Published in final edited form as:

Wiley Interdiscip Rev RNA. 2014 September ; 5(5): 637–646. doi:10.1002/wrna.1235.

THE ROLE OF SON IN SPLICING, DEVELOPMENT AND DISEASE

Xinyi Lu,

Gene Regulation Laboratory, Genome Institute of Singapore

Huck-Hui Ng, and

Gene Regulation Laboratory, Genome Institute of Singapore, Dept of Biological Sciences, National University of Singapore; Dept of Biochemistry, National University of Singapore; Graduate School for Integrative Sciences & Engineering, National University of Singapore; School of Biological Sciences, Nanyang Technological University

Paula A. Bubulya

Dept of Biological Sciences, Wright State University, Dayton, Ohio 45435, USA

Paula A. Bubulya: paula.bubulya@wright.edu

Abstract

SON is a nuclear protein involved in multiple cellular processes including transcription, pre-mRNA splicing and cell cycle regulation. Although SON was discovered 25 years ago, the importance of SON's function was only realized recently when its roles in nuclear organization and pre-mRNA splicing as well as the influence of these activities in maintaining cellular health were unveiled. Furthermore, SON was implicated to have a key role in stem cells as well as during the onset of various diseases such as cancer, influenza, and hepatitis. Here we review the progress that has been made in studying this multi-functional protein and discuss questions that remain to be answered about SON.

DISCOVERY OF SON: A HISTORICAL PERSPECTIVE

Structure and conservation of SON

The *SON* gene spans 34,463bp on human chromosome 21 and is composed of 23 exons encoding transcripts of various lengths ranging from ~1.5kb to ~8kb. The protein encoded by *SON* contains a unique long stretch of repetitive nucleotide sequences, which were initially identified in human and later found to be conserved in *SON* across multiple species. Discovery of the *SON* gene came in 1988 when a partial clone called *SON3* was isolated from a human embryonic cDNA library and noted as having unusual sequence structure¹. *SON3* showed homology to the chicken *gallin* as well as to human *C-MYC* and *MOS* oncogenes¹. Due to this relatedness to DNA binding proteins, the protein encoded by *SON* gene was named 'SON DNA binding protein'. Continuation of this work to isolate *SON* from a human placenta cDNA library demonstrated in 1992 that multiple mRNAs having alternate 5' ends were produced from the *SON* gene².

Another group subsequently mapped the genomic location of the human *SON* gene to a region neighboring *GART* in the D142H8 YAC clone³. Meanwhile, the mouse homolog of *SON* was located in a 4-gene cluster with *GART*, *INFAR* and *CFR2-4*⁴. Interestingly, it was shown that gene pairs (*CRF2-4*, *INFAR*) and (*GART*, *SON*) maintained synteny in human and mouse⁴. *SON* was later found to be highly conserved, and related sequences are present in multiple species. *SON* transcripts of various lengths are expressed differentially in cell lines⁵. Although multiple other *SON* clones have been documented, including BASS1 (Bax antagonist selected in *Saccharomyces* 1)⁶ and KIAA1019⁷, the full length gene structure of *SON* was not documented until 2000, when Wynn et al. mapped the *SON* gene locus and its neighboring genes (*GART* and *DONSON*) in both mouse and human⁸. The mouse *SON* gene is composed of 23 exons encoding a 7215 nt-long ORF⁸. A later study mapped human chromosome 21 transcripts and found that human *SON* encodes 6 alternative transcripts with the longest of these containing a 7,281 nt-long open reading frame (ORF)⁹.

Determining the structure of the *SON* gene was met with challenges due to the length of this gene and its repeat-rich nucleotide sequence. The short transcripts initially identified from the *SON* gene could be attributed to fragmentation of RNA or difficulty in cloning. It will be interesting to use contemporary technology to confirm the existence of only a single full length transcript from the *SON* locus or to reveal bona-fide shorter transcripts that may have specific roles in human development and/or diseases.

Expression and function of SON

Hints about the potential functions of *SON* were initially based on its primary sequence and cellular expression. Pioneering studies revealed the ubiquitous but variable expression pattern and putative molecular roles for *SON*^{5, 10}. Notably, a high expression level in fetal tissue implicated a role of *SON* in proliferation and/or differentiation of cells during embryonic development⁴. Although *SON* was initially considered to be a DNA binding protein, the demonstration of its specific localization in nuclear speckles (also called splicing factor compartments) later implicated a role for *SON* in pre-mRNA processing. Moreover, the similarity of motif structures among *SON* and serine/arginine-rich (SR) splicing factors further supports that *SON* plays a role in splicing.

The first piece of experimental evidence that *SON* acts as a DNA binding protein came from a screening of factors that interact with the X-box consensus sequence (CCTAGCAACAGATGCGT) of a **human leukocyte antigen (HLA) Class II promoter**¹⁰ where partial *SON* was cloned from cDNA library as *DBP-5*. It was noted that the ~7.5 kb ubiquitously expressed *DBP-5* transcript encodes a protein containing multiple repetitive motifs¹⁰. Immunostaining of HeLa cells with an antibody raised against *DBP-5* revealed its punctate nuclear localization, suggesting a role of *DBP-5* in transcription¹⁰. In 1994, Khan et al.⁵ isolated a *SON* cDNA clone during a screen to identify DNA binding proteins that regulate involucrin expression in keratinocytes⁵. This isolated cDNA clone encoded a partial sequence of *SON* protein that included a 260 a.a. basic region⁵ containing Ser-Pro-Lys(Arg)-Lys(Arg) repeats found in a variety of DNA binding proteins such as mouse c-Myc and Histone H1¹¹.

Additional motifs later found in both human and mouse SON proteins⁸ pointed toward a role for Son in pre-mRNA processing. Perhaps the most interesting feature noted in the primary sequence of SON is its multiple unique repetitive sequence motifs of unknown function that span nearly 1,200 amino acids. SON also contains an RS domain that is enriched in serine (S)/arginine (R) dipeptides⁸ and typically serves to localize SR protein splicing factors to nuclear speckles¹², that are nucleoplasmic clusters enriched with pre-mRNA splicing complexes^{13, 14}. The presence of an RS domain⁸, the subnuclear localization reported earlier for DBP-5¹⁰, and the confirmed localization of SON in nuclear speckles with pre-mRNA processing factors^{8, 15, 16} all hinted at pre-mRNA splicing functions for SON. Further support for this hypothesis came from other studies that identified a double-stranded RNA binding domain and G-patch domain in SON^{9, 15}, both of which are frequently found in pre-mRNA processing proteins¹⁷.

Taken together, important insights gained from resolving the gene structure of *SON* included its putative molecular functions in pre-mRNA processing as well as the potential structural organization or scaffolding role of novel repeat motifs not previously found in any other mammalian proteins. Uncovering the functions of these repeats and how they are post-translationally regulated is a current challenge. While SON was initially described as DNA-binding protein, we discuss below how it was later shown that SON has RNA binding ability and is involved in splicing. It is possible that both RNA and DNA binding activities are important for SON's role in regulating gene expression, since transcription and splicing are coupled and co-regulated¹⁸.

SON REGULATES PRE-MESSENGER RNA SPLICING

SON in nuclear organization

Proteomic analysis of nuclear speckles isolated from mouse liver nuclei identified SON among nearly 200 other proteins^{15, 19}. The majority of factors identified in purified nuclear speckles were pre-mRNA processing proteins, but also included subunits of RNA polymerase II and several transcription factors^{15, 19}. Later, a study that aimed to identify 'nuclear architecture' components for nuclear organization and higher order chromatin structure by mass spectrometry, found SON among 502 proteins in insoluble nuclear fractions from HeLa extracts²⁰. Nuclear speckle localization and motif structure of SON described by these studies further suggested that SON may play a role in pre-mRNA splicing. The first comprehensive investigation on the role of SON in nuclear speckles was done by Sharma et al²¹, who assembled the first full-length *SON* cDNA encoding all predicted motifs, and confirmed the nuclear speckle localization of yellow fluorescent protein (YFP)-SON in HeLa cells²¹. Two antibodies raised against SON showed the co-localization of SON with the splicing factor SRSF1 (previously called SF2/ASF)²¹. In addition, other groups found that SON could directly interact with spliceosome components SRSF1 and PRP8¹⁶, and with SRp53²². These suggest a direct role of SON in nuclear speckle organization and pre-mRNA splicing. The absence of SON in HeLa cells led to reorganization of many nuclear speckle constituents including the U1 snRNP protein U1-70K, splicing factor SRSF1, exon junction complex (EJC) components, as well as nuclear polyadenylated RNA²¹. Disrupting nuclear speckle organization, either by SON

depletion²¹ or by hyperphosphorylation of SR proteins²³ did not alter global transcription. Expression of siRNA-resistant *SON* rescued nuclear speckle organization in *SON*-depleted HeLa cells²¹. Expression of various partial *SON* constructs to rescue depletion of the endogenous protein showed that the unique repetitive domain of *SON* is essential for its role in organizing splicing factors and polyadenylated RNA in nuclear speckles²¹.

Interestingly, *SON*'s impact on a nuclear-retained long noncoding RNA called *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1; also called *NEAT2*) was somewhat different than its effect on other nuclear speckle components. *MALAT1* is proposed to be a 'molecular sponge' that interacts with SR splicing factors and helps to regulate their local availability and activity during alternative splicing of nascent transcripts²⁴. Interestingly, *SON* repeats are not sufficient for the organization of *MALAT1* in nuclear speckles, since *SON* deletion mutants having the repeats but missing the G-patch and RS domain did not retain *MALAT1* in nuclear speckles²⁴. Altogether these data revealed a critical role for *SON* in organizing not only the protein components but also the RNA components of nuclear speckles, and further implied that *SON* has a central role in regulating pre-mRNA splicing.

How *SON* depletion results in dramatic reorganization of pre-mRNA processing factors within mammalian cell nuclei is not known, but it was initially proposed that *SON* repeats provides a network or scaffold within nuclear speckles to provide a landing pad for RNA processing proteins in these nuclear domains²¹. Additionally, the interactions of *SON* with EJC proteins reveal a potentially broad interactome for *SON* as part of mRNP complexes²⁹. Recent evidence that overexpression or knockdown of other nuclear speckle proteins (in particular, certain EJC proteins) leads to nuclear speckle reorganization necessitates a somewhat revised model for *SON*'s role in nuclear speckle assembly. For example, exogenous expression of EJC factors show similar localization of nuclear speckles as seen in *SON*-depleted nuclei^{25, 26}. A well-established model for nuclear speckle organization is based on self-organization of components via self-interaction domains²⁷. Self-organizing complexes might include higher-order assembly onto docking sites at extensive repeat motifs in proteins such as *SON*. This process could scaffold or distribute processing factor complexes throughout speckles via multi-molecular binding. However, intermolecular interactions within pre-mRNA processing complexes could potentially become unbalanced by increased or decreased abundance of individual components, leading to reorganization of pre-mRNA processing factors within nuclear speckles as seen following *SON* depletion. Absence of such molecular "glue" (i.e. *SON*) would presumably dislodge self-organized complexes without causing their complete disassembly, and may explain why enhanced distribution of EJC components and other splicing factors is observed at perispeckle regions following *SON* depletion²¹.

SON in pre-mRNA splicing

SON's role in pre-mRNA splicing has been demonstrated in multiple studies^{28–30}. By profiling the expression change after *SON* knockdown in HeLa cells, microarray analysis by Ahn et al. uncovered 659 genes whose expression was affected²⁸. Examination of downregulated transcripts showed splicing defects at specific introns. Through ultraviolet crosslinking and immunoprecipitation (CLIP), it was demonstrated that *SON* directly

interacts with these transcripts²⁸. Moreover, this study discovered that the RS and G-patch domains are responsible for the splicing activity of SON, whereby the unique repetitive motifs of SON further enhanced the activity of SON in splicing²⁸. This study also demonstrated that SON-regulated introns contain weak or dual-specificity splice sites. Since the absence of SON selectively disrupts the association of splicing factors SC35 with U1-70K and U2AF65, as well as Ser2-phosphorylated RNAP II, SON may regulate co-transcriptional splicing²⁸.

SON is not only involved in constitutive splicing but also plays a role in alternative splicing of exons. The first link between SON and alternative splicing was established by Moore et al³¹, who discovered that SON was a candidate for regulators of Bcl-x and Mcl1 alternative splicing in a genome-wide RNA interference (RNAi) screen. The relationship between SON and alternative splicing was more comprehensively studied later by Sharma et al.²⁹, who showed that SON accumulated at constitutively active β -tropomyosin minigene reporter transcription sites together with other pre-mRNA splicing factors such as SRSF1 and U1-70K further supporting a co-transcriptional role of SON in splicing. In addition, the scaffolding domain of SON is essential for the localization of SON to β -tropomyosin reporter transcripts²⁹. These results imply that although the repeat domains of SON have only a minor role during constitutive co-transcriptional splicing, they may be essential for SON in mediating alternative splicing. Furthermore, a genome-wide exon array study revealed that *SON* depletion caused defects in all types of alternative splicing²⁹. *SON* depletion caused exon skipping at transcription loci *in situ*, but did not affect the recruitment of SRSF1 or U1-70K to the transcription site. One of the most interesting aspects of SON splicing regulation is that it targets only a subset of human transcripts encoding proteins involved in several major biological processes. These proteins participate in a variety of pathways including cell cycle regulation, apoptosis, integrin mediated cell adhesion, smooth muscle contraction, G protein signaling, Wnt mediated signaling, TGF beta signaling, translation factors, glycogen metabolism, and cholesterol biosynthesis. Sharma et al. (2011) validated altered abundance of transcripts that were upregulated (*GEMIN5*, *IL1A*, *TNFRSF21*, *CCNG1* and *CYP11B1*) or downregulated (*DNAH2*, *TNNC1* and *CDK5*), as well as altered splicing for several transcripts encoding epigenetic regulators HDAC6, SetD8 and ADA whose processing requires SON²⁹.

Another common thread observed by many groups is that *SON* depletion causes defects in cell survival^{16, 21, 28, 29, 32, 33}. Examination of mitotic cells showed that *SON* depletion causes spindle defects and mitotic arrest at metaphase²¹. During mitosis, SON colocalizes with pre-mRNA splicing factors in mitotic interchromatin granule clusters (MIGs) (Figure 1), but localization in other mitotic structures such as the mitotic spindle has not been observed. This supports the hypothesis that the mitotic defects resulting from the loss of SON following knockdown result from the impact of SON on splicing of mitotic regulator transcripts, rather than loss of SON from mitotic structures. In HeLa cells, SON is required for proper splicing of a subset of human transcripts that are necessary for maintaining proper cell growth, explaining cell cycle defects seen after *SON* knockdown in several studies^{16, 21, 28, 29, 32, 33}. Besides transcripts encoding cell cycle proteins, SON also regulates splicing of a subset of transcripts from DNA repair pathways²⁸. Later studies by

Lu et al.³⁰ (discussed below) firmly support a model that SON specifically associates with only a subset of human transcripts. While the mechanisms underlying each type of splicing defect remain to be elucidated, these findings taken altogether implicate a role for SON to work in concert with other splicing regulators such as other SR proteins, EJC components, and *MALAT1* in coupling transcription with splicing (Figure 2a). The repetitive nature and serine-rich sequence of SON's repeat motifs is intriguingly analogous to the heptad repeats of RNA polymerase II (RNAPII) carboxy-terminal domain (CTD) that are responsible for loading RNA processing factors onto nascent transcripts^{34–36}. It is therefore possible that SON regulates pre-mRNA processing at specific splice sites by controlling the availability of factors in the vicinity. We provide a model illustrating how SON could modulate local recruitment of SR proteins to nascent transcripts to affect splicing outcomes (Figure 2b). Work from many groups supports a role for SON as an essential co-transcriptional splicing regulator in both constitutive and alternative splicing. The mechanism of SON-dependent splicing of specific primary transcripts is still unclear. Thus, by continuing to systematically analyze affected splice sites, we will gain a much fuller picture regarding how SON mediates splicing outcomes.

Importantly, SON not only takes part in pre-mRNA splicing but also plays a part in other steps of mRNA maturation. Knockdown of *SON* in HeLa cells led to a redistribution of nuclear speckle-localized exon junction complex (EJC)²¹, which marks the position of exon–exon junctions and participate in mRNA export and decay³⁷. Together with other SR proteins, SON was associated with EJC proteins, suggesting a role in downstream mRNA metabolism after splicing³⁸. Furthermore, SON could also directly regulate splicing of transcripts encoding regulators of the nonsense-mediated mRNA decay (NMD) pathway, such as UPF1, SMG1 or DHX34³⁰. The interplay of SON in EJC complex distribution, mRNP regulation and the NMD pathway infers a sophisticated role of SON in the co-regulation between mRNA splicing and post-splicing RNA metabolism.

ROLE OF SON DURING DEVELOPMENT

SON transcripts have been detected in multiple tissues including human embryo¹, placenta², B cell lymphocytes¹⁰, skin⁵, liver⁵, and brain⁷. Studies on SON function during development were done in stem cells isolated from tissues and embryos^{30, 39}. Stem cells are undifferentiated cells with unlimited potential to self-renew and the ability to differentiate into more specialized cells. Pluripotent stem cells, such as embryonic stem cells, can generate cells of all three germ layers, whereas adult stem cells, such as hematopoietic stem cells, can give rise to mature differentiated cells of a specific lineage⁴⁰. Identification and characterization of new regulators in stem cells will be key to deciphering the mechanisms behind the potency of these cells.

SON in embryonic stem cells

A genome-wide RNAi screen revealed SON as a regulator of human embryonic stem cell (hESC) pluripotency and survival⁴¹. *SON* is also upregulated in both induced pluripotent cells (iPSCs) and hESCs when compared to MRC-5 fibroblasts³⁰. Moreover, hESC differentiation induced by *PRDM14* or *NFRKB* depletion is accompanied by downregulation of *SON*, suggesting a specific role of SON in the regulation of pluripotency in hESCs³⁰. In

addition to pluripotency maintenance, SON is also essential for the acquisition of pluripotency by somatic cells³⁰. RNA-seq analysis revealed that *SON* deficiency led to an almost exclusive upregulation of intronic expression, suggesting that SON acts as an intron splicing activator³⁰ (Figure 3a). These introns regulated by SON in hESCs have shorter length and weaker splicing strength than non-targeted introns³⁰. Furthermore, the splice sites flanking these SON target introns are GC-enriched³⁰. These suggest the splicing regulation by SON is dependent on intron size, strength of splice sites and nucleotide content of pre-mRNA transcripts.

Splice site strength is positively correlated with intron length⁴². Short introns may undergo different splicing mechanisms than long introns; therefore, the strength of splice sites alone may not be sufficient for the recognition by SON. It was suggested that short introns directly utilize splice donor and acceptor as a splicing entity (intron-based splicing recognition)^{43, 44}, while long introns are excised upon the recognition and pairing of exon splice sites (exon-based splicing recognition)⁴⁵. These two splicing mechanisms cause different effects during splice site mutation. Particularly, mutation of splice sites flanking small introns will cause intron retention whereas mutation of splice sites flanking long introns leads to exon skipping^{46–49}. These observations may be partially attributed to the distinct sequence composition of small and large introns, as the polypyrimidine tract between 3' splice site and the branch point are absent in small introns⁴⁷. Since the splicing of short introns is also regulated by SON, it may mainly recognize the splicing signals near splice junction regions and within the intron. SON could be one of the factors responsible for the splicing mechanism differences between short intron and long introns. Beside intron size and splicing strength, signals from splicing enhancers and silencers as well as RNA secondary structure can also affect splice site selection by splicing factors^{50, 51}. A high GC content around splice sites of SON target genes could influence splicing by altering mRNA secondary structure. Furthermore, the presence of other factors interacting with SON, such as transcription factors or splicing factors, may contribute to the selection of splice site as well. It will be intriguing to investigate the SON-splicing mechanism in hESCs in the future.

Interestingly, SON not only regulates the splicing of cell survival genes such as *TUBG1*, *TUBGCP2*, *AURKB*, and *FANCG*, but also binds to and regulates the splicing of hESC maintenance genes such as *OCT4*, *MED24* and *E4F1*, and the affected transcripts may be targeted by NMD pathway for degradation³⁰. Besides intron splicing, SON also regulates alternative splicing in hESCs. 940 splicing events in 694 genes were found to be alternatively spliced when *SON* is depleted and most of these events induce the emergence of minor splice variants with single exon skipping³⁰. Strikingly, these alternative splicing events significantly overlap with SON-regulated introns³⁰, implicating a coordinated regulation of intron splicing and exon skipping. Moreover, *SON* deficiency induces alternative splicing in hESC specific genes³⁰. This is exemplified by *PRDM14* isoform switch after *SON* depletion. *PRDM14* is a pluripotency regulator in mouse⁵² and human⁴¹. In the absence of *SON*, a previously un-annotated isoform of *PRDM14* lacking exon2 was upregulated while the known isoform of *PRDM14* was downregulated³⁰. It was further found that this shorter isoform of *PRDM14* could not activate the enhancer region or promote pluripotency induction during reprogramming, raising the possibility that altering

the splicing of *PRDM14* is one of the main events driving hESC differentiation after SON knockdown³⁰ (Figure 3b).

SON in hematopoietic stem cells

Besides hESCs, SON also has an important role in hematopoietic stem cells (HSCs). SON is highly expressed in HSCs and downregulated during hematopoietic differentiation³⁹. Knockdown of *SON* in these cells causes downregulation of GATA2 protein expression without affecting its mRNA level, suggesting post-transcriptional regulation of GATA2 by SON³⁹. Examination of GATA2 3'-UTR revealed *microRNA-27a* (*miR-27a*) and *microRNA-24* (*miR-24*) as potential GATA2 regulators³⁹. Further analysis identified *miR27a* as the functional miRNA behind GATA2 regulation. It was also found that the promoter of *miR-23a-27a-24-2* cluster, which encodes *miR-27a*, is regulated by SON, indicating a role of SON in transcription repression in addition to its splicing activity³⁹ (Figure 3c). However, it still remains unclear whether transcription regulation by SON is direct versus indirect due to altered abundance and mRNA processing for transcripts encoding other transcription regulators.

ROLE OF SON IN DISEASES

Role of SON in cancer

SON participates in cellular processes important for cancer development such as cell proliferation, apoptosis, as well as DNA damage and repair^{16, 21, 28, 29, 32}, therefore, activity of SON has been linked to tumorigenesis. Early studies found that overexpression of SON fragments could transform mammalian cells to an epithelial-like morphology⁵³ or inhibit cell apoptosis⁶. In leukemia cells, SON interacts with growth arrest inducing protein AML1-ETO instead of its tumorigenic isoform AML1-ETO(C663S)^{32, 54}. In addition, SON translocates from nuclear speckles to the cytoplasm in t(8;21)+ leukemia cells. SON also interacts with tumor suppressor protein von Hippel-Lindau in 786-O renal carcinoma cells³². The expression of *SON* has been detected in various cancer cell lines such as leukemia-derived, HeLa, and pancreatic cancer cell lines^{16, 21, 28, 29, 32, 33, 54}. *SON* deficiency inhibited the tumorigenicity of pancreatic tumor cells *in vivo*³³. While depletion of *SON* induced significant growth arrest and apoptosis in cancer cell lines, these effects were less pronounced in normal cells^{16, 21, 28, 29, 32, 33}. Taken together, these data implicate SON as a key factor in cell growth that could impact tumorigenesis. Given the specific role of SON in cancer cell survival, SON may serve as a novel target for cancer treatment.

Role of SON in virus infection

Besides cancer development, SON is also involved in pathways regulating virus infection. Transcription repression by SON was not only observed in HSCs but also on viral promoters⁵⁵. SON binds to a negative regulatory element (NRE) of human hepatitis B virus (HBV) via a GA(G/T)AN(C/G)(A/G)CC motif in the NRE, inhibiting transcription of HBV genes and virion production through regulating its core promoter⁵⁵. In addition to its role in HBV infection, SON is a regulator of influenza virus infection as discovered by a genome-wide RNAi screen for host factors critical for influenza virus replication⁵⁶. Reduced influenza viral RNA levels and decreased viral infection were seen following *SON*

knockdown⁵⁶; suggesting that SON is needed for influenza virus replication. However, it still remains unclear if these observations are direct effects or secondary effects of the knockdown due to unsuccessful transcription or splicing of transcripts encoding other cellular proteins needed for viral replication and trafficking. Sorting out primary (direct role in process e.g. mitosis) versus secondary effects (due to misprocessing of other transcripts) under *SON* knockdown conditions is a serious challenge. Additional knowledge derived from the mapping of direct interfaces between RNA and RNA binding proteins will provide new insights into the rules that govern selective and specific interactions. Techniques such CLIP-seq (cross-linking immunoprecipitation coupled with high-throughput sequencing) will undoubtedly be used more widely to study RNA binding proteins^{57, 58}.

Conclusion

SON is preferentially expressed in undifferentiated stem cells and downregulated during stem cell differentiation^{30, 39}. SON regulates the potency of stem cells through both transcription and splicing control. It remains to be answered whether *SON* is equally important to all stem cells or only specific types of stem cells. All studies thus far are based on knockdown or overexpression of SON in cell lines, so generation of *SON* knockout mice will be key for studying its role during development. It will be interesting to learn if SON helps to scaffold the assembly of mature splicing complexes in nuclear speckles. How this would in turn promote the overall stability of RNP complexes, the dynamic exchange of complexes among different nuclear compartments, or the precise mechanisms for SON-mediated splicing are all questions for future study.

Acknowledgments

The authors wish to thank Labib Rouhana and Friedrich Sachs for critical reading of the manuscript. Work discussed here was supported by funding from the National Institutes of Health (GM084407–02 to P.A.B.), and from the Agency for Science, Technology and Research (A*STAR) to H.-H.N. and X.L.

References

1. Berdichevskii FB, Chumakov IM, Kiselev LL. Decoding of the primary structure of the son3 region in human genome: identification of a new protein with unusual structure and homology with DNA-binding proteins. *Mol Biol (Mosk)*. 1988; 22:794–801. [PubMed: 3054499]
2. Bliskovskii VV, Kirillov AV, Zakhar'ev VM, Chumankov IM. The human son gene: the large and small transcripts contains various 5'-terminal sequences. *Mol Biol (Mosk)*. 1992; 26:807–812. [PubMed: 1435774]
3. Tassone F, Cheng S, Gardiner K. Analysis of chromosome 21 yeast artificial chromosome (YAC) clones. *Am J Hum Genet*. 1992; 51:1251–1264. [PubMed: 1463009]
4. Cheng S, Lutfalla G, Uze G, Chumakov IM, Gardiner K. GART, SON, IFNAR, and CRF2-4 genes cluster on human chromosome 21 and mouse chromosome 16. *Mamm Genome*. 1993; 4:338–342. [PubMed: 8318737]
5. Khan IM, Fisher RA, Johnson KJ, Bailey ME, Siciliano MJ, Kessling AM, Farrer M, Carritt B, Kamalati T, Buluwela L. The SON gene encodes a conserved DNA binding protein mapping to human chromosome 21. *Ann Hum Genet*. 1994; 58:25–34. [PubMed: 8031013]
6. Greenhalf W, Lee J, Chaudhuri B. A selection system for human apoptosis inhibitors using yeast. *Yeast*. 1999; 15:1307–1321. [PubMed: 10509013]
7. Kikuno R, Nagase T, Ishikawa K, Hirosawa M, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O. Prediction of the coding sequences of unidentified human genes. XIV. The complete

- sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.* 1999; 6:197–205. [PubMed: 10470851]
8. Wynn SL, Fisher RA, Pagel C, Price M, Liu QY, Khan IM, Zammit P, Dadrah K, Mazrani W, Kessling A, et al. Organization and conservation of the GART/SON/DONSON locus in mouse and human genomes. *Genomics.* 2000; 68:57–62. [PubMed: 10950926]
  9. Reymond A, Friedli M, Henrichsen CN, Chapot F, Deutsch S, Ucla C, Rossier C, Lyle R, Guipponi M, Antonarakis SE. From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map. *Genomics.* 2001; 78:46–54. [PubMed: 11707072]
  10. Mattioni T, Hume CR, Konigorski S, Hayes P, Osterweil Z, Lee JS. A cDNA clone for a novel nuclear protein with DNA binding activity. *Chromosoma.* 1992; 101:618–624. [PubMed: 1424986]
  11. Churchill ME, Suzuki M. ‘SPKK’ motifs prefer to bind to DNA at A/T-rich sites. *EMBO J.* 1989; 8:4189–4195. [PubMed: 2556263]
  12. Long JC, Cáceres JF. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J.* 2009; 417:15–27. [PubMed: 19061484]
  13. Han J, Xiong J, Wang D, Fu XD. Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* 2011; 21:336–343. [PubMed: 21514162]
  14. Spector DL, Lamond AI. Nuclear speckles. *Cold Spring Harb Perspect Biol.* 2011;3.
  15. Saitoh N, Spahr CS, Patterson SD, Bubulya P, Neuwald AF, Spector DL. Proteomic analysis of interchromatin granule clusters. *Mol Biol Cell.* 2004; 15:3876–3890. [PubMed: 15169873]
  16. Huen MS, Sy SM, Leung KM, Ching YP, Tipoe GL, Man C, Dong S, Chen J. SON is a spliceosome-associated factor required for mitotic progression. *Cell Cycle.* 2010; 9:2679–2685. [PubMed: 20581448]
  17. Korneta I, Magnus M, Bujnicki JM. Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res.* 2012; 40:7046–7065. [PubMed: 22573172]
  18. Shukla S, Oberdoerffer S. Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochim Biophys Acta.* 2012; 1819:673–683. [PubMed: 22326677]
  19. Mintz PJ, Patterson SD, Neuwald AF, Spahr CS, Spector DL. Purification and biochemical characterization of interchromatin granule clusters. *EMBO J.* 1999; 18:4308–4320. [PubMed: 10428969]
  20. Takata H, Nishijima H, Ogura S, Sakaguchi T, Bubulya PA, Mochizuki T, Shibahara K. Proteome analysis of human nuclear insoluble fractions. *Genes Cells.* 2009; 14:975–990. [PubMed: 19695025]
  21. 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]
  22. Cazalla D, Newton K, Cáceres JF. A novel SR-related protein is required for the second step of Pre-mRNA splicing. *Mol Cell Biol.* 2005; 25:2969–2980. [PubMed: 15798186]
  23. Sacco-Bubulya P, Spector DL. Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. *J Cell Biol.* 2002; 156:425–436. [PubMed: 11827980]
  24. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell.* 2010; 39:925–938. [PubMed: 20797886]
  25. Daguenet E, Baguet A, Degot S, Schmidt U, Alpy F, Wendling C, Spiegelhalter C, Kessler P, Rio MC, Le Hir H, et al. Perispeckles are major assembly sites for the exon junction core complex. *Mol Biol Cell.* 2012; 23:1765–1782. [PubMed: 22419818]
  26. Schmidt U, Richter K, Berger AB, Lichter P. In vivo BiFC analysis of Y14 and NXF1 mRNA export complexes: preferential localization within and around SC35 domains. *J Cell Biol.* 2006; 172:373–381. [PubMed: 16431928]
  27. Misteli T. The concept of self-organization in cellular architecture. *J Cell Biol.* 2001; 155:181–185. [PubMed: 11604416]

28. Ahn EY, DeKolver RC, Lo MC, Nguyen TA, Matsuura S, Boyapati A, Pandit S, Fu XD, Zhang DE. SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol Cell*. 2011; 42:185–198. [PubMed: 21504830]
29. Sharma A, Markey M, Torres-Munoz K, Varia S, Kadakia M, Bubulya A, Bubulya PA. Son maintains accurate splicing for a subset of human pre-mRNAs. *J Cell Sci*. 2011; 124:4286–4298. [PubMed: 22193954]
30. Lu X, Goke J, Sachs F, Jacques PE, Liang H, Feng B, Bourque G, Bubulya PA, Ng HH. SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat Cell Biol*. 2013; 15:1141–1152. [PubMed: 24013217]
31. Moore MJ, Wang Q, Kennedy CJ, Silver PA. An alternative splicing network links cell-cycle control to apoptosis. *Cell*. 2010; 142:625–636. [PubMed: 20705336]
32. Ahn EY, Yan M, Malakhova OA, Lo MC, Boyapati A, Ommen HB, Hines R, Hokland P, Zhang DE. Disruption of the NHR4 domain structure in AML1-ETO abrogates SON binding and promotes leukemogenesis. *Proc Natl Acad Sci U S A*. 2008; 105:17103–17108. [PubMed: 18952841]
33. Furukawa T, Tanji E, Kuboki Y, Hatori T, Yamamoto M, Shimizu K, Shibata N, Shiratori K. Targeting of MAPK-associated molecules identifies SON as a prime target to attenuate the proliferation and tumorigenicity of pancreatic cancer cells. *Mol Cancer*. 2012; 11:88. [PubMed: 23227827]
34. Bentley DL. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol*. 2005; 17:251–256. [PubMed: 15901493]
35. Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell*. 2007; 26:867–881. [PubMed: 17588520]
36. Zhong XY, Wang P, Han J, Rosenfeld MG, Fu XD. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell*. 2009; 35:1–10. [PubMed: 19595711]
37. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J*. 2001; 20:4987–4997. [PubMed: 11532962]
38. Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell*. 2012; 151:750–764. [PubMed: 23084401]
39. Ahn EE, Higashi T, Yan M, Matsuura S, Hickey CJ, Lo MC, Shia WJ, DeKolver RC, Zhang DE. SON protein regulates GATA-2 through transcriptional control of the microRNA 23a~27a~24-2 cluster. *J Biol Chem*. 2013; 288:5381–5388. [PubMed: 23322776]
40. Schöler, HR. The Potential of Stem Cells: An Inventory. In: Knoepffler, Nikolaus; Schipanski, Dagmar; Sorgner, Stefan Lorenz, editors. *Human biotechnology as Social Challenge*. Ashgate Publishing, Ltd; 2007. p. 28
41. Chia NY, Chan YS, Feng B, Lu X, Orlov YL, Moreau D, Kumar P, Yang L, Jiang J, Lau MS, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*. 2010; 468:316–320. [PubMed: 20953172]
42. Dewey CN, Rogozin IB, Koonin EV. Compensatory relationship between splice sites and exonic splicing signals depending on the length of vertebrate introns. *BMC Genomics*. 2006; 7:311. [PubMed: 17156453]
43. Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci U S A*. 2005; 102:16176–16181. [PubMed: 16260721]
44. Guthrie C. Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science*. 1991; 253:157–163. [PubMed: 1853200]
45. Robberson BL, Cote GJ, Berget SM. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol*. 1990; 10:84–94. [PubMed: 2136768]
46. Grabowski PJ, Nasim FU, Kuo HC, Burch R. Combinatorial splicing of exon pairs by two-site binding of U1 small nuclear ribonucleoprotein particle. *Mol Cell Biol*. 1991; 11:5919–5928. [PubMed: 1834932]

47. Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res.* 1992; 20:4255–4262. [PubMed: 1508718]
48. Talerico M, Berget SM. Effect of 5' splice site mutations on splicing of the preceding intron. *Mol Cell Biol.* 1990; 10:6299–6305. [PubMed: 2247057]
49. Talerico M, Berget SM. Intron definition in splicing of small *Drosophila* introns. *Mol Cell Biol.* 1994; 14:3434–3445. [PubMed: 8164690]
50. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev.* 2003; 17:419–437. [PubMed: 12600935]
51. Zhang J, Kuo CC, Chen L. GC content around splice sites affects splicing through pre-mRNA secondary structures. *BMC Genomics.* 2011; 12:90. [PubMed: 21281513]
52. Ma Z, Swigut T, Valouev A, Rada-Iglesias A, Wysocka J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat Struct Mol Biol.* 2011; 18:120–127. [PubMed: 21183938]
53. Chumakov IM, Berdichevskii FB, Sokolova NV, Reznikov MV, Prasolov VS. Identification of a protein product of a novel human gene SON and the biological effect upon administering a changed form of this gene into mammalian cells. *Mol Biol (Mosk).* 1991; 25:731–739. [PubMed: 1944255]
54. DeKolver RC, Yan M, Ahn EY, Shia WJ, Speck NA, Zhang DE. Attenuation of AML1-ETO cellular dysregulation correlates with increased leukemogenic potential. *Blood.* 2013; 121:3714–3717. [PubMed: 23426948]
55. Sun CT, Lo WY, Wang IH, Lo YH, Shiou SR, Lai CK, Ting LP. Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. *J Biol Chem.* 2001; 276:24059–24067. [PubMed: 11306577]
56. Karlas A, Machuy N, Shin Y, Pleissner KP, Artarini A, Heuer D, Becker D, Khalil H, Ogilvie LA, Hess S, et al. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature.* 2010; 463:818–822. [PubMed: 20081832]
57. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 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]
58. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature.* 2008; 456:464–469. [PubMed: 18978773]

FIGURE 1.

SON localization throughout the cell cycle. SON colocalizes with splicing factor SRSF1 in HeLa cells throughout the cell cycle. Immunofluorescence localization of SON and of SRSF1 show colocalization as well as reorganization of cellular distribution of these splicing factors during mitosis. Specific localization is observed in mitotic interchromatin granules (MIGs; arrow) during mitosis. DNA was stained with DAPI. Panels in order from left to right: interphase, prophase (before nuclear envelope breakdown, prophase (after nuclear envelope breakdown), metaphase, anaphase, telophase, and G1. Bar = 5 μ m. Images provided by Keshia Torres-Munoz and Paula Bubulya.

FIGURE 2.

A model for SON as a multifunctional scaffold. (a) SON organizes nuclear speckle components through its repetitive domains while the C-terminus of SON including the G patch is required for its interaction with the nuclear-retained long noncoding RNA *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1). Expression and distribution of components impacts the overall subnuclear organization of RNA processing factors. DNA is stained by Hoechst in blue and SON is stained red to mark nuclear speckles. (b) SON and *MALAT1* recruit SR proteins to mediate co-transcriptional splicing of sub-optimal splice sites and alternative splicing of exons. Son may remain associated with mature mRNPs²⁹. CTD: c-terminal domain of RNA polymerase II. * represents weak/dual-specificity splice sites.

FIGURE 3.

SON regulates gene expression in stem cells. (a) SON promotes intron removal from hESC maintenance and survival genes³⁰. (b) SON facilitates correct isoform expression of hESC specific genes³⁰. (c) SON may directly or indirectly inhibit expression of *miR23a-27a-24-2* cluster to repress *GATA2* expression in HSCs³⁹. SON may interact with other factor to directly inhibit transcription of *miR23a-27a-24-2* cluster or act indirectly through regulating splicing of other factors responsible for *miR23a-27a-24-2* expression.