Published in final edited form as:

Nat Cell Biol. 2013 October ; 15(10): 1141–1152. doi:10.1038/ncb2839.

SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells

Xinyi Lu^{#1,2}, Jonathan Göke^{#1}, Friedrich Sachs^{1,3}, Pierre-Étienne Jacques⁴, Hongqing Liang¹, Bo Feng⁵, Guillaume Bourque⁶, Paula A. Bubulya⁷, and Huck-Hui Ng^{1,2,3,8,9,11}

¹Gene Regulation Laboratory, Genome Institute of Singapore, 138672, Singapore

²Department of Biological Sciences, National University of Singapore, 117543, Singapore

³Department of Biochemistry, National University of Singapore, 117597, Singapore

⁴Department of Biology, Université de Sherbrooke, 2500 boulevard de l'Université Sherbrooke, Québec J1K 2R1, Canada

⁵School of Biomedical Sciences, Lo Kwee Seong Integrated Biomedical Sciences Building, Chinese University of Hong Kong, Hong Kong 999077, China

⁶McGill University & Genome Quebec Innovation Center, 740 Dr Penfield Avenue, Montréal, Québec H3A 1A4, Canada

⁷Department of Biological Sciences, Wright State University, Dayton, Ohio 45435, USA

⁸Graduate School for Integrative Sciences & Engineering, National University of Singapore, 117456, Singapore

⁹School of Biological Sciences, Nanyang Technological University, 637551, Singapore

[#] These authors contributed equally to this work.

Abstract

Human embryonic stem cells (hESCs) harbour the ability to undergo lineage-specific differentiation into clinically relevant cell types. Transcription factors and epigenetic modifiers are known to play important roles in the maintenance of pluripotency of hESCs. However, little is known about regulation of pluripotency through splicing. In this study, we identify the spliceosome-associated factor SON as a factor essential for the maintenance of hESCs. Depletion of *SON* in hESCs results in the loss of pluripotency and cell death. Using genome-wide RNA

© 2013 Macmillan Publishers Limited. All rights reserved

¹¹Correspondence should be addressed to H-H.N. nghh@gis.a-star.edu.sg.

Accession numbers. The primary ArrayExpress accession number for the RNA-seq data generated in this study is E-MTAB-1687 (primary accession), which is available at the ArrayExpress. Public data for H3K27ac (referenced accession) were downloaded from the UCSC ENCODE portal (<http://encodeproject.org/ENCODE/>).

AUTHOR CONTRIBUTIONS X.L.: conception, design, collection of data, data analysis and manuscript writing; J.G.: design, collection of data, data analysis and manuscript writing; F.S., H.L. and B.F.: collection and data analysis; P-É.J. and G.B.: collection and data analysis; P.A.B.: design, and provision of study material; H-H.N.: conception, design, supervision, data interpretation and manuscript writing.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Supplementary Information is available in the online version of the paper

profiling, we identified transcripts that are regulated by SON. Importantly, we confirmed that SON regulates the proper splicing of transcripts encoding for pluripotency regulators such as OCT4, PRDM14, E4F1 and MED24. Furthermore, we show that SON is bound to these transcripts *in vivo*. In summary, we connect a splicing-regulatory network for accurate transcript production to the maintenance of pluripotency and self-renewal of hESCs.

Human embryonic stem cells (hESCs) have the unique ability to self-renew and differentiate into cells of the three embryonic lineages^{1,2}. As such, hESCs possess clinical potential, whereby clinically relevant cell types can be derived and potentially used for cell-based therapy^{3,4}. To fully harness the potential of hESCs, it is important to understand the molecular mechanisms that govern hESC identity.

We have previously performed a genome-wide RNA interference (RNAi) screen to identify candidate genes that are important for the maintenance of the hESC identity⁵. Transcription factors and transcription co-factors are over-represented among the top candidates, and this is consistent with their important functions in embryonic stem cells^{6–14} (ESCs). One of the top ten candidates is the spliceosome-associated factor SON, which was initially proposed to be a potential DNA-binding protein^{15,16}. Recent studies suggest that SON is a protein localized in nuclear speckles^{17,18}, and that it is involved in cell-cycle progression and pre-messenger RNA processing^{19–22}. Furthermore, SON was shown to physically interact with the spliceosome and deficiency of *SON* compromises spliceosome function in HeLa cells^{19,21}.

In this study, we examined the function of *SON* in hESCs and found that SON is essential for hESC survival and pluripotency. We show that SON is required for proper splicing of transcripts encoding the cell-cycle protein TUBG1 and the pluripotency regulators OCT4, PRDM14, MED24 and E4F1 in hESCs. Using an RNA immunoprecipitation assay, we further confirm that SON binds to these transcripts. Moreover, we show that *SON* depletion induces the expression of a previously un-annotated isoform of *PRDM14* that lacks the reprogramming capacity of the known isoform. Taken together, our study expands our understanding of hESC pluripotency beyond the transcriptional level and demonstrates that specific regulation of RNA splicing serves as an important regulatory mechanism in hESCs.

RESULTS

SON regulates cell survival and maintenance of pluripotency in hESCs

Our previous genome-wide RNAi screen had identified a repertoire of regulators essential for the maintenance of hESC identity⁵. Although transcription factors and transcription co-factors are enriched among the top candidate genes, we were particularly interested in a candidate (*SON*) that is ranked ninth among the 21,000 analysed genes. *SON* is also ranked the highest among all splicing-related factors. *SON* is expressed in multiple cell types at varying levels (Supplementary Fig. S1a). Importantly, induction of pluripotency in fibroblasts (MRC5) using the 4 Yamanaka factors (OCT4, SOX2, KLF4 and C-MYC) results in increased expression of *SON*, comparable to that observed in hESCs (Supplementary Fig. S1b,c). In addition, hESC differentiation induced by *PRDM14* or

NFRKB depletion or removal of FGF is accompanied by downregulation of *SON* (Supplementary Fig. S1b,d), further highlighting the particular relevance of *SON* in the context of pluripotency. To confirm the finding of our previous RNAi screen result⁵, we designed four short hairpin RNAs (shRNAs) that efficiently deplete *SON* at both transcript and protein levels (Fig. 1a,b and Supplementary Fig. S2a,b). Remarkably, depletion of *SON* resulted in the loss of alkaline phosphatase staining and induced hESC differentiation into fibroblast-like cells (Fig. 1c). The differentiation phenotype of *SON*-knockdown cells was also supported by a downregulation of pluripotency-associated genes (Fig. 1d and Supplementary Fig. S2b) and hESC surface markers (Supplementary Fig. S2c,d). On the other hand, expression of differentiation markers was upregulated whereas the expression of housekeeping genes remained unchanged (Fig. 1e and Supplementary Fig. S2b–e). Interestingly, depletion of *SON* during reprogramming in the presence of 4 Yamanaka factors (OCT4, SOX2, KLF4 and C-MYC) leads to a reduction in the efficiency of induced pluripotency (Supplementary Fig. S2f). Furthermore, co-expression of an RNAi-resistant form of *SON* in hESCs could rescue the effect of *SON* knockdown, confirming that the observed hESC differentiation phenotype and concomitant loss of pluripotency marker expression are *SON*-depletion specific (Fig. 2a,b and Supplementary Fig. S2g). Together, these results show that expression of *SON* is enriched in hESCs and confirm that *SON* is important for the maintenance of the pluripotent state as well as induction of pluripotency.

Besides the differentiation phenotype, *SON* depletion induces a marked reduction of cell number. TUNEL (TdT-mediated dUTP nick end labelling) assay reveals an increased incidence of DNA fragmentation, suggesting that this reduction in cell number is due to cell death (Supplementary Fig. S3a). Notably, despite the cell death and differentiation phenotype of *SON* knockdown in hESCs, knockdown of *SON* in MRC5 fibroblast cells slows down the growth of the cells but does not lead to massive apoptosis (Supplementary Fig. S3b–d), suggesting that *SON* has a cell-type-specific role in hESCs. Moreover, in agreement with previous findings in HeLa cells^{19,21,22}, *SON* depletion in hESCs leads to cell-cycle defects (Supplementary Fig. S3e–g). It has been suggested that cell-cycle arrest may initiate differentiation in hESCs (ref. 23). To exclude the possibility that the observed cell differentiation phenotype is a secondary effect of apoptosis induced by cell-cycle defects, we depleted the cell survival genes *TUBA1C* and *CCNA2* as a control. *TUBA1C* is a major constituent of microtubules, which are crucial for cell-cycle progression and cell survival²⁴. *CCNA2* is a protein important to cell-cycle progression through G1/S and G2/M phase²⁵. Depletion of *TUBA1C* or *CCNA2* in hESCs resulted in increased cell death, but failed to induce fibroblast-like morphology (Supplementary Fig. S3h–i). Furthermore, expression of pluripotency markers remained unchanged following *TUBA1C* or *CCNA2* depletion (Supplementary Fig. S3j–k), indicating that cell death and differentiation in hESCs could occur independently.

To further distinguish the cell death phenotype from the differentiation phenotype, we first analysed the kinetics following *SON* knockdown in hESCs. However, the kinetics of marker induction and cell-cycle changes is similar (Supplementary Fig. S4a–e). Therefore, to investigate the two processes (cell survival versus pluripotency), we inhibited cell apoptosis with the ROCK inhibitor²⁶. Despite a reduction in apoptosis, hESCs underwent

differentiation after depletion of *SON* with concomitant downregulation of pluripotency markers and upregulation of differentiation markers (Fig. 2c–e and Supplementary Fig. S4f–g). Moreover, fluorescence-activated cell sorting analysis following TUNEL assay revealed that only ~50% cells undergo apoptosis after *SON* depletion, whereas *TUBA1C* depletion led to massive cell death (~85%) in hESCs (Supplementary Fig. S4h). Notably, addition of ROCK inhibitor reduced the number of cells undergoing apoptosis after *SON* depletion from 50% to ~4%. Furthermore, staining of the pluripotency marker SOX2 together with the apoptotic marker γ -H2AX (ref. 27) reveals that apoptotic cells expressed the pluripotency gene at a similar level to living cells after *SON* depletion (Supplementary Fig. S4i). In addition, pluripotency marker TRA-1-60 expression is downregulated in cells at all cell-cycle stages examined after *SON* knockdown (Supplementary Fig. S4j). These observations indicate that cell death and differentiation occur concurrently but independently after *SON* depletion.

Genome-wide profiling identifies *SON*-regulated transcripts

SON is known to regulate splicing of transcripts encoding cell-cycle proteins in HeLa cells^{21,22}. To further investigate the function of *SON* in hESCs, we performed paired-end mRNA sequencing (RNA-seq) after a 3-day depletion of *SON* (Fig. 3a and Supplementary Fig. S5a). First, we calculated differential gene expression between control knockdown and *SON* knockdown in hESCs. Genes that show increased expression after depletion of *SON* are significantly enriched in developmental gene ontology (GO) terms, confirming that cells start to undergo differentiation (Fig. 3b and Supplementary Fig. S5b). When we examined the RNA-seq data for *TUBG1*, a known target of *SON*, we found that *SON* depletion indeed results in severe splicing defects (Fig. 3c). Most notably, the data show increased levels of transcripts with intronic sequences, indicating that *SON* depletion induces intron inclusion in mature mRNA. Importantly, the data also confirm the specificity of *SON*-regulated splicing, as only introns 7–10 of *TUBG1* were affected (Fig. 3c).

To identify splicing defects following *SON* depletion, we first identified all splice events using the set of mapped RNA-seq reads that span a splice junction (spliced reads, Fig. 3d, in total 466,085 junctions). We further restricted our analysis to high-confidence junctions that can be associated with known exons (Methods for details). The reduced set covers 173,874 introns and 99% of all spliced reads. We then calculated the number of spliced and unspliced reads that span exon–intron boundaries to estimate the rate of intron inclusion (Supplementary Information for results from alternative estimation of intron expression). Genes that show increased intron inclusion following *SON* depletion are involved in different biological processes, such as the cell cycle, as has been reported previously^{19–21} (Fig. 3e). To identify the most severely affected introns after *SON* depletion, we calculated a *t*-test statistic score and selected a test-score cutoff of –8 (Supplementary Fig. S5c); and we further required that at least 5 spliced reads support the splice site. Using these thresholds, we obtained a set of 1,994 introns from 1,127 genes whose splicing was impaired following *SON* depletion (Supplementary Table S1). From this set, we validated 30 intron inclusion events using PCR with reverse transcription (RT–PCR), confirming the robustness of our method to correctly identify splicing defects after depletion of *SON* (Supplementary Fig. S5d,e). Reassuringly, our analysis identifies previously known targets of *SON* (*TUBG1*,

TUBGCP2, *AURKB*, *FANCG* and *HDAC6*, Supplementary Table S1). Consistent with what we observed for *TUBG1*, most of these introns show a significant increase of intronic signal when compared with exonic signal after the knockdown of *SON*, confirming intron inclusion as a major splicing defect (Fig. 3f,g and Supplementary Fig. S5f).

SON regulates splicing of specific short introns with weak splice sites

To better understand the mechanism behind SON-regulated splicing, we investigated whether SON-regulated introns exhibit specific properties. For this analysis we selected introns from the same genes, but which did not show significant inclusion, as a control to correct for gene to gene variations. Interestingly, we find that SON-regulated introns are significantly shorter than unaffected introns from the same genes (Fig. 4a, Wilcoxon test $P = 3.85 \times 10^{-131}$; Supplementary Fig. S6a,b). Furthermore, we found that genes with introns regulated by SON had a significantly weaker splicing strength when compared with unaffected introns (Fig. 4b,c, Wilcoxon test $P = 1.62 \times 10^{-16}$ (5'); $P = 3.51 \times 10^{-11}$ (3'); Supplementary Fig. S6c,d; ref. 28). These observations are consistent, as splicing strength tends to positively correlate with intron length²⁹. As splicing strength is a property of the splice junction sequence, we calculated the nucleotide frequency for SON-targeted introns. Strikingly, SON-targeted introns show a very distinct sequence from the genome-wide average (Fig. 4d,e). Even when we considered only SON-targeted introns and non-targeted introns from the same genes, we still observed a difference in sequence composition, most notably a higher GC content in SON-targeted introns (Fig. 4f and Supplementary Fig. S6e–g). Introns with high levels of GC content have been associated with intron inclusion events³⁰. In summary, these results highlight the large-scale splicing deficiency at preferentially short introns with weak splice sites following *SON* knockdown in hESCs, and indicate that sequence-specific properties might be involved in SON-mediated regulation of splicing.

As the activity of SON could be due to a general targeting effect of SON on all weak spliced introns, we determined the expression of other introns with weak splicing sites after *SON* depletion in hESCs to exclude this possibility. We found that splicing of weak introns from *CEBPZ*, *SPAG5* and *SMARCE1* is not affected by *SON* depletion (Supplementary Fig. S6h). In addition, we examined the effect of depletion of another weak splicing regulator (*TIA1*) to determine whether it functions the same as SON in hESCs. Knockdown of *TIA1* does not cause morphological changes in hESCs (Supplementary Fig. S6i), and the expression of pluripotency markers as well as differentiation markers remains unchanged (Supplementary Fig. S6j). There are also no changes in splicing for introns regulated by SON after *TIA1* knockdown (Supplementary Fig. S6k). These data suggest that SON exerts a specific function in regulating splicing of transcripts.

SON regulates pluripotency by promoting splicing of pluripotency genes

As *SON* expression is required for maintaining pluripotency in hESCs, we investigated whether splicing of pluripotency-associated genes is affected by *SON* depletion. By overlapping SON-regulated transcripts with genes essential for hESC identity maintenance⁵, we identified 41 candidates where the observed splicing defects might directly affect pluripotency (Fig. 5a, $P = 0.0093$). Using an independent set of hESC-specific genes, we

confirmed the significant enrichment of pluripotency-associated genes in the set of SON-targeted genes (Supplementary Fig. S7a, P value = 8.8×10^{-5}). Moreover, by overlapping SON-targeted genes with genes essential for hESC cell survival⁵, we found that SON regulates splicing of 45 genes essential for hESC survival (Fig. 5b, $P = 0.00016$). These introns regulated by SON from pluripotency genes and cell survival genes are also shorter than non-targeted introns (Supplementary Fig. S7b–d). Notably, *OCT4*, *PRDM14*, *MED24* and *E4F1* are among the pluripotency genes targeted by SON (Fig. 5a,c). Whereas all four genes are important regulators of the hESC identity, both *OCT4* and *PRDM14* are also key transcription factors involved in induced pluripotency^{5,31}.

To confirm that *SON* depletion affects splicing of these key factors, we validated these predictions using quantitative PCR (qPCR). Indeed, SON-targeted introns from *OCT4*, *PRDM14*, *MED24* and *E4F1* transcripts were upregulated (Fig. 5d,e), whereas non-targeted introns from the same genes were not affected following knockdown of *SON* (Fig. 5d,e). A time-course analysis confirms that intron inclusion for these transcripts increased with extended time of *SON* depletion (Supplementary Fig. S7e). This may explain why significant cell differentiation occurs only at a certain time point after a sufficient amount of unspliced transcripts accumulated at day 3.5. By comparing the expression of all the introns and exons in the *OCT4* transcript, we found that only specific introns are regulated by SON (Supplementary Fig. S7f). Notably, another important pluripotency regulator, *NANOG*, is not affected by the depletion of *SON* (Supplementary Fig. S7g). This suggests that SON might be governing hESC pluripotency in a specific manner.

To address whether SON is directly bound to these transcripts in hESCs, we performed RNA crosslinking and immunoprecipitation (RNA-CLIP) assays. SON was first crosslinked to RNA by ultraviolet irradiation. The SON–RNA complexes were then purified using SON-specific antibodies. Real-time PCR analysis showed enrichment of unspliced transcripts of *OCT4*, *PRDM14*, *MED24*, *E4F1* and *TUBG1* for SON RNA-CLIP, but not for control antibody RNA-CLIP (Fig. 5f). The intronic sequence for the negative control transcript *TUBA1B* was not enriched by SON RNA-CLIP (Fig. 5f). Hence, the RNA-CLIP results indicate that SON physically interacts with specific introns from *OCT4*, *PRDM14*, *MED24*, *E4F1* and *TUBG1*. In summary, these data indicate that SON regulates hESC pluripotency and cell survival by promoting specific splicing of introns from transcripts encoding pluripotency regulators and cell-cycle proteins.

SON-regulated transcripts are subjected to nonsense-mediated decay

SON depletion in hESCs leads to the accumulation of a number of transcripts containing unspliced introns; these unspliced transcripts, which may contain a premature stop codon, could be regulated by the nonsense-mediated mRNA decay (NMD) pathway^{32–34}. Interestingly, the expression of cytoplasmic proteins essential for the NMD pathway, such as UPF1, is affected by *SON* depletion as well (Fig. 6a–c). This might explain why intron retention is observed in poly(A) mRNA after *SON* depletion. The degradation of transcripts through the NMD pathway eventually leads to the reduction of corresponding protein expression (Fig. 6c). Depletion of *SON* together with *UPF1* and *UPF2* leads to increased

accumulation of unspliced transcripts (Fig. 6d), confirming that unspliced transcripts are regulated by the NMD pathway.

SON regulates the alternative splicing in hESCs

As SON has an important splicing regulatory function, we studied whether alternative splicing can be observed following *SON* depletion in hESCs besides intron inclusion. To detect alternative splicing events, we calculated the fold change of the read count at every exon–exon junction (Fig. 7a). As we were interested in a change of junction usage relative to all other junctions of the same gene, we calculated differential splicing for every gene separately, thereby correcting for the overall change of expression. In total, we identified 940 splicing events in 694 genes that show a change after depletion of *SON* (Supplementary Table S2). Strikingly, most of these alternative splicing events are skipped exons after knockdown of *SON* (Fig. 7b) and most encode minor isoforms (Fig. 7c). Interestingly, many alternative splicing events overlap SON-regulated introns (Fig. 7d, Fisher's test $P = 5.3 \times 10^{-123}$), suggesting that exon skipping is indeed a result of loss of the splicing regulatory function of SON.

SON knockdown induces the alternative splicing of *PRDM14*

As for intron inclusion events, we find that *SON*-knockdown-induced alternative splicing events significantly overlap with hESC-specific genes (Fig. 8a, Fisher's test $P = 9.66 \times 10^{-8}$) as well as genes regulating pluripotency and survival of hESCs (Supplementary Fig. S8a,b). One of the top candidates of SON-regulated alternative splicing events was again from the gene coding for the pluripotency regulator *PRDM14*. The RNA-seq data show that the expression of exon 2 but not other exons from *PRDM14* was specifically downregulated after *SON* depletion (Fig. 8b). At the same time, there was a significant increase in transcripts that connect exons 1 and 3, corresponding to a previously un-annotated shorter isoform of *PRDM14* (Fig. 8b). Therefore, we validated the isoform expression of *PRDM14* after *SON* depletion using RT–PCR. The results strikingly show the isoform switch following *SON* depletion (Fig. 8c). We identified a shorter isoform of *PRDM14* present in hESCs and this isoform is upregulated after *SON* depletion whereas the full-length known isoform of *PRDM14* is downregulated (Fig. 8c). By sequencing the RT–PCR product, we confirmed that exon 2 of *PRDM14* was skipped in the *PRDM14* short isoform (Fig. 8c). By co-expressing either *PRDM14* isoform with a reporter carrying *PRDM14* motifs, we found that the *PRDM14* long isoform but not the short isoform is functional as an activator of gene expression (Fig. 8d). Moreover, only the long *PRDM14* isoform could enhance reprogramming efficiency, whereas the short isoform failed, suggesting that the long isoform is the functional isoform in the maintenance and gain of pluripotency (Fig. 8e). Previously it was shown that *PRDM14* is essential for silencing the expression of developmental genes³⁵. Thus, this isoform switch of pluripotency factor *PRDM14* may potentially serve as a main downstream event driving the subsequent differentiation of hESCs. Together, these data indicate that SON links splicing with the core regulatory network that maintains pluripotency by regulating correct splicing of transcripts encoding pluripotency factors.

DISCUSSION

We propose a model (Fig. 8f) whereby SON interacts with other splicing factors in the spliceosome^{19,21} to promote the splicing of introns and inclusion of alternative exons from genes essential for hESC identity. SON may provide a control switch for appropriate processing of a specific set of transcripts that alter cell fate. Hence, splicing regulation by SON is important for normal expression of these pluripotency genes and cell survival genes in the maintenance of hESC identity.

Transcript splicing is a key regulatory step for the proper expression of intron-containing genes^{36,37}. Although it is known that profound changes in alternative splicing occur during ESC differentiation^{38,39}, the precise mechanism(s) that regulate this process is not well studied. In mouse ESCs, alternative splicing has been found to play a major role in cell fate determination during differentiation⁴⁰. In hESCs, a deficit in the splicing factor FOX2 leads to cell death⁴¹. An ESC-specific splicing switch in *FOXP1* transcripts has been reported to play a role in the maintenance of pluripotency⁴². More recently, MBNL proteins were found to inhibit ESC-specific alternative splicing⁴³.

Apart from cell survival and cell fate determination, SON-regulated splicing is also essential for the maintenance of pluripotency in hESCs. Our study has demonstrated that depletion of the splicing factor SON led to both cell death and differentiation of hESCs. The differentiation phenotype observed after knockdown of *SON* is conversely unique to hESCs. Hence, we propose that SON regulates hESC identity by influencing the splicing of transcripts encoding pluripotency factors such as OCT4, PRDM14, MED24 and E4F1. Notably, OCT4 is a member of core transcriptional circuitry in hESCs, and mis-regulation of *OCT4* expression causes cell differentiation^{31,44}. PRDM14 was found to play a central role in the maintenance and gain of pluripotency^{5,35}. Mediator complexes were also found to play an important role in maintaining ESC identity⁴⁵, and *MED24* depletion in hESCs similarly causes cell differentiation⁵. E4f1, a transcription factor that was also previously identified in a hESC RNAi screen⁵, has been shown to be essential to mouse ESC maintenance⁴⁶. Hence, the inappropriate splicing of *OCT4*, *MED24* and *E4F1* transcripts and isoform switch of *PRDM14* may contribute to the differentiation phenotype induced by *SON* deficiency in hESCs. Together, our study links the splicing-regulatory network with pluripotency through SON, which targets genes that encode for key hESC regulators, demonstrating how accurate splicing is required to maintain hESC identity.

METHODS

hESC culture and transfection

The hESC line H1 (WA-01, passage 30) was cultured feeder-free in conditioned medium on Matrigel (BD; ref. 48). Conditioned medium contains 20% Knock-out serum replacement (Gibco), 1 mM L-glutamine (Gibco), 1% non-essential amino acids (Gibco) 0.1 mM 2-mercaptoethanol (Gibco) and 4 ng ml⁻¹ human basic fibroblast growth factor (bFGF; Invitrogen), and medium was supplemented with additional bFGF (8 ng ml⁻¹) before use. The hESCs were passaged for expansion with 1 mg ml⁻¹ collagenase IV (Gibco) every 5–7 days. For transfection, the hESCs were passaged with 0.25% trypsin (Invitrogen) for

dissociation and transfected with 1.5 μ g shRNA and 4.5 μ l Mirus LT1 transfection reagents the following day. Cells were selected under 1 μ g ml⁻¹ puromycin after transfection and collected 4 days after transfection for RNA. For RNA-seq, cells were collected for mRNA 3 days after transfection of *SON* shRNA1. For rescue experiments, CAG-*SON* was transfected into hESCs to establish hESCs overexpressing *SON*. shRNA (1 μ g) targeting the UTR region of *SON* or control shRNA was transfected into hESCs. Cells were collected 4 days after transfection for RNA or cell imaging. For inhibition of apoptosis, ROCK inhibitor was added 3 days after depletion of *SON* and incubated for 1 day before collection. For depletion of *SON*, *UPF1* and *UPF2* at the same time, 1.2 μ g *SON* shRNA was transfected together with 1.2 μ g of *UPF1* and *UPF2* shRNA. Sequences of shRNAs: 5'-GATGAAATGGGTAAAGTACA-3' (control shRNA), 5'-GCTGAGCGCTCTATGATGT-3' (*SON* shRNA1), 5'-GATACAGAACTACGATATA-3' (*SON* shRNA2), 5'-GGTCTTTCGTGGTCAGTAA-3' (*SON* shRNA3), 5'-AATGTCAGTGGAGTATCA-3' (*SON* shRNA4), 5'-GGACTAGCGAGGAGGAGTT-3' (*SON* shRNA UTR), 5'-GGCTGCCCTTGAGAAGGATTA-3' (*TUBA1C*), 5'-CCATTGGTCCCTCTTGATT-3' (*CCNA2*), 5'-GATGCAGTTCGCTCCATT-3' (*UPF1*; ref. 49), 5'-GAAGTTGGTACGGGCATC-3' (*UPF2*; ref. 50), 5'-GCTCTAATTCTGCAACTCTTT-3' (*TIA1* shRNA1; ref. 51), 5'-GCTCTAATTCTGCAACTCTTT-3' (*TIA1* shRNA2; ref. 51). The cell line used has not been authenticated recently but it has been tested for mycoplasma contamination.

Virus production and reprogramming

SON shRNAs, *PRDM14* or the *PRDM14* isoform was cloned into pMX vector for retrovirus-mediated knockdown. Retroviruses were packaged using the Pantropic Retroviral Expression System (Clontech) and concentrated with centrifugal filter devices (Millipore). MRC-5 human lung primary fibroblast cells obtained from ATCC were cultured in 15% FBS/DMEM. Confluent MRC-5 cells from a 6-well plate were split into 24 wells one day before and then transduced with retroviruses containing 4 Yamanaka factors (OCT4, SOX2, KLF4 and C-MYC) with or without custom viruses in the presence of 4 μ g ml⁻¹ Polybrene (Sigma). After 24 h, the cells were changed to fresh 15% FBS/DMEM medium. MRC-5 cells were passaged on CF-1 feeder cells and maintained in conditioned hESC medium for 3–4 weeks before collection for immunostaining.

Reverse transcription and qPCR

Total RNA was extracted from cells using the TRIzol reagent (Invitrogen) and was then treated with DNase (Promega) for 1 h. The concentration of RNA was determined by NanoDrop 2000 (ThermoScientific). DNase-treated total RNA (500 ng) was reverse transcribed with oligo(dT) and SuperScript II (Invitrogen). qPCR was performed with 2 μ l 5 $\times$ diluted cDNA per 10 μ l reaction and KAPA SYBR Green (KapaBiosystems). Quantitative real-time PCR was performed on the PRISM 7900HT system (Applied Biosystem). Gene expression levels were then normalized to that of β -actin. For semi-quantitative RT-PCR, 2 μ g total RNA was reversed transcribed. For downstream PCR, 1/20 of the reverse transcription product was used. Band intensity was determined by ImageJ and the percentage of unspliced transcript was calculated. Primers used for detection of intron inclusion are listed in Supplementary Table S3.

Alkaline phosphatase staining

The alkaline phosphatase detection kit (Chemicon) was used according to the manufacturer's instructions to determine alkaline phosphatase activity. H1 hESCs were cultured in a 6-well plate, and then fixed with 4% formaldehyde at 25 °C for 30 min. Cells were then permeabilized in PBS containing 0.1% Tween-20 for 10 min. Staining solution was later added to cells and incubated for 30 min. Cells were washed with PBS 3 times and visualized by microscopy. Images of $\times 200$ magnification were captured by the Leica Application Suite using a Leica DM IRD microscope. Experiments were repeated 3 times.

TUNEL assay and fluorescence-activated sorting analysis

TUNEL assays were performed using the *in situ* Cell Death Detection Kit (Roche). H1 hESCs were cultured in a 6-well plate, fixed with 4% formaldehyde at 4 °C overnight and incubated with the TUNEL labelling mixture for 1 h at 37 °C. Plates were washed with PBS and imaged at $\times 200$ magnification. Experiments were repeated 3 times. To quantify the TUNEL assay, fluorescence-activated sorting analysis was used to determine the composition of the cell population after staining. H1 hESCs were cultured in a 10 cm plate and treated with *SON* shRNA and control shRNA. After 4-day treatment, cells were detached from the plate by trypsin treatment and dissociated into single cells. TUNEL assay was performed with these cells, and cells were analysed at 530 nm. For cell-cycle analysis, *SON* was depleted from 2 to 3.5 days. Cells were then collected and fixed in ethanol overnight. Propidium iodide was used to stain the cells for cell-cycle analysis by fluorescence-activated sorting analysis.

Immunostaining

Human ESCs were fixed with 4% formaldehyde in PBS. After permeabilization in 1% Triton X-100/PBS for 30 min, cells were blocked in 1% BSA for 30 min. Immunostaining was performed using the following primary antibodies: SSEA-4 (clone 813–70, sc-21704, Santa Cruz), GATA6 (sc-7244, Santa Cruz), MSX1 (sc-17726, Santa Cruz), Tra-1-81 (clone TRA-1–80, sc-21706, Santa Cruz), TRA-1-60 (clone TRA-1-60, sc-21705, Santa Cruz), OCT4 (clone C30A3, #2840, Cell Signaling), Gamma-H2AX (clone JBW301, 05-636, Millipore), AURORA A (clone 35C1, ab13824, Abcam), α -tubulin (clone B-5-1-2, T5168, Sigma). Secondary antibodies used were Alexa Fluor 546 anti-mouse IgG (A11003, Invitrogen), Alexa Fluor 546 anti-rabbit IgG (A11010, Invitrogen), Alexa Fluor 546 anti-goat IgG (A11056, Invitrogen), Alexa Fluor 488 anti-rabbit IgG (A21441, Invitrogen), Alexa Fluor 488 anti-goat IgG (A21467, Invitrogen), Alex Fluor 488 (A21200, Invitrogen) anti-mouse (A21200, Invitrogen), and Alexa Fluor 647 anti-mouse IgG (A21463, Invitrogen). Hoechst (Invitrogen) was used for staining the nuclei. Immunostaining experiments were repeated 3 times.

Western blot

Protein concentration was measured with a Bradford assay kit (Bio-Rad). Cell lysate (50 μ g) was resolved on a 10% SDS–polyacrylamide gel and transferred to a polyvinylidenedifluoride membrane (Millipore). After blocking in 5% skimmed milk for 1 h, the appropriate primary antibodies were added: anti-SON (rabbit polyclonal antibody

generated against the amino-terminal 250 amino acids of SON), anti-Oct4 (1:5,000, ab19857, Abcam), anti-NANOG (1:800, AF1997, R&D), anti-UPF1 (1:1,000, sc-48802, Santa Cruz), anti-DNMT1 (1:1,000, sc-20701, Santa Cruz), anti-PRDM14 (generated against the N-terminus of PRDM14)⁵, anti-SF2 (1:1,000, sc-28724, Santa Cruz), anti-GAPDH (1:1,000, sc-25778, Santa Cruz) or anti- β -actin (1:1,000, clone C4, sc-47778, Santa Cruz) primary antibodies for 1 h. After washing with PBS/0.1% Tween-20, horse-radish peroxidase (HRP)-conjugated anti-rabbit IgG (1:5,000, NA934V, Amersham), HRP-conjugated anti-goat IgG (1:5,000, sc-2020, Santa Cruz) or HRP-conjugated anti-mouse IgG (1:5,000, NA931V, Amersham), secondary antibodies were added for 1 h. After washing with PBST, signals were then detected using the West Dura Extender Duration Substrate (ThermoScientific). Western blot experiments were repeated twice.

RNA preparation and sequencing

Trizol (Invitrogen) was used for extraction of total RNA from hESCs treated with control shRNA or *SON*shRNA for 3 days. RNA (4 μ g) extracted from biological duplicate samples was poly(A) enriched by incubating with oligo(dT)-coated Sepharose beads for two rounds (Ambion). Of the resultant enriched mRNA, 500 ng was fragmented, size selected and reverse transcribed according to the IlluminaTruSeq RNA Sample Preparation Guide. Multiplexed samples were sequenced with the IlluminaHighSeq system. Each library was sequenced for 2×80 million of 75-nucleotide reads.

Bioinformatics analysis of RNA-sequencing data

The reads of *SON*-shRNA-treated and control-shRNA-treated samples were mapped to the human genome assembly hg19 using TopHat⁵² version 1.4.1 with parameters -M -butterfly-search -microexon-search -n 2 -m 1 -N 3 -i 30 -coverage search, we also used the -GTF option with transcriptome annotations from Ensembl (GRCh37.69). Differential expression was estimated using cufflinks version 2.02 (ref. 53). Genes with increased expression in Fig. 3b were selected on the basis of the expression log-fold change between control and knockdown (cutoff 0.5). Intron inclusion was estimated by two different scores. As a first score we calculated the ratio of spliced over unspliced reads. Reads were counted that either mapped to the two nucleotides 5' and 3' of the splice site (unspliced reads) or that mapped only to the exonic nucleotide and to a second exon for the rest of the read (spliced read). For every sample we then divided the number of spliced reads for every junction by the number of unspliced reads that cover the 3' or 5' splice site. A pseudo-count of 1 was added both for spliced and unspliced reads. As the intron inclusion score we calculate the *t*-test statistic for the ratio of spliced over unspliced reads between control and knockdown. To obtain a set of high-confidence intron inclusion events, we required a test score below -8, a minimum number of 5 spliced reads that support the junction, and a minimum number of at least 3 unspliced reads for the 3' and 5' splice sites (Supplementary Table S1). For unaffected introns we required a test score between -5 and 1.5 and a minimum number of 5 reads that support the splice junction. To control for gene to gene variation we compared introns from the set of genes that contain both affected and unaffected introns. As a second score, we estimated intron expression. We counted reads that map into introns divided by the length of the introns and normalized this to the number of reads mapping into the adjacent exons (divided by the length of the exons). MaxEntScan was used to estimate splicing strength for

all splice sites detected in our RNA-seq data using the Maximum Entropy model²⁹. To estimate alternative splicing, we calculated the read count for all junctions used in this study (see main text). We associated every junction with the gene that was annotated for the adjacent exons. The log-fold change and significance were estimated using DESeq2 (ref. 54) for every gene separately to get estimates independent of the change in expression. Box plots were generated in R using standard settings. All bioinformatics analyses were carried out using R version 3.0 and Bioconductor version 2.12.

RNA crosslinking and immunoprecipitation

RNA immunoprecipitation experiments were performed according to a published protocol with modifications⁵⁵. Briefly, hESCs were ultraviolet-irradiated at 254 nm, 600 mJ cm⁻² (Stratagene Stratalinker) and cells were lysed in Triton buffer with short disruptive sonication. DNA was removed from nuclear lysate by 15 min DNase treatment and RNA was fragmented by 5 min RNase A treatment at 37 °C. Nuclear lysates were precleared with Protein G Dynabeads (Invitrogen) for 2 h at 4 °C, and then incubated along with antibodies overnight. RNA was immunoprecipitated with Protein G Dynabeads. Proteins were removed by Proteinase K (Fermentas) treatment after RNA elution. RNA was extracted using Trizol (Invitrogen) and RT-qPCR was performed as described above.

Luciferase assay

For the *Luciferase* assay, 75 ng pGL4 reporter with minimal promoter or plus 5xPRDM14-binding motifs was transfected into HEK293T cells (ATCC) together with 75 ng *PRDM14* or *PRDM14* isoform and 3 ng SV40-*Renilla* expression plasmid. Transfected cells were maintained in culture for 3 days. Cells were collected and the *Luciferase* assay was performed according to the Dual-Luciferase Reporter Assay System Protocol (Promega). The activity of the Luciferase gene is normalized against the activity of *Renilla*. The effect of PRDM14 or the PRDM14 isoform on reporter activity was normalized against minimal promoter or promoter with 5xPRDM14-binding motifs. The cell line used has not been authenticated recently but has been tested for mycoplasma contamination.

Statistical analysis

Statistical analysis was performed using R. Data are presented as mean±s.e.m. from biological triplicate data unless noted otherwise. Two-tailed Student's *t*-test was appropriately used to estimate the difference between two groups unless noted otherwise. The data are expected to be normally distributed and the variance is expected to be similar between the groups that are being statistically compared.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the Agency for Science, Technology and Research (A*STAR) for their generous funding, and also for NIH R15 GM084407 to P.A.B. We acknowledge T. C. Peow, L-P. Yaw, T. Tng and G. K. Jee for their technical assistance. We acknowledge the Genome Technology Biology (GTB) group from the Genome Institute of

Singapore (GIS) for their help and expertise in performing the RNA sequencing. We thank T-L. Huber for critical comments on the manuscript. J. Göke is supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service, DAAD.

References

1. Thomson JA, et al. Embryonic stem cell lines derived from human blastocysts. *Science*. 1998; 282:1145–1147. [PubMed: 9804556]
2. Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A. Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nat. Biotechnol.* 2000; 18:399–404. [PubMed: 10748519]
3. Wobus AM, Boheler KR. Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol. Rev.* 2005; 85:635–678. [PubMed: 15788707]
4. Kimbrel EA, Lu SJ. Potential clinical applications for human pluripotent stem cell-derived blood components. *Stem Cells Int.* 2011; 2011:273076. [PubMed: 21437192]
5. Chia NY, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*. 2010; 468:316–320. [PubMed: 20953172]
6. Chen X, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008; 133:1106–1117. [PubMed: 18555785]
7. Göke J, Chan YS, Yan J, Vingron M, Ng HH. Genome-wide kinasechromatin interactions reveal the regulatory network of ERK signaling in human embryonic stem cells. *Mol. Cell*. 2013; 50:844–855. [PubMed: 23727019]
8. Göke J, et al. Combinatorial binding in human and mouse embryonic stem cells identifies conserved enhancers active in early embryonic development. *PLoS Comput. Biol.* 2011; 7:e1002304. [PubMed: 22215994]
9. Jiang J, et al. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* 2008; 10:353–360. [PubMed: 18264089]
10. Loh YH, et al. Genomic approaches to deconstruct pluripotency. *Annu. Rev. Genom. Hum. Genet.* 2011; 12:165–185.
11. Macarthur BD, Ma'ayan A, Lemischka IR. Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* 2009; 10:672–681. [PubMed: 19738627]
12. Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. *Cell*. 2011; 145:835–850. [PubMed: 21663790]
13. Young RA. Control of the embryonic stem cell state. *Cell*. 2011; 144:940–954. [PubMed: 21414485]
14. Yuan P, et al. Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev.* 2009; 23:2507–2520. [PubMed: 19884257]
15. Khan IM, et al. The SON gene encodes a conserved DNA binding protein mapping to human chromosome 21. *Ann. Hum. Genet.* 1994; 58:25–34. [PubMed: 8031013]
16. Sun CT, et al. Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. *J. Biol. Chem.* 2001; 276:24059–24067. [PubMed: 11306577]
17. Wynn SL, et al. Organization and conservation of the GART/SON/DONSON locus in mouse and human genomes. *Genomics*. 2000; 68:57–62. [PubMed: 10950926]
18. Saitoh N, et al. Proteomic analysis of interchromatin granule clusters. *Mol. Biol. Cell*. 2004; 15:3876–3890. [PubMed: 15169873]
19. Huen MS, et al. SON is a spliceosome-associated factor required for mitotic progression. *Cell Cycle*. 2010; 9:2679–2685. [PubMed: 20581448]
20. Sharma A, Takata H, Shibahara K, Bubulya A, Bubulya PA. Son is essential for nuclear speckle organization and cell cycle progression. *Mol. Biol. Cell*. 2010; 21:650–663. [PubMed: 20053686]
21. Ahn EY, et al. SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol. Cell*. 2011; 42:185–198. [PubMed: 21504830]
22. Sharma A, et al. Son maintains accurate splicing for a subset of human pre-mRNAs. *J. Cell Sci.* 2011; 124:4286–4298. [PubMed: 22193954]

23. Ruiz S, et al. A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr. Biol.* 2011; 21:45–52. [PubMed: 21167714]
24. Sorger PK, Dobles M, Tournéize R, Hyman AA. Coupling cell division and cell death to microtubule dynamics. *Curr. Opin. Cell Biol.* 1997; 9:807–814. [PubMed: 9425345]
25. Pagano M, Pepperkok R, Verde F, Ansorge W, Draetta G. Cyclin A is required at two points in the human cell cycle. *EMBO J.* 1992; 11:961–971. [PubMed: 1312467]
26. Watanabe K, et al. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* 2007; 25:681–686. [PubMed: 17529971]
27. Rogakou EP, Nieves-Neira W, Boon C, Pommier Y, Bonner WM. Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139. *J. Biol. Chem.* 2000; 275:9390–9395. [PubMed: 10734083]
28. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* 2004; 11:377–394. [PubMed: 15285897]
29. Dewey CN, Rogozin IB, Koonin EV. Compensatory relationship between splice sites and exonic splicing signals depending on the length of vertebrate introns. *BMC Genom.* 2006; 7:311–319.
30. Amit M, et al. Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* 2012; 1:543–556. [PubMed: 22832277]
31. Hay DC, Sutherland L, Clark J, Burdon T. Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells.* 2004; 22:225–235. [PubMed: 14990861]
32. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 2007; 76:51–74. [PubMed: 17352659]
33. Ishigaki Y, Li X, Serin G, Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell.* 2001; 106:607–617. [PubMed: 11551508]
34. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* 2004; 5:89–99. [PubMed: 15040442]
35. Chan YS, et al. A PRC2-dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem cell reprogramming. *Stem Cells.* 2012; 31:682–692. [PubMed: 23280602]
36. Fedor MJ. Alternative splicing minireview series: combinatorial control facilitates splicing regulation of gene expression and enhances genome diversity. *J. Biol. Chem.* 2008; 283:1209–1210. [PubMed: 18024424]
37. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev.* 2003; 17:419–437. [PubMed: 12600935]
38. Yeo GW, et al. Alternative splicing events identified in human embryonic stem cells and neural progenitors. *PLoS Comput. Biol.* 2007; 3:1951–1967. [PubMed: 17967047]
39. Salomonis N, et al. Alternative splicing in the differentiation of human embryonic stem cells into cardiac precursors. *PLoS Comput. Biol.* 2009; 5:e1000553. [PubMed: 19893621]
40. Salomonis N, et al. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc. Natl Acad. Sci. USA.* 2010; 107:10514–10519. [PubMed: 20498046]
41. Yeo GW, et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* 2009; 16:130–137. [PubMed: 19136955]
42. Gabut M, et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell.* 2011; 147:132–146. [PubMed: 21924763]
43. Han H, et al. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature.* 2013; 498:241–245. [PubMed: 23739326]
44. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell.* 2005; 122:947–956. [PubMed: 16153702]
45. Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature.* 2010; 467:430–435. [PubMed: 20720539]

46. Le Cam L, Lacroix M, Ciemerych MA, Sardet C, Sicinski P. The E4F protein is required for mitotic progression during embryonic cell cycles. *Mol. Cell Biol.* 2004; 24:6467–6475. [PubMed: 15226446]
47. ENCODE Project Consortium. Myers RM, et al. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 2011; 9:e1001046. [PubMed: 21526222]
48. Xu C, et al. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* 2001; 19:971–974. [PubMed: 11581665]
49. Paillusson A, Hirschi N, Vallan C, Azzalin CM, Muhlemann O. A GFP-based reporter system to monitor nonsense-mediated mRNA decay. *Nucleic Acids Res.* 2005; 33:e54. [PubMed: 15800205]
50. Wittmann J, Hol EM, Jack HM. hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol. Cell. Biol.* 2006; 26:1272–1287. [PubMed: 16449641]
51. Izquierdo JM, Alcalde J, Carrascoso I, Reyes R, Ludena MD. Knockdown of T-cell intracellular antigens triggers cell proliferation, invasion and tumour growth. *Biochem. J.* 2011; 435:337–344. [PubMed: 21284605]
52. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25:1105–1111. [PubMed: 19289445]
53. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
54. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010; 11:R106. [PubMed: 20979621]
55. Zhao J, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell.* 2010; 40:939–953. [PubMed: 21172659]

Figure 1.

The effects of *SON* depletion in hESCs. **(a)** Expression level of *SON* transcript after *SON* shRNA treatment as determined by qPCR. Four shRNAs were designed to target *SON* independently. shRNA against the *Luciferase* gene was used as the control. Samples were analysed 4 days after transfection. Mean $\pm$ s.e.m. from biological triplicate data is shown in all graphs. **(b)** Immunostaining of SON in hESCs. DAPI positively stains the nucleus. **(c)** Alkaline phosphatase staining of hESCs after *SON* depletion. **(d)** Expression of pluripotency markers as determined by qPCR after *SON* depletion. Biological triplicate data for qPCR are presented as mean $\pm$ s.e.m. **(e)** Expression of differentiation markers as determined by qPCR after *SON* depletion. Biological triplicate data are presented as mean $\pm$ s.e.m. Scale bars, 20 μ m.

Figure 2.

Rescue of *SON* depletion phenotype in hESCs. **(a)** Expression of *SON* and pluripotency markers in hESCs expressing *SON* shRNA against *SON* UTR, *SON* shRNA+RNAi immune *SON* or control shRNA. Mean±s.e.m. of biological triplicate data is shown in all graphs. The *P* value was determined by two-tailed Student's *t*-test. **(b)** Immunostaining of pluripotency cell surface marker after transfecting hESCs carrying RNAi immune *SON* with shRNA targeting the UTR of the *SON* transcript. **(c)** Cell morphology of hESCs transfected by *SON* shRNA and treated with ROCK inhibitor (inh.). **(d)** Expression of *SON* and pluripotency markers after depletion of *SON* in the presence of ROCK inhibitor. Mean±s.e.m. from biological triplicate data is shown. **(e)** Expression of differentiation markers after depletion of *SON* in the presence of ROCK inhibitor. Mean±s.e.m. from biological triplicate data is shown. Scale bars, 20 μ m.

Figure 3.

RNA-seq analysis of transcripts regulated by SON. **(a)** Scheme of the RNA-seq analysis. Cells were transfected with constructs expressing *SON* shRNA (shRNA1) or control-shRNA-expressing constructs for 3 days. RNA was extracted from pooled samples. Next, poly(A)-tailed mRNA was enriched and fragmented for sequencing. Data analysis was carried out for three aspects: gene expression, intron expression and alternative splicing. **(b)** Gene ontology analysis of upregulated genes after 3-day *SON* depletion according to RNA-seq data. **(c)** Fold change of RNA-seq reads from control and *SON* knockdown at the *TUBG1* locus (hg19). Shown are fold changes with $P < 0.4$. Introns that are known *SON* targets are marked by asterisks. **(d)** Intron inclusion is estimated using the ratio of spliced over unspliced reads that cover exon–intron boundaries. See Supplementary Information for alternative estimation of intron inclusion. **(e)** Gene ontology analysis of genes with intron inclusion after *SON* depletion. **(f)** Number of events with increased and decreased intron inclusion after depletion of *SON*. **(g)** Log ratio of spliced over unspliced reads at *SON*-targeted introns for control and *SON* knockdown. Significance was estimated using the Wilcoxon signed-rank test.

Figure 4.

SON regulates splicing of small introns with weak splice sites. **(a)** Size distribution (box plot) of SON-targeted introns, and non-target introns from the same genes. Significance was estimated using the Wilcoxon signed-rank test. **(b)** Splice strength for the 5' splice sites, estimated using MaxEntScan; only introns from genes that had at least one affected intron are compared to account for local properties. **(c)** Splice strength for the 3' splice sites, estimated using MaxEntScan; only introns from genes that had at least one affected intron are compared to account for local properties. **(d)** Nucleotide composition of 5' and 3' splice sites of all introns. **(e)** Nucleotide composition of 5' and 3' splice sites of introns affected by *SON* depletion. **(f)** GC content for all introns (black), and introns from genes that have at least one intron affected after depletion of *SON* (red: affected introns, green: unaffected introns from the same genes).

Figure 5.

SON targets specific introns of transcripts that encode pluripotency regulators. **(a)** Venn diagram showing the overlap of hESC identity maintenance genes (green) with genes targeted by SON. Positive control genes are highlighted in blue. Genes essential for hESC identity maintenance were defined as genes with average pluripotency F score < -2 (Supplementary Table S1 from genome-wide RNAi screening in hESCs (ref. 5)). Significance was estimated using Fisher's exact test. **(b)** Venn diagram showing the overlap of self-renewal genes (red) with genes regulated by SON. Genes essential for hESC survival were defined as genes with average cell survival N score < -2 (Supplementary Table S2 from genome-wide RNAi screening in hESCs (ref. 5)). The P value was calculated using Fisher's test. **(c)** Plot of quantile-normalized H3K27 acetylation at promoters in H1 hESCs and 20 non-pluripotent cell lines⁴⁷. SON-targeted genes are highlighted in red. Examples of SON-targeted genes with a hESC-specific epigenetic profile are highlighted. **(d)** Schematic illustration of the structural features of genes regulated by SON, showing: control introns (line, grey label), introns regulated by SON (line, red label), and control exons (filled rectangle, green label). Primer locations are labelled as arrows. **(e)** qPCR analysis of intron inclusion in transcripts of pluripotency genes (*PRDM14*, *OCT4*, *E4F1* and *MED24*), positive control gene *TUBG1* and negative control gene *TUBA1B* after 3-day depletion of *SON* with *SON* shRNA1. Mean $\pm$ s.e.m. from biological triplicate data is shown. The P value was determined by Student's t -test. **(f)** RNA-CLIP analysis on interaction of SON with intronic sequences of pluripotency-associated transcripts. Biological triplicate data are presented as mean $\pm$ s.e.m.

Figure 6.

Unspliced transcripts induced by *SON* depletion are potentially targeted by the NMD pathway. **(a)** The fold change of RNA-seq reads from control and *SON* knockdown at the *UPF1* locus (hg19). Shown are fold changes with $P < 0.4$. **(b)** Expression of *UPF1* and *UPF2* after transfection of shRNAs according to qPCR. Mean $\pm$ s.e.m. from biological triplicate data is shown. **(c)** Expression of proteins regulated by *SON* depletion determined by western blotting. **(d)** Expression of introns regulated by *SON* after depletion of *SON* together with *UPF1* and *UPF2* depletion according to qPCR. Mean $\pm$ s.e.m. from biological triplicate data is shown. Uncropped images of blots are shown in Supplementary Fig. S9.

Figure 7.

SON regulates alternative splicing in hESCs. **(a)** Scheme to illustrate our approach to identify splicing differences between control and *SON* knockdown. **(b)** Histogram showing the number of skipped exons for *SON*-knockdown-induced alternative splicing events and the genome-wide distribution. Significance was estimated using Fisher's exact test. **(c)** Fraction of splice junction usage induced by *SON* depletion and the genome-wide distribution. **(d)** Percentage of exon-skipping events induced by *SON* depletion that overlap intron inclusion events induced by *SON* depletion. Significance was estimated using Fisher's exact test.

Figure 8.

SON knockdown induces an isoform switch of *PRDM14*. **(a)** Overlap of SON-regulated transcripts with hESC-specific transcripts. Significance was estimated using Fisher's exact test. **(b)** Top: the fold change of RNA-seq reads from control and SON knockdown at the *PRDM14* locus (hg19). Shown are fold changes with $P < 0.4$. Bottom: fold change of normalized splice junction RNA-seq reads from control and *SON* knockdown; splicing events are highlighted in red (increased in *SON* knockdown) and green (decreased in *SON* knockdown). **(c)** RT-PCR analysis of *PRDM14* isoform expression after *SON* depletion. Primers were placed at the beginning and end of the transcript. The exon-skipping event was confirmed by sequencing. **(d)** *Luciferase* assay analysis of the gene activation potential of the *PRDM14* isoform. The reporter structure is shown at the top of the figure. Biological triplicate data are presented as mean $\pm$ s.e.m. **(e)** Reprogramming in the presence of *PRDM14* isoforms. OSKC: the 4 Yamanaka reprogramming factors: OCT4 (O), SOX2 (S), KLF4 (K) and c-MYC (C). Biological triplicate data are presented as mean $\pm$ s.e.m. **(f)** Schematic of SON-regulated splicing in hESCs. In the presence of SON, introns are properly spliced out and correct isoforms of transcripts are expressed to maintain hESC pluripotency and self-renewal. In the absence of SON, exon skipping and intron retention occur, subsequently leading to differentiation and cell death of hESCs.