Please cite this article in press as: Ohta et al., Global Splicing Pattern Reversion during Somatic Cell Reprogramming, Cell Reports (2013), http://dx.doi.org/10.1016/j.celrep.2013.09.016

Cell Reports
Report

Global Splicing Pattern Reversion during Somatic Cell Reprogramming

Sho Ohta,[1,2] Eisuke Nishida,[2,3] Shinya Yamanaka,[1,4,5] and Takuya Yamamoto[1,4,*]
[1]Department of Reprogramming Science, Center for iPS Cell Research and Application, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan
[2]Department of Cell and Developmental Biology, Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
[3]JST, CREST, Chiyoda-ku, Tokyo 102-0075, Japan
[4]Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan
[5]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA
*Correspondence: takuya@cira.kyoto-u.ac.jp
http://dx.doi.org/10.1016/j.celrep.2013.09.016

SUMMARY

Alternative splicing generates multiple transcripts from a single gene, and cell-type-specific splicing profiles are important for the properties and functions of the cells. Recently, somatic cells have been shown to undergo dedifferentiation after the forced expression of transcription factors. However, it remains unclear whether somatic cell splicing is reorganized during reprogramming. Here, by combining deep sequencing with high-throughput absolute qRT-PCR, we show that somatic splicing profiles revert to pluripotent ones during reprogramming. Remarkably, the splicing pattern in pluripotent stem cells resembles that in testes, and the regulatory regions have specific characteristics in length and sequence. Furthermore, our siRNA screen has identified RNA-binding proteins that regulate splicing events in iPSCs. We have then demonstrated that two of the RNA-binding proteins, U2af1 and Srsf3, play a role in somatic cell reprogramming. Our results indicate that the drastic alteration in splicing represents part of the molecular network involved in the reprogramming process.

INTRODUCTION

Alternative splicing is a posttranscriptional process in which premature transcripts are selectively cut and joined in more than one way to generate multiple mRNA forms from a single gene. These different mRNAs can produce different proteins with different functions and properties (Nilsen and Graveley, 2010). Therefore, alternative splicing is one of the main sources of proteomic diversity; such diversity may allow complex and flexible intracellular molecular networks. It has been suggested that the transcripts from more than 90% of human genes are alternatively spliced in a variety of adult tissues (Pan et al., 2008; Wang et al., 2008). Tissue- and cell-type-specific splicing patterns have been shown to play critical roles in the intrinsic properties and functions of the tissues and cells. In fact, it has been shown that the disruption of splicing or its regulation can cause disease and that at least 15% of all human disease mutations have an impact on splice site selection (Cooper et al., 2009). In addition, in embryonic stem cells (ESCs), several genes have been shown to undergo ESC-specific splicing to ensure the specific functions (Atlasi et al., 2008; Cheong and Lufkin, 2011; Gabut et al., 2011; Gopalakrishnan et al., 2009; Rao et al., 2010; Salomonis et al., 2010; Yeo et al., 2007). These findings demonstrate the physiological importance of precise splicing regulation.

Induced pluripotent stem cells (iPSCs) were originally generated from somatic cells by the expression of just four transcription factors (Takahashi and Yamanaka, 2006), and their morphology, differentiation capacity, and gene expression profile are almost exactly equivalent to those of ESCs (Mikkelsen et al., 2008; Okita et al., 2007). However, it remains unclear whether somatic splicing is reorganized during the process.

To examine changes in genome-wide splicing profiles during the cellular reprogramming process, we combined deep sequencing with high-throughput quantitative RT-PCR (qRT-PCR). Our results demonstrated that iPSCs closely resemble ESCs in their splicing profiles and, thus, indicated that somatic splicing profiles revert to undifferentiated ones by cell reprogramming. Then, to obtain mechanistic insights into the splicing regulation in iPSCs, we performed siRNA screen for RNA-binding proteins that are responsible for the splicing regulation. Consequently, several RNA-binding proteins that could function as splicing regulators in pluripotent stem cells have been identified. Moreover, we have shown that downregulation of two RNA-binding proteins, U2af1 and Srsf3, significantly suppressed the efficiency of somatic cell reprogramming. Thus, our results identify splicing regulation as part of the molecular network involved in the reprogramming process and the maintenance of pluripotency.

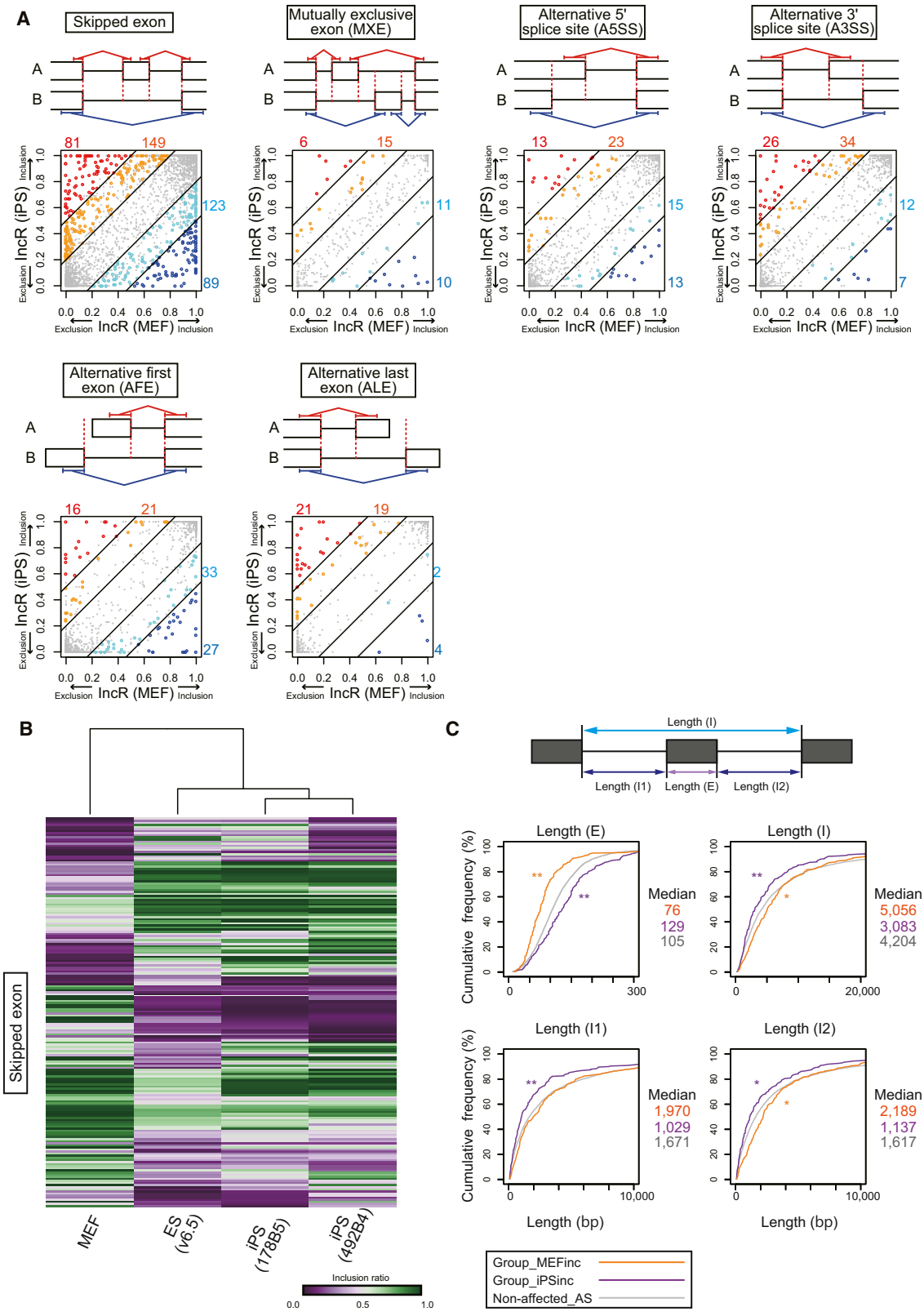
RESULTS

Global Analysis of Changes in Splicing Patterns during Reprogramming

To analyze the mRNA splicing differences before and after reprogramming, we performed RNA sequencing (RNA-seq) on

Please cite this article in press as: Ohta et al., Global Splicing Pattern Reversion during Somatic Cell Reprogramming, Cell Reports (2013), http://dx.doi.org/10.1016/j.celrep.2013.09.016

Cell PRESS
Open ACCESS



(legend on next page)

mRNA from mouse embryonic fibroblasts (MEFs) and mouse iPSCs. As a result, more than 70% of sequence reads were mapped to the reference sequences. In this mapping, we obtained approximately ten million junction reads, which spanned exon splice junction sites, from each sample (Table S1). We used these reads to analyze the splicing patterns (Wang et al., 2008). Alternative splicing events were subdivided into several types (Figure 1A). We examined six splicing event types by comparing the "inclusion ratio," which was defined as the ratio of the read number derived from the "inclusion" (red) junction region, to the total number derived from the inclusion and the "exclusion" (blue) junction region (Figure 1A). There were 770 splicing events (587 genes), whose inclusion ratios were significantly different (FDR <0.01) between the MEFs and iPSCs by a minimum of 0.2 absolute change (Figure 1A; Table S2). We next performed clustering analysis based on the inclusion ratios and found that two iPSC lines and one ESC line showed similar patterns for all of the splice types examined (Figures 1B and S1A). This result indicates that the global splicing pattern of MEF is dramatically converted into that of ESCs during iPSC induction. We analyzed which biological processes and pathways are relevant to the change in splicing during iPSC induction by Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, http://www.ingenuity.com) (Figure S1B). As a result, several biological functions and pathways including "Embryonic Development" and "RNA Posttranscriptional Modification" are significantly enriched in the genes whose splicing patterns are changed during iPSC induction (Figure S1B). These results imply that splicing switches during reprogramming are relevant to pluripotency and its acquisition and that RNA posttranscriptional regulators including splicing factors affect themselves by autoregulation (Lareau et al., 2007; Ni et al., 2007) and convert posttranscriptional regulatory networks into the pluripotent state.

Splicing Patterns in Multiple iPSC Lines and ESC Lines

We focused on exon-skipping events to further analyze the splicing regulation during iPSC induction because this type of splicing occurs most frequently in higher eukaryotes. To validate the alternative exon data obtained from the RNA-seq analysis, we performed qRT-PCR. For this analysis, we used the BioMark System (Fluidigm), which enables us to perform high-throughput real-time qRT-PCR. We also used digital PCR to determine the absolute expression level of each splicing variant. By combining qRT-PCR and digital PCR ("absolute qRT-PCR"), we were able to calculate the exact inclusion ratio (see Figure S2A) accurately and reliably. Here, we focused on the "DNA-binding protein"- and "Protein kinase"-encoding genes because many studies have shown that these types of genes are involved in the reprogramming process. In the skipped exons whose inclusion ratio in MEFs is different from that in iPSCs, there were 45 and 24 genes that are annotated in the Gene Ontology database as "DNA-binding" and "Protein kinase," respectively. Although several targets could not be detected, we could examine the splicing patterns of 38 genes by the absolute qRT-PCR method. As a result, we could confirm that the inclusion ratio in MEFs is significantly different from that in iPSCs ($p < 0.05$) in 36 out of 38 genes. Then we examined the inclusion ratio of the 36 genes in an additional two MEFs and six iPSC lines as well as three partial iPSC (piPSC) lines (Okita et al., 2007) and three ESC lines to identify the genes that show a pluripotent stem cell-specific splicing pattern. Consequently, we found that in 27 out of 36 genes, the inclusion ratio in MEFs is significantly different from that in iPS/ESC lines (FDR <0.05, Figures 2A and S2B). Although the inclusion ratios of some genes varied among pluripotent stem cells, our clustering analysis showed that the overall splicing patterns of the 27 genes are similar among iPSC and ESC lines and different from those in piPSCs and MEFs (Figure S2C; see Supplemental Discussion). Furthermore, we confirmed that other genes with different function (eight genes) were also alternatively spliced as predicted in the RNA-seq analysis and showed similar splicing patterns in all the iPSC/ESC lines examined (Figure S2D).

Characteristics in Sequences and Lengths in and Adjacent to Alternative Exons

Next, we set out to explore genomic signatures associated with the splicing changes during reprogramming. First, we searched for consensus motifs in the exons and introns within and around the exons because specific splicing factors are thought to recognize specific RNA sequences to regulate the splicing patterns. According to our RNA-seq data, approximately 400 skipped exons are alternatively spliced between the MEFs and iPSCs (Figure 1A, skipped exon). We subdivided these exon-skipped regions into two groups and investigated enriched sequences. Each group includes exons that are preferentially expressed in MEFs (referred to as Group_MEFinc) or in iPSCs (referred to as Group_iPSinc). As a result, we identified several overrepresented motifs within the alternatively skipped exons and the adjacent introns in Group_MEFinc or Group_iPSinc (Figure S1C). Thus, particular RNA-binding proteins might affect exon reorganization through binding to specific RNA sequences.

We next investigated the length distribution of the skipped exons and their surrounding introns of the two groups because exon-intron structure can also be important for splicing regulatory mechanisms (Fox-Walsh et al., 2005; Sterner et al., 1996; Yeo et al., 2007). Interestingly, the lengths of the introns surrounding the Group_iPSinc exons (median, 3,083 bp) are significantly shorter, and the lengths of the introns surrounding the

Figure 1. Identification of Gene Sets Whose Splicing Patterns Are Changed during the Somatic Cell Reprogramming Process

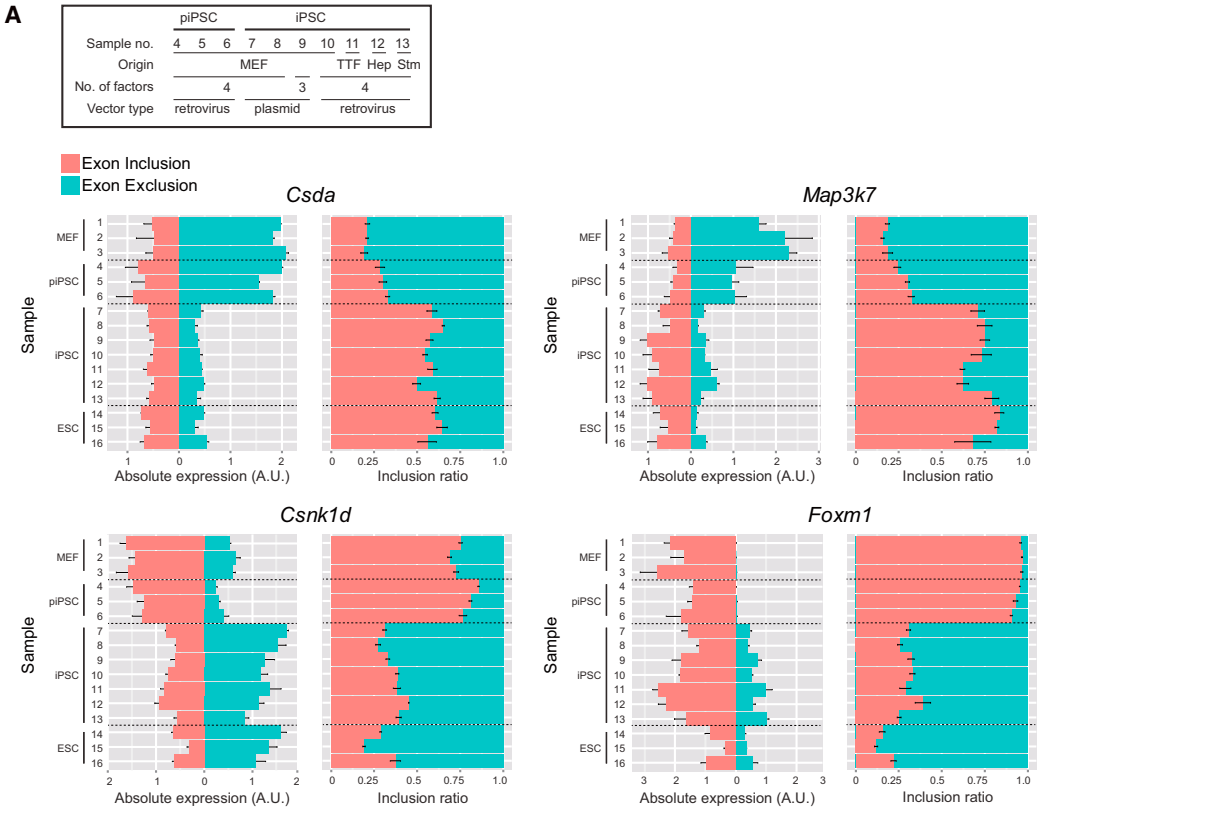
(A) Scatterplots show inclusion ratios (IncRs) of iPSCs (492B4) and MEFs. Each open circle indicates a splicing event whose inclusion ratio is changed more than 0.5 (red and blue) or 0.2–0.5 (orange and cyan) (Fisher's exact test, FDR <0.01). The numbers of open circles in each fraction are shown.

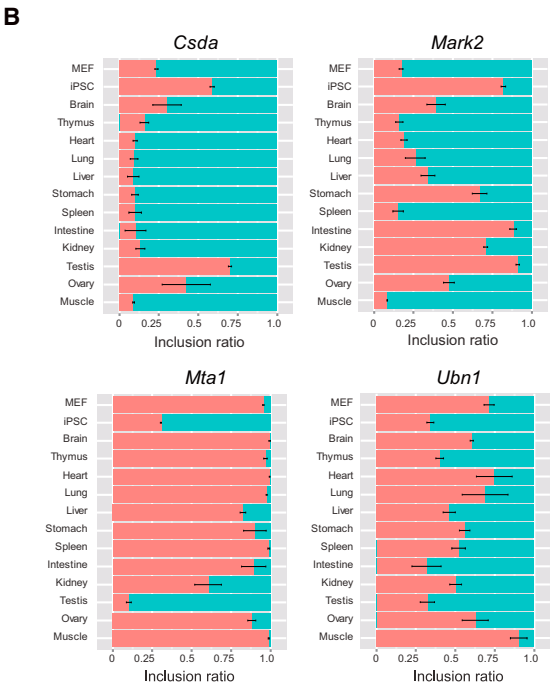
(B) Clustering analysis of the skipped exon inclusion ratios is illustrated. The skipped exons that differ in inclusion ratio between MEFs and ESCs (V6.5) by more than 0.2 (Fisher's exact test, FDR <0.01) were used.

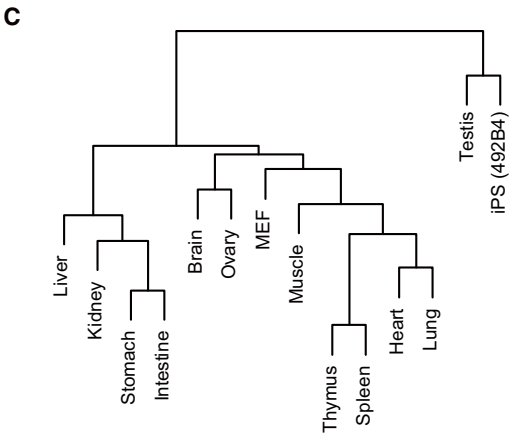
(C) The cumulative frequencies of the lengths of the skipped exons and their surrounding introns are shown. *$p < 0.05$, **$p < 0.005$ for Mann Whitney U test comparing to nonaffected AS (alternative splicing), respectively.

See also Figure S1.

Please cite this article in press as: Ohta et al., Global Splicing Pattern Reversion during Somatic Cell Reprogramming, Cell Reports (2013), http://dx.doi.org/10.1016/j.celrep.2013.09.016







(legend on next page)

Group_MEFinc exons (median size, 5,056 bp) are significantly longer than those of the expressed but nonaffected exons (median size, 4,204 bp) (Figure 1C). In contrast, the lengths of the Group_iPSinc exons (median size, 129 bp) are longer, and the lengths of the Group_MEFinc exons (median size, 76 bp) are shorter than those of nonaffected exons (median size, 105 bp) (Figure 1C). Thus, the fundamental mechanisms for the splice site selection in pluripotent stem cells may be different from those in somatic cells, which might contribute to generating cell identities such as the stemness of pluripotent stem cells.

The Splicing Pattern in iPSCs Is Most Similar to that in Testes

We identified the 27 genes by the absolute qRT-PCR data, which undergo alterations in splicing pattern during the reprogramming process. Next, to determine whether the alterations in splicing patterns have specificity for pluripotent stem cells, we characterized the splicing profiles of the 27 genes across multiple tissues by utilizing the absolute qRT-PCR. Interestingly, we found that the splicing patterns in the iPSCs are similar to those in the testes (Figure 2B). To confirm this observation, we performed clustering analysis based on the inclusion ratio values from the absolute qRT-PCR shown in Figures 2B and S2E. This analysis demonstrated that the splicing patterns of the iPSCs were most similar to those of the testes (Figure 2C). This result suggests that pluripotent stem cells may use the same mechanisms to regulate alternative splicing as the testes do.

Splicing Transitions during Reprogramming Are Temporally Coordinated and Can Be Divided into Subgroups

Previous studies demonstrated that pluripotent markers or other genes are sequentially expressed during the reprogramming of MEFs (Brambrink et al., 2008; Chen et al., 2010; Stadtfeld et al., 2008). To determine whether splicing switches occur concurrently or in sequential steps, we collected cells by FACS (BD Biosciences) during different phases of cellular reprogramming using anti-Thy1 and anti-SSEA-1 antibodies (Stadtfeld et al., 2008). In this experiment, we used Nanog-GFP reporter MEFs that express GFP when the endogenous Nanog promoter is induced (Okita et al., 2007). First, according to the established method, we fractionated Thy-1-positive Nanog-GFP reporter MEFs 1 day before the infection with viral vectors encoding reprogramming factors (Day −1) to isolate a population homogeneous for Thy-1 expression as the starting material; these cells accounted for approximately 20% of the MEF population (Figure 3A). Analyses were then conducted on day 4, 8, and 14 after infection. qRT-PCR for Thy-1 and Nanog expression confirmed that the sorted fractions contained correct populations (Figure S3A). The splicing transitions of the gene set were analyzed by examining the inclusion ratios in each sorted fraction using the absolute qRT-PCR. The result showed that the splicing pattern switches occur stepwise at particular points in time. The gene set could be divided into three subgroups according to the switch times (Figures 3B and S3B). We also compared the splicing patterns in the sorted cells with those in piPSCs (Figures 2A and S2B). The clustering analysis showed that the splicing patterns in piPSCs are the intermediate state in the reprogramming process (Figure S3C). Collectively, these data indicate that, similar to gene expression, alternative splicing is sequentially regulated, suggesting that splicing regulation may be tightly linked to the transcriptional regulation during somatic reprogramming.

siRNA Screen Identifies Splicing Regulators that Play a Role in Somatic Cell Reprogramming

To obtain mechanistic insights into the distinct splicing regulation in iPSCs, we performed an RNAi screen in pluripotent stem cells. Tissue-specific alternative splicing can be mediated by the tissue specifically expressed RNA-binding proteins that bind to regulatory *cis* elements (Licatalosi and Darnell, 2010). Therefore, we focused our attention on the RNA-binding protein-encoding genes. According to the NEXTBIO database, there were 501 genes that belong to the "RNA-binding" category in the whole mouse genome. To identify RNA-binding protein-encoding genes that are involved in the reprogramming process by positively regulating alternative splicing specific for pluripotent stem cells, we performed gene expression profiling by GeneChip microarray (Affymetrix) and selected 92 genes, based on the criterion that their expression level is at least 2-fold higher in iPSCs/ESCs than in MEFs (Figure S4A). We then examined the expression profile of the gene set (92 RNA-binding protein-encoding genes) in 96 mouse tissues and cell lines using the BioGPS public data (http://biogps.gnf.org). Consequently, the expression profile of the gene set in ESCs was found to be most similar to that in testes among various tissues and cell lines (Figure S4B), as in the case of the similarity in the splicing regulation between pluripotent stem cells and testes (Figures 2B and 2C), suggesting that these 92 RNA-binding protein-encoding genes might be involved in the splicing regulation in pluripotent stem cells. We thus generated an siRNA library that targets the 92 genes to explore siRNAs that induce differences in alternative splicing. In our RNAi screen, each gene was targeted by a pool of three different siRNA duplexes, and each splicing pattern was analyzed in duplicate in both ESCs and iPSCs. First, we confirmed that our experimental system is appropriate for assessing the effects of RNA-binding protein depletion (Figure S4C). We next examined the effect of each siRNA pool on the splicing pattern (the ratio between the expression levels of inclusion forms and exclusion forms) of the genes that were

Figure 2. Characterization of Splicing Patterns in MEFs and iPSCs

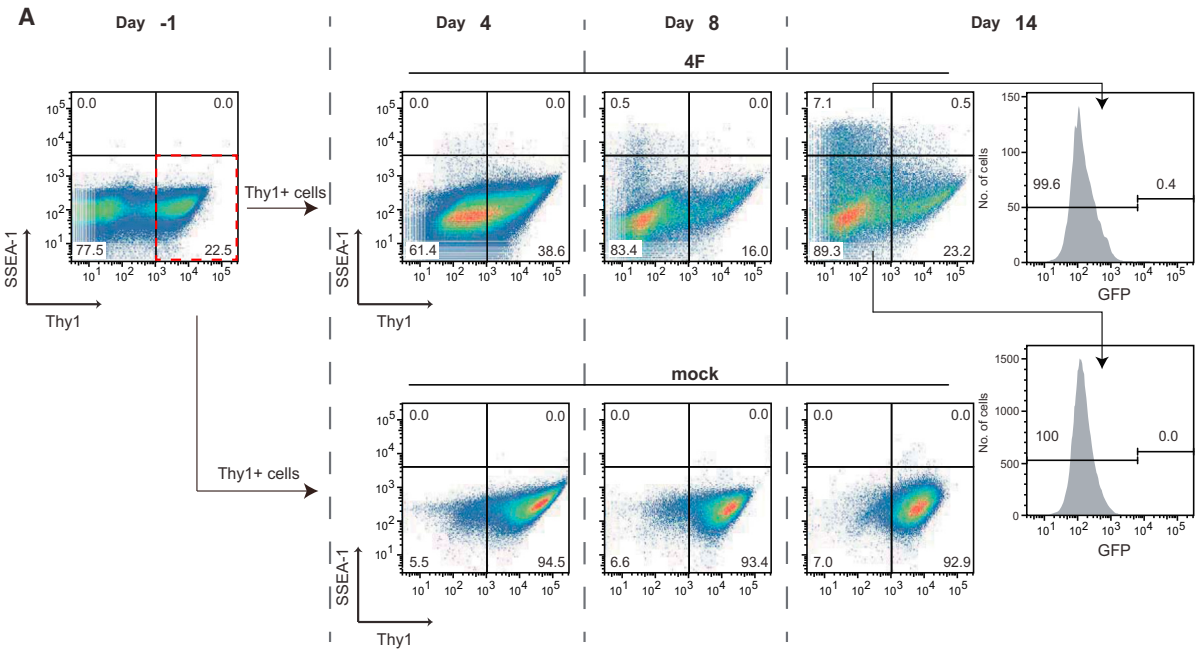
(A) Splicing patterns in MEFs, three piPSC lines, seven iPSC lines, and three ESC lines using absolute qRT-PCR are shown. The upper table represents the cell origin, the number of reprogramming factors, and the type of vectors that were used for inducing each piPSC/iPSC: hepatocyte (Hep), stomach (Stm; gastric epithelial cell), and tail tip fibroblast (TTF). For each gene, absolute expression levels (left) and inclusion ratios (right) are presented. Mean ± SD (n = 3).

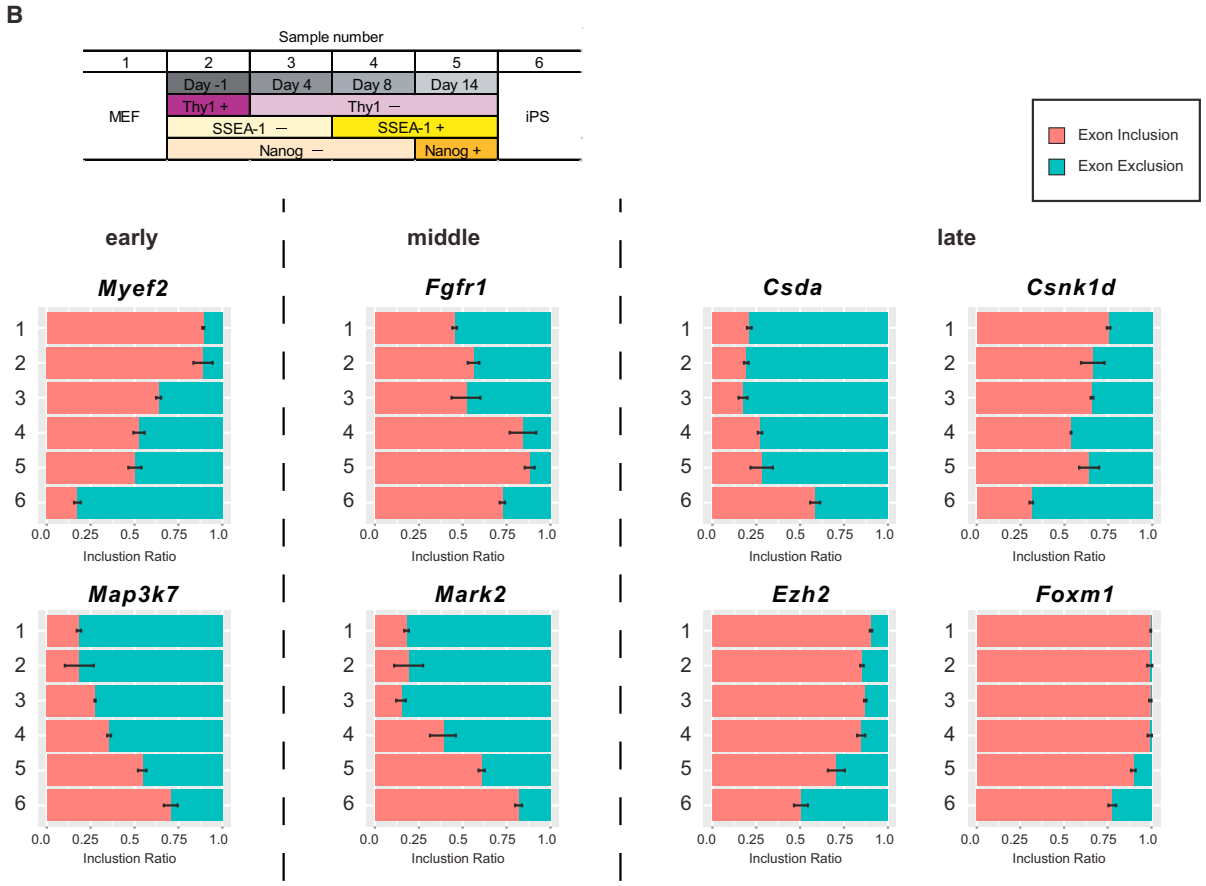
(B) Splicing patterns across multiple adult mouse tissues using absolute qRT-PCR are shown. The graphs indicate the inclusion ratios. Mean ± SD (n = 3).

(C) Clustering analysis of splicing patterns across multiple tissues is presented. This analysis was based on the inclusion ratios obtained in (B) and Figure S2E. See also Figure S2.

Please cite this article in press as: Ohta et al., Global Splicing Pattern Reversion during Somatic Cell Reprogramming, Cell Reports (2013), http://dx.doi.org/10.1016/j.celrep.2013.09.016

Cell PRESS
Open ACCESS





(legend on next page)

alternatively spliced between MEFs and iPSCs and found that about one-third of our siRNA pools influenced the splicing patterns (Figures 4B and 4C; data not shown). These siRNA target genes would be strong candidates for the splicing regulators in pluripotent stem cells. For nine RNA-binding protein-encoding genes that affected the splicing patterns effectively in our initial RNAi screen (Figure 4C), we produced retroviral-mediated small hairpin RNAs (shRNAs) (Figures 4D, 4E, and S4I) and examined the impact of the shRNAs on somatic cell reprogramming. We coinfected Nanog-GFP reporter MEFs with retroviruses expressing the reprogramming factors (Oct4, Sox2, Klf4, and c-Myc) and the shRNAs. As a result, downregulation of U2af1 or Srsf3 suppressed both the efficiency of AP-positive colony formation (Figure 4F) and the extent of Nanog-GFP expression (Figure 4G), although there were no obvious phenotypes such as reduced cell division and change in morphology as long as we observed MEFs 2 days after the shRNA infection. We also confirmed that U2af1 and Srsf3 were upregulated at the protein level in iPSCs compared to MEFs (Figure 4E) and that downregulation of U2af1 and Srsf3 by each of the siRNA pools altered the splicing patterns without affecting the iPSC proliferation and the expression of housekeeping genes (Figures S4E–S4H; data not shown). Taken together, these data suggest that U2af1 and Srsf3 play a role in somatic cell reprogramming.

DISCUSSION

In this study, we performed global analysis of alternative splicing and identified several hundred genes whose splicing patterns are changed during the reprogramming process. Moreover, our data indicate that molecular properties of somatic cells revert to those of ESCs in terms of isoform expression. Although the functional significance of each splicing variant in the reprogramming process remains to be elucidated, our analysis reveals that cellular reprogramming accompanies the drastic changes in splicing regulation.

Our siRNA screen experiment has identified candidate RNA-binding proteins that function as splicing regulators in pluripotent stem cells. Moreover, our analysis showed that U2af1 and Srsf3 play a role in somatic cell reprogramming. After completion of our work, there appeared a paper of Han et al. (2013) reporting that two RNA-binding proteins, MBNL1 and MBNL2, whose expression levels are much lower in pluripotent stem cells than in many differentiated cells, play an important role in the maintenance of pluripotency and cellular reprogramming by negatively regulating alternative splicing specific for pluripotent stem cells. In addition, our splicing analyses in this study identified the RNA-binding protein-encoding genes whose splicing patterns are changed during the reprogramming process (Figure S1B). Taken together, the dynamic reorganization of the alternative splicing profile, which occurs during somatic cell reprogramming, might be achieved by the quantitative and qualitative changes in the molecular repertoire of RNA-binding proteins. In summary, our study describes the drastic change in splicing isoform expression and its regulatory mechanisms during reprogramming and suggests that alternative splicing regulation represents part of the mechanisms of cellular reprogramming and has important roles in pluripotency, although the functional relevance of splicing during cellular reprogramming remains to be elucidated.

EXPERIMENTAL PROCEDURES

Cell Culture and iPSC Generation

Mouse iPSCs, mouse ESCs, and MEFs were cultured as previously described (Takahashi and Yamanaka, 2006). piPSCs were cultured in the same condition as iPSCs. See Supplemental Experimental Procedures for details. The generation of mouse iPSCs with retroviruses was performed as previously described (Takahashi and Yamanaka, 2006) with some modifications, which are detailed in the Supplemental Experimental Procedures.

RNA-Seq

RNA was isolated using the RNeasy Mini Kit (QIAGEN). The polyA fraction was selected using a magnetic-based purification kit (Dynabeads mRNA Purification Kit; Invitrogen). The cDNA libraries for MEFs and iPSCs were generated and sequenced with the SOLiD System (Life Technologies) according to the instructions of the manufacturer. See Supplemental Experimental Procedures for details.

Data Analysis

The RNA-seq data sets were mapped to both the mouse exon-exon junction sequences that were defined by three transcript databases (i.e., RefSeq, UCSC Known Genes, and Ensemble::Transcript) and the mouse reference genome (mm9) using BioScope v.1.3 (Life Technologies) with default mapping parameters. The calculation of the inclusion ratio, the statistical analysis, the clustering analysis, and Motif analysis were performed using Microsoft SQL server and R software with custom-made programs. The functional analyses were generated through the use of IPA (Ingenuity Systems, http://www.ingenuity.com).

Flow Cytometry

Cells were incubated with PE-conjugated rat anti-Thy1 (sc-52616 PE; Santa Cruz Biotechnology) and Alexa Flour 647-conjugated mouse anti-SSEA-1 (sc-21702 AF647; Santa Cruz Biotechnology) antibodies and analyzed on a FACSAria II instrument (BD Biosciences). Dead cells were excluded by staining with DAPI. The data were analyzed using FlowJo software (Tree Star).

RNA Isolation, qRT-PCR, Digital PCR, and Microarray Analysis

RNA was isolated using the RNeasy Mini Kit following the manufacturer's instructions, and cDNA was produced using the QuantiTect Reverse Transcription Kit (QIAGEN). qRT-PCRs were set up with SYBR Premix Ex TaqII (Perfect Real Time) (TaKaRa) and run on a StepOne Plus QPCR System (Applied Biosystems). For high-throughput qRT-PCR, qRT-PCRs were run using the BioMark System (Fluidigm) following the manufacturer's instructions. For digital PCR, cDNA was diluted on dqPCR 37K chips, and PCRs were run using the BioMark System following the manufacturer's instructions. See Table S3 for the primer sequences for qRT-PCR. For expression profiling, RNAs derived from MEFs and iPSCs (492B4) were reverse transcribed, labeled, and analyzed on the Affymetrix microarray (mouse Gene 1.0 ST arrays). Further

Figure 3. Splicing Pattern Transitions during Somatic Cell Reprogramming

(A) FACS plots of Thy1, SSEA-1, and Nanog expression on day 4, 8, and 14 after infection are shown. A cutoff was set using unstained control cells. Retroviruses expressing Oct4, Sox2, Klf4, and c-Myc were used to infect Nanog-GFP MEFs on day 0.

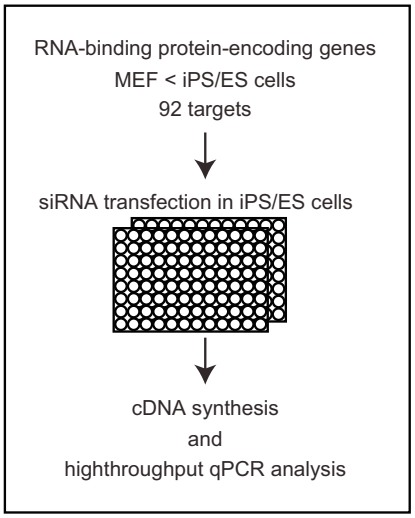
(B) Determination of splicing switch times during reprogramming using absolute qRT-PCR is presented. The graphs indicate the inclusion ratios. Larger sample numbers correspond to more reprogrammed samples. The iPS492B4 cell line was used as a mature iPSC sample. Mean ± SD (n = 3).

See also Figure S3.

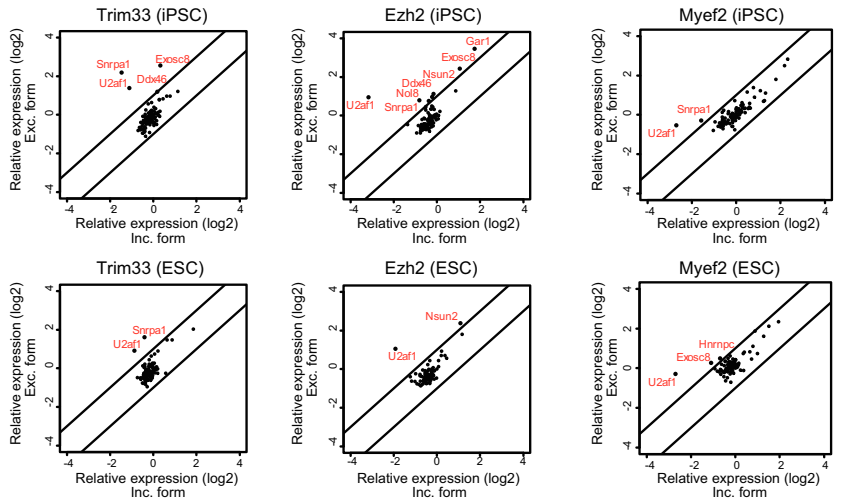
Please cite this article in press as: Ohta et al., Global Splicing Pattern Reversion during Somatic Cell Reprogramming, Cell Reports (2013), http://dx.doi.org/10.1016/j.celrep.2013.09.016

Cell PRESS
Open ACCESS

A



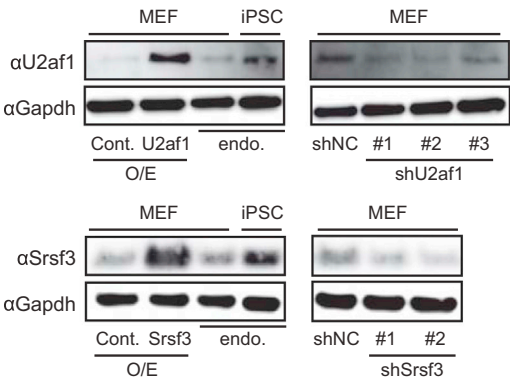
B



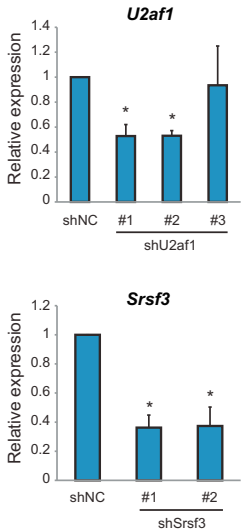
C

Target gene by siRNA	Inc / Exc ↑	Inc / Exc ↓
U2af1	*Foxm1, Mark3, Mta1*	*Ezh2, Map3k7, Map4k4, Maz, Myef2, Nrf1, Tfdp2, Trim33*
Srsf3	*Hmgxb4, Rnps1, Tbx3*	*Map4k4, Mark2, Maz*
Snrpa1		*Fgfr1, Map3k7, Trim33, Ubn1*
Nsun2		*Ezh2, Rnps1, Tbx3*
Ddx46	*Prpf4b*	*Cenpa, Mlh3*
Hnrnpa1	*Dclk2*	*Cenpa, Hmgxb4*
Celf1	*Map3k7, Ubn1*	*Mta1*
Nhp2		*Dclk2, Map4k4*
Larp4		*Hsf2, Map4k4*

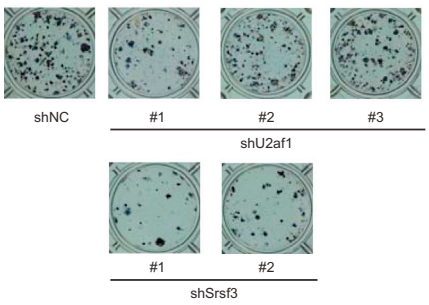
E

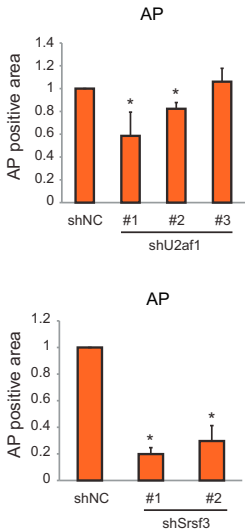


D

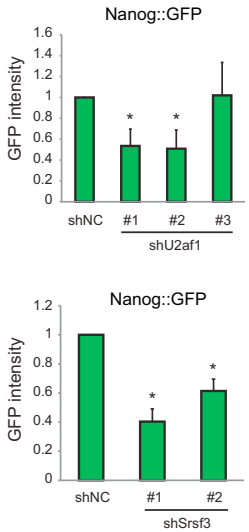


F





G



(legend on next page)

data analysis was performed using GeneSpring GX software (Agilent Technologies). See Supplemental Experimental Procedures for details.

siRNA Screen

Silencer siRNAs (Life Technologies) for a single target gene were pooled. Then the siRNA pools were transfected into murine iPSCs and ESCs by using Lipofectamine 2000 (Life Technologies) by reverse transfection according to manufacturers' instructions. After siRNA treatment for 48 hr, cDNA was synthesized using the Cells to CT Kit (Life Technologies). Then, real-time quantitative PCRs were run on the BioMark System following the manufacturer's instructions. See Supplemental Experimental Procedures for details.

Immunoblotting

See Supplemental Experimental Procedures for details and antibodies used.

Public Microarray Data Analysis

The expression profiles of the 92 RNA-binding genes in tissues and cell lines were obtained from the BioGPS public database (http://biogps.gnf.org).

ACCESSION NUMBERS

All sequencing data were deposited in the Gene Expression Omnibus under the accession number GSE45916.

All DNA microarray data were deposited in the Gene Expression Omnibus under the accession number GSE46149.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Discussion, Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.09.016.

ACKNOWLEDGMENTS

We thank the members of our laboratory for valuable scientific discussions and administrative support. We thank M. Kabata, T. Satoh, and S. Sakurai for technical assistance; T. Aoi, K. Okita, and M. Koyanagi for the mouse iPSCs and ESCs; Y. Yamada for the mouse ESCs; M. Maekawa for help with cell culture; K. Iwabuchi for help with flow cytometry; T. Nakamura for help with mouse tissue isolation; T. Satoh for iPSC induction; K. Takahashi for vectors; H. Saito and S. Kashida for shRNA design; S. Matsumura and F. Toyoshima for western blot; and M. Sone, Y. Yamada, H. Saito, K. Woltjen, and S. Honjoh for careful reading of the manuscript and making helpful suggestions. This work was supported by Yamanaka iPS Cell Special Project and CREST of Japan Science and Technology Agency; the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) of the Japanese Society for the Promotion of Science (JSPS); and Grants-in-Aid for Scientific Research from JSPS and MEXT. S.Y. is a member without salary of the scientific advisory boards of iPierian, iPS Academia Japan, Megakaryon Corporation, and HEALIOS K. K. Japan. The iCeMS is supported by World Premier International Research Center Initiative, MEXT, Japan.

Received: June 9, 2013
Revised: August 10, 2013
Accepted: September 11, 2013
Published: October 17, 2013

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.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151–159.

Chen, T., Yuan, D., Wei, B., Jiang, J., Kang, J., Ling, K., Gu, Y., Li, J., Xiao, L., and Pei, G. (2010). E-cadherin-mediated cell-cell contact is critical for induced pluripotent stem cell generation. Stem Cells *28*, 1315–1325.

Cheong, C.Y., and Lufkin, T. (2011). Alternative splicing in self-renewal of embryonic stem cells. Stem Cells Int. *2011*, 560261.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. Cell *136*, 777–793.

Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc. Natl. Acad. Sci. USA *102*, 16176–16181.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell *147*, 132–146.

Gopalakrishnan, S., Van Emburgh, B.O., Shan, J., Su, Z., Fields, C.R., Vieweg, J., Hamazaki, T., Schwartz, P.H., Terada, N., and Robertson, K.D. (2009). A novel DNMT3B splice variant expressed in tumor and pluripotent cells modulates genomic DNA methylation patterns and displays altered DNA binding. Mol. Cancer Res. *7*, 1622–1634.

Han, H., Irimia, M., Ross, P.J., Sung, H.K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241–245.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature *446*, 926–929.

Figure 4. siRNA Screen for RNA-Binding Proteins that Regulate Splicing Patterns in Pluripotent Stem Cells

(A) Schematic of siRNA screen to identify splicing regulators in pluripotent stem cells is presented. siRNAs were transfected into iPSCs (492B4) and ESCs (RF8) in 96-well plates. Splicing patterns were assessed by high-throughput qRT-PCR 48 hr after transfection.

(B) Scatterplots show fold change of each inclusion (Inc.; horizontal) and exclusion (Exc.) form (vertical) relative to negative control siRNA-treated cells. Isoform expressions of Trim33, Ezh2, and Myef2 are shown. siRNAs that changed the ratio by >2-fold are labeled in red.

(C) Targets of siRNAs that affected splicing patterns are shown. Inc/Exc represents the ratio of the inclusion form expression to the exclusion form expression. The upward (↑) and downward (↓) arrows indicate upregulation and downregulation, respectively.

(D) qRT-PCR analysis for knockdown efficiencies by each shRNA in MEFs is presented. The expression levels were normalized to Gapdh. The expression levels in the shNC-treated sample were set to 1. shNC, negative control shRNA. Mean ± SD (n = 3). *$p < 0.05$ for Student's t test comparing to control shRNA-expressing MEFs.

(E) Immunoblotting analysis for endogenous U2af1 and Srsf3 in MEFs and iPSCs (left) and for knockdown efficiencies by each shRNA in MEFs (right) is shown. GAPDH was used as a loading control. endo., endogenous expression; Cont., control vector; O/E, overexpression.

(F) AP staining at day 14 after retroviral infection of the four reprogramming factors and shRNAs is shown. Representative AP-stained reprogramming wells are shown (left). The areas of AP-positive cells per well were scored (right). The AP-positive area in the shNC-treated sample was set to 1. Mean ± SD (n = 3). *$p < 0.05$ for Student's t test comparing to control shRNA-expressing MEFs.

(G) Nanog-GFP intensity at day 14 after retroviral infection of the four reprogramming factors and shRNAs is shown. Nanog-GFP reporter activity was measured by a microplate reader on day 14 after infection. Noninfected MEF was used as a GFP-negative control. The GFP intensity in the shNC-treated sample was set to 1. Mean ± SD (n = 3). *$p < 0.05$ for Student's t test comparing to control shRNA-expressing cells.

See also Figure S4.

Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet. *11*, 75–87.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. *21*, 708–718.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

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.

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

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.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230–240.

Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. Proc. Natl. Acad. Sci. USA *93*, 15081–15085.

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, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

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.