PROSPECTS
Journal of Cellular Biochemistry 115:224–231 (2014)

Journal of Cellular Biochemistry

New Discoveries of Old SON: A Link Between RNA Splicing and Cancer

Christopher J. Hickey,[1] Jung-Hyun Kim,[1] and Eun-Young Erin Ahn[1,2]*

[1]*Mitchell Cancer Institute, University of South Alabama, Mobile, Alabama 36604*

[2]*Department of Biochemistry and Molecular Biology, College of Medicine, University of South Alabama, Mobile, Alabama 36688*

ABSTRACT

The SON protein is a ubiquitously expressed DNA- and RNA-binding protein primarily localized to nuclear speckles. Although several early studies implicated SON in DNA-binding, tumorigenesis and apoptosis, functional significance of this protein had not been recognized until recent studies discovered SON as a novel RNA splicing co-factor. During constitutive RNA splicing, SON ensures efficient intron removal from the transcripts containing suboptimal splice sites. Importantly, SON-mediated splicing is required for proper processing of selective transcripts related to cell cycle, microtubules, centrosome maintenance, and genome stability. Moreover, SON regulates alternative splicing of RNAs from the genes involved in apoptosis and epigenetic modification. In addition to the role in RNA splicing, SON has an ability to suppress transcriptional activation at certain promoter/enhancer DNA sequences. Considering the multiple SON target genes which are directly involved in cell proliferation, genome stability and chromatin modifications, SON is an emerging player in gene regulation during cancer development and progression. Here, we summarize available information from several early studies on SON, and highlight recent discoveries describing molecular mechanisms of SON-mediated gene regulation. We propose that our future effort on better understanding of diverse SON functions would reveal novel targets for cancer therapy. J. Cell. Biochem. 115: 224–231, 2014.

KEY WORDS: SON; GENE EXPRESSION; RNA SPLICING; TRANSCRIPTION; CANCER

The product encoded by the *SON* locus is a large protein composed of 2,426 amino acids primarily localized to nuclear speckles [Wynn et al., 2000; Reymond et al., 2001; Ahn et al., 2011]. Although the *SON* cDNA fragment was partially cloned more than two decades ago, the functional importance of this protein has been behind the veil for long time. Recently, SON was identified as a new member of the SR-like family of proteins, which are involved in pre-mRNA splicing [Ahn et al., 2011; Sharma et al., 2011]. As emerging evidence has shown that aberrant RNA splicing and alternative splicing are significant events regulating gene expression for both physiological and pathological conditions [Faustino and Cooper, 2003; Garcia-Blanco et al., 2004; Kornblihtt et al., 2013], the discovery of SON as an RNA-binding protein and a novel splicing regulator is an exciting addition to our understanding of gene expression. Recent works on the function of SON demonstrated that depletion of SON leads to improper RNA splicing of genes associated with cell-cycle progression and DNA repair, thereby negatively impacting cell proliferation and genome stability [Huen et al., 2010; Ahn et al., 2011; Sharma et al., 2011]. In addition, SON has a DNA-binding ability and functions in transcriptional repression at the promoter region of selective genes [Sun et al., 2001; Komori et al., 2010; Ahn et al., 2013]. Given the diverse functions of SON in multiple cellular processes, this protein offers exciting avenues for future exploration and understanding of gene expression. Despite the potential functions of SON in multiple cellular processes, there has been a lack of informative summary on this protein. In this review, we concisely summarize the available literature and recent findings regarding SON and the contributions it plays in normal cellular processes and cancer.

DISCOVERY AND EARLY STUDIES ON SON

Early reports have described several different partial sequences of full-length *SON*, while using a variety of different names based on

Christopher J. Hickey and Jung-Hyun Kim contributed equally to this work.

The authors have no conflict of interest to declare.

Grant sponsor: Mitchell Cancer Institute, University of South Alabama.

*Correspondence to: Dr. Eun-Young Erin Ahn, Mitchell Cancer Institute, University of South Alabama, 1660 Springhill Ave, Mobile, AL 36604. E-mail: eahn@health.southalabama.edu

Manuscript Received: 29 August 2013; Manuscript Accepted: 6 September 2013

Accepted manuscript online in Wiley Online Library (wileyonlinelibrary.com): 12 September 2013

DOI 10.1002/jcb.24672 • © 2013 Wiley Periodicals, Inc.

functionality (summarized in Table I and Fig. 1). The first description of the *SON* gene was reported from the screening of human embryonic cDNA library [Berdichevskii et al., 1988]. The authors gave the translated protein from the identified cDNA the name SON3, which had striking homology to *c-myc* and *mos*. Identified within the translated amino acid sequences were tandem repeats which are similar to a DNA binding protein expressed in chicken, called gallin. A short time later, the functional activity of SON was described when the 3′ end of *SON* cDNA was cloned and ectopically expressed in transformed murine cells. Upon overexpression of the SON fragment, the transformed cells were found to take on morphology more similar to normal epithelial cells [Chumakov et al., 1991]. In this study, antiserum was raised to synthetic peptides corresponding to the C-terminus of SON, and two different sized bands, 120 and 92 kDa, were detected in the lysates from different cell lines [Chumakov et al., 1991]. A year later, several overlapping cDNA clones for the smaller transcripts were isolated from human placenta cDNA library, and identified a longer transcript of *SON*, *SON-a* (5,686 kb), with repetitive sequences [Bliskovskii et al., 1992]. From the translated amino acid sequences, four types of tandem repeats were predicted. The authors suggested the presence of these repetitive amino acids might provide an important structure for the function of SON.

Around the same time, another group described a novel DNA-binding protein identified from a screen for factors that bind to the HLA class II promoter in normal human B cells, and named it DBP-5 [Mattioni et al., 1992]. The DBP-5 gene was a partial sequence of the previously described *SON* gene, and the DBP-5 transcript was ubiquitously expressed among various types of human cells. Upon close analyses of the translated product, several protein motifs were described for the 1,179 amino acid product. These motifs include multiple tandem repeats such as an N-terminal proline-rich region and the C-terminal serine/arginine-rich domain. Antisera generated against DBP-5 detected 180 and 42kDa bands in Western blotting, and stained the nuclei of HeLa cells revealing a punctate pattern [Mattioni et al., 1992]. Based on its nuclear localization and DNA-binding ability, DBP-5 was speculated to have a role in chromatin organization and/or transcriptional regulation. This report also suggested the homology of DBP-5 with gallin and mos as described in a previous study by Berdichevskii et al. [Berdichevskii et al., 1988]. However, the authors claimed that this finding may not be informative,

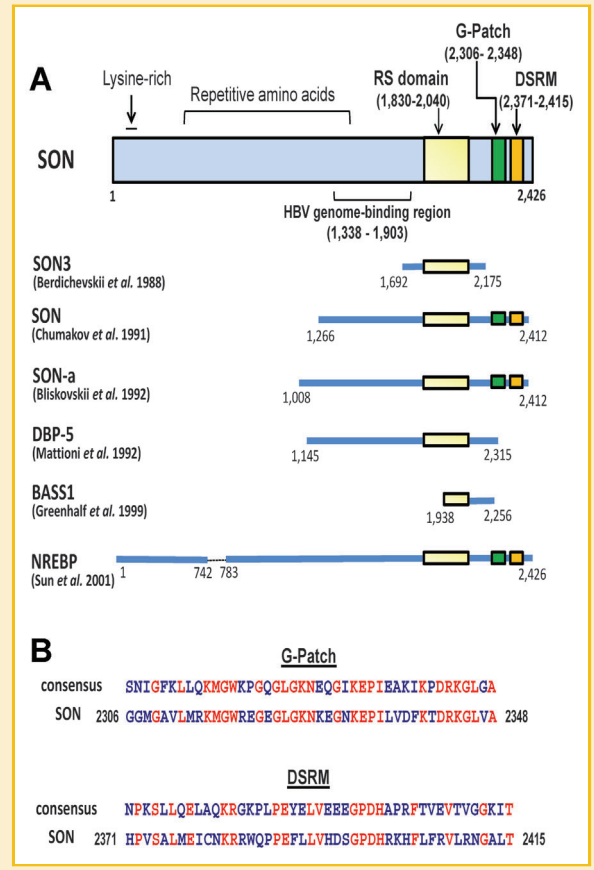

Fig. 1. Schematic structure of the SON protein and various partial fragments. A: Full-length SON consists with 2,426 amino acids, and contains RS domain (serine/arginine-rich domain), G-patch and DSRM (double stranded RNA binding motif). Various partial SON fragments reported in early studies are presented. Numbers are based on full-length SON amino acid sequences (NCBI Reference Sequence NP_620305.2). B: Sequence alignments of SON with representative G-patch and DSRM consensus sequences. Conversed amino acids are marked in red.

TABLE I. Early Studies on SON

Name designated	Suggested functions/findings	Size of the protein examined	Sources	GenBank Acc. No. (#nucleotide, *protein)	References
SON3	Cloned from the human embryonic cDNA library contains tandem repeats	483 a.a.	Human embryonic cDNA library	#M36428.1, *AAA36624.1	Berdichevskii et al. [1988]
SON	Changes morphology of transformed cells to epithelial-like shape	1,265 a.a.	Human liver cells, Human primary embryonic cells	Not available	Chumakov et al. [1991]
SON-a	Contains four types of amino acids tandem repeats	1,523 a.a.	Human placenta cDNA	#A63753.1, *CAA45282.1	Bliskovskii et al. [1992]
DBP-5	Binds to the promoter of the HLA class II gene. Localized in the nucleus, ubiquitously expressed	1,179 a.a.	Normal human B cells	#X63071.1, *CAA44793.1	Mattioni et al. [1992]
BASS1	Protects cells from Bax-mediated apoptosis	345 a.a.	Human brain tissue	#AF139897.1, *AAD50078.1	Greenhalf et al. [1999]
NREBP	Represses transcription of human hepatitis B virus genes	2,386 a.a.	HepG2 hepatoma cells	#AY026895.1, *AAK07692.1	Sun et al. [2001]

since the homology was found within the repetitive amino acid sequences [Mattioni et al., 1992].

In 1999, a cDNA library screen using human brain tissue identified *BASS1* (Bax antagonist selected in *saccharomyces 1*) as being similar to *SON* [Greenhalf et al., 1999]. BASS1 was found to overcome Bax-mediated growth arrest in yeast, and was also able to protect mammalian cells (COS-1) from apoptosis induced by Bax over-expression and staurosporine treatment. In Northern blot with the *BASS1* probe, strong expression of *BASS1* was observed in the heart, placenta, skeletal muscle, and pancreas. Weak expression was seen in the brain, lung, liver and kidney [Greenhalf et al., 1999].

Adopting a different name for a protein with homology to SON, Sun et al. [2001] described a gene encoding a cDNA sequence originating from the HepG2 hepatoma cells. The protein product of this cDNA interacted with the negative regulatory element (NRE) located in the core promoter and the enhancer II (ENII) of hepatitis B viral genome. The authors named this protein NREBP (NRE binding protein). The interaction of NREBP to NRE was demonstrated by gel-shift experiments, and overexpression of NREBP was able to repress NRE in a reporter assay.

In summary, early studies recognized several overlapping cDNA fragments of SON which are the partial regions of the full-length SON (Table I and Fig. 1). It is not clear whether these cDNAs merely represent partial sequences of full-length SON, or they are indeed transcription/splicing variants. It is likely that different stop codon locations were predicted due to errors in sequencing of these repetitive DNAs. Taken together, early reports on various functions of the partial SON fragments have implied unique functions of each domains/regions as well as functional significance of full-length SON in multiple cellular processes.

THE SON GENE

The *SON* locus and the *SON* gene were described in greater detail following the isolation of a cDNA clone from a human keratinocyte cDNA library and subsequent investigation to assign its chromosomal location using human/murine somatic cell hybrid library [Khan et al., 1994]. The genetic locus harboring the *SON* is found on human chromosome 21, in the distal arm region encompassing bands q22.1–q22.2. Sun et al. [2001] reported that human SON (NREBP) gene contains 13 exons and 12 introns. Later, it is more widely accepted that exons 3 and 4 designated by Sun et al. are processed as one exon, consisting total 12 exons for full-length [Reymond et al., 2001]. The murine locus for *Son* has been described in detail by Wynn et al. [2000]. *Son* spans over 35 kilobases (kb) on mouse chromosome 16, encompassing 12 exons and 11 introns. An obvious feature of the SON gene structure is the massive size of exon 3, relative to other exons of this same gene. Exon 3 was reported to contain 70% of the coding sequence for Son [Wynn et al., 2000]. This exon harbors the coding sequence for long stretches of repetitive amino acids as well as the serine–arginine (SR)-rich domain which is significant for this protein classification among SR family members [Khan et al., 1994; Wynn et al., 2000; Saitoh et al., 2004]. While comparing the genomic sequence to the Expressed Sequence Tags database (dbEST), they identified a rare, novel Son cDNA sequence containing a potential alternative exon, giving rise to a truncated Son [Wynn et al., 2000]. Additionally, six alternative transcripts of human SON have been predicted based on exon–intron junction sequence analyses as forms A–F [Reymond et al., 2001], suggesting the presence of multiple splice variants.

PROTEIN STRUCTURE

The SON protein is a multi-domain protein made up of 2,426 amino acids (full-length, Fig. 1) [Ahn et al., 2011]. The N-terminal region of SON contains a lysine-rich region, and multiple tandemly repeated amino acid sequences cover the central portion of SON [Wynn et al., 2000; Sun et al., 2001; Saitoh et al., 2004; Sharma et al., 2010; Ahn et al., 2011]. Extensive amino acid repeats, including MDSQMLASST and SMMSSAYERS, might serve as a scaffold for accessory protein loading similar to the function of the heptad-repeat sequences (YSPTSPS) identified at the C-terminal domain of RNA polymerase II. Centrally located but toward the C-terminus is the region shown to interact with an NRE of the HBV genome [Sun et al., 2001], representing a potential DNA-binding region. Adjacent to this region (C-terminal) contains a serine/arginine-rich domain, called RS domain. The RS domain has been reported to mediate protein-protein interactions of SR protein family members, known to have a role in RNA processing, including alternative splicing [Birney et al., 1993; Graveley, 2000; Sacco-Bubulya and Spector, 2002; Zhong et al., 2009]. Similar to other SR protein members, SON contains RNA-binding motifs in addition to the RS domain. While most of the splicing factors bear the RRM (RNA recognition motif), SON has two distinct RNA binding motifs, G-patch and the double stranded RNA-binding motif (DSRM) at its C-terminus (Fig. 1, G-patch motif at residues 2,306–2,348 and double-stranded RNA-binding motif (DSRM) at residues 2,371–2,415, numbered according to NCBI Reference Sequence NP_620305.2) [Reymond et al., 2001]. The G-patch module is enriched with glycine residues across a 48-amino acid stretch and is widespread among eukaryotes. The presence of six highly conserved glycine residues together with bulky, hydrophobic residues (I, L, V, M) and aromatic residues (F, Y, W) is the characteristic feature of G-patch. Since the G-patch motif is often combined with other well-known RNA-binding domains, it has been predicted that this domain may function in RNA interaction [Aravind and Koonin, 1999]. The DSRM domain has been found in diverse proteins involved in RNA metabolisms and RNA interference (RNAi), such as PKR (dsRNA-dependent protein kinase) and Dicer [Saunders and Barber, 2003; Masliah et al., 2013]. The presence of both G-patch and DSRM suggests the role of SON in multiple processes of RNA regulation.

SEQUENCE CONSERVATION AND EXPRESSION

The *SON* gene has been found to be highly conserved in mammals, and the SON protein sequence showed about 84% of similarity between human and mouse [Wynn et al., 2000]. Interestingly, the genes for SON and SON homologues found in mammals (human, rodents, primates, dog, and cattle) encode the large-sized proteins (over 2,000 amino acids), other non-mammal vertebrates and lower organisms (zebrafish, *Drosophila*, *Caenorhabditis elegans*) have a

SON homologue with smaller size (700–1,000 amino acids). While the C-terminal domains of SON, G-patch and DSRM, seem to be highly conserved across many species, the N-terminus and the region containing the multimeric amino acid repeats appear only in mammals (Refer to NCBI HomoloGene, hgid 10551 and hgid 124220). This leads to an interesting speculation that higher organisms may require the long, repetitive sequences in the SON protein to carry out specific cellular functions.

While SON is ubiquitously expressed in human tissues [Mattioni et al., 1992; Sun et al., 2001], higher expression levels have been found in heart, placenta, and pancreas as well as hematopoietic organs/cells such as thymus, spleen, lymph node, and peripheral blood leukocytes [Sun et al., 2001; Ahn et al., 2013]. Interestingly, high levels of SON expression have been observed in developing, undifferentiated tissues. A higher SON level was described in hematopoietic stem cells compared to total bone marrow cells or differentiated macrophages [Ahn et al., 2013]. It has also been shown that Son expression is higher in developing pancreas in embryo compared to fully differentiated adult islet cells [Hoffman et al., 2008]. Thus, the underlying theme suggests that SON expression may facilitate the maintenance of cells in an undifferentiated state at least in some cell/tissue types.

SON IN GENE EXPRESSION AND CANCER

CRITICAL ROLES OF SON IN CELL CYCLE, MICROTUBULE ORGANIZATION AND GENOME STABILITY

Early studies and screenings have inferred SON functions in DNA-binding, and analyses of its amino acid composition and motifs have raised a possibility that SON may be involved in RNA metabolisms. However, the function of full-length SON had not been characterized until recent exploration with RNA interference (RNAi) approach. A connection between SON and cell cycle has been documented in several reports following the depletion of SON expression by SON siRNA transfection into various cell lines [Ahn et al., 2008, 2011; Huen et al., 2010; Sharma et al., 2010; Sharma et al., 2011]. The most prominent phenotype after SON knockdown was mitotic arrest. Following the knockdown of SON, cell cycle arrest at prometaphase/metaphase of mitosis was observed [Huen et al., 2010; Sharma et al., 2010; Ahn et al., 2011]. Microtubule staining of the cells in mitotic arrest revealed severe defects in mitotic spindle organization and a failure in chromosome alignment during metaphase was observed in the absence of SON [Huen et al., 2010; Ahn et al., 2011]. This cellular arrest was shown to involve the activation of MAD2, a spindle assembly checkpoint protein [Huen et al., 2010]. Furthermore, a limited number of cells that went through the mitosis showed significant abnormalities including aneuploidy/polyploidy [Ahn et al., 2011]. Centrosome abnormalities, including centrosome amplification and impaired centrosome separation during mitosis, were also found after SON depletion [Ahn et al., 2011]. It is worth noting that SON has been identified while screening the phosphoproteome of the mitotic spindle complex, thereby suggesting a direct role in mitosis [Nousiainen et al., 2006]. However, the capacity to which SON plays in this cellular location remains unknown. In the absence of SON expression, microtubule formation in the interphase cells was also significantly aberrant as depicted within the cytoskeletal framework showing a dense aster pattern of microtubules originating from the centrosome and devoid of forming proper microtubule networks [Ahn et al., 2011].

In addition, SON depletion caused compromised DNA integrity which was identified by increased γH2A.X, a marker of double-stranded DNA breaks [Ahn et al., 2011]. SON depletion also caused micronuclei, nucleoplasmic bridges, and nuclear buds [Huen et al., 2010; Ahn et al., 2011], which indicate chromosomal break and instability. Consisting with these observations, SON was found as one of the factors associated with genome stabilization in a genome-wide siRNA screen [Paulsen et al., 2009]. These results demonstrated pivotal roles of SON in maintenance of genome stability.

SON AND RNA SPLICING

As described above, SON depletion leads to defects in mitotic apparatus and chromosome stability. A recent work by Ahn et al. provided the explanation for how the deficiency of the SON protein causes these multiple defects. Microarray and extensive RT-PCR analyses demonstrated that quantitative loss of SON results in increased level of multiple intron retention during RNA splicing of diverse genes related to cell cycle, DNA repair, and cell signal pathway. The affected genes include microtubule/cell cycle-related genes such as gamma-tubulin (*TUBG1*), gamma-tubulin complex protein 2 (*TUBGCP2*), pericentrin (*PCNT*), katanin p80 (*KATNB1*), and aurora kinase B (*AURKB*), as well as DNA repair genes, such as *RAD23A* and a fanconi anemia gene (*FANCG*) [Ahn et al., 2011]. Furthermore, RNA immunoprecipitation experiments demonstrated that SON indeed interacts with RNAs of those affected genes, which was the first confirmation of SON′s ability to interact with RNAs. Interestingly, SON depletion did not affect processing of all exon–intron junctions of these genes, and only certain splice junctions were susceptible to SON depletion [Ahn et al., 2011]. More details on specificity of SON regulation of RNA splicing was examined by minigene assay. Using minigenes containing specific exon–intron junction sequences, Ahn et al. demonstrated that SON regulates mRNA splicing of specific target genes, which contain weak splice sites or dual specific splice sites where the sequences are not optimal for the splicing machinery to recognize. Interestingly, when the splicing sites were modified to resemble strong splicing sites at both the 5′ and 3′ exon/intron junctions, splicing was no longer SON-dependent. The exon–intron junctions examined in this study were constitutive splice sites where the intron between the flanking exons should be removed for proper mRNA processing [Ahn et al., 2011]. These results uncovered a novel SON function as a co-activator for RNA splicing of specific target genes bearing weak constitutive splice sites (Fig. 2).

In addition to SON function in constitutive splicing, the role of SON in alternative splicing has also been demonstrated. The first clue was shown in the report by Moore et al. which identified SON as one of the factors controlling alternative splicing of apoptosis regulators, *Bcl-x* and *Mcl1* in a whole-genome siRNA screen [Moore et al., 2010]. Later, Sharma et al. also demonstrated that SON is involved in alternative splicing of multiple genes. According to the exon array results in this report, absence of SON leads to exon skipping at specific RNA regions from the genes of chromatin modifying enzymes including adenosine deaminase (*ADA*), histone deacetylase 6 (*HDAC6*) and histone lysine

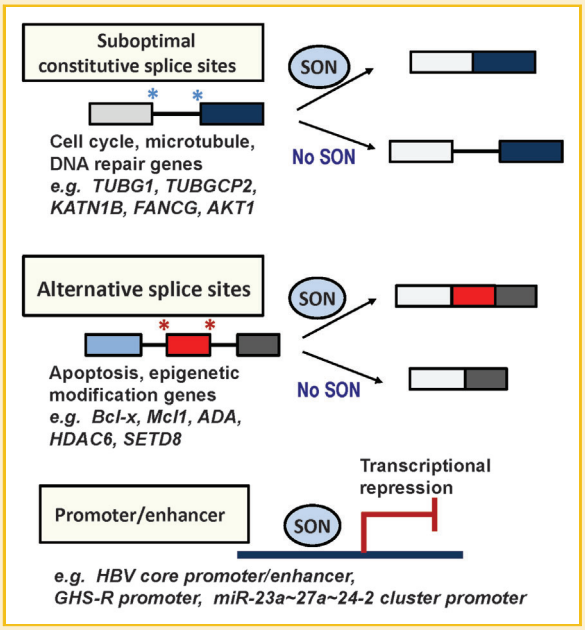

Fig. 2. Functions of SON in RNA splicing and transcription. At the pre-mRNA containing suboptimal constitutive splice sites, SON is required for recognizing these weak splice sites (blue asterisks) as exon–intron boundaries in order to remove the intron. Without SON, intron retention occurs, which results in downregulation of properly processed mRNAs. Multiple genes involved in cell cycle, microtubule biogenesis/organization and DNA repair are found to contain suboptimal constitutive splice sites, and their RNA splicing is affected by the SON level. At the alternative splice sites (red asterisks), SON functions to include alternative exons, which suggests SON is required for recognition of the 5′ and 3′ end of the alternative exon as splice sites. Skipping of the alternative exon occurs in the absence of SON. Several genes involved in apoptosis and epigenetic modification have been found as affected genes. SON also suppresses gene transcription at the promoter/enhancer by unknown mechanisms, and only a few target genes have been identified.

N-methyltransferase 8 (*SETD8*) [Sharma et al., 2011]. These results have uncovered SON′s function in alternative splicing by dictating inclusion and skipping of alternative exons, especially for the genes associated with apoptosis and epigenetic regulation (Fig. 2).

Immunostaining experiments revealed that the SON protein co-localizes with a well-known SR protein, SC35, and small nuclear RNAs (snRNAs), the core components of the spliceosome [Sharma et al., 2010; Ahn et al., 2011]. These results were also confirmed by RNA-FISH combined with SON immunostaining, which demonstrated that endogenous SON and splicing factors co-exist at the target gene transcription site [Sharma et al., 2011]. Interestingly, depletion of SON influences SC35 localization, turning the irregular nuclear speckle pattern of SC35 staining into round shaped dots. In addition, the interaction of SC35 with RNA polymerase II, U1-70K, and $U2AF_{65}$ was weakened by SON knockdown [Ahn et al., 2011], indicating that SON facilitates physical coupling between transcriptional machinery and RNA splicing factors. SON-mediated co-transcriptional RNA splicing may be particularly important in weak splice site processing, since the C-terminal tandem repeat domain (CTD) of RNA polymerase II could efficiently recruit splicing co-factors, such as SR proteins, necessary for enhancing splice site recognition (Fig. 3 and [Ahn et al., 2011]). Transiently overexpressed N-terminus of SON, which contains amino acid repeats and the RS domain, was also localized with target transcription site [Sharma et al., 2011]. These data suggest that the N-terminal region of SON may function as a core protein for organizing spliceosome complexes. However, Ahn et al. [2011] demonstrated that in addition to the RS domain, the G-patch motif present at the C-terminus of SON is required for its RNA splicing function. Therefore, it is likely that the N-terminus of SON provides docking sites for the recruited splicing factors while the G-patch plays a role in holding the RNA strand in the spliceosome complex to achieve efficient RNA splicing.

Recently, proteomics-based native RNA-protein immunoprecipitation in tandem (RIPiT) provided several new associated proteins with Exon Junction Complex (EJC) which is a large multiprotein complex deposited upstream of exon–exon junctions after splicing. The EJC proteins protect spliced mRNA from nuclease digestion, control mRNA export and decay, and enhance translation [Le Hir et al., 2000; Nott et al., 2004; Tange et al., 2004; Gudikote et al., 2005;

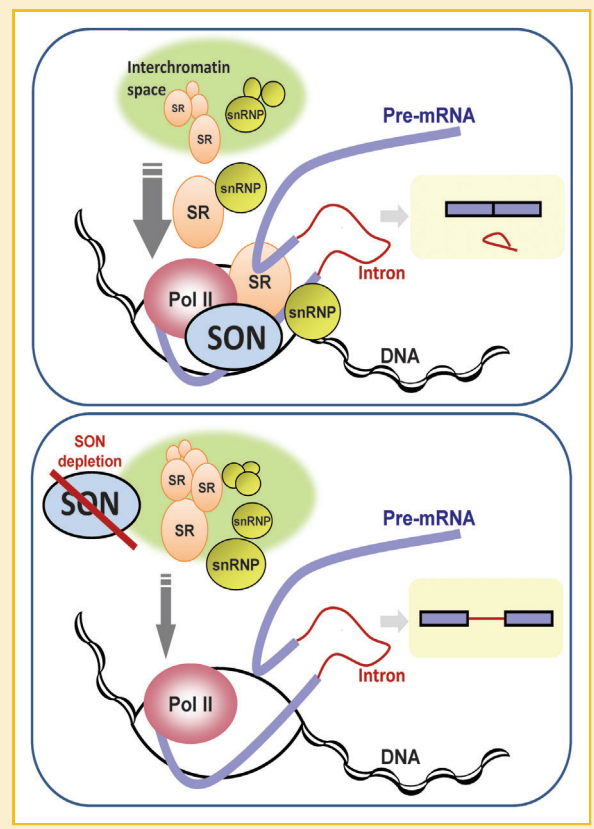

Fig. 3. SON-mediated co-transcriptional RNA splicing. SON facilitates recruitment of SR proteins (such as SC35) and snRNPs (small nuclear ribonucleic particles, such as U1 complex) from the interchromatin space (green area) to the RNA polymerase II complex at the site of active transcription. This efficient recruitment (thick gray arrow) enhances the processing of the splice sites, especially the weak splice sites or alternative splice sites which required SR proteins for exon–intron boundary recognition. Without SON, the recruitment of these factors is not efficient (thin gray arrow), resulting in delayed or impaired processing of the transcripts.

Giorgi and Moore, 2007; Ma et al., 2008]. Surprisingly, SON was found to be enriched in EJC together with multiple SR proteins (SRSF1, 3, 7, 9, 10, and 12) [Singh et al., 2012]. This finding suggests that SON is not only involved in the RNA splicing process, but also remains on spliced mRNAs to control further steps in RNA metabolisms. Based on its large size and ability to recruit SR proteins, SON may work as a factor bridging protein–protein interactions in EJC [Singh et al., 2012].

SON AND TRANSCRIPTION

A partial sequence of SON was initially identified in a screening for DNA-binding factors specific to HLA class II promoters [Mattioni et al., 1992]. Later, SON was named as negative regulatory element binding protein (NREBP), since it was identified to interact with the NRE DNA sequence of human hepatitis B virus (HBV) genome. SON interaction with viral NRE downregulates viral gene expression and replication, suggesting SON has a repressive effect at the regulatory elements. In this study, a pool of random oligonucleotides were incubated with a partial fragment of NREBP, and a potential SON-interacting, NRE-like sequence, (G/T)AN(C/G)(A/G)CC, was determined by PCR. In the mouse hypothalamus, SON was upregulated by leptin injection and SON expression affected the promoter activity of growth hormone secretagogue-receptor (GHS-R) [Komori et al., 2010]. Recently, SON was shown to regulate the promoter of the microRNA (miR) cluster *miR-23a∼27a∼24-2* in human and mouse hematopoietic cells. Upon knockdown of *SON*, miR-27a expression is increased and miR-27a targets GATA-2, a potent transcriptional regulator for the maintenance of hematopoietic stem cells [Ahn et al., 2013]. Taken together, accumulating evidence has shown that SON is involved in transcriptional regulation, especially in suppression of promoter activity (Fig. 2). However, whether SON recognizes only NRE-like sequence, or there are more other factors determining the specificity of SON-mediated transcriptional repression remains unclear.

SON AND TUMORIGENESIS

As described above, SON regulates multiple cellular processes critical for cancer development and progression, such as cell cycle, cell proliferation, apoptosis, chromatin remodeling and genome stability. Although its exact contributions still need to be defined, SON has been implicated in cancer cell proliferation. The first report investigating the role SON plays in tumor development found that it had tumor-suppressor qualities when partial SON sequence was expressed in NIH3T3 cells with enhanced metastatic potential (information based on English translation of the report Chumakov et al. [1991], written in Russian). However, another partial fragment of SON, BASS1, was later found to inhibit Bax-mediated apoptosis in yeast and mammalian cells [Greenhalf et al., 1999]. Recently, SON was shown to be highly expressed in the bone marrow cells from acute myeloid leukemia patients [Ahn et al., 2013]. In addition, SON interacts with a leukemia-causing fusion protein, AML1-ETO, probably with multiple contact points within AML1-ETO, including the region critical for leukemogenesis [Ahn et al., 2008; DeKelver et al., 2013]. The exact role of SON in leukemia development and/or progression still needs further investigation.

It has been confirmed that the absence of SON results in growth arrest and apoptosis in several human cancer cell lines [Ahn et al., 2008, 2011; Sharma et al., 2010; Furukawa et al., 2012]. In addition, knockdown of SON showed the remarkable anti-proliferative phenotypes in an siRNA screen using a pancreatic cancer cell line, MIA PaCa-2 [Furukawa et al., 2012]. When aggressive pancreatic cancer cells were depleted of SON expression and inoculated into nude mice, tumor growth was significantly diminished. The authors noted that cells depleted of SON exhibited cell cycle arrest (G2/M arrest) and apoptosis. Importantly, in pancreatic ductal adenocarcinoma patients, the expression of SON was indeed strongly increased compared to non-neoplastic ducts or pancreatic intraepithelial neoplasia (a precursor lesion of ductal adenocarcinoma) [Furukawa et al., 2012]. Thus, overexpression of SON may play a critical role in cell proliferation and survival during pancreatic cancer progression. Besides cell cycle/proliferation-related genes, SON regulates RNA splicing of several epigenetic modification factors which are closely associated with tumorigenesis [Sharma et al., 2011]. SON regulates proper RNA splicing for *HDAC6*, *DNMT1*, and *SETD8*, whose overexpressions have been shown to promote cancer development and progression. HDAC6 overexpression has been linked in cancer cell motility/metastasis [Tran et al., 2007; Aldana-Masangkay and Sakamoto, 2011], and is a target of leukemia therapy [Hackanson et al., 2012]. Mutations/overexpression of *DNMT1* and *SETD8* has also been shown to be critical for cancer cell proliferation and survival [Chen et al., 2007; Takawa et al., 2012]. Given the significance of SON target genes in cancer, examining the SON expression pattern in various types of cancer and the role(s) SON plays in tumorigenesis would be promising future investigation.

FUTURE PERSPECTIVES

After being untouched for almost two decades, the SON protein is now showing off its critical functions in gene regulation, which is directly related to cell cycle, genome stability, and cell survival. As described above, recent findings on the molecular mechanisms of SON function in RNA splicing and transcriptional control have highlighted SON as a novel multifunctional player in gene expression. While several early studies implied SON′s function in tumorigenesis and apoptosis, most of these works were performed with partial sequences of SON with inconsistent nomenclature, which delayed our understanding of the *SON* gene and the SON protein. Recent development and progression in RNAi approaches, cloning of full-length cDNA and generation of antibodies would be able to facilitate further characterization of SON.

For our complete understanding of this protein, there are still many questions waiting to be answered. Although SON has been found to work for RNA splicing, its domains and structures are unique when compared to those of other well-known splicing factors. While most of the SR protein family members possess the RRM (RNA-recognition motif) as a domain responsible for RNA interaction, SON instead has G-patch and DSRM as RNA binding domains, suggesting SON may have a unique role in selecting target RNAs. In SON-mediated minigene splicing, the RS domain and G-patch were necessary [Ahn et al., 2011]. It is still unclear whether the DSRM, another RNA-binding motif present in SON, is also necessary for splicing certain RNAs, or if DSRM is totally dispensable for RNA splicing function of SON. Since DSRM is often found in other RNA binding proteins

involved in RNA transport and miR biogenesis [Saunders and Barber, 2003; Masliah et al., 2013], it would be interesting to delineate SON functions mediated by its DSRM. In addition, detail mechanisms of SON-mediated transcriptional control are largely unknown. So far, only a few target promoters have been reported, and certainly genome-wide screening of target promoters affected by SON would be one of the interesting assignments for us. Although a potential SON binding motif has been proposed in an early study [Sun et al., 2001], this sequence was drawn from incubation of partial SON fragment with a pool of random oligonucleotides. Therefore, we cannot rule out that full-length SON might recognize different or more diverse sequences for DNA binding. Identification of SON domains/structure necessary for transcriptional control and *cis*-/*trans*-elements conferring SON-dependency should provide a clue to the novel mechanism of SON function.

Besides full-length SON, several splice variants of SON have been predicted by sequence analyses [Reymond et al., 2001]. Interestingly, some of these predicted variants include the isoforms missing critical functional domains such as RNA-binding motifs. Investigation of isoform expression in different tissues and cell types as well as the functional importance of these isoforms will expand our understanding of the SON protein. Several studies have found that antisera or purified antibodies against SON detected multiple bands in Western blot [Chumakov et al., 1991; Mattioni et al., 1992; Sun et al., 2001; Ahn et al., 2008], suggesting the presence of various forms of SON. Finally, the most exciting study would be to elucidate the role of SON in human diseases. In addition to potential impacts of SON overexpression in cancer cell proliferation/survival, there are a few more indications of abnormalities of SON localization in cancer, such as cytoplasm-localized SON in t(8;21)-positive acute myeloid leukemia patient cells [Ahn et al., 2008], and secreted SON detected in the plasma of lung adenocarcinomas mouse model [Taguchi et al., 2011]. How the SON protein was detected in these aberrant cellular locations is completely unknown. Certainly, there are more stories hidden behind SON. Our future work and effort on this large protein may uncover new aspects of gene regulation and reveal potential cancer therapeutic targets.

ACKNOWLEDGMENTS

We thank Erika Tumulak for her help in editing the manuscript. This work was supported by the start-up fund from Mitchell Cancer Institute, University of South Alabama (to E.Y.A).

REFERENCES

Ahn EY, Yan M, Malakhova OA, Lo MC, Boyapati A, Ommen HB, Hines R, Hokland P, Zhang DE. 2008. Disruption of the NHR4 domain structure in AML1-ETO abrogates SON binding and promotes leukemogenesis. Proc Natl Acad Sci USA 105:17103–17108.

Ahn EY, DeKelver RC, Lo MC, Nguyen TA, Matsuura S, Boyapati A, Pandit S, Fu XD, Zhang DE. 2011. SON controls cell-cycle progression by coordinated regulation of RNA splicing. Mol Cell 42:185–198.

Ahn EE, Higashi T, Yan M, Matsuura S, Hickey CJ, Lo MC, Shia WJ, DeKelver RC, Zhang DE. 2013. SON protein regulates GATA-2 through transcriptional control of the microRNA 23a∼27a∼24-2 cluster. J Biol Chem 288:5381–5388.

Aldana-Masangkay GI, Sakamoto KM. 2011. The role of HDAC6 in cancer. J Biomed Biotechnol 2011:875824.

Aravind L, Koonin EV. 1999. G-patch: A new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. Trends Biochem Sci 24:342–344.

Berdichevskii FB, Chumakov IM, Kiselev LL. 1988. [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) 22:794–801.

Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res 21:5803–5816.

Bliskovskii VV, Berdichevskii FB, Tkachenko AV, Belova ME, Chumankov IM. 1992. Coding part of the son gene small transcript contains four areas of complete tandem repeats. Mol Biol (Mosk) 26:793–806.

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E. 2007. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet 39:391–396.

Chumakov IM, Berdichevskii FB, Sokolova NV, Reznikov MV, Prasolov VS. 1991. 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) 25:731–739.

DeKelver RC, Yan M, Ahn EY, Shia WJ, Speck NA, Zhang DE. 2013. Attenuation of AML1-ETO cellular dysregulation correlates with increased leukemogenic potential. Blood 121:3714–3717.

Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and human disease. Genes Dev 17:419–437.

Furukawa T, Tanji E, Kuboki Y, Hatori T, Yamamoto M, Shimizu K, Shibata N, Shiratori K. 2012. Targeting of MAPK-associated molecules identifies SON as a prime target to attenuate the proliferation and tumorigenicity of pancreatic cancer cells. Mol Cancer 11:88.

Garcia-Blanco MA, Baraniak AP, Lasda EL. 2004. Alternative splicing in disease and therapy. Nat Biotechnol 22:535–546.

Giorgi C, Moore MJ. 2007. The nuclear nurture and cytoplasmic nature of localized mRNPs. Semin Cell Dev Biol 18:186–193.

Graveley BR. 2000. Sorting out the complexity of SR protein functions. RNA 6:1197–1211.

Greenhalf W, Lee J, Chaudhuri B. 1999. A selection system for human apoptosis inhibitors using yeast. Yeast 15:1307–1321.

Gudikote JP, Imam JS, Garcia RF, Wilkinson MF. 2005. RNA splicing promotes translation and RNA surveillance. Nat Struct Mol Biol 12:801–809.

Hackanson B, Rimmele L, Benkisser M, Abdelkarim M, Fliegauf M, Jung M, Lubbert M. 2012. HDAC6 as a target for antileukemic drugs in acute myeloid leukemia. Leuk Res 36:1055–1062.

Hoffman BG, Zavaglia B, Witzsche J, Ruiz de Algara T, Beach M, Hoodless PA, Jones SJ, Marra MA, Helgason CD. 2008. Identification of transcripts with enriched expression in the developing and adult pancreas. Genome Biol 9:R99.

Huen MS, Sy SM, Leung KM, Ching YP, Tipoe GL, Man C, Dong S, Chen J. 2010. SON is a spliceosome-associated factor required for mitotic progression. Cell Cycle 9:2679–2685.

Khan IM, Fisher RA, Johnson KJ, Bailey ME, Siciliano MJ, Kessling AM, Farrer M, Carritt B, Kamalati T, Buluwela L. 1994. The SON gene encodes a conserved DNA binding protein mapping to human chromosome 21. Ann Hum Genet 58:25–34.

Komori T, Doi A, Furuta H, Wakao H, Nakao N, Nakazato M, Nanjo K, Senba E, Morikawa Y. 2010. Regulation of ghrelin signaling by a leptin-induced gene, negative regulatory element-binding protein, in the hypothalamic neurons. J Biol Chem 285:37884–37894.

Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. 2013. Alternative splicing: A pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol 14:153–165.

Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. EMBO J 19:6860–6869.

Ma XM, Yoon SO, Richardson CJ, Julich K, Blenis J. 2008. SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. Cell 133:303–313.

Masliah G, Barraud P, Allain FH. 2013. RNA recognition by double-stranded RNA binding domains: A matter of shape and sequence. Cell Mol Life Sci 70:1875–1895.

Mattioni T, Hume CR, Konigorski S, Hayes P, Osterweil Z, Lee JS. 1992. A cDNA clone for a novel nuclear protein with DNA binding activity. Chromosoma 101:618–624.

Moore MJ, Wang Q, Kennedy CJ, Silver PA. 2010. An alternative splicing network links cell-cycle control to apoptosis. Cell 142:625–636.

Nott A, Le Hir H, Moore MJ. 2004. Splicing enhances translation in mammalian cells: An additional function of the exon junction complex. Genes Dev 18:210–222.

Nousiainen M, Sillje HH, Sauer G, Nigg EA, Korner R. 2006. Phosphoproteome analysis of the human mitotic spindle. Proc Natl Acad Sci USA 103:5391–5396.

Paulsen RD, Soni DV, Wollman R, Hahn AT, Yee MC, Guan A, Hesley JA, Miller SC, Cromwell EF, Solow-Cordero DE, Meyer T, Cimprich KA. 2009. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Mol Cell 35:228–239.

Reymond A, Friedli M, Henrichsen CN, Chapot F, Deutsch S, Ucla C, Rossier C, Lyle R, Guipponi M, Antonarakis SE. 2001. From PREDs and open reading frames to cDNA isolation: Revisiting the human chromosome 21 transcription map. Genomics 78:46–54.

Sacco-Bubulya P, Spector DL. 2002. Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. J Cell Biol 156:425–436.

Saitoh N, Spahr CS, Patterson SD, Bubulya P, Neuwald AF, Spector DL. 2004. Proteomic analysis of interchromatin granule clusters. Mol Biol Cell 15:3876–3890.

Saunders LR, Barber GN. 2003. The dsRNA binding protein family: Critical roles, diverse cellular functions. FASEB J 17:961–983.

Sharma A, Takata H, Shibahara K, Bubulya A, Bubulya PA. 2010. Son is essential for nuclear speckle organization and cell cycle progression. Mol Biol Cell 21:650–663.

Sharma A, Markey M, Torres-Munoz K, Varia S, Kadakia M, Bubulya A, Bubulya PA. 2011. Son maintains accurate splicing for a subset of human pre-mRNAs. J Cell Sci 124:4286–4298.

Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ. 2012. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell 151:750–764.

Sun CT, Lo WY, Wang IH, Lo YH, Shiou SR, Lai CK, Ting LP. 2001. Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. J Biol Chem 276:24059–24067.

Taguchi A, Politi K, Pitteri SJ, Lockwood WW, Faca VM, Kelly-Spratt K, Wong CH, Zhang Q, Chin A, Park KS, Goodman G, Gazdar AF, Sage J, Dinulescu DM, Kucherlapati R, Depinho RA, Kemp CJ, Varmus HE, Hanash SM. 2011. Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. Cancer Cell 20:289–299.

Takawa M, Cho HS, Hayami S, Toyokawa G, Kogure M, Yamane Y, Iwai Y, Maejima K, Ueda K, Masuda A, Dohmae N, Field HI, Tsunoda T, Kobayashi T, Akasu T, Sugiyama M, Ohnuma S, Atomi Y, Ponder BA, Nakamura Y, Hamamoto R. 2012. Histone lysine methyltransferase SETD8 promotes carcinogenesis by deregulating PCNA expression. Cancer Res 72:3217–3227.

Tange TO, Nott A, Moore MJ. 2004. The ever-increasing complexities of the exon junction complex. Curr Opin Cell Biol 16:279–284.

Tran AD, Marmo TP, Salam AA, Che S, Finkelstein E, Kabarriti R, Xenias HS, Mazitschek R, Hubbert C, Kawaguchi Y, Sheetz MP, Yao TP, Bulinski JC. 2007. HDAC6 deacetylation of tubulin modulates dynamics of cellular adhesions. J Cell Sci 120:1469–1479.

Wynn SL, Fisher RA, Pagel C, Price M, Liu QY, Khan IM, Zammit P, Dadrah K, Mazrani W, Kessling A, Lee JS, Buluwela L. 2000. Organization and conservation of the GART/SON/DONSON locus in mouse and human genomes. Genomics 68:57–62.

Zhong XY, Wang P, Han J, Rosenfeld MG, Fu XD. 2009. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. Mol Cell 35:1–10.