Cell

An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming

Mathieu Gabut,[1,2] Payman Samavarchi-Tehrani,[3,5] Xinchen Wang,[1,2] Valentina Slobodeniuc,[1,2] Dave O'Hanlon,[1,2] Hoon-Ki Sung,[4] Manuel Alvarez,[2,6] Shaheynoor Talukder,[1,2] Qun Pan,[1,2] Esteban O. Mazzoni,[7] Stephane Nedelec,[7] Hynek Wichterle,[7] Knut Woltjen,[4] Timothy R. Hughes,[1,2] Peter W. Zandstra,[2,6] Andras Nagy,[4,5] Jeffrey L. Wrana,[3,5] and Benjamin J. Blencowe[1,2,5,*]

[1]Banting and Best Department of Medical Research
[2]The Donnelly Centre
University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada
[3]Center for Systems Biology, Samuel Lunenfeld Research Institute
[4]Center for Stem Cells and Tissue Engineering, Samuel Lunenfeld Research Institute
Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada
[5]Department of Molecular Genetics, University of Toronto, 1 Kings College Circle, Toronto, Ontario M5S 1A8, Canada
[6]Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Toronto, Ontario M5T 1P7, Canada
[7]Columbia University Medical Center, 630 West 168th Street, New York, NY 10032, USA
*Correspondence: b.blencowe@utoronto.ca
DOI 10.1016/j.cell.2011.08.023

SUMMARY

Alternative splicing (AS) is a key process underlying the expansion of proteomic diversity and the regulation of gene expression. Here, we identify an evolutionarily conserved embryonic stem cell (ESC)-specific AS event that changes the DNA-binding preference of the forkhead family transcription factor FOXP1. We show that the ESC-specific isoform of FOXP1 stimulates the expression of transcription factor genes required for pluripotency, including *OCT4*, *NANOG*, *NR5A2*, and *GDF3*, while concomitantly repressing genes required for ESC differentiation. This isoform also promotes the maintenance of ESC pluripotency and contributes to efficient reprogramming of somatic cells into induced pluripotent stem cells. These results reveal a pivotal role for an AS event in the regulation of pluripotency through the control of critical ESC-specific transcriptional programs.

INTRODUCTION

During the past several years, great strides have been made in our understanding of the regulatory processes responsible for maintenance of the pluripotent state of embryonic stem cells (ESCs) and for the reprogramming of somatic cells to induced pluripotent stem cells (iPSCs). A core set of transcription factors that includes Oct4, Nanog, Sox2, and Tcf3 functions in ESC maintenance, with the first three of these factors cross-regulating each other's expression, as well as genes that stabilize the ESC state (Chen et al., 2008; Kim et al., 2008; Silva et al., 2009). Indeed, exogenous Oct4 and Sox2, together with Klf4 and c-Myc, reprogram somatic cells to iPSCs (Takahashi and Yamanaka, 2006) by remodeling the transcriptome through successive stages (Samavarchi-Tehrani et al., 2010) that culminate in activation of the core pluripotency transcriptional regulatory network.

In contrast to our understanding of transcriptional networks regulating pluripotency, the role of alternative splicing (AS) in this process is not well understood. Recent studies have identified AS differences between ESC and differentiated cell populations (Atlasi et al., 2008; Kunarso et al., 2008; Pritsker et al., 2005; Rao et al., 2010b; Salomonis et al., 2010; Wu et al., 2010; Yeo et al., 2007), and two such events have been implicated in changing the activities of Tcf3 and Sall4, transcription factors that function in pluripotency (Rao et al., 2010b; Salomonis et al., 2010). Therefore, specific AS events may modulate transcriptional networks involved in pluripotency maintenance versus cell-type specification.

Forkhead box (FOX) transcription factors regulate a large number of genes involved in cell proliferation, differentiation, and development (Wijchers et al., 2006). The forkhead box forms a winged helix domain of 80 to 100 amino acids that binds to DNA (Li et al., 2004). FOXP1 is one of four FOXP subfamily members that contain a C-terminal forkhead domain together with N-terminal zinc finger and leucine zipper domains. FOXP1 is widely expressed, and its loss or fusion with other proteins through chromosomal translocations is associated with several cancers (Koon et al., 2007). Knockout of murine *Foxp1* disrupts the establishment of specific cell types (Dasen et al., 2008; Zhang et al., 2010) and results in early embryonic lethality (Wang et al., 2004). Several splice variants of FOXP1 have been identified (Brown et al., 2008), yet the functions of these are not well understood.

In this study, we identify a highly conserved AS event in FOXP1 transcripts that is activated in ESCs and silenced during cell

differentiation. This AS event modifies critical amino acid residues within the forkhead domain and alters its DNA-binding specificity. In ESCs this switches the transcriptional output of FOXP1 such that the pluripotency genes *OCT4, NANOG, GDF3*, and *NR5A2* are stimulated, while genes involved in cell-lineage specification and differentiation are repressed. Induced expression of the ESC-specific isoform of FOXP1 promotes self-renewal and the maintenance of pluripotency, whereas its silencing inhibits iPSC programming. An evolutionarily conserved AS event thus reconfigures transcriptional regulatory networks required for transitions between ESC pluripotency maintenance and differentiation.

RESULTS

An Embryonic Stem Cell-Specific Splice Variant from the *FOXP1* Gene

To identify AS events that might control stem cell pluripotency, we used microarray profiling to compare patterns of AS in undifferentiated and differentiated H9 human (h)ESCs (Extended Experimental Procedures and Figure S1 available online). Whereas few AS changes were detected at day 2, ~165 (2.85%) of the profiled exons were predicted to undergo inclusion level changes (Table S1 and data not shown) at day 10 following neural lineage induction. Genes containing these predicted AS changes were represented by diverse functional Gene Ontology (GO) categories. In this study, we focus on a previously unidentified AS change detected in transcripts from the *FOXP1* gene.

Our analysis indicated that FOXP1 exon 18 had increased inclusion in day 10 neural progenitor-enriched cells compared to undifferentiated H9 hESCs (~96% versus 79% inclusion; Table S1). Reverse-transcription-polymerase chain reaction (RT-PCR) assays confirmed this but also detected two unexpected additional bands that are ~50 nt and ~170 nt longer than the transcripts containing exon 18 (Figure 1A). Sequencing of the +50 nt band revealed the inclusion in hESCs of a previously uncharacterized exon, which we refer to as exon 18b, in FOXP1 transcripts in place of exon 18, whereas the +170 nt band (asterisk in Figure 1) contained exons 18 and 18b. Consistent with low or undetectable (see below) expression of this isoform, inclusion of both exons introduces a termination codon 121 nt downstream of exon 18b that likely elicits nonsense-mediated mRNA decay. However, inclusion of exon 18b instead of exon 18 preserves the open reading frame but modifies the forkhead domain (see below, Figure S2A).

Exon 18b is efficiently included in undifferentiated H9 hESCs (>64%, Figure 1B, lanes 1 and 2) and in H9 cells 2 days after differentiation induction (>58%, Figure 1B, lanes 3 to 5), relative to the neural lineage-enriched cell population at day 10 (11%, Figure 1B, lane 6). Consistent with this observation, a high proportion of H9 hESCs still expressed pluripotency markers at day 2 compared to day 10 post-induction of differentiation (Figure S1D). Next, we used RT-PCR to investigate exon 18 and 18b inclusion levels in another hESC line, CA1 (Figure 1B, lane7), and in a panel of partially or fully differentiated human cell lines (Figure 1B lanes 8–15; refer to legend). Similar to H9 hESCs, exon 18b was highly included in CA1 (62%), whereas exon 18 was the only exon detected in differentiated cell lines. Immunoblotting confirmed that FOXP1 protein containing exon 18b is more highly expressed in hESCs and is not expressed in differentiated cells (see below and Figure S4A). These results show that FOXP1 exon 18b switches from efficient inclusion in hESCs to almost complete skipping in differentiated cells.

To further confirm that exon 18b is specifically included in self-renewing, pluripotent ESCs, we assessed exon 18b and 18 inclusion levels in H9 cells sorted for expression of the pluripotency markers TRA1-81 and SSEA-3 (Figure 1C). Partially differentiated H9 hESCs were used, such that reduced exon 18b inclusion was present (compare Figure 1C, lane 1 with lanes 1 and 2 in Figure 1B). However, in sorted cells expressing TRA1-81 and SSEA-3, exon 18b inclusion was high (Figure 1C, lane 3), whereas only minor levels of exon 18b inclusion were detected in the TRA1-81/SSEA-3-negative population (Figure 1C, lane 2). Thus, inclusion of FOXP1 exon 18b is specific to self-renewing, pluripotent hESCs. Hereafter we refer to the exon 18b splice variant as "FOXP1-ES" and the exon 18 variant as "FOXP1."

Evolutionary Conservation of FOXP1-ES Regulation

Comparison across species reveals that human FOXP1 exons 18 and 18b are located within an ~1000 nt genomic region that is highly conserved (PhastCons mean 0.959, variance 0.029) in 46 vertebrates (Figure 1D). This region includes ~120 nt upstream of exon 18, 373 nt between exons 18 and 18b, and ~205 nt downstream of exon 18b. This observation suggests that exons 18 and 18b likely have conserved patterns of AS in diverse vertebrate species. To test this, we analyzed the regulation of the orthologous exons (exons 16 and 16b) in mouse Foxp1 transcripts (Figure 1D and Figure S2B).

The AS levels of exons 16 and 16b were analyzed in three undifferentiated mouse (m)ESCs lines, CGR8, Hb9, and R1, and following induction into different lineages (Figure 1E and Figure S2C). Similar to human cells, exon 16b displayed the highest inclusion in undifferentiated mESCs (Figure 1E, lanes 1 and 7; Figure S2C, lane 1), and its inclusion level progressively decreased when CGR8- or R1-derived embryoid bodies (EBs) were induced to form cardiomyocytes over a 14 day period (Figure 1E, lanes 3 to 6; Figure S2C, lanes 2 to 4). Furthermore, exon 16b inclusion decreased in day 14 CGR8- or R1-derived neural and glial progenitor-enriched neurospheres (Figure 1E, lane 2; Figure S2C, lane 5), or when Hb9 mESCs were induced to form motor neuron (MN) precursors (Figure 1E, lane 8), and was almost entirely skipped in sorted, differentiated MNs and in the neuroblastoma cell line Neuro2A (Figure 1E, lanes 9 and 10). Similar to human exon 18, mouse exon 16 displayed inclusion in all of the samples but at reduced levels relative to exon 16b in undifferentiated mESCs. Consistent with the high sequence conservation associated with exons 18b/16b and 18/16 and the surrounding intronic regions, these exons thus display conserved patterns of regulation.

FOXP1 and FOXP1-ES Have Distinct DNA-Binding Specificities

The forkhead domain of human FOXP1 overlaps exons 16 to 19. FOXP forkhead domains are highly homologous and bind

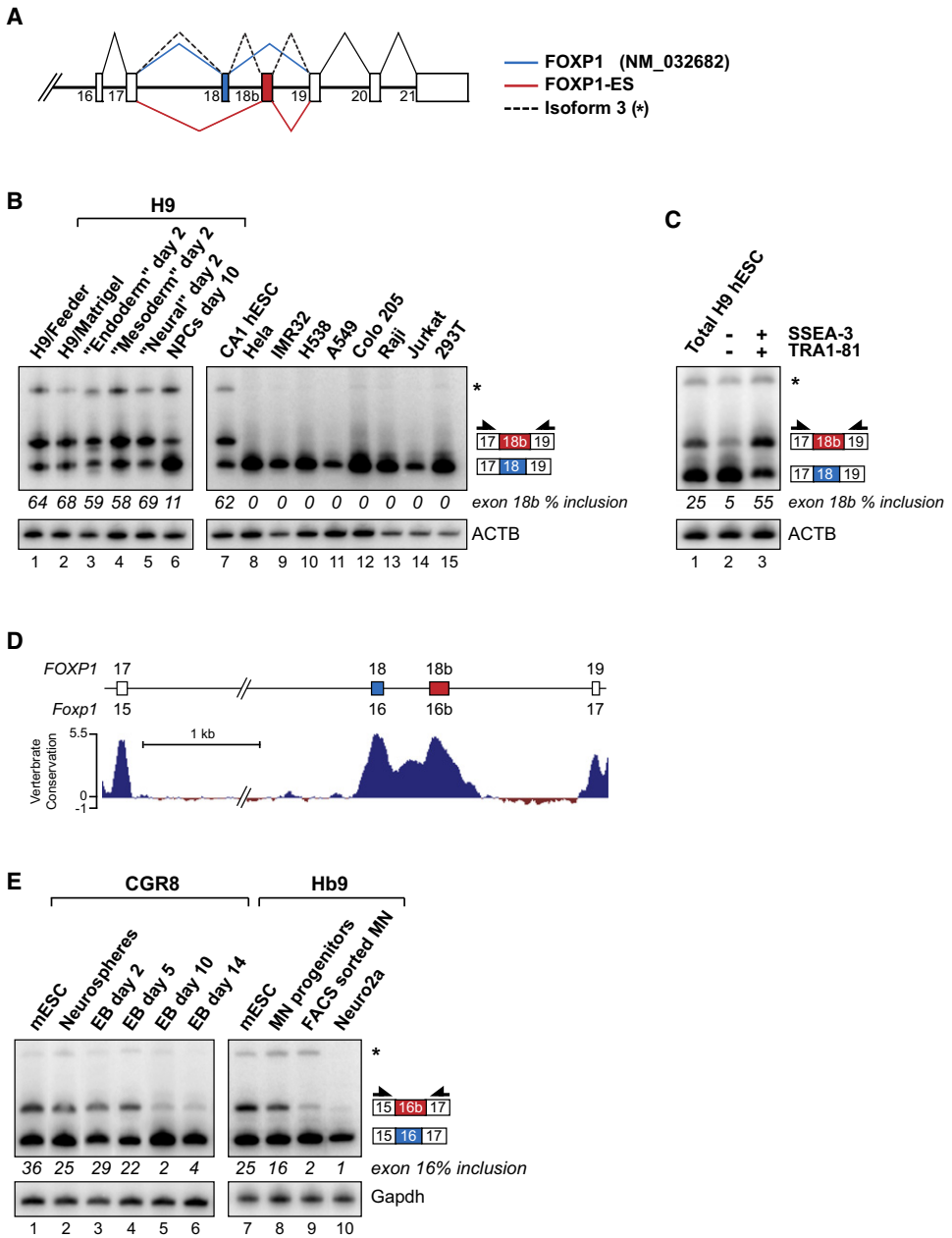

Figure 1. Identification of an Embryonic Stem Cell-Specific Splice Variant from the Human and Mouse *FOXP1*/*Foxp1* Genes

(A) Schematic representation of exons 16 to 21 of the human *FOXP1* gene. Transcripts including alternative exon 18 (blue; "FOXP1"; NM_032682) encode the widely expressed, canonical form of FOXP1, and transcripts including alternative exon 18b (red; "FOXP1-ES") are specifically detected in hESCs. Transcripts simultaneously including exons 18 and 18b (indicated by an asterisk) are predicted to be targeted by nonsense-mediated mRNA decay and are detected at low levels in hESCs. See also Figure S2A.

(B) RT-PCR assays using primers annealing to FOXP1 exons 17 and 19 (arrows) were used to analyze FOXP1 splice isoform levels in H9 hESCs grown in the presence of MEF feeder cells (lane 1) or matrigel (lane 2), H9 hESCs induced to differentiate for 2 days toward primitive endoderm (lane 3), primitive mesoderm (lane 4), neural lineages (lane 5), or neural progenitor cells (NPCs) at day 10 (lane 6) post-induction (see also Figure S1). FOXP1 splice isoforms were also analyzed in a second hESC line (CA1, lane 7) and in eight human immortalized cell lines of diverse origin as indicated in lanes 8–15. HeLa, cervical carcinoma; IMR32, neuroblastoma; H538, lung carcinoma; A549, lung adenocarcinoma; Colo 205, colorectal carcinoma; Raji, B lymphoblastoma; Jurkat, T lymphoblastoma; 293T, "embryonic kidney." *, isoform containing both exons 18 and 18b. ACTB mRNA levels are shown for comparison.

(C) RT-PCR analysis (as performed in panel B) of FOXP1 splice isoform levels in unsorted H9 hESCs (lane 1) and, following fluorescence-activated cell sorting (FACS), in H9 hESCs that are either double negative (lane 2) or double positive (lane 3) for the cell surface-expressed pluripotency markers TRA1-81 and SSEA-3. *, isoform containing both exons 18 and 18b. ACTB mRNA levels are shown for comparison.

(D) Conservation analysis of sequences surrounding FOXP1 human exons 18 and 18b (orthologous to exons 16 and 16b in mouse Foxp1) across 46 vertebrate species. The conservation plot was generated from the UCSC browser using the hg19 genome assembly. See also Figure S2B.

a canonical consensus motif GTAAACA as monomers and homo- and/or heterodimers (Koh et al., 2009; Li et al., 2004), and a FOXP2-DNA (Stroud et al., 2006) costructure reveals that residues directly contacting DNA are conserved in FOXP1 (highlighted in green in Figure 2A). Exon 18b is predicted to substitute 35 residues (highlighted in red in Figure 2A), none of which are predicted to alter secondary structure or dimerization (black dots show residues involved in dimerization in Figure 2A). However, of four residues that contact DNA in FOXP2, two (Asn510 and His514) are substituted in FOXP1-ES. These residues form critical hydrogen bonds with the adenine-thymine (A-T) base pair at the fourth position in the canonical FOXP site (underlined in GTAAACA), and their substitution may therefore affect binding affinity and/or specificity. We therefore investigated the DNA-binding properties of FOXP1 and FOXP1-ES using protein-binding microarrays (PBMs; Berger et al., 2006, 2008).

The PBM analysis revealed that FOXP1 and FOXP1-ES forkhead domains fused to glutathione S transferase (GST) preferentially recognize distinct DNA-binding motifs (Figure 2B and Figure S3A). The canonical binding motif GTAAACAA was represented by the majority of the highest-scoring GST-FOXP1-bound sequences (blue dots), whereas GST-FOXP1-ES preferentially bound CGATACAA or closely related sequences (red dots). Other sequences preferentially bound by FOXP1-ES contained specific C/A-rich motifs (orange dots), whereas other C/A-rich motifs were bound by both proteins (green dots) (Figure 2B and Figure S3A). We confirmed these binding preferences by gel mobility shift assays (Figure 2C and Figure S3B). For example, GST-FOXP1-ES, when compared to GST-FOXP1, preferentially bound dsDNA probes containing AATAAACA and CGATACAA (orange and red dots in Figure 2B, respectively), whereas GST-FOXP1 preferentially bound the consensus GTAAACAA. Furthermore, GST-FOXP1 and GST-FOXP1-ES did not bind mutant versions of each of the analyzed PBM-derived binding sites (Figure 2C and Figure S3B; mutant positions underlined).

These results show that hESC-specific inclusion of exon 18b changes the DNA-binding specificity of FOXP1. Moreover, consistent with the prediction that substitution of Asn510 and His514 in FOXP1-ES would affect recognition of the fourth A-T base pair in the consensus site, FOXP1-ES bound a T-A base pair at this position in a subset of the preferentially bound PBM sequences. Additional substitutions of conserved residues at the DNA-binding interface of FOXP1-ES presumably account for other changes in the DNA-binding properties of this splice isoform, including its preferential binding to specific C/A-rich motifs. Additionally, the results from the PBM experiments and gel mobility shift assays reveal that GST-FOXP1-ES binds a broader spectrum of sequences than does GST-FOXP1, although with apparent reduced affinity, as at similar concentration ranges GST-FOXP1-ES bound less efficiently to its high-scoring PBM sequences than did GST-FOXP1 (Figure 2 and Figure S3A). Collectively, these findings suggests that FOXP1-ES and FOXP1 direct different gene expression programs in ESCs.

FOXP1 and FOXP1-ES Regulate Distinct Programs of Gene Expression in hESCs

To investigate whether FOXP1 and FOXP1-ES control different sets of genes, we performed knockdowns using custom siRNA pools targeting either exon 18 or exon 18b in undifferentiated H9 cells, followed by RNA-Seq profiling. Relative to a control siRNA pool (Figure 3A, lane 1), each siRNA pool resulted in efficient (>80%) knockdown of only the expected FOXP1 isoform (RT-PCR in Figure 3A, lanes 2 and 3, and immunoblotting in Figure S4A, lanes 4 and 5). RNA-Seq reads from each sample were then mapped to RefSeq cDNAs to establish counts of unique-mapping reads per kb per million mapped reads (RPKM; Mortazavi et al., 2008), and genes with at least 2-fold differences were further analyzed. Knockdown of FOXP1 caused changes in expression of 153 genes, whereas FOXP1-ES knockdown was more dramatic, resulting in altered expression of 472 genes, 76 of which overlapped with the FOXP1-dependent gene set (Figure 3B; Table S2 for a full analysis). Analysis by qRT-PCR of a representative set of 19 genes with predicted changes ranging from 2- to 20-fold agreed well with the RNA-Seq-derived estimates ($r = 0.941$; Figure S4B; see below). Of the affected genes, a significantly higher proportion showed increased expression upon FOXP1-ES knockdown versus FOXP1 (86% versus 58.2%; p = 1.63E-05, Chi-square test). Moreover, of the 76 genes affected in both knockdowns (Figure 3B), 61 (80.3%) displayed increased expression. These results suggest that in undifferentiated hESCs, FOXP1 and FOXP1-ES control distinct but overlapping sets of genes, with a substantially larger set of genes controlled by FOXP1-ES compared to FOXP1 in hESCs. Moreover, FOXP1-ES predominantly acts to suppress gene expression.

A GO enrichment analysis of genes decreased upon knockdown of FOXP1 or FOXP1-ES revealed significant enrichment of terms related to early development ($p < 1.21E-05$, Figure 3C; Table S3 for a full analysis). Interestingly, a subset of the FOXP1-ES-dependent genes are involved in ESC pluripotency maintenance (see below and Figure 3D). Genes upregulated upon knockdown of FOXP1 were not significantly enriched in any GO category, whereas genes upregulated upon knockdown of FOXP1-ES were highly significantly enriched in GO annotations associated with development, transmembrane receptor activity, and cell differentiation ($p < 2.24E-06$, Figure 3C; Table S3).

qRT-PCR validation confirmed that knockdown of FOXP1-ES results in an ~2-fold or greater decrease in the expression of the pluripotency genes *OCT4, NANOG, NR5A2, GDF3*, and *TDGF1* and an ~2-fold or greater increase in expression of differentiation-associated genes including *GAS1, HESX1, SFRP4*, and

(E) RT-PCR analysis of Foxp1 splice isoforms in self-renewing CGR8 and Hb9 mouse (m)ESC lines (lanes 1 and 7), in CGR8 mESCs differentiated toward neural and glial progenitors (lane 2), in CGR8 mESCs aggregated to form embryoid bodies (EB) grown in conditions favoring differentiation into cardiomyocytes (EB days 2–10, lanes 3–5), and in beating cardiomyocytes (EB day 14, lane 6). Hb9 mESCs were differentiated into motor neuron (MN) progenitors (lane 8) and into mature MNs, which were FACS sorted (lane 9). Analysis of Neuro2a cells is shown in lane 10. *, isoform containing both exons 16 and 16b. Gapdh mRNA levels are shown for comparison. See also Figure S2C.

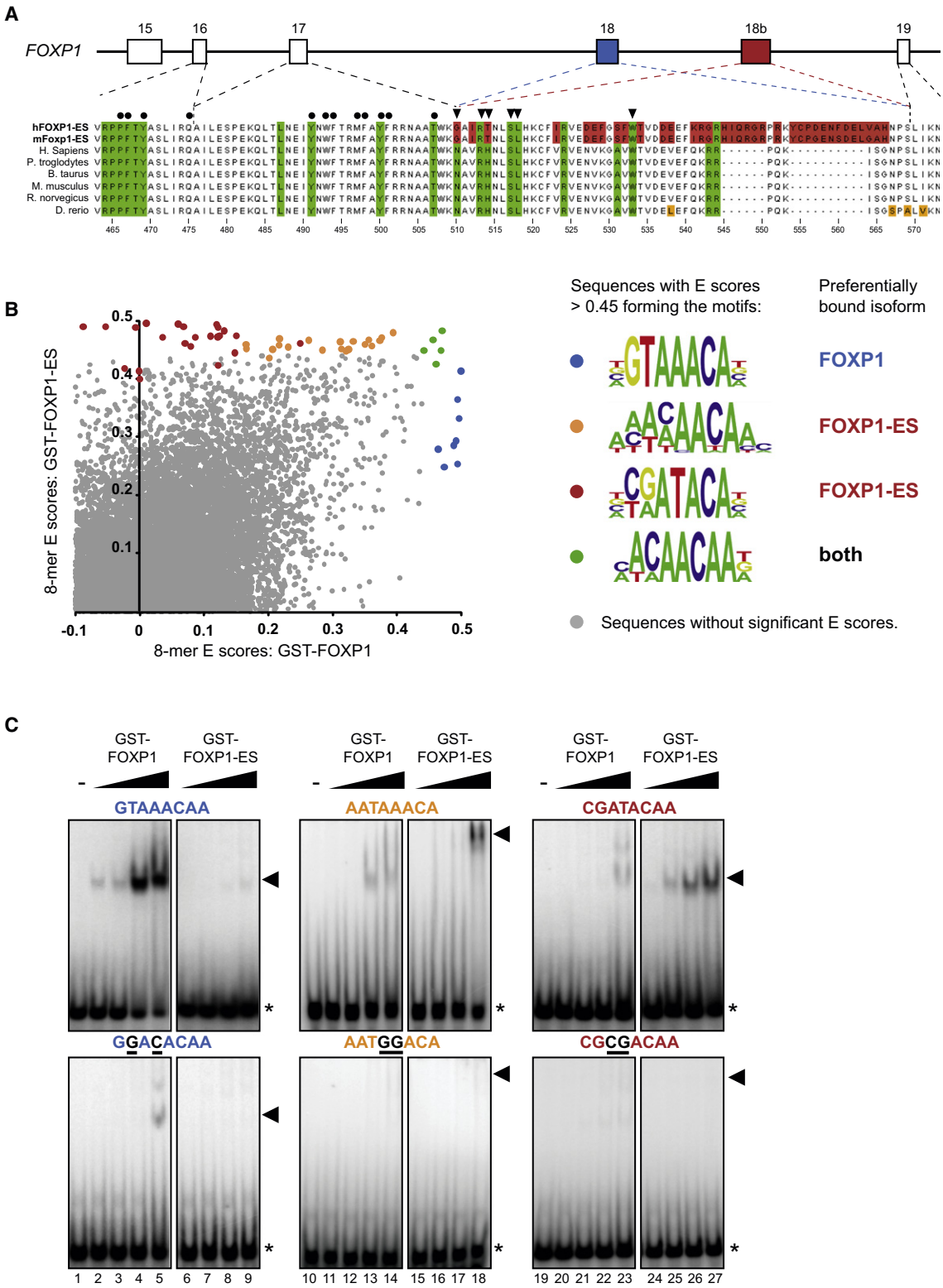

Figure 2. FOXP1-ES Has a Distinct DNA-Binding Specificity Compared to the Canonical Form of FOXP1

(A) Multiple alignment of amino acid sequences encoding the FOXP1 and FOXP1-ES forkhead DNA-binding domains from different vertebrate species. Amino acid sequences conserved across all species analyzed are indicated in white. Amino acid changes introduced as a consequence of splicing of exon 18b in FOXP1-ES are highlighted in red. Species-specific amino acid differences occurring in the canonical form of FOXP1 are highlighted in orange. Amino acids

WNT1 (Figure 3D). Several other genes that function in pluripotency maintenance and reprogramming, including *KLF4, KLF5, SOX2, C-MYC, ZSCAN10, ESRRB, REXO1*, and *TBX3*, displayed negligible or less pronounced changes in mRNA expression upon FOXP1-ES knockdown (Figure S4C and data not shown), indicating that the decreases in *OCT4*, *NANOG*, *NR5A2*, *GDF3*, and *TDGF1* expression are a specific consequence of reduced FOXP1-ES rather than an indirect effect arising from induction of differentiation. Further, consistent with the RNA-Seq analysis, knockdown of FOXP1 resulted in negligible (<1.5-fold) changes in the expression levels of these and many other FOXP1-ES-regulated genes (Figure 3D and Figure S4C).

These results thus provide evidence that expression of FOXP1-ES in hESCs suppresses a large number of genes with important functions in cell differentiation and development, while promoting the expression of a specific subset of genes that support pluripotency.

Direct Binding of FOXP1-ES and FOXP1 to Regulated Target Genes

We next performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) to identify genes that are potentially directly regulated by FOXP1-ES and FOXP1 in H9 ESCs. Using an antibody that efficiently immunoprecipitates both isoforms, >3,400 significant ChIP-Seq peaks were detected across the human genome (Tables S4A and S4B). To assess whether these peaks are sites of FOXP1 and FOXP1-ES occupancy, we determined whether they are significantly enriched in individual PBM-derived 8-mers (Figure 2) that bind to either or both isoforms.

Scatterplots directly comparing the relative under-peak enrichment and PBM scores for individual 8-mers are shown in Figure 4A (see Tables S5A and S5B for a full analysis). PBM 8-mers that bind preferentially to FOXP1-ES (orange dots) or FOXP1 (blue dots), and other 8-mers that bind to both proteins (green dots), are significantly enriched under the ChIP-Seq peaks. In contrast, the CGATACA consensus and closely related sequences preferentially bound by FOXP1-ES in vitro do not appear to be widely utilized by this factor in vivo (Figure 4A). Previous studies have revealed examples of transcription factors that preferentially bind lower-affinity sites in vivo (Jaeger et al., 2010; Rowan et al., 2010), and this property may be important to facilitate dynamic changes in transcriptional output mediated by FOXP1-ES and FOXP1 upon induction of ESC differentiation.

Together with the analysis of genes differentially expressed upon knockdown of FOXP1 and FOXP1-ES (Table S2), the ChIP-Seq analysis provides a list of 116 candidate direct in vivo targets of these proteins (Table S5C for a full list). GO enrichment analysis showed that these target genes are significantly enriched in terms associated with early development ($p < 7.4E-11$) and cell differentiation ($p < 1.7E-06$) (Table S5D). Examples of such direct target candidate genes are shown in Figure 4B. Importantly, the data support *OCT4* and *NANOG* as possible direct targets of FOXP1-ES, as the promoters of these genes are proximal to peaks containing 8-mers that preferentially bind FOXP1-ES in vitro.

Further supporting a possible direct role for FOXP1-ES in regulating OCT4, we observed a statistically significant overlap between the RNA-Seq-profiled genes that are dependent on FOXP1-ES for expression in H9 hESCs (Figure 3D) and a set of genes previously reported (Kunarso et al., 2010) to be both directly bound and regulated by OCT4 in H1 hESCs (Figure 4C; $p = 0.0016$, Chi-square test). The majority (26/33) of these overlapping genes show changes in the same direction upon knockdown of either factor (data not shown). In contrast, genes stimulated or repressed by FOXP1 in H9 hESCs did not significantly overlap with the previously reported OCT4 target genes (Figure 4C). Collectively, the results suggest that FOXP1-ES may regulate ESC self-renewal and pluripotency maintenance by directly controlling the expression of a subset of key pluripotency genes.

Foxp1-ES Expression Promotes mESC Self-Renewal and Pluripotency

Mouse Foxp1-ES, like human FOXP1-ES, is specifically expressed in mESCs and stimulates the expression of Oct4 and Nanog (Figure 1 and data not shown). Therefore, we hypothesized that Foxp1-ES is required for mESC self-renewal and pluripotency. To test this, we first asked whether ectopic Foxp1-ES expression suppresses mESC differentiation. CGR8 mESC lines stably expressing 3×Flag-Foxp1-ES or 3×Flag-Foxp1 isoforms at levels comparable to endogenous protein and under Doxycycline (Dox)-inducible control (Figure S5A)

predicted to contact DNA (based on the cocrystal structure of FOXP2 bound to its recognition site; Stroud et al., 2006) are indicated in green, residues that are the most highly conserved across Forkhead protein family members are indicated by black arrowheads above the alignment, and residues indicated with a black dot are involved in dimerization.

(B) Protein-binding microarray (PBM) analysis of the DNA-binding preferences of GST-FOXP1 and GST-FOXP1-ES forkead domain fusion proteins. Relative binding preferences measured as anti-GST fluorescence signal intensity are represented as "E scores" (Berger et al., 2006). The scatterplot directly compares E scores for GST-FOXP1 and GST-FOXP1-ES, after averaging data from two independent experiments. Sequences of probes with E scores > 0.45 in at least one of the two repeat experiments were clustered to derive consensus binding sites. Blue dots represent all probe sequences containing the consensus GTAAACA, which is preferentially bound by GST-FOXP1; red dots represent all probe sequences containing the consensus sequences CGATACA, CAATACA or TGATACA, which are preferentially bound by GST-FOXP1-ES; orange and green dots represent C/A-rich motifs preferentially bound by GST-FOXP1-ES or by both isoforms, respectively; gray dots indicate all other probe sequences with E scores < 0.45. A full version of the scatterplot is shown in Figure S3A.

(C) Electrophoretic mobility shift assay (EMSA) validating PBM-derived consensus DNA-binding sites for FOXP1 and FOXP1-ES. Radiolabeled dsDNA probes containing two copies of GTAAACAA (top left panel), AATAAACA (top middle panel), or CGATACAA (top right panel) or two copies of mutant versions of these sequences (bottom panels) were incubated in the absence (lanes 1, 10, and 19) or in the presence of increasing amounts (0.2 to 3.2 pmol) of recombinant GST-FOXP1 or GST-FOXP1-ES proteins, as indicated. Positions mutated in the probe sequences are highlighted in black and underlined. Shifted protein-dsDNA complexes are indicated by arrows, and free dsDNA probe is indicated by an asterisk. Additional EMSA experiments assaying other PBM-derived preferred binding sites for FOXP1 and FOXP1-ES are shown in Figure S3B.

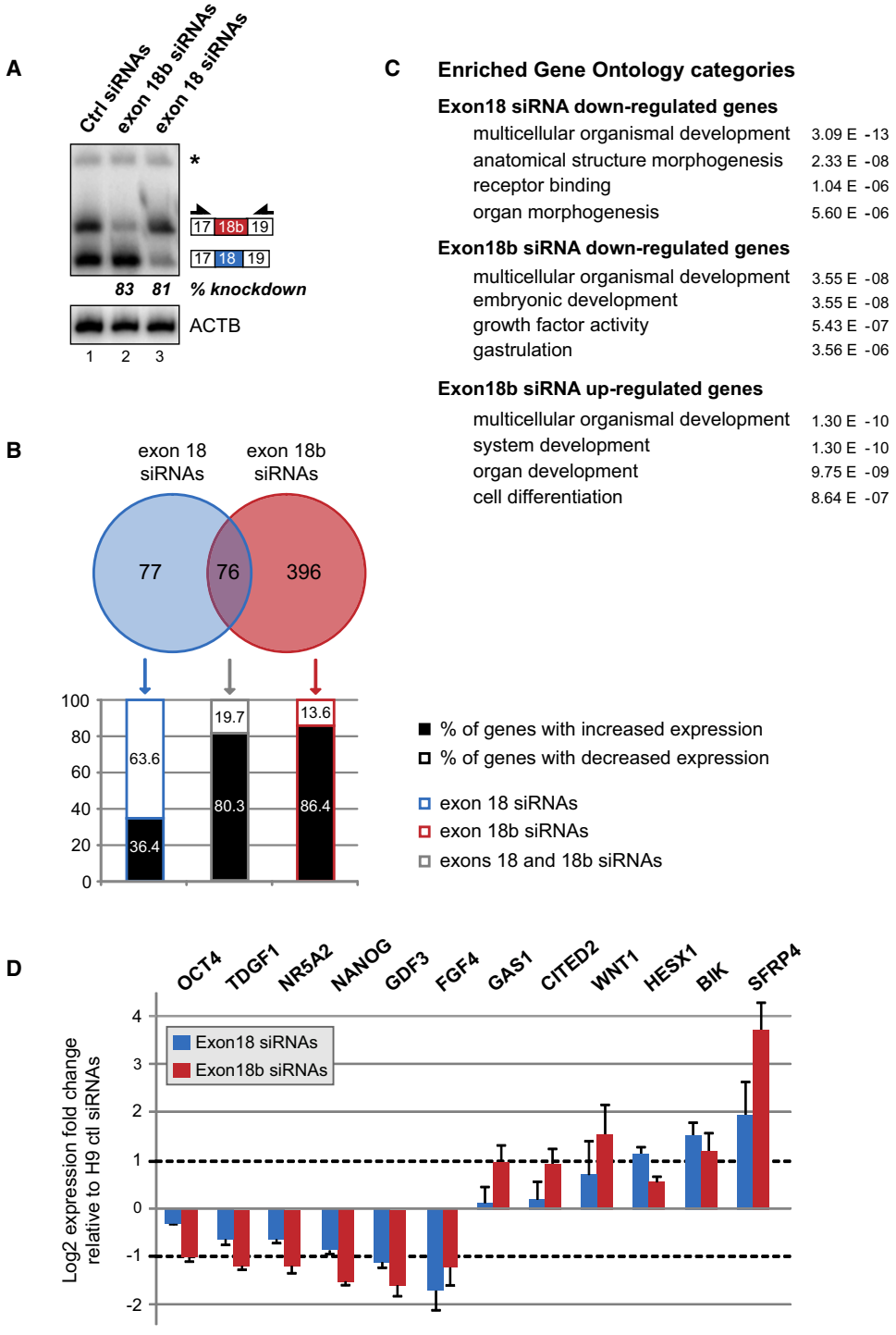

Figure 3. Knockdown of FOXP1 and FOXP1-ES Affects the Expression of Distinct Sets of Genes in hESCs

(A) RT-PCR analysis of FOXP1 and FOXP1-ES splice isoforms in H9 hESCs transfected with a control, nontargeting siRNA pool (lane 1), an siRNA pool targeting exon 18b (lane 2), and an siRNA pool targeting exon 18 (lane 3). ACTB mRNA levels are shown as a loading/recovery control. The corresponding western blot analysis is shown in Figure S4A.

(B) Top: Venn diagram showing numbers of genes with estimated 2-fold to 10.8-fold transcript level changes between the FOXP1 (blue circle) or FOXP1-ES (red circle) knockdowns and the control knockdown samples shown in (A). Bottom: Bar graph showing proportions of genes with up- (black fill) or downregulation (white fill) in the gene sets affected by siRNA knockdown of exon 18- or exon 18b-containing transcripts. Genes with transcript changes affected in both knockdowns are also indicated (bar with gray outline).

were aggregated to form EBs under conditions that favor neural cell differentiation. In the absence of Dox, all three cell lines supported neural differentiation, as revealed by the appearance of cells with neuronal morphology that immunostained with an antibody to the neuronal marker β-III tubulin (Figures 5Aa, 5Ac, and 5Ae). β-III tubulin-positive neurons were also observed in the control line and in 3×Flag-Foxp1-expressing cells after Dox stimulation (Figures 5Ab and 5Ad, respectively). In marked contrast, overexpression of 3×Flag-Foxp1-ES almost completely abolished neural cell differentiation (Figure 5Af), and only the 3×Flag-Foxp1-ES cells showed prominent Oct4 immunostaining (compare Figure 5Al with Figures 5Ag–5Ak). Furthermore, knockdown of Foxp1 did not significantly impact proliferation, whereas knockdown of Foxp1-ES reduced formation of CGR8 mESC colonies by ~3-fold (Figures S5B–S5D). Altogether, these results provide evidence that Foxp1 promotes mESC differentiation, whereas expression of Foxp1-ES prevents differentiation and is required for mESC self-renewal.

To further establish whether Foxp1-ES expression is required for the maintenance of stem cell identity, the 3×Flag-Foxp1- or 3×Flag-Foxp1-ES-expressing CGR8 cell lines were cultured in the presence of different amounts of leukemia inhibitory factor (LIF), a cytokine that is required for pluripotency maintenance of mESCs. Both cell lines were cultured with excess LIF, which supports mESC self-renewal (Figure 5B, LIF 1:1, continuous lines), or with 10% of this amount, which is insufficient to prevent cell differentiation (Figure 5B, LIF 1:10, dashed lines). In the absence of Dox, reduced LIF led to a decrease (~50% of total) in the number of Oct4-positive cells after four cell passages (Figure 5B, right panels) and reduced cell division rates (Figure 5B, white dashed lines in left panels). However, Dox-induced overexpression of either Foxp1 isoform in the presence of standard LIF concentrations resulted in increased rates of cell division (Figure 5B, solid black lines), and the majority (>80%) of cells remained Oct4 positive. In contrast, in reduced LIF, 3×Flag-Foxp1 expression did not prevent cell differentiation, with cell division declining after two passages, and only ~40% of the cells remaining Oct4 positive after four passages (Figure 5B, blue panel, black dashed lines). Under the same reduced LIF conditions, overexpression of 3×Flag-Foxp1-ES prevented loss of pluripotency characteristics, as more than 90% of the cells remained Oct4 positive after four passages and cell division rates were comparable to controls grown in standard amounts of LIF (Figure 5B, red panel, black dashed lines).

To further assess whether Foxp1-ES but not Foxp1 promotes pluripotency, we cultured the 3×Flag-Foxp1- and 3×Flag-Foxp1-ES-expressing cell lines in the absence of exogenous LIF, with or without Dox. As expected, the two cell lines rapidly differentiated in the absence of Dox, and the 3×Flag-Foxp1 line could not be maintained in culture beyond five or six passages even in the presence of Dox (data not shown). Strikingly, the 3×Flag-Foxp1-ES-expressing cells continued to grow for over 30 passages in the absence of LIF. We refer to these cells as 3×Flag-Foxp1-ESΔLIF. qRT-PCR analysis confirmed that these cells express Oct4, Nanog, and Nr5a2 at levels comparable to the parental CGR8 cells, but they display reduced levels of Sox2, Klf4, and LifR (Figure 5C). The 3×Flag-Foxp1-ESΔLIF cells were then aggregated to form EBs and cultured under conditions that favor neural differentiation. As before, in absence of Dox, the cells adopted neuronal morphology, expressed β-III tubulin, and displayed negligible Oct4 expression (Figure S5E, left panel). Finally, when injected subcutaneously in mice, the 3×Flag-Foxp1-ESΔLIF CGR8 mESCs formed teratomas that reproduce all three germ cell types in vivo (Figure 5D and Figure S5F). Collectively, these results support the conclusion that increased expression of Foxp1-ES, but not of Foxp1, promotes the maintenance of CGR8 mESCs in a pluripotent state.

Foxp1-ES Is Required for Efficient iPSC Formation

We next asked whether Foxp1-ES expression is important for the formation of iPSCs from mouse embryonic fibroblasts (MEFs). For this experiment, secondary mouse MEFs (2°–6C MEFs) were employed that contained integrated *piggyBac* transposons expressing, under Dox-inducible control, the four Yamanaka transcription factors "OKMS" (Oct4, Klf4, c-Myc, and Sox2) required for iPSC reprogramming (Takahashi and Yamanaka, 2006; Woltjen et al., 2009). In the presence of Dox, the 2°–6C MEFs efficiently form secondary iPSCs (2°–6C iPSCs) that are pluripotent (Woltjen et al., 2009). Consistent with a key role for Foxp1-ES in the maintenance of mESC pluripotency, RT-PCR assays showed that Foxp1 exon 16b is almost completely skipped in primary MEFs but is included to ~32% in iPSCs, which is comparable with its inclusion level in mESCs (Figure 6A and Figure S6A, lanes 1 and 2). During 2°–6C MEF reprogramming, Foxp1 exon 16 is predominantly included at early stages but displays progressively decreased inclusion toward the end of reprogramming (compare days 2–21 in Figures 6A and 6B and Figure S6A). Conversely, exon 16b is weakly included (<4%) at the earliest stages of reprogramming (lanes 3–5) but is efficiently included at later stages (days 5–16), reaching the highest level of inclusion (~37%) in 2°–6C iPSCs (Figures 6A and 6B and Figure S6A).

We next investigated whether Foxp1 and Foxp1-ES are important for iPSC formation. Each isoform was selectively knocked down using siRNA pools specific for either exon 16b or exon 16 sequences, and siRNAs comprising these pools that produced the most efficient isoform-specific knockdown (Figure S6B) were then used in pairs to validate results. Each

(C) Gene Ontology (GO) annotations enrichment analysis performed on sets of genes displaying increased or decreased transcript levels following siRNA knockdown of exon 18- and exon 18b-containing splice isoforms. The top four most enriched annotations are shown for each gene set with corresponding p values, corrected using the Benjamini false-discovery rate. The full analysis is shown in Table S3.

(D) qRT-PCR assays validating RNA-Seq predictions (see Table S2) of ~2-fold or greater changes in transcript levels from the pluripotency-associated genes (*OCT4*, *TDGF1*, *NR5A2*, *NANOG*, *GDF3*, and *FGF4*) and differentiation-associated genes (*GAS1*, *CITED2*, *WNT1*, *HESX1*, *BIK*, and *SFRP4*) following siRNA knockdown of FOXP1-ES and FOXP1 in H9 hESCs. Measurements are relative to levels detected with a control siRNA pool and represent averages from three independent analyses; standard deviations (SDs) are indicated. See also Figures S4B and S4C.

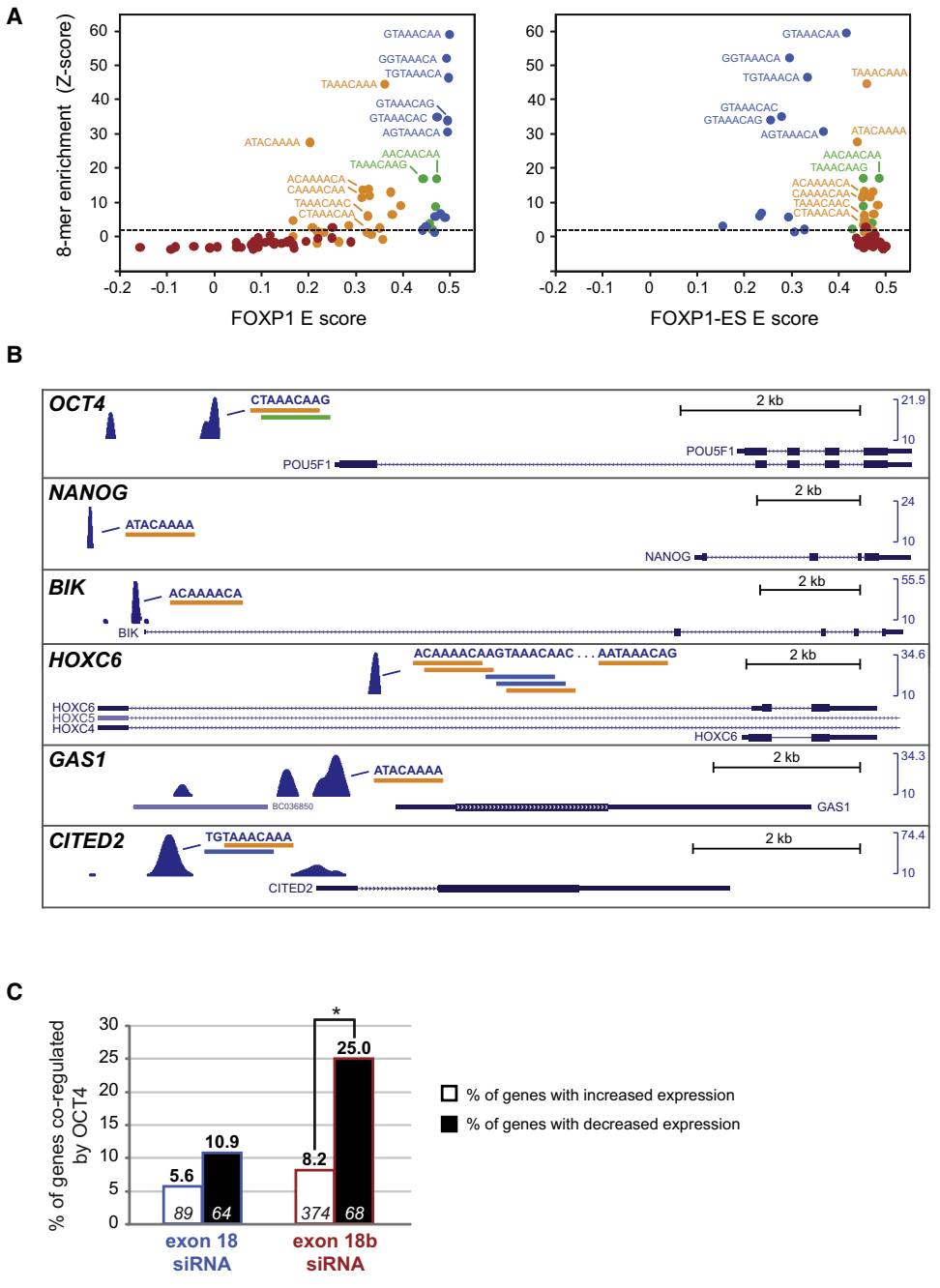

Figure 4. Chromatin Immunoprecipitation and High-Throughput Sequencing Analysis of FOXP1/FOXP1-ES Target Genes in hESCs

(A) ChIP-Seq analysis in H9 hESCs was performed using a pan-FOXP1 isoform-specific antibody. The scatterplots compare relative enrichment scores for PBM-derived FOXP1 and FOXP1-ES 8-mer binding sequences under ChIP-Seq peaks and PBM-derived binding strengths. Z scores were calculated by counting motif occurrences in peak sequences relative to occurrences after randomizing the same peak sequences 100,000 times. PBM 8-mer sequences that bind preferentially to FOXP1, FOXP1-ES, or both proteins are colored as in Figure 2B. See Table S5A for a full analysis.

(B) Representative tracks showing locations of FOXP1/FOXP1-ES ChIP-Seq peaks proximal (± 20 kb of the transcription start site) to genes that display an ~2-fold or greater change in mRNA expression upon knockdown of FOXP1 isoforms. See also Table S4 and Table S5.

(C) Bar graph representing the percentage of genes up- or downregulated in response to FOXP1 or FOXP1-ES siRNA knockdown in H9 hESCs, which are experimentally supported (based on combined ChIP and knockdown-expression analysis; Kunarso et al., 2010) targets of OCT4. OCT4 target genes only significantly overlap those genes showing decreased but not increased expression following knockdown of FOXP1-ES ($p = 0.0016$; Chi-square test).

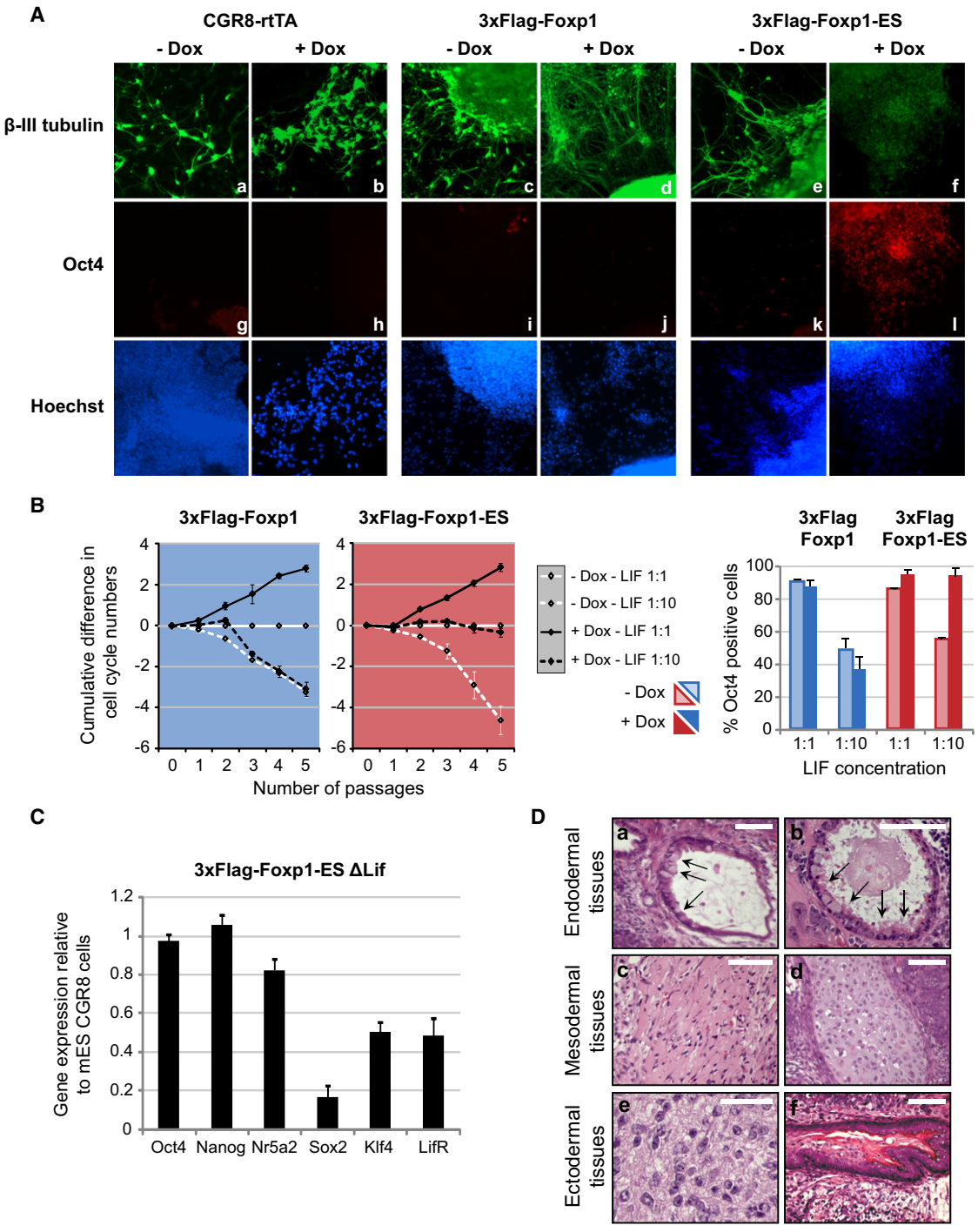

Figure 5. Expression of Foxp1-ES but Not Foxp1 Promotes Pluripotency Maintenance of mESCs

(A) CGR8 mESC lines expressing 3×Flag-Foxp1 or 3×Flag-Foxp1-ES under Doxycycline (Dox)-inducible control (see Figure S5A) and the parental line used to generate these two cell lines (CGR8-rtTA) were aggregated to form embryoid bodies (EBs) and then cultured under conditions promoting neural differentiation. The cultured EBs were treated with or without Dox and then immunostained for β-III tubulin (neural marker) or Oct4 (pluripotency marker). Nuclei were stained with Hoechst.

(B) Quantification of CGR8 mESC proliferation in response to Dox-induced expression of 3×Flag-Foxp1 or 3×Flag-Foxp1-ES in the presence of excess LIF (LIF 1:1), which promotes mESC self-renewal, or in the presence of concentrations of LIF that are insufficient for promoting mESC self-renewal (LIF 1:10). Left panels show cell growth rates calculated as the cumulative difference in cell-cycle numbers relative to the control condition (LIF1:1) without Dox-induced expression of the 3×Flag-Foxp1(-ES) transgenes. Right panel: Quantification of the proportions of cells expressing Oct4 under the different growth conditions indicated after four cell passages. Quantifications represent four independent analyses and SDs are indicated.

isoform-specific pool resulted in a selective reduction (>65%) of only its respective target mRNA isoform (Figure 6C). Subsequently, siRNAs were transfected into 2°–6C MEFs at day 0 or day 13 during reprogramming, then harvested 5 and 3 days later, respectively. Reprogramming colonies were either collected and analyzed by flow cytometry to quantify cells positive for both SSEA-1 and GFP, which marks the reprogramming population (Figure 6D), or fixed and imaged by confocal microscopy (Figure 6E). Although not all SSEA-1-expressing cells eventually progress to iPSCs, SSEA-1 expression during the initiation phase of reprograming made it ideal for assessing the early effects of knockdown of Foxp1-ES and Foxp1. This also afforded more reliable quantification of reprogramming initiation compared to markers such as Nanog, which are expressed at later stages (data not shown). As expected, in the absence of OKMS expression, no SSEA-1/GFP-positive cells formed compared to Dox-induced cultures, which showed robust initiation of SSEA-1 expression. However, transfection of 2°–6C MEFs with siRNAs targeting Oct4 reduced the population of SSEA-1/GFP-positive cells by ~5 fold (Figures 6D and 6E). Importantly, knockdown of Foxp1-ES resulted in a comparable (~4-fold) reduction in SSEA-1/GFP-positive cells, whereas knockdown of Foxp1 had little to no effect (Figures 6D and 6E). Knockdown of Foxp1-ES (or Oct4) at day 13 also significantly reduced the proportion of SSEA-1/GFP-positive cells (Figure S6C). Finally, we asked whether overexpression of either Foxp1-ES or Foxp1, together with OKMS, differentially affects primary MEF reprogramming. Although overexpression of Foxp1-ES with OKMS factors did not substantially alter the efficiency of formation of SSEA-1-positive colonies, overexpression of Foxp1 completely blocked OKMS induction of AP- and SSEA-1-positive colonies (Figure S6D; Extended Experimental Procedures and data not shown).

Taken together with the results described earlier, these data provide evidence that the AS-mediated switch controlling Foxp1-ES expression is critical for efficient iPSC formation, as well as for the maintenance of ESC self-renewal and pluripotency.

DISCUSSION

Previous investigations of gene regulatory networks that control ESC self-renewal and pluripotency and iPSC reprogramming have largely focused on the roles of transcription factors, chromatin remodeling, and noncoding RNAs in these processes. An important aspect of our findings is the observation that an AS switch controlling the expression of the FOXP1-ES splice isoform is integral to the control of the highly interconnected transcriptional regulatory network required for ESC pluripotency and iPSC reprogramming (Figure 7; refer to Introduction). This regulatory paradigm is reminiscent of AS events with pivotal roles in the control of transcription factors involved in *Drosophila* sex determination, courtship behavior, and eye development (Demir and Dickson, 2005; Fic et al., 2007; Förch and Valcárcel, 2003). Moreover, additional AS events have recently been reported to influence the activity of transcription or signaling factors implicated in the control of pluripotency genes (Mayshar et al., 2008; Rao et al., 2010b; Salomonis et al., 2010; see Introduction). Thus, a small number of AS events have the capacity to dramatically impact the wiring of transcriptional networks and other processes with critical regulatory functions in pluripotency and early development.

Our results extend recent reports establishing critical roles for FOXP1/Foxp1 in the specification of cell lineages in early development. Foxp1 has been reported to coordinate the balance between cardiomyocyte proliferation and differentiation through lineage-specific regulation of Fgf ligands and the Hox protein Nkx2.5 (Zhang et al., 2010) and to promote midbrain identity in mESC-derived dopamine neurons through direct regulation of the homeobox protein Pitx3 (Konstantoulas et al., 2010). It also coordinates the expression of other Hox proteins required for columnar organization of spinal motor neurons (Rousso et al., 2008). Interestingly, FOXP1 together with several other transcription factors has been reported to promote the self-renewal and differentiation potential of mesenchymal stem cells (Kubo et al., 2009), and it has also been implicated in the transition between pro- and pre-B cells during B cell maturation (Hu et al., 2006; Rao et al., 2010a). Taken together with our findings, it is apparent that the differential regulation of FOXP1/Foxp1 and its isoforms can have a profound impact on transitions between cell proliferation, lineage specification, and differentiation in multiple biological contexts.

In future studies, it will be of considerable interest to elucidate the mechanisms responsible for the regulation of FOXP1-ES expression. In particular, it will be important to establish which splicing factors control the inclusion of FOXP1/Foxp1 exons 18/16 and 18b/16b, and how these factors themselves are differentially regulated in ESCs and differentiated cells, so as to govern the transcriptional networks that regulate ESC self-renewal and pluripotency.

EXPERIMENTAL PROCEDURES

Microarray Hybridization, Data Extraction, and Analysis

Total RNA was extracted from ESCs using TRI reagent (Sigma-Aldrich) as per the manufacturer's recommendations. Poly(A)$^+$ mRNA was purified using Nucleotrap Midiprep kits (Clonetech). cDNA was synthesized using the WT-Ovation RNA Amplification System (Nugen) and was hybridized to custom AS microarrays as described previously (Pan et al., 2004). Data analysis was performed essentially as described previously (Pan et al., 2004) (S. Mavadadi, J. Calarco, X.W., B.J.B., Q.P., and Q. Morris, unpublished data).

(C) qRT-PCR analysis of transcript expression from genes involved in pluripotency maintenance in Dox-treated CGR8 mESCs expressing 3×Flag-Foxp1-ES and grown in absence of LIF (ΔLIF) for 24 passages. Average expression levels of *Oct4*, *Nanog*, *Nr5a2*, *Sox2*, *Klf4*, and *LifR* in CGR8 3×Flag-Foxp1-ES ΔLIF cells are shown relative to the average expression levels of the same genes in the parental CGR8 mESCs, cultured in parallel in the presence of 1:1 LIF. Expression ratios are average measurements from three independent analyses; positive SDs are indicated.

(D) Teratoma assay assessing the pluripotency potential of mouse CGR8 3×Flag-Foxp1-ES ΔLIF cells (see panel C). Hematoxylin and eosin staining of teratoma sections detected all three embryonic germ layer-derived tissues. Endodermal derivatives: ciliated respiratory (a) and intestine-like epithelium (b); mesodermal derivatives: muscle (c) and cartilage (d); ectodermal derivatives: neuronal (e) and skin epithelial cell (f).

Bar = 50 μm. See also Figures S5E and S5F.

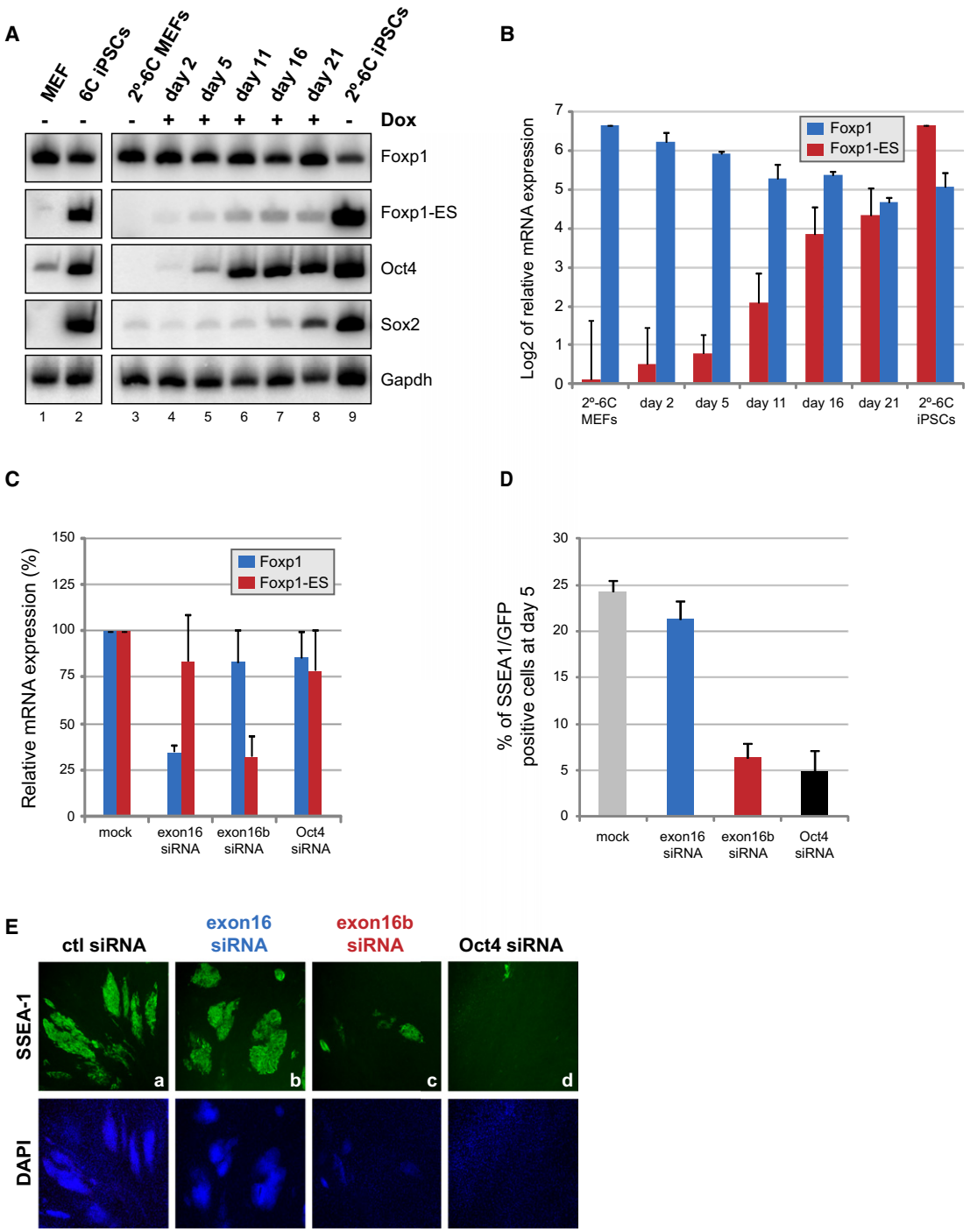

Figure 6. Foxp1-ES Is Required for Efficient Reprogramming of MEFs to iPSCs

(A) Semiquantitative RT-PCR analysis of the endogenous expression levels of Foxp1, Foxp1-ES, Oct4, and Sox2 during the course of reprogramming of secondary MEF-6C cells into secondary iPSC colonies (2°–6C iPSCs). Induction of Oct4, Klf4, cMyc, and Sox2 transcription factors by addition of Dox at day 0 (2°–6C MEFs) was followed by monitoring transcript levels 2, 5, 11, 16, 21, and 30 days (2°–6C iPSCs) post Dox induction. Gapdh mRNA levels are shown as a loading/recovery control.

(B) Bar graph showing the relative levels of expression of endogenous transcripts encoding Foxp1 and Foxp1-ES during reprogramming of 2°–6C MEFs. The levels of expression of Foxp1 and Foxp1-ES were normalized to Gapdh expression levels at each time point and represented as log2 ratios relative to the levels of Foxp1 and Foxp1-ES detected in 2°–6C MEFs and 2°–6C iPSCs, respectively. Positive SDs are indicated.

(C) Bar graph showing the relative expression of Foxp1 and Foxp1-ES isoforms following transfection of siRNA pools. Cells were either mock-transfected or transfected with siRNA pools specific for Foxp1 exon 16, Foxp1-ES exon 16b, or siRNA pools specific for Oct4. Expression levels were determined by

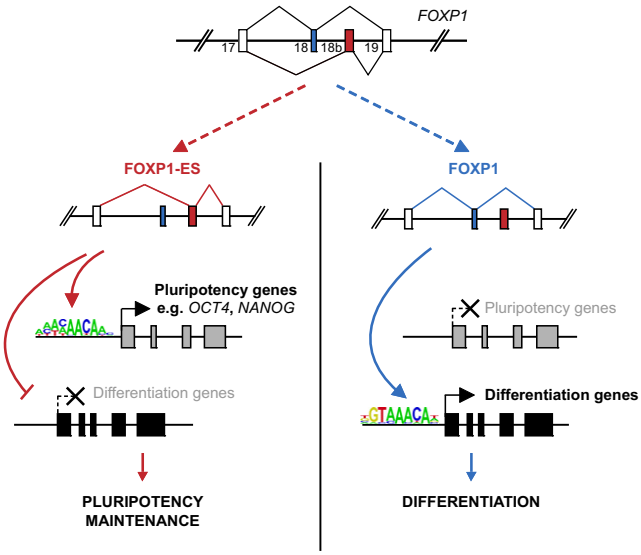

Figure 7. Model for the Role of an AS Switch in Controlling Transcriptional Networks Required for the Regulation of ESC Pluripotency and Differentiation

In pluripotent ESCs or iPSCs, inclusion of FOXP1 exon 18b results in the expression of FOXP1-ES, which preferentially binds to a distinct set of DNA motifs. This event promotes the expression of key transcription factors including OCT4 and NANOG required for the maintenance of pluripotency, and it represses genes required for ESC differentiation. During differentiation, exon 18b is entirely skipped, resulting in the exclusive inclusion of exon 18 and expression of the "canonical" FOXP1 isoform. This leads to a change in DNA recognition, a consequence of which is reduced expression of pluripotency genes and increased expression of genes required for differentiation.

RNA-Seq Data Generation and Analysis

H9 hESCs were transfected with siRNA pools (Dharmacon) using DharmaFECT as per the manufacturer recommendations, transfected again after 2 days, then harvested 2 days later. Total RNA from two independent transfections was pooled and submitted to Illumina Inc. for mRNA sequencing. RNA-Seq analysis was performed essentially as described previously (Pan et al., 2008).

Reverse-Transcription-Polymerase Chain Reaction Assays

RT-PCR assays were performed using the OneStep kit (QIAGEN) as described previously (Calarco et al., 2007). For qRT-PCR assays, cDNA from 2 μg total RNA was synthesized using SuperScript III Reverse Transcriptase (Invitrogen) as per manufacturer recommendations. Reactions were performed in a 384-well format using 20 ng of cDNA and FastStart Universal SYBR Green Master (Rox) (Roche Applied Science). Primer sequences are available upon request.

Protein-Binding Microarrays and Data Analysis

GST-FOXP1 and GST-FOXP1-ES were analyzed on PBMs as described previously (Badis et al., 2009), and the resulting data were processed as described in Lam et al. (2011). 8-mers with an E score > 0.45 in at least one of the two experimental repeats were considered significant (Berger et al., 2008) and were aligned to generate consensus sequences using enoLOGOS (Workman et al., 2005).

Gel Mobility Shift Assays

dsDNA probes contained two copies of representative PBM-derived binding sequences separated by two cytosines, or mutated derivatives of these sequences. Gel shift assays were performed as described in Hellman and Fried (2007).

ChIP-Seq

ChIP-Seq experiments were performed as described previously (Schmidt et al., 2009), using an anti-FOXP1 (Abcam) antibody and 5 × 10^7 H9 hESCs per sample. Genomic DNA Sample Prep Kits (Illumina) were used to prepare dsDNA libraries from fragmented immunoprecipitated and total DNA as per the manufacturer's protocol, and libraries were sequenced using a HiSeq machine (Illumina).

Immunofluorescence Microscropy

CGR8 cells were analzyed by immunofluoresence microscopy using polyclonal anti-β-III tubulin (Sigma-Aldrich) and murine monoclonal anti-Oct4 (Pierce), were stained with Hoechst dye (Sigma-Aldrich), then mounted with Aqueous mounting Medium (Permafluor). Images were acquired by epi-fluorescence imaging as previously described (Samavarchi-Tehrani et al., 2010).

iPSC Reprogramming Assays and Imaging

Secondary (6C) MEFs harboring OKMS transgenes under tetracycline-inducible control were derived using the *piggyBac* system as previously described (Woltjen et al., 2009). In brief, 2°–6C MEFs were cultured on collagen-coated plates and expression of the transgenes was induced on day 0 of reprogramming using 1.5 μg/ml Dox in standard mouse ESC media. For knockdown experiments, single or pooled siRNAs (Dharmacon/ThermoFisher) were transfected into the 2°–6C MEFs at day 0 or at day 13 after induction, as previously described (Samavarchi-Tehrani et al., 2010). Cells were then cultured for another 3 or 5 days prior to analysis by flow cytometry, immunostaining, or isolation of total RNA for qPCR analysis, also as described in Samavarchi-Tehrani et al. (2010).

ACCESSION NUMBERS

Preprocessed probe intensity scores for all AS and PB microarray data, and short read sequence data, are available from the GEO database under accession number GSE30992.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and five tables and can be found with this article online at doi:10.1016/j.cell.2011.08.023.

ACKNOWLEDGMENTS

B.J.B. dedicates this paper to the memory of S. Levine (1911–2011). We are grateful to D. Schmidt, D. Odom, Q. Morris, N. Barbosa-Morais, S. Mavadadi, and H. van Bakel for advice on data analysis and to A. Golipour for assistance with the iPSC reprogramming experiments. L. Attisano, D. Geschwind, and

semiquantitative RT-PCR assays, normalized to Gapdh levels and relative to the expression levels of the same transcripts in the mock-transfected control. Positive SDs are indicated. See also Figure S6B.

(D) Bar graph showing relative proportions of flow cytometry-sorted, reprogramming 2°–6C MEFs that are double-positive for GFP and the ESC/iPSC marker SSEA-1. 2°–6C MEFs were Dox treated to induce OKMS factors and transfected with siRNA pools indicated in (C) at day 0, then were analyzed by flow cytometry and immunostaining 5 days later. Results from analyzing the effects of transfecting the same siRNA pools at day 13 of reprogramming are shown in Figure S6C. Range over mean values for two independent analyses are indicated.

(E) Representative images of SSEA-1- and DAPI-stained cells at day 5 following Dox induction of OKMS factors and post-transfection of siRNA pools as described in (C) and (D).

K.-H. Krause kindly provided valuable reagents, and A. Saltzman, B. Raj, J. Calarco, J. Ellis, H. Han, N. Barbosa-Morais, and M. Q.-Vallières provided helpful comments on the manuscript. Our research was supported by grants from the Canadian Institutes for Health Research (CIHR) to B.J.B., J.L.W., A.N., P.W.Z., and T.R.H., a grant from the Canadian Cancer Society to B.J.B., a grant from Genome Canada (through the Ontario Genomics Institute) to B.J.B. and others, a grant from the Ontario Ministry of Research and Innovation to A.N., and a grant from the Ontario Research Fund to J.L.W., B.J.B., and others. E.O.M., S.N., and H.W. were supported by NIH grant P01 NS055923. E.O.M. is the David and Sylvia Lieb Fellow of the Damon Runyon Cancer Research Foundation (DRG-1937-07). M.G. was supported by postdoctoral fellowships from the C.H. Best Foundation and CIHR. J.L.W. is an International Scholar of the HHMI.

Received: February 8, 2011
Revised: June 10, 2011
Accepted: August 4, 2011
Published online: September 15, 2011

REFERENCES

Atlasi, Y., Mowla, S.J., Ziaee, S.A., Gokhale, P.J., and Andrews, P.W. (2008). OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. Stem Cells *26*, 3068–3074.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720–1723.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., 3rd, and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. *24*, 1429–1435.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

Brown, P.J., Ashe, S.L., Leich, E., Burek, C., Barrans, S., Fenton, J.A., Jack, A.S., Pulford, K., Rosenwald, A., and Banham, A.H. (2008). Potentially oncogenic B-cell activation-induced smaller isoforms of FOXP1 are highly expressed in the activated B cell-like subtype of DLBCL. Blood *111*, 2816–2824.

Calarco, J.A., Xing, Y., Cáceres, M., Calarco, J.P., Xiao, X., Pan, Q., Lee, C., Preuss, T.M., and Blencowe, B.J. (2007). Global analysis of alternative splicing differences between humans and chimpanzees. Genes Dev. *21*, 2963–2975.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Dasen, J.S., De Camilli, A., Wang, B., Tucker, P.W., and Jessell, T.M. (2008). Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. Cell *134*, 304–316.

Demir, E., and Dickson, B.J. (2005). fruitless splicing specifies male courtship behavior in Drosophila. Cell *121*, 785–794.

Fic, W., Juge, F., Soret, J., and Tazi, J. (2007). Eye development under the control of SRp55/B52-mediated alternative splicing of eyeless. PLoS ONE *2*, e253.

Förch, P., and Valcárcel, J. (2003). Splicing regulation in Drosophila sex determination. Prog. Mol. Subcell. Biol. *31*, 127–151.

Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat. Protoc. *2*, 1849–1861.

Hu, H., Wang, B., Borde, M., Nardone, J., Maika, S., Allred, L., Tucker, P.W., and Rao, A. (2006). Foxp1 is an essential transcriptional regulator of B cell development. Nat. Immunol. *7*, 819–826.

Jaeger, S.A., Chan, E.T., Berger, M.F., Stottmann, R., Hughes, T.R., and Bulyk, M.L. (2010). Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. Genomics *95*, 185–195.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

Koh, K.P., Sundrud, M.S., and Rao, A. (2009). Domain requirements and sequence specificity of DNA binding for the forkhead transcription factor FOXP3. PLoS ONE *4*, e8109.

Konstantoulas, C.J., Parmar, M., and Li, M. (2010). FoxP1 promotes midbrain identity in embryonic stem cell-derived dopamine neurons by regulating Pitx3. J. Neurochem. *113*, 836–847.

Koon, H.B., Ippolito, G.C., Banham, A.H., and Tucker, P.W. (2007). FOXP1: a potential therapeutic target in cancer. Expert Opin. Ther. Targets *11*, 955–965.

Kubo, H., Shimizu, M., Taya, Y., Kawamoto, T., Michida, M., Kaneko, E., Igarashi, A., Nishimura, M., Segoshi, K., Shimazu, Y., et al. (2009). Identification of mesenchymal stem cell (MSC)-transcription factors by microarray and knockdown analyses, and signature molecule-marked MSC in bone marrow by immunohistochemistry. Genes Cells *14*, 407–424.

Kunarso, G., Wong, K.Y., Stanton, L.W., and Lipovich, L. (2008). Detailed characterization of the mouse embryonic stem cell transcriptome reveals novel genes and intergenic splicing associated with pluripotency. BMC Genomics *9*, 155.

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. *42*, 631–634.

Lam, K.N., van Bakel, H., Cote, A.G., van der Ven, A., and Hughes, T.R. (2011). Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. Nucleic Acids Res. *39*, 4680–4690.

Li, S., Weidenfeld, J., and Morrisey, E.E. (2004). Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol. Cell. Biol. *24*, 809–822.

Mayshar, Y., Rom, E., Chumakov, I., Kronman, A., Yayon, A., and Benvenisty, N. (2008). Fibroblast growth factor 4 and its novel splice isoform have opposing effects on the maintenance of human embryonic stem cell self-renewal. Stem Cells *26*, 767–774.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., et al. (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol. Cell *16*, 929–941.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. *40*, 1413–1415.

Pritsker, M., Doniger, T.T., Kramer, L.C., Westcot, S.E., and Lemischka, I.R. (2005). Diversification of stem cell molecular repertoire by alternative splicing. Proc. Natl. Acad. Sci. USA *102*, 14290–14295.

Rao, D.S., O'Connell, R.M., Chaudhuri, A.A., Garcia-Flores, Y., Geiger, T.L., and Baltimore, D. (2010a). MicroRNA-34a perturbs B lymphocyte development by repressing the forkhead box transcription factor Foxp1. Immunity *33*, 48–59.

Rao, S., Zhen, S., Roumiantsev, S., McDonald, L.T., Yuan, G.C., and Orkin, S.H. (2010b). Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. Mol. Cell. Biol. *30*, 5364–5380.

Rousso, D.L., Gaber, Z.B., Wellik, D., Morrisey, E.E., and Novitch, B.G. (2008). Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. Neuron *59*, 226–240.

Rowan, S., Siggers, T., Lachke, S.A., Yue, Y., Bulyk, M.L., and Maas, R.L. (2010). Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. Genes Dev. *24*, 980–985.

Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. Proc. Natl. Acad. Sci. USA *107*, 10514–10519.

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.

Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods *48*, 240–248.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Stroud, J.C., Wu, Y., Bates, D.L., Han, A., Nowick, K., Paabo, S., Tong, H., and Chen, L. (2006). Structure of the forkhead domain of FOXP2 bound to DNA. Structure *14*, 159–166.

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.

Wang, B., Weidenfeld, J., Lu, M.M., Maika, S., Kuziel, W.A., Morrisey, E.E., and Tucker, P.W. (2004). Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. Development *131*, 4477–4487.

Wijchers, P.J., Burbach, J.P., and Smidt, M.P. (2006). In control of biology: of mice, men and Foxes. Biochem. J. *397*, 233–246.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., et al. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766–770.

Workman, C.T., Yin, Y., Corcoran, D.L., Ideker, T., Stormo, G.D., and Benos, P.V. (2005). enoLOGOS: a versatile web tool for energy normalized sequence logos. Nucleic Acids Res. *33*(Web Server issue), W389–W392.

Wu, J.Q., Habegger, L., Noisa, P., Szekely, A., Qiu, C., Hutchison, S., Raha, D., Egholm, M., Lin, H., Weissman, S., et al. (2010). Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. Proc. Natl. Acad. Sci. USA *107*, 5254–5259.

Yeo, G.W., Xu, X., Liang, T.Y., Muotri, A.R., Carson, C.T., Coufal, N.G., and Gage, F.H. (2007). Alternative splicing events identified in human embryonic stem cells and neural progenitors. PLoS Comput. Biol. *3*, 1951–1967.

Zhang, Y., Li, S., Yuan, L., Tian, Y., Weidenfeld, J., Yang, J., Liu, F., Chokas, A.L., and Morrisey, E.E. (2010). Foxp1 coordinates cardiomyocyte proliferation through both cell-autonomous and nonautonomous mechanisms. Genes Dev. *24*, 1746–1757.