfering TF genes resulted in only a minor effect on the global expression (Fig. S3 *D* and *E*) [which may reflect redundancy or that the repression of core TFs in mature cell types often results in a minor effect on the transcriptional profile (1), possibly due to the stable nature of transcriptional network that consists of a large number of TFs (14)], the GO term enrichment showed that the repression of the interfering TF genes tended to affect the expression of genes involved in developmental processes (Fig. S3*F*). Moreover, these affected genes may also be involved in the interference with iPSC induction because the repression of the interfering TF genes increased the iPSC induction efficiency (Fig. S3*G*), as seen in a report on fibroblasts (15). These data showed that the overexpression of interfering TFs maintains the proliferation rate and global transcriptional profile of the NSEB5-2C line, whereas the repression of them enhances the iPSC induction efficiency, suggesting that the interfering TFs maintain the NSEB5-2C–specific transcriptional profile.

To address the maintenance activity of the interfering TFs further, we next focused on the gene expression changes in the initial phase of iPSC induction. We used an NSEB5-2C–derivative cell line carrying a Dox-dependent, secondary iPSC induction system, designated N31 (Fig. S4 *A*–*C*). The N31 line exhibited the three hallmark characteristics of NSEB5-2C cells (Figs. S2*L* and S4 *D* and *E*). Upon the addition of Dox, OSKM expression was induced, *Nanog* expression was gradually up-regulated, and alkaline phosphatase (AP)-positive colonies were appeared (Fig. S4 *F*–*H*). We noticed that, at 24 h after Dox addition, the N31 cells overexpressing *EGFP* as a control began to show a relatively packed morphology (Fig. 4*A*), which was consistent

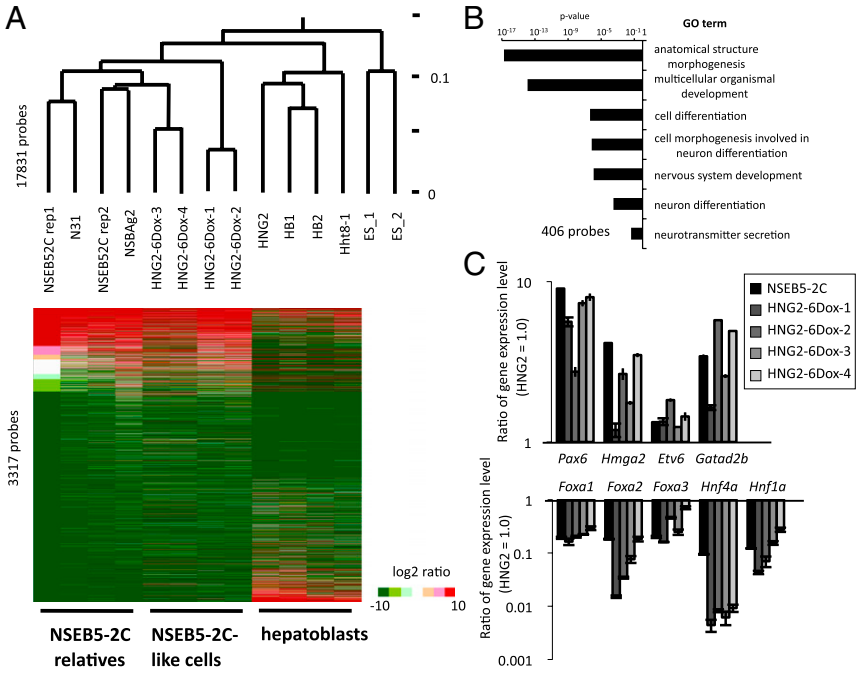

Fig. 3. NSEB5-2C–like cells have similar transcriptional profile to that of NSEB5-2C. (*A Upper*) Unsupervised hierarchical clustering. (*A Lower*) Heat map using probes showing differential expression levels between NSEB5-2C and hepatoblasts. See *SI Materials and Methods* for cell lines in each cell type. (*B*) GO terms enriched in probes up-regulated in HNG2-6Dox clones compared with hepatoblasts. Corrected *P* values are shown. (*C*) Quantitative RT-PCR (RT-qPCR) of endogenous genes for interfering TFs (*Upper*) and core TFs in hepatic lineage cells (*Lower*) ($n = 3$), compared with HNG2. Error bars indicate SD.

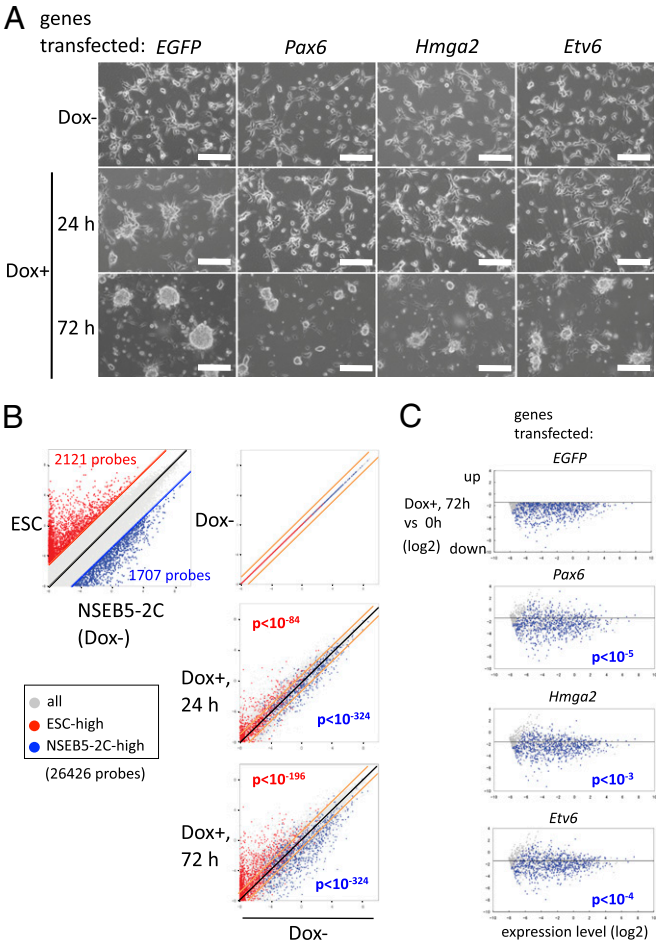

Fig. 4. Interfering TFs maintain the transcriptional profile. (*A*) Morphologies of N31 and the derivatives in the presence or absence of Dox. (Scale bars: 200 μm.) (*B*) Scatter plots of averaged gene expression of duplicated samples of N31. Red and blue lines indicate eightfold changes, whereas Green lines indicate twofold changes. Fisher's exact test. (*C*) Scatter plots of averaged gene expression of duplicated samples of N31 and the derivatives. Lines indicate threefold down-regulation. Only down-regulated probes in control cells were shown. Genes expressed higher in NSEB5-2C (more than eightfold) are highlighted in blue. Gray dots show probes down-regulated at 72 h after iPSC induction in control cells, compared with cells overexpressing EGFP at 0 h. χ^2 test.

with neural progenitors showing faster responses to iPSC induction than fibroblasts (7). A microarray analysis indicated, that within 72 h, down-regulated and up-regulated genes (more than twofold) enriched genes that are differentially expressed between NSEB5-2C and ESCs (more than eightfold) (Fig. 4*B* and Datasets S2 and S3). A similar expression pattern was also observed in another cell type, MEFs (Fig. S5*B*) (6). These expression changes were consistent with previous reports in which initial phase of iPSC induction involved down-regulation of cell type-specific genes and up-regulation of a part of ESC-specific genes (16–18). The numbers of genes expressed higher in NSEB5-2C and down-regulated at 72 h of iPSC induction were substantially reduced by overexpression of *Pax6* or *Etv6*, and modestly by *Hmga2*, accompanied by delayed the morphological change (Fig. 4 *A* and *C* and Dataset S4). Overexpression of these TFs did not significantly reduced the numbers of genes expressed higher in ESCs and up-regulated at 72 h of iPSC induction. This finding suggested that interfering TFs preferentially maintain expression of genes expressed higher in NSEB5-2C that were to be down-regulated during iPSC induction. The genes regulating cell type specificity (also known as developmental regulators) are generally repressed by polycomb repressor complexes (PRCs) in ESCs (19). ESCs lacking *Ring1A* and *Ring1B* (also known as *Ring1* and *Rnf2*, respectively; PRC1 components) derepress target genes and cannot maintain pluripotency, indicating that the PRC targets are able directly or indirectly to repress ESC-specific genes (20). The genes expressed at higher levels in NSEB5-2C cells and genes down-regulated during iPSC induction in these cells contained an enriched amount of PRC targets in ESCs (genes derepressed in ESCs lacking *Ring1A* and *Ring1B* or H3K4me3/H3K27me3-modified genes repressed by PRC2) (Fig. 5*A* and Fig. S5*A* and Dataset S3; this was also observed in MEFs, Fig. S5*C*). The PRC targets in ESCs expressed in

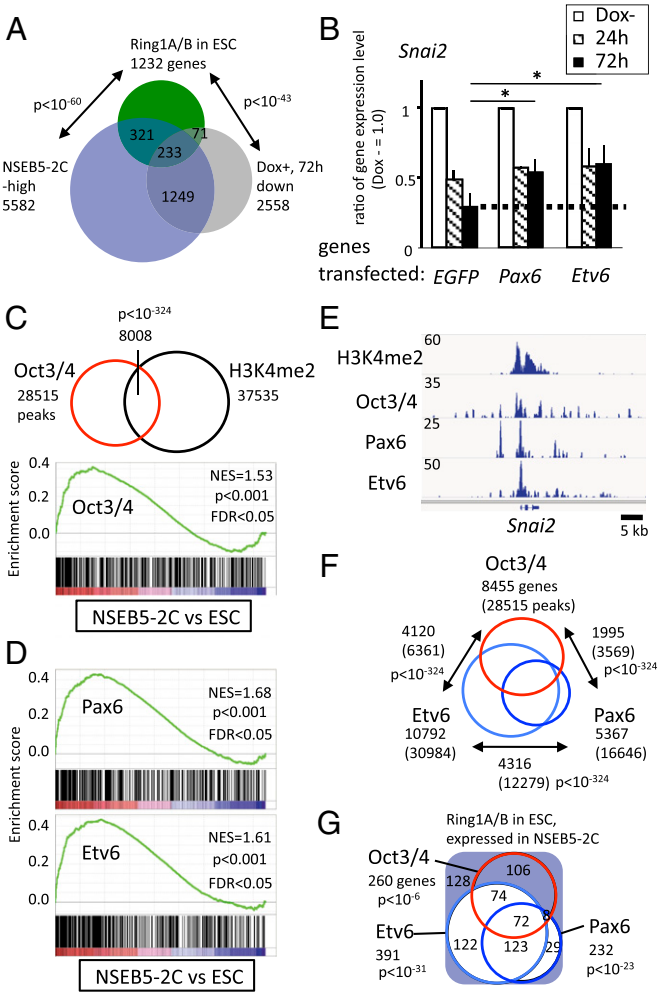

Fig. 5. Interference may involve maintained expression of developmental regulators. (*A*) Venn diagram of Ring1A/B-repressed genes in ESC that overlap with genes expressed higher in NSEB5-2C (than in ESC, greater than twofold) and with genes down-regulated at 72 h after iPSC induction. (*B*) RT-qPCR of a mesenchymal gene *Snai2* in N31 overexpressed indicated genes during iPSC induction ($n = 3$). Error bars indicate SD. *$P < 0.05$ (Student *t* test). (*C Upper*) Venn diagram showing the overlap of Oct3/4-bound regions in N31 and H3K4me2-enriched regions in N31. Fisher's exact test. (*C Lower*) Concordance between Oct3/4 occupied genes in N31 and expression level by GSEA; NES, normalized enrichment score; p, nominal *P* value; FDR, false discovery rate. Red and blue bars indicate genes expressed higher and lower in NSEB5-2C compared with ESCs. (*D*) GSEA showing concordance between Pax6 (*Upper*) or Etv6 (*Lower*) occupied genes in N31 and relative expression level among indicated cell types. (*E*) Binding profiles for H3K4me2, Oct3/4, Pax6, and Etv6, at *Snai2* loci in N31. ChIP-Seq data are shown in reads per million. (*F*) Venn diagram showing the overlap of Oct3/4-bound regions with those bound by Pax6 or Etv6. Fisher's exact test. (*G*) Venn diagram showing the overlap of Pax6, Etv6, or Oct3/4 occupied genes with those repressed by Ring1A/B in ESC, but expressed in NSEB5-2C. Fisher's exact test.

the NSEB5-2C cells contained many mesenchymal regulators, such as *Snai2* and *Twist1* (Dataset S3), whereas the NSEB5-2C cells expressed lower levels of epithelial genes compared with the ESCs (Fig. S6*A*); the expression status of these genes were oppositely regulated upon iPSC induction (i.e., approaching that of ESCs) (Fig. 5*B* and Fig. S6 *B* and *C*). This finding suggested that mesenchymal regulators repress iPSC induction through a mechanism known as epithelial-to-mesenchymal transition (EMT) in which multiple mesenchymal regulators simultaneously act to repress epithelial genes, as reported previously (6, 21). The mesenchymal regulators snail homolog 2 (Drosophila)(*Snai2*); twist basic helix-loop-helix transcription factor 1 (*Twist1*); and transforming growth factor, beta 1 (*Tgfb1*) (22) were down-regulated within 72 h of Dox addition in the control cells, whereas this down-regulation was inhibited in the cells overexpressing the interfering TFs (Fig. 5*B* and Fig. S6*B*). Conversely, the up-regulation of an epithelial gene, cadherin 1 (*Cdh1*) (also known as *E-cadherin*), during iPSC induction was repressed by the overexpression of *Pax6* or *Hmga2* (consistent with its known activity to repress epithelial genes; ref. 23) (Fig. S6*C*). Other epithelial genes were also repressed by the overexpression of the interfering TFs (Fig. S6*C*). The overexpression of *Twist1* modestly interfered with iPSC induction (Dataset S1), suggesting the involvement of MET in iPSC induction in these cells. These results indicated that the interfering TFs maintained the expression of genes expressed at higher levels in the NSEB5-2C line, including the PRC targets in ESCs, at least a portion of which need to be repressed for iPSC induction.

We next sought to gain mechanistic insight on interference using ChIP-seq analysis. We first examined Oct3/4 occupation in the NSEB5-2C cells (at 24 h after Dox addition in N31) (Dataset S5). As reported in fibroblasts (17), widespread occupation was observed (Fig. 5*C*). Gene set enrichment analyses (GSEA) indicated that Oct3/4 preferentially occupied genes that were more highly expressed in the NSEB5-2C cells than in ESCs (Fig. 5*C*). Consistently, Oct3/4-bound regions (ChIP-seq peaks) significantly overlapped with H3K4me2-enriched regions (Fig. 5*C*), which primarily mark the promoter and enhancers of active genes and, thus, are thought to be relatively accessible to exogenously expressed TFs (24, 25). The same tendency was observed using published data of fibroblasts (17) (Fig. S7*A*). In accordance with this finding, the Oct3/4 targets in ESCs (26, 27) showed enriched overlapping with genes more highly expressed in ESCs (Fig. S7 *B* and *C*), indicating that Oct3/4 occupation depended on the cellular (perhaps chromatin) context. These results suggested that Oct3/4 targets the accessible regions of genes expressed at higher levels in NSEB5-2C cells and that would be repressed in iPSCs. Pax6 and Etv6 also preferentially occupied genes that were more highly expressed in NSEB5-2C cells (Fig. 5*D* and Datasets S6 and S7). A substantial part of the respective target region (more than one third) was cobound with Oct3/4 (Fig. 5 *E* and *F*). Moreover, the target regions included the large portion of PRC targets that would be repressed in iPSCs (Fig. 5*G*). These results suggested that the interference occurred through a functional conflict on the target regions, such as *Snai2* (Fig. 5*E*). Pax6 and Etv6 regulated genes that were more highly expressed in the NSEB5-2C cells than in hepatoblasts, which included PRC targets that are substantially different from those in hepatoblasts (Fig. S7 *E* and *F*), suggesting that occupancy on PRC targets may be responsible for induction of the NSEB5-2C–specific transcriptional profile.

We interpreted the above data as follows. Oct3/4 triggers dedifferentiation by repressing genes more highly expressed in NSEB5-2C cells, including such PRC targets as mesenchymal genes that would be repressed in iPSCs, by targeting the accessible regions (Fig. S7*D*). However, by maintaining expression of genes expressed at higher levels in NSEB5-2C cells, Pax6 and Etv6 interfere with the repression. Indeed, in other cell types, the maintained expression of cell type-specific transcription factors inhibits iPSC induction (6, 8, 28), a process that can involve functional conflicts on the enhancers of PRC targets (25).

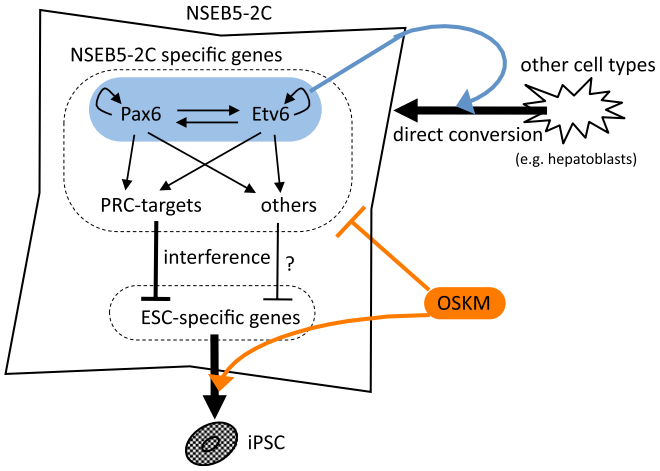

Fig. 6. Interfering TFs maintain and induce cell type-specific transcriptional profiles. Model of how Pax6 and Etv6 maintain NSEB5-2C–specific transcriptional profile and interfere with iPSC induction. Arrows in NSEB5-2C–specific genes are supported by occupation (Datasets S6 and S7).

Because PRC targets primarily regulate cell type specificity (29), Pax6 and Etv6, as identified through the maintenance of such cell type-specific genes as PRC targets, tended to have the ability to induce the cell type-specific transcriptional profile of other cell types, such as hepatoblasts (Fig. 6). Consistently, Pax6 and Etv6 preferentially occupied genes that were more highly expressed in the NSEB5-2C cells than in hepatoblasts (Fig. S7*F*). Because of the strong effect on the transcriptional profile, these two TFs might be the core TFs of NSEB5-2C cells. Indeed, the occupation pattern suggested that they were core TFs: A large portion of the Pax6-bound regions were cobound with Etv6 (Fig. 5*F*), with Pax6 bound to Etv6 and vice versa, and the two bound to the respective endogenous locus (Datasets S6 and S7). These patterns, extensive co-occupation and reciprocal- and auto-regulation, are observed in core TFs (26, 27, 30).

To examine the applicability of above scheme in another cell type, we next examined the correlation between interference with iPSC induction and induction of the cell type-specific transcriptional profile in HNG2 hepatoblasts (Fig. S7*G*), a hepatic-lineage cell type that was thought to be distant from the NSEB5-2C line (presumably neural lineage). By microarray, the TFs that were expressed at higher levels in HNG2 cells than in both MEFs and ESCs were selected using the GO terms "DNA binding" and "transcription activity," resulting in 40 TFs (Fig. S7*H* and Dataset S8). We constructed retroviruses with these 40 TFs and performed the interference assay (Fig. S7 *I* and *J*). As a result, *Foxa2*, *Hnf4a*, *Foxa3*, and *Hnf1a* (Fig. S7*J* and Dataset S8), which play pivotal roles in the maintenance and induction of the transcriptional profile of hepatic-lineage cells (31–33), were significantly enriched with respect to other TFs, as based on the interference scores ($P < 10^{-4}$, Student t test). The interfering TFs in the hepatoblasts did not show interference in the NSEB5-2C cells, and the interfering TFs in the NSEB5-2C cells also did not show interference in hepatoblasts (Fig. S7*K*), suggesting the specificity of the interference assay. We briefly examined the induction activity (for induced hepatocyte-like cells, iHep; refs. 31 and 32) of these TFs in our hands. We introduced either the two interfering TFs (*Foxa2* and *Hnf4a*, reported to be sufficient for iHep induction) (31) or the 10 strongest interfering TFs [forkhead box G1 (*Foxg*1); ligand dependent nuclear receptor corepressor (*Lcor*); *Foxa2*; *Hnf4a*; forkhead box O6 (*Foxo6*); caudal type homeobox 2 (*Cdx2*); HNF1 homeobox A (*Hnf1a*); forkhead box A3 (*Foxa3*); HNF1 homeobox B (*Hnf1b*); and *Hnf6* (also known as Onecut1, one cut domain, family member 1)] into a related NSEB5-2C line, NSBAg2, carrying a neomycin resistance gene in the alpha fetoprotein (*Afp*) locus, which is expressed higher in hepatoblasts than in NSEB5-2C cells (Figs. S2*L* and S8 *A*–*C*). As a result,

CELL BIOLOGY

neomycin-resistant cells with flattened morphology were obtained with the expression of both 2 TFs (NSBAg2-2F) and 10 TFs (NSBAg2-10F), and we also obtained similar cells (hereafter iHep) from NSEB5-2C cells using a Dox-regulated retrovirus (NSEB52C-10FDox) at an efficiency >0.01% comparable to the published data (31, 32). Similar to the iHep cells derived from fibroblasts (31), the iHep cells in our study actively proliferated while expressing Alb in the absence of Dox (Fig. S8 *D* and *E*) and showed the accumulation of a glycogen-like store (Fig. S8*H*). A microarray analysis and unsupervised hierarchical clustering indicated that the global transcriptional profile of the iHep clones clustered with that of the hepatoblasts (Fig. S8*F*). In iHep clones, many of the marker genes for NSEB5-2C cells were down-regulated, whereas many of the hepatoblast marker genes were up-regulated (Fig. S8*G*). These results indicated that, through their maintenance activity, the TFs interfering with iPSC induction in hepatic-lineage cells tended to have induction activity for a hepatic-lineage cell-specific transcriptional profile. Taken together with the previous reports (6, 8, 21, 28, 34), the correlation between interference and induction activity might be applicable in various cell types.

The TFs regulating the differentiation status form hierarchy (1), and those that lie most upstream of it (core TFs), are thought to reciprocally activate one another to organize a stable "core network" for maintaining a cell type-specific transcriptional profile (1, 26, 27, 30). Our data suggest that the stable nature of core networks inhibits dedifferentiation and induces the cell type-specific transcriptional profile, because TFs interfered with iPSC induction tend to have induction activity (i.e., the interfering TFs are core TFs or their activators). This antagonism might be partly due to the maintained expression of genes for developmental regulators that are repressed by PRCs in ES/iPSCs (19, 20). Because of this mechanism is likely to be applicable in many cell types, each differentiation status might be maintained by stable core networks.

Maintenance of differentiation status may be involved in suppressing tumorigenesis. Malignant tumors tend to express TFs for pluripotency, with enhanced self-renewal ability (35). Thus, core TFs may be able directly or indirectly to suppress tumorigenesis, as observed in Pax5 in B cells (36). Our data provide a clear example that core TFs directly or indirectly inhibit the function of TFs for pluripotency; thus, core TFs in various cell types could have similar roles in protecting against other types of cancers.

The identification of core TFs promote the progress of many studies, including direct conversion, which is expected to be an alternative way to prepare cell types for clinical use. In addition, because cancer cells may also have specific sets of core TFs to induce and maintain them (37), the identification of those TFs may potentially provide pharmaceutical targets. However, the identification of core TFs in a given cell type normally requires a priori functional information of the TFs (e.g., knockout mice), which slows the progress of research (4). Using interference of iPSC induction, our data showed that TFs with induction activity can be identified without a priori functional information, thereby providing an approach to identify those TFs systematically.

Materials and Methods

Full methods are provided in *SI Materials and Methods*. For interference assay, retrovirus mixtures were added to each well containing 2×10^4 cells. Presumptive iPSC colonies were stained using an AP kit. For induction of NSEB5-2C–like cells, 5×10^4 hepatoblasts were infected with retroviruses. N31 was established by Piggybac vectors. Primers are listed in Dataset S9. Microarray data analysis was mainly done by GeneSpring. The Gene Expression Omnibus accession number for the array data is GSE29875; the accession number for ChIP-seq data is GSE41252.

ACKNOWLEDGMENTS. We thank Kazutoshi Takahashi, Kiyoko Nashiro, Satsuki Fukuda, Shigeharu Yabe, Noriko Ohtsuki, Daiki Taneichi, Yukihito Ishizaka, and Haruhisa Inoue for their technical assistance. This work was supported in part by the PRESTO program of the Japan Science and Technology Agency. This work was also supported in part by JSPS KAKENHI Grant Numbers 20052032 and 24615004, by the Takeda Science Foundation, by the Kanae Foundation for the promotion of medical science, by The NOVARTIS Foundation (Japan) for the Promotion of Science, and by the Foundation for Promotion of Cancer Research (to S.M.).

1. Holmberg J, Perlmann T (2012) Maintaining differentiated cellular identity. *Nat Rev Genet* 13(6):429–439.
2. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663–676.
3. Sancho-Martinez I, Baek SH, Izpisua Belmonte JC (2012) Lineage conversion methodologies meet the reprogramming toolbox. *Nat Cell Biol* 14(9):892–899.
4. Cohen DE, Melton D (2011) Turning straw into gold: Directing cell fate for regenerative medicine. *Nat Rev Genet* 12(4):243–252.
5. Neph S, et al. (2012) Circuitry and dynamics of human transcription factor regulatory networks. *Cell* 150(6):1274–1286.
6. Samavarchi-Tehrani P, et al. (2010) Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 7(1):64–77.
7. Silva J, et al. (2008) Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* 6(10):e253.
8. Hanna J, et al. (2008) Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* 133(2):250–264.
9. Conti L, et al. (2005) Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol* 3(9):e283.
10. Han DW, et al. (2012) Direct reprogramming of fibroblasts into neural stem cells by defined factors. *Cell Stem Cell* 10(4):465–472.
11. Lee CS, Friedman JR, Fulmer JT, Kaestner KH (2005) The initiation of liver development is dependent on Foxa transcription factors. *Nature* 435(7044):944–947.
12. Li J, Ning G, Duncan SA (2000) Mammalian hepatocyte differentiation requires the transcription factor HNF-4alpha. *Genes Dev* 14(4):464–474.
13. Plath K, Lowry WE (2011) Progress in understanding reprogramming to the induced pluripotent state. *Nat Rev Genet* 12(4):253–265.
14. Biggin MD (2011) Animal transcription networks as highly connected, quantitative continua. *Dev Cell* 21(4):611–626.
15. Yang CS, Lopez CG, Rana TM (2011) Discovery of nonsteroidal anti-inflammatory drug and anticancer drug enhancing reprogramming and induced pluripotent stem cell generation. *Stem Cells* 29(10):1528–1536.
16. Polo JM, et al. (2012) A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151(7):1617–1632.
17. Soufi A, Donahue G, Zaret KS (2012) Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 151(5):994–1004.
18. Buganim Y, et al. (2012) Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150(6):1209–1222.
19. Boyer LA, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441(7091):349–353.
20. Endoh M, et al. (2008) Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* 135(8):1513–1524.
21. Li R, et al. (2010) A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* 7(1):51–63.
22. Thiery JP, Acloque H, Huang RY, Nieto MA (2009) Epithelial-mesenchymal transitions in development and disease. *Cell* 139(5):871–890.
23. Thuault S, et al. (2006) Transforming growth factor-beta employs HMGA2 to elicit epithelial-mesenchymal transition. *J Cell Biol* 174(2):175–183.
24. Dunham I, et al.; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
25. Taberlay PC, et al. (2011) Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* 147(6):1283–1294.
26. Kim J, Chu J, Shen X, Wang J, Orkin SH (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132(6):1049–1061.
27. Chen X, et al. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133(6):1106–1117.
28. Watanabe S, et al. (2011) MyoD gene suppression by Oct4 is required for reprogramming in myoblasts to produce induced pluripotent stem cells. *Stem Cells* 29(3):505–516.
29. Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* 469(7330):343–349.
30. Novershtern N, et al. (2011) Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144(2):296–309.
31. Sekiya S, Suzuki A (2011) Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* 475(7356):390–393.
32. Huang P, et al. (2011) Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* 475(7356):386–389.
33. Kyrmizi I, et al. (2006) Plasticity and expanding complexity of the hepatic transcription factor network during liver development. *Genes Dev* 20(16):2293–2305.
34. Onder TT, et al. (2012) Chromatin-modifying enzymes as modulators of reprogramming. *Nature* 483(7391):598–602.
35. Ben-Porath I, et al. (2008) An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* 40(5):499–507.
36. Cobaleda C, Jochum W, Busslinger M (2007) Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* 449(7161):473–477.
37. Carro MS, et al. (2010) The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463(7279):318–325.