Supplemental Information

SON and its alternatively spliced isoforms control MLL complex-mediated H3K4me3 and transcription of leukemia-associated genes

Jung-Hyun Kim, Melody C. Baddoo, Eun Young Park, Joshua K. Stone, Hyeonsoo Park, Thomas W. Butler, Gang Huang, Xiaomei Yan, Florencia Pauli-Behn, Richard M. Myers, Ming Tan, Erik K. Flemington, Ssang-Taek Lim, and Eun-Young Erin Ahn

- **SUPPLEMENTAL DATA:** Supplemental Figures 1 - 7

  Supplemental Tables 1 - 3
- **SUPPLEMENTAL EXPERIMENTAL PROCEDURES**
- **SUPPLEMENTAL REFERENCES**

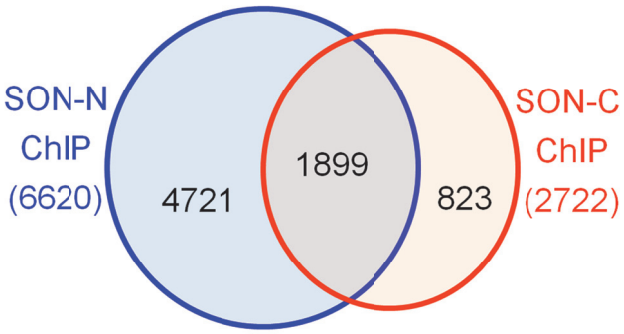
SUPPLEMENTAL DATA

Supplemental Figures

A

	SON-N Ab	SON-C Ab
Positive peaks (Enriched peaks)	6,620	2,722
Negative peaks (Peaks enriched in the background library versus the ChIP library)	59	77

B



C

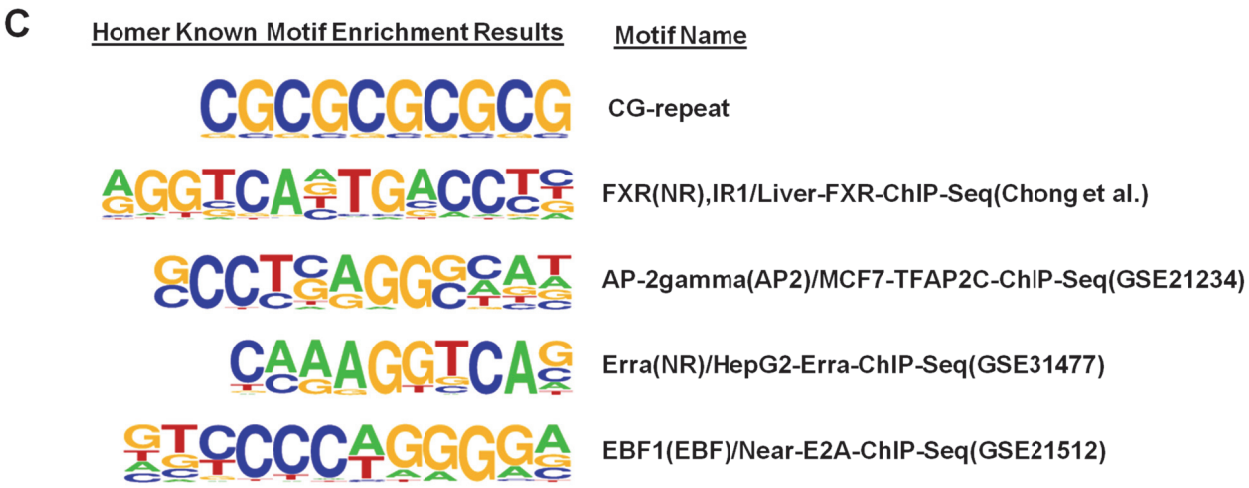


Figure S1. Whole Genomic Distribution of SON-Binding Sites (Related to Figure 1).

(A) The result of a pilot ChIP-seq study to verify specificity of two SON antibodies, SON-N Ab and SON-C Ab. The numbers of ChIP-seq peaks are summarized.

(B) Venn diagram showing number of SON peaks determined by a pilot study ChIP-seq using SON-N Ab and SON-C Ab.

(C) Top five “known motifs” enriched in SON-binding sites analyzed by HOMER.

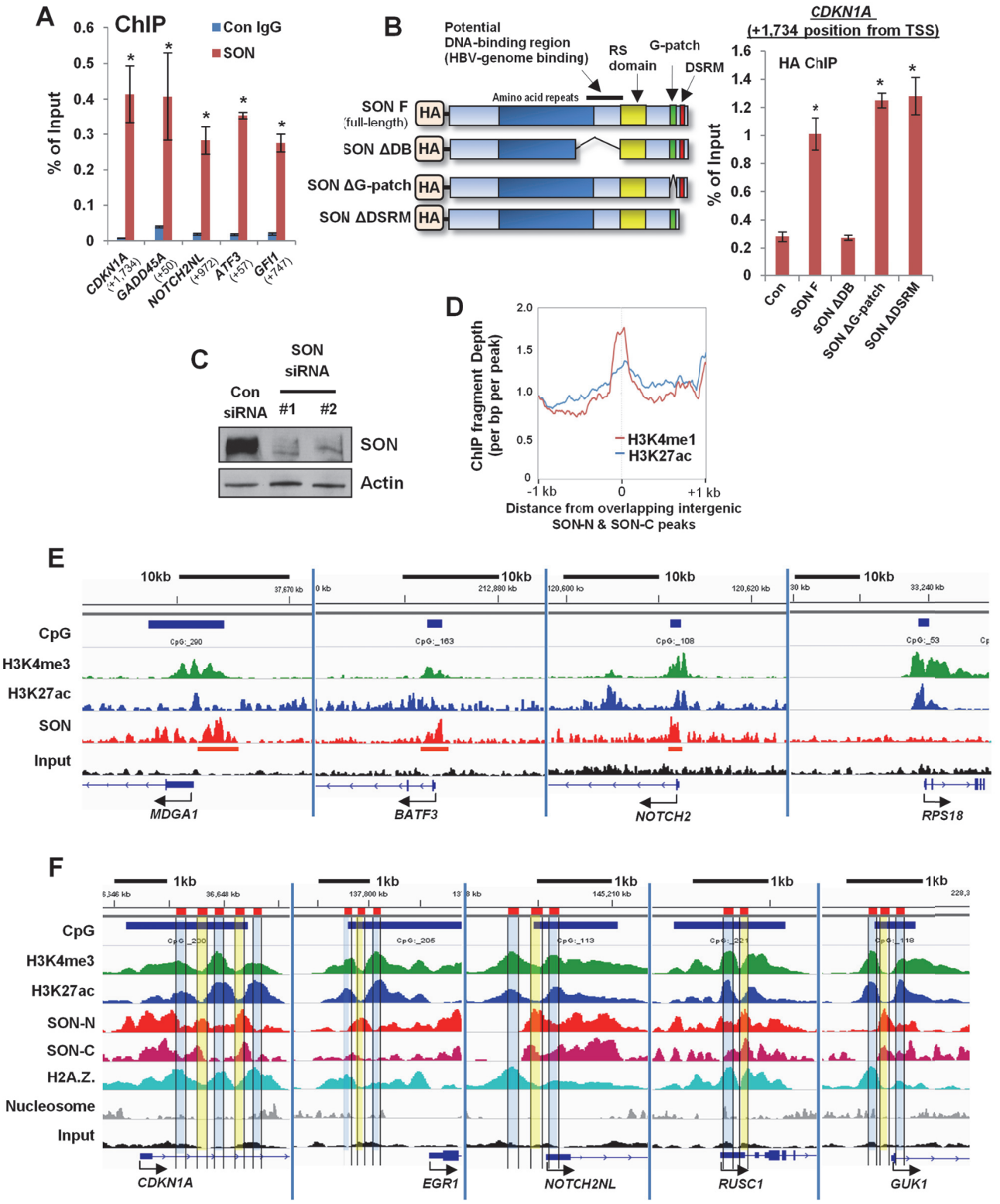


Figure S2. Validation of SON-DNA Binding and Various Analyses of SON ChIP-seq Results (Related to Figures 1 and 2).

(A) SON ChIP-qPCR analysis (with SON-N Ab) confirming SON enrichment near the transcription start site of indicated genes in K562 cells. The number in the parenthesis indicates the base-pair counts from the transcription start site of each gene, and the primers for qPCR (Table S3) were designed to detect SON enrichment around these positions. *p-values <0.01.

(B) HA ChIP-qPCR analysis measuring DNA-binding ability of HA-tagged full-length SON and several deletion mutants in K562 cells; Full-length SON (SON F), potential DNA-binding region-deleted SON (SON ΔDB, amino acids 1,263 - 1,818 deleted), G-patch-deleted SON (SON ΔG-patch), and double-stranded RNA binding motif-deleted SON (SON ΔDSRM). Potential DNA-binding region that has been shown to interact with human hepatitis B virus genome (Sun *et al*., 2011) is indicated. Empty vector transfected sample was used as a control. *p-values <0.01.

(C) Western blot confirmed knockdown efficiency of two different siRNAs (#1 and #2) against *SON* in K562 cells.

(D) Average signal profiles of H3K4me1 and H3K27ac around the intergenic SON-binding sites.

(E) IGV browser image of density profiles of SON (SON-N Ab ChIP), CpG islands, H3K4me3 and H3K27ac at several examples of SON target genes. SON-binding regions (called peaks) are depicted as red bars at the bottom of SON ChIP-seq results. *RPS18* promoter region is an example of the area where H3k4me3 is present, but SON peak is absent (non-target). Scale bar (top, black), 10kb.

(F) Close-up images of ChIP-seq peaks of SON-N, SON-C, H3K4me3, H3K27ac, H2A.Z and MNase-seq (nucleosome) peaks in representative SON target genes. The areas with high H3K4me3 level (H3K4me3 peaks) and low SON level (SON valleys) are indicated in blue, and the areas with low H3K4me3 (H3K4me3 valleys) and high SON (SON peaks) are indicated in yellow. Scale bar (top, black),1kb.

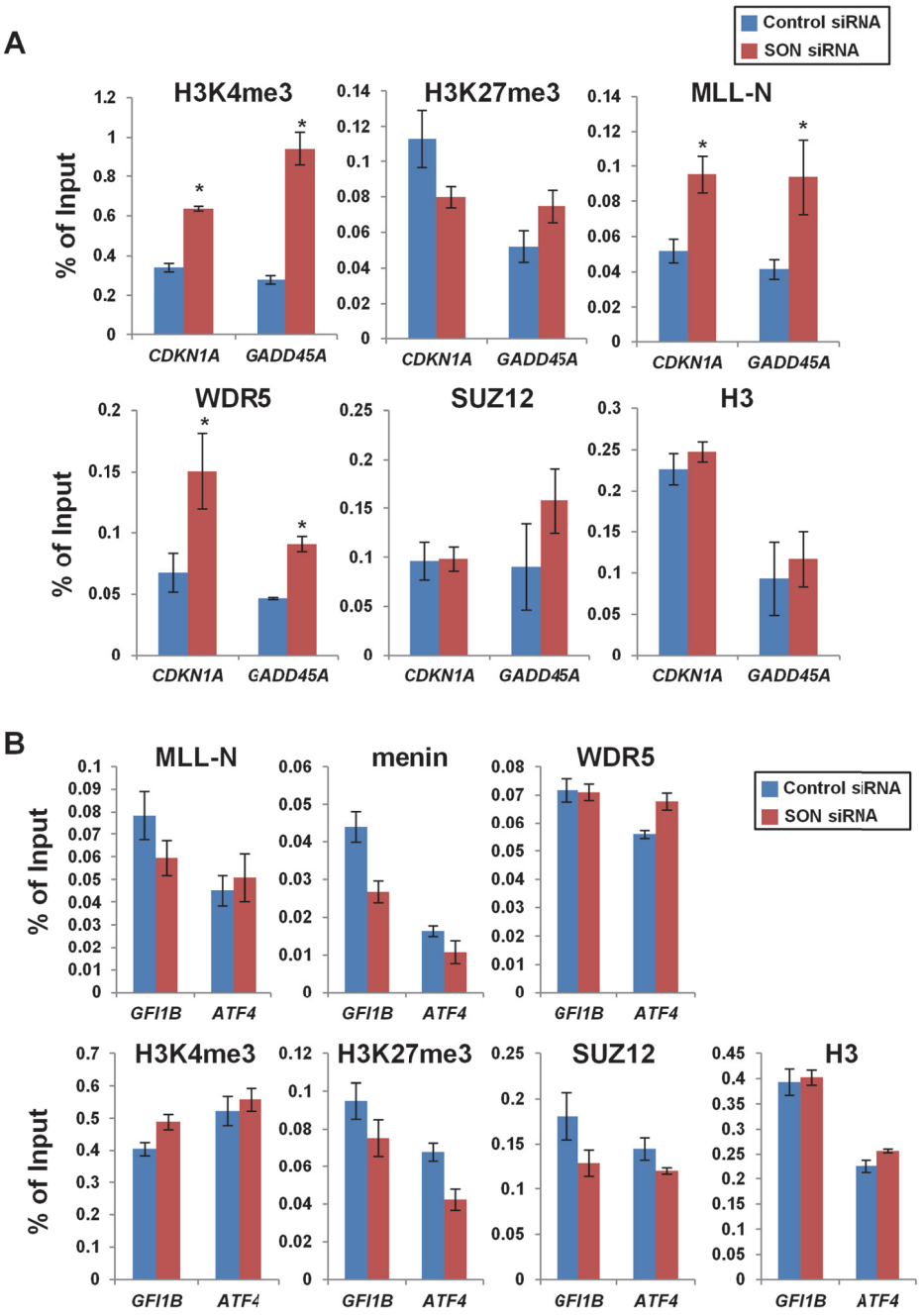


Figure S3. SON Depletion Leads to H3K4me3 Modification and MLL Complex Recruitment (Related to Figure 3). ChIP-qPCR analysis of two SON target genes *(CDKN1A* and *GADD45A*; A) and two non-targets (*GFI1B* and *ATF4*; B), were conducted using indicated antibodies in K562 cells transfected with control or SON siRNA-#2 (A) and #1 (B). ChIP-qPCR results are plotted as percentage of input DNA. *p-values <0.05.

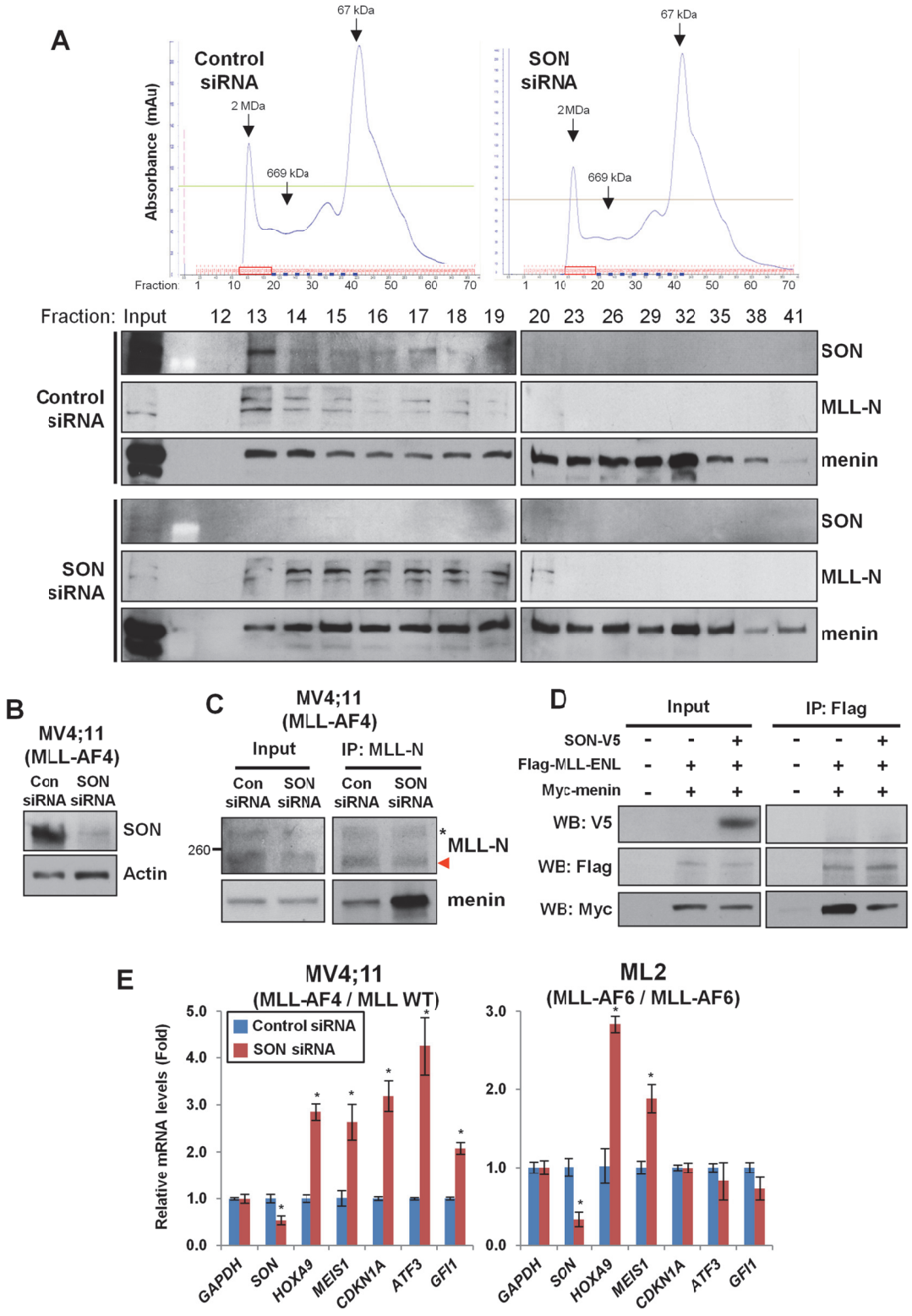


Figure S4. SON Inhibits Menin Interaction with Both Wild-type MLL and MLL-Fusion Protein and Regulates Their Target Gene Expression (Related to Figure 4).

(A) Nuclear extracts from control and SON knockdown-K562 cells were size-fractionated on FPLC and analyzed for SON, MLL1 N-terminus (MLL-N) and menin distribution by Western blotting using indicated antibodies. Fraction numbers in red box on elution profile (top) indicate MLL-N-enriched fractions (left panel of WB, fractions 12-19) and blue line on elution profile (top) indicates further downstream fractions (right panel of WB, fractions 20-41). Arrows indicate the approximate molecular mass of eluted protein complex.

(B) Western blot confirmed knockdown efficiency of SON siRNA in MV4;11 cells.

(C) Nuclear extracts from MV4;11 transfected by control and SON siRNA were used for immunoprecipitation with MLL-N antibody and Western blotted with indicated antibodies. The N-terminus of wild-type MLL1 is marked by an asterisk and the MLL-AF4 fusion protein is marked by a red arrow head.

(D) 293T cells transiently co-expressing Flag-MLL-ENL and Myc-menin were transfected with control vector or SON-V5 construct and subjected to co-immunoprecipitation assays using a Flag antibody. Immunoprecipitated Flag-MLL-ENL and Myc-menin were analyzed by Western blotting.

(E) Expression of target genes of MLL fusion proteins (*HOXA9* and *MEIS1*) or SON protein (*CDKN1A*, *ATF3*, and *GFI1*) measured by real-time qPCR in the control and SON siRNA-transfected MLL-rearranged leukemic cell lines, MV4;11 and ML-2. Note that the genes identified by our SON ChIP-seq in K562 (e.g. *CDKN1A*, *ATF3* and *GFI1*) were also upregulated upon SON knockdown in MV4;11 cells (heterozygous for MLL-AF4) which retain one copy of wild-type MLL. However, these genes were not changed upon SON knockdown in ML2 cells which do not have wild-type MLL (homozygous for MLL-AF6), suggesting that the negative effect of SON on those target genes is exerted through inhibition of the MLL wild-type complex. Data represent three replicates from three independent experiments. *p *values* < 0.01,

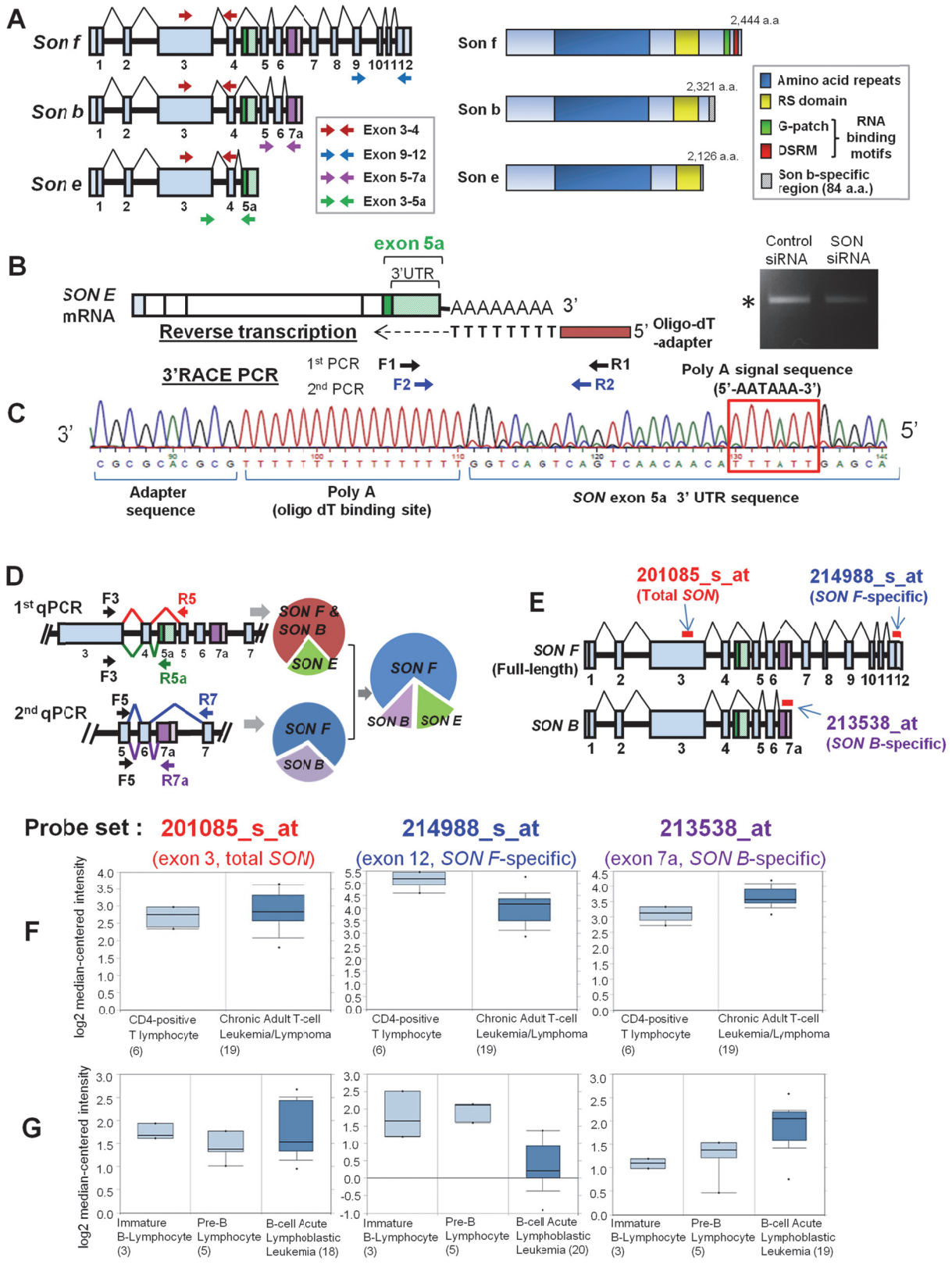


Figure S5. Expression of Short SON Isoforms, SON B and SON E. (Related to Figure 5)

(A) Schematic representation of the mouse *Son* genes and the Son proteins. Similar to human SON, full-length mouse Son (Son f; mouse Son isoform 1) is generated by alignment of 12 constitutive exons (sky blue). Inclusion of alternative exons (exon 7a, labeled in purple; exons 5a, labeled in green) produces two different alternatively spliced isoforms, Son b (a predicted isoform) and Son e (mouse Son isoform 2). The primer sets used for mouse *Son* qPCR (presented in Figure 5F) are indicated with horizontal arrows.

(B) Structure of the *SON E* mRNA and position of PCR primers for 3' RACE PCR to confirm the 3'UTR sequence of *SON E*. Black arrows indicate primers for 1st RACE PCR and blue arrows indicate primers for 2nd RACE PCR. EtBr-stained agarose gel showed 3′ RACE-amplified PCR products (asterisk) that were used for sequencing. Decreased amount of the 3' RACE PCR product in total SON siRNA-transfected K562 cells, compared to control K562 cells, indicates the specificity of 3'RACE PCR.

(C) Sequencing result of the 3' RACE PCR product from human *SON E* transcripts.

(D) A schematic strategy for measurement of relative ratio of SON isoforms presented in Figures 5D, 5E and 5G. To calculate relative expression level of SON isoforms, each qPCR analysis was performed using same forward primer and two different reverse primers. At the 1st qPCR, the ratio of exon 5a-included transcripts (SON E) and exon 5-included transcripts (total of SON F and SON B) was determined using a common forward primer (F3) and exon-specific reverse primers (R5 or R5a). At the 2nd qPCR, the ratio of exon 7a-included transcripts (SON B) and exon 7-included transcript (SON F) was determined using a common forward primer (F5) and exon-specific reverse primers (R7 or R7a). Based on two qPCR analyses, relative levels of SON F, SON B and SON E were determined.

(E) Schematic of genomic structure of full-length *SON* and *SON B*. Each red bar with the probe set number indicates the specific position of DNA probes used in microarray analysis (Affymetrix U133A microarrays).

(F and G) Relative expression levels of SON detected by three different probe sets. The microarray data were from (F) Stegmaier Leukemia Dataset (Stegmaier et al., 2004) and (G) Maia Leukemia

Dataset (Maia et al., 2005) from Oncomine database. The values of log2 median-centered intensity detected by indicated probe sets were displayed as a boxplot according to Oncomine output.

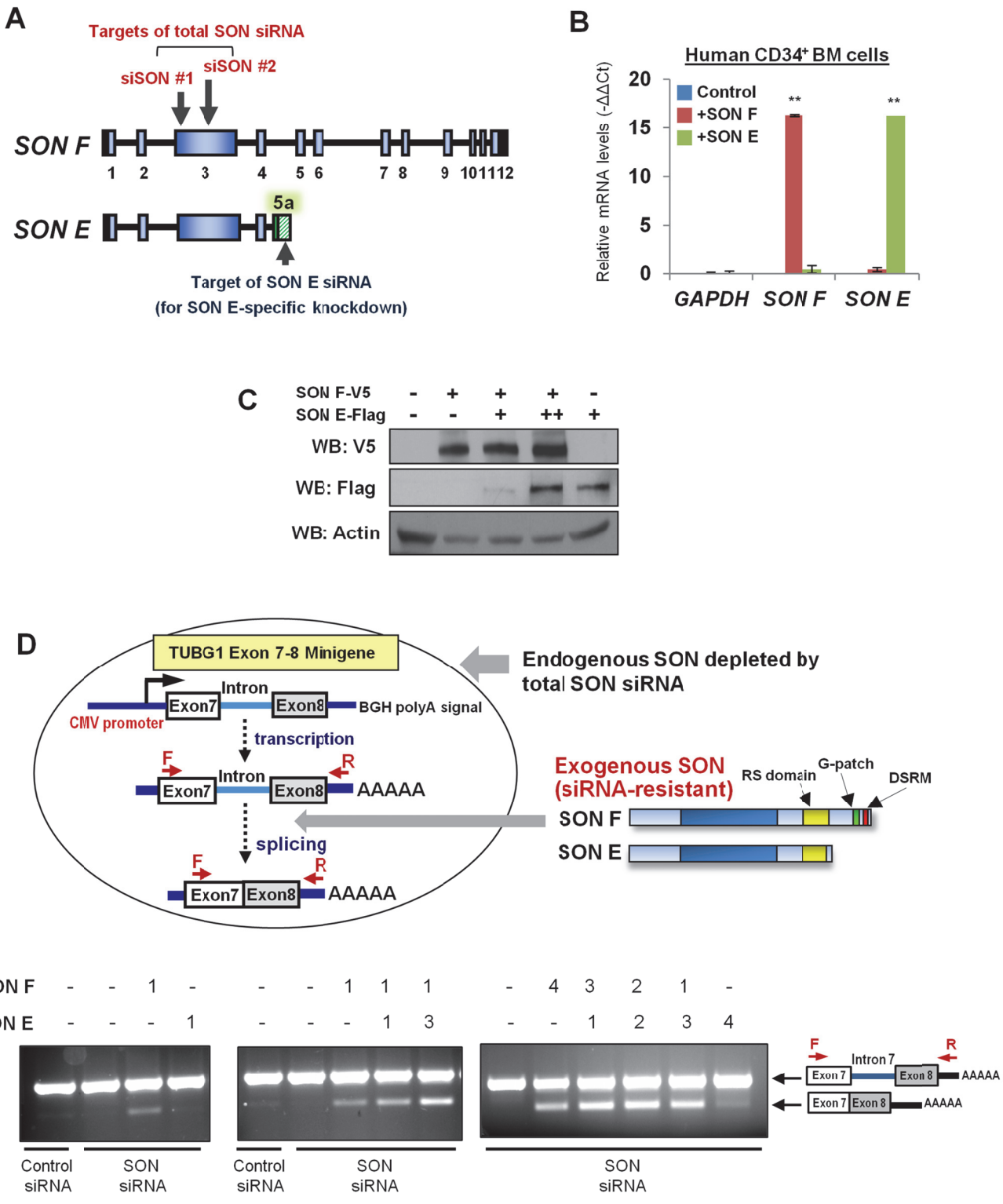


Figure S6. SON F and SON E Overexpression Conditions and Minigene Assays Determining RNA Splicing Ability of SON F and SON E (Related to Figure 6).

(A) Location of the target regions of total SON siRNAs (siRNAs #1 and #2, targeting exon 3) and SON E siRNA (targeting exon 5a). SON siRNA sequences are described in Supplemental Experimental Procedures.

(B) Verification of SON F and SON E overexpression in primary human $CD34^+$ bone marrow cells after nucleofection of the expression constructs. Exogenous expression of SON was determined by qPCR using the forward primers (F-hSON-Exon12 for SON F, F-hSON-Exon4 for SON E) together with a reverse primer targeting the V5 sequence (R-V5-exo) in the plasmid. ***p*-values <0.01.

(C) The expression level of SON F-V5 and SON E-Flag in the K562 cells used for V5-ChIP and Flag-ChIP presented in Figure 6E.

(D) Strategies of minigene assay with the *TUBG1* exon7- intron 7 - exon8 minigene model (containing SON-dependent splicing sites; Ahn et al., 2011; Lu et al., 2013). Splicing efficiencies of transfected SON F and SON E (expressed by the siRNA-resistant form of cDNA; Ahn et al., 2011) were accessed by RT-PCR using a forward primer (F) and a reverse primer (R) to detect unspliced and spliced RNA.

(E) RT-PCR results demonstrating minigene splicing efficiencies. The numbers above the each lane indicates the relative amount of exogenous SON F and/or SON E transfected together with the minigene.

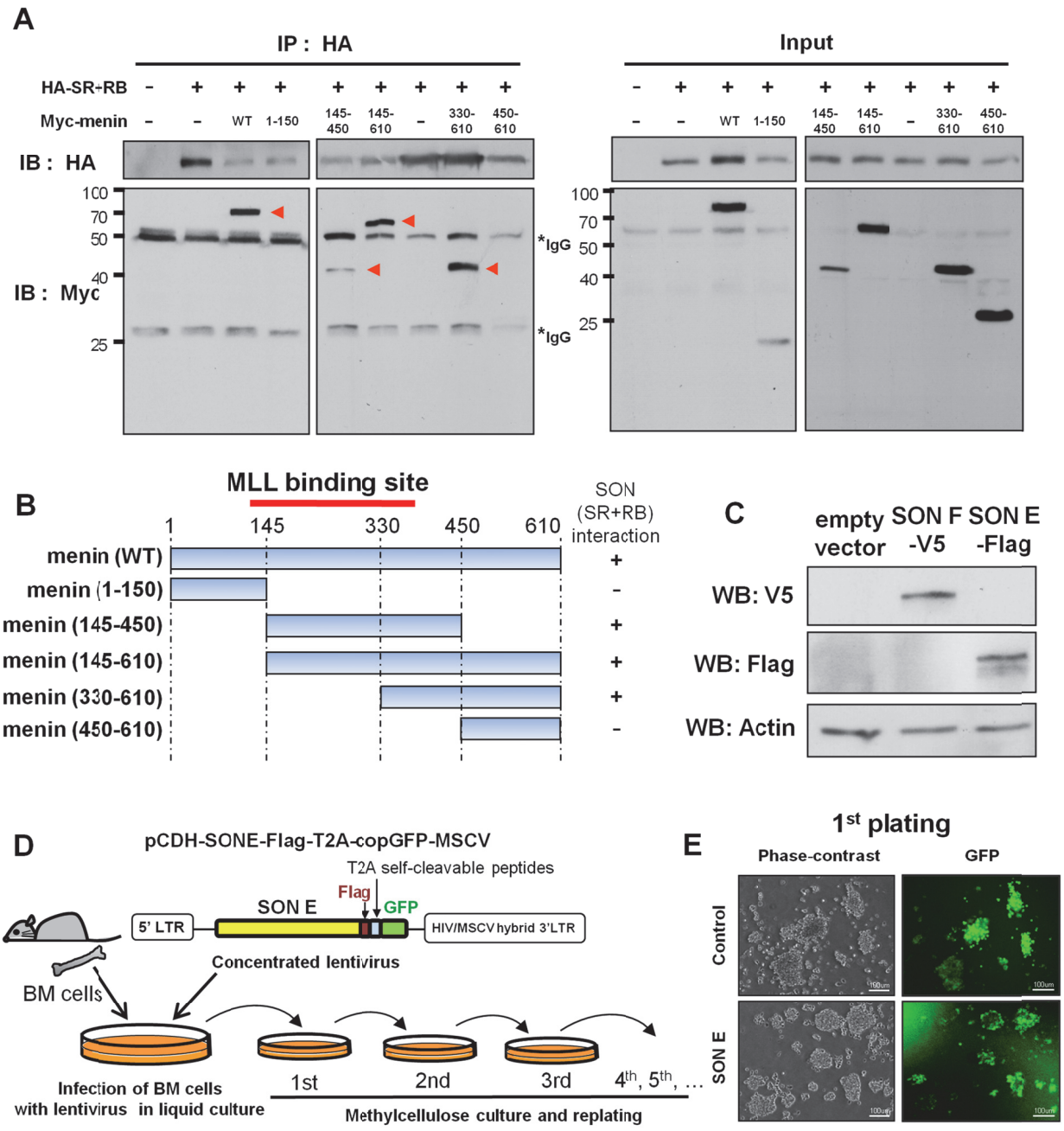


Figure S7. The C-terminal Region of SON (SR+RB) Directly Interacts with the MLL-Binding Site of Menin and Overexpression of a Short Isoform of SON (SON E) Enhances Replating Potential of Hematopoietic Progenitors (Related to Figure 7).

(A) Co-immunoprecipitation of HA-tagged SON SR+RB with Myc-tagged full-length/wild type (WT) or several partial fragments of menin after transient transfections of the constructs into 293T cells as

indicated. Immunoprecipitations were performed using an HA antibody and analyzed by Western blotting using a Myc antibody. Wild type and several partial fragments of menin were detected in the precipitates (upper panel, lane 3, 5, 6, and 8; red arrowheads). Asterisks indicate IgG.

(B) The schematic diagram shows the structure of the menin and its deletion constructs used in Figure S7A. The region previously shown to interact with MLL is indicated with a red bar.

(C) Western blot verified expression of V5-tagged SON F (SON F-V5) and Flag-tagged SON E (SON E-Flag) in K562 cells used for the experiment presented in Figure 7E.

(D) Schematic structures of the lentiviral construct for SON E overexpression and the illustration of the experimental design. Bone marrow (BM) cells were isolated from mice, infected with empty lentivirus (Control) or Flag-tagged SON E–expressing lentivirus (SON E) and subjected to the colony forming and serial replating assays. For each round of plating, a total of 2×10^4 BM cells were cultured in MethoCult (M3434).

(E) Representative photomicrographs of colonies from control and SON E lentivirus-infected BM cells from the first plating, demonstrating the high efficiency of infection (GFP-positive cells). White scale bars on pictures represent 100 μm.

Supplemental Tables

Table S1. Nomenclature of the SON isoforms. Related to Figures 5 and S5.

Species	Name[1]	NCBI database name	NCBI accession number (Gene)	NCBI accession number (Protein)	Ensembl database name	Ensembl transcript ID	CDS length (bp)	Amino acid length
Human	SON F	SON transcript variant F	NM_138927	NP_620305	SON-001	ENST00000356577	7,280	2,426
	SON B	SON transcript variant B	NM_032195	NP_115571	SON-003	ENST00000300278	6,911	2,303
	SON E	SON transcript variant E	NM_001291411	NP_001278340	SON-006	ENST00000381679	6,326	2,108
Mouse	Son f	SON transcript variant 1	NM_178880	NP_849211	Son-013	ENSMUST00000114037	7,334	2,444
	Son b	SON transcript variant x10	XM_006522973	XP_006523036	Son-003	ENSMUST00000119368	6,965	2,321
	Son e	SON transcript variant 2	NM_019973	NP_064357	Son-002	ENSMUST00000114036	6,380	2,126

[1]Name used for the SON isoforms in this manuscript

CDS: Coding sequence

Table S2. Patient sample information. Related to Figures 5 and 6.

Sample No.	Registry ID	Collection ID	Sample Type	Diagnosis	Subtype	WBC (%)	Blasts (%)	Cytogenetics Abnormality	FLT3	NPM	ckit
P1	**R1111**	1762	BM-MNC	**AML**	M2 without t(8;21) [with maturation]	78.2	61	46,XY,del(11)(p11.2p15)[1]/46,XY[13]	WT		
		1766	PB-MNC								
P2	**R1117**	1777	BM-MNC	**AML**	M2 without t(8;21) [with maturation]	1.1	90	48,XX,+13,+2149,+15[11]	WT		
P3	**R2029**	4006	BM-MNC	**AML**	M2 without t(8;21) [with maturation]	3.5	51	46,XX,add(6)(p23)[14] /46,XX,add(2)(q37)[5]/46,XX[1]	mut	mut	
P4	**R1006**	1856	BM-MNC	**AML**	M2 without t(8;21) [with maturation]	5.4	71	43-44,XX,t(1;4)(p32;p12),del(5)(q13q33), del(7)(q11.2q32), -8,del(12)(p11.2),+del(12)(q15q24.1), add(13)(p11.2),-15, -17,del(20)(q11.2),+mar[cp14]/46,XX[6]	WT		
P5	**R1296**	2281	PB-MNC	**AML**	AML with t(8;21)(q22;q22)	5.9	75	46,XX,t(8;21)(q22;q22),del(9)(q22)[20]	WT		pos
		2282	PB-MNC								
P6	**R1379**	2523	BM-MNC	**AML**	AML with t(8;21)(q22;q22)	4.8	35	46,XY,t(8;21)(q22;q22)[20]	WT		
P7	**R1650**	3164	BM-MNC	**AML**	AML with t(8;21)(q22;q22)	6.9	90	47,XX,t(8;21)(q22;q22),del(9)(q13q22), +der(21)t(8;21)[20]	WT	WT	
P8	**R1575**	2985	BM-MNC	**AML**	AML with t(8;21)(q22;q22)	4.1	84		WT	WT	
P9	**R1890**	3672	BM-MNC	**AML**	AML with t(8;21)(q22;q22)	1.8	80	45,XY,t(8;21)(q22;q22)[5]/45,idem,del(9) (q22q34),?del(17)(p12),-20,+mar[cp15]			
		4001	PB-MNC								
P10	**R1823**	3521	BM-MNC	**AML**	AML with t(8;21)(q22;q22)	59.5	68	46,XY,t(8;21)(q22;q22)[20]	WT	WT	
		3520	PB-MNC								
P11	**R0538**	553	PB-MNC	**AML**	M2 without t(8;21) [with maturation]	71.5	92	46,XX[25], FISH RARA NEG IN 200 CELLS			
P12	**1453**		PB-MNC	**AML**			36	t(9;22)	WT	WT	
P13	**1424**		PB-MNC	**AML**				del17q			
P14	**1431**		PB-MNC	**AML**				t(16;16)			
P15	**1488**		PB-MNC	**MDS**				del5q, der(7;17)t(7;17) -7q,-17q			
P16	**1544**		PB-MNC	**MDS**							

BM-MNC, bone marrow mononuclear cell; PB-MNC, peripheral blood mononuclear cell; WT, wild-type; mut, mutation.

Table S3. Primer sequences. Related to Figures 2, 3, 5, 6, 7, S2, S3, S4, S5 and S6

Name	Oligo sequence
- SON qPCR (Related to Figures 5 and S6) -	
F-hSON Exon1	CAGATTTTTAGGTCTTTCGTGGT
R-hSON Exon3	TTTTTCTGGAGCCCTCTTTC
F-mSON Exon 3	GGCGGAAAAGATCCAGGT
R-mSON Exon 4	GCTCCTCCAGACTTTTTAGCAA
F-hSON Exon3	GCTTCAGTATCTCACCAGTC
R-hmSON Exon5a	GTTTGGCTTCTAAAACTCAG
R-hSON Exon5	GCGATCTGTTTACATTTCTCAG
F-h/mSON Exon 5	TCGGATGAAGAGGAAGAAGAAC
R-hSON Exon7	AGATGTCCACAGGCTCTGA
R-h/mSON Exon7a	GTTTAACCCGGCCCTCTG
F-mSON Exon3	TATATCACCGGTCAGATTAAGG
R-mSON Exon5	CTGTTTGCATTTCTCAGTTAGT
R-mSON Exon7	GAAATGTCCACAGGCTCTGA
F-hSON Exon9	GCAATAAGGAACCCATCCTAGT
R-hSON Exon12	AGGGCTCCATTTCTCAATACC
F-mSON Exon9	GGAGAGAAGGAGAAGGATTAGGA
R-mSON Exon12	CTGGTAAGGGCTTCCATTTCT
F-hSON Exon12-exo	GGCAACCACCTGAATTTCTATTG
F-hSON Exon4	GGTGTCCCTTTACCACCAAA
R-V5-exo	AGAGGGTTAGGGATAGGCTTAC
F-hGAPDH	GGCGCTGAGTACGTCGTGGAGTCCA
R-h/m GAPDH	AAAGTTGTCATGGATGACCTTGG
F-mGAPDH	GGTGCTGAGTATGTCGTGGAGTCTA
- Target gene qPCR (Related to Figures 2, 6 and S4) -	
F-hCDKN1A qPCR	GGACAGCAGAGGAAGACCATGT
R-hCDKN1A qPCR	GGCGTTTGGAGTGGTAGAAATC
F-hGADD45A qPCR	TGCTGGTGACGAATCCACATT
R-hGADD45A qPCR	AGCGACTTTCCCGGCAAAA
F-hb-ACTIN qPCR	AGCCTCGCCTTTGCCGA
R-hb-ACTIN qPCR	CTGGTGCCTGGGGCG
F-hFOXO3A qPCR	GAGTGAGAGGCAATAGCATACA
R-hFOXO3A qPCR	AGCACCTATACAGCACCATAAC
F-hNOTCH2NL qPCR (3-4)	CAAGTCGGGTTTACAGGTAAGG
R-hNOTCH2NL qPCR (3-4)	GAGGCATTTGCAGGAGAACT
F-hATF3 qPCR (3-4)	CTGCAGAAAGAGTCGGAGAAG
R-hATF3 qPCR (3-4)	CCGATGAAGGTTGAGCATGTA
F-hGFI1 qPCR (4-5)	GGCTCCTACAAGTGCATCAA
R-hGFI1 qPCR (4-5)	ACATCTCGCAGGCAAAGG
F-hEGR1 qPCR (1-2)	CAGCACCTTCAACCCTCAG
R-hEGR1 qPCR (1-2)	AGCACCTTCTCGTTGTTCAG
F-hSRC qPCR (5-6)	CTTTGTGGCCCTCTATGACTATG
R-hSRC qPCR (5-6)	CAGTCTCCCTCTGTGTTGTTG
F-GFI1B qPCR	CCAGAAGTCCGACATGAAGAAG
R-GFI1B qPCR	CTGTGGGTGATGAGGTTGG
F-hHOXA9 qPCR (1-2)	CCCATCGATCCCAATAACCC
R-hHOXA9 qPCR (1-2)	GTTCCAGGGTCTGGTGTTT
F-hMEIS1 qPCR (6-7)	TCGATTTGGTGATAGACGATAGAG
R-hMEIS1 qPCR (6-7)	GTTCCTCCTGAACGAGTAGATG

Table S3. Primer sequences *(continued)*

Name	Oligo sequence
F-TUBG1 qPCR	AGCTGGTGTCTACCATCATGT
R-TUBG1 qPCR	CGTAGTGAGAGGGGTGTAGC
F-HDAC6 qPCR	AAGAAGACCTAATCGTGGGACT
R-HDAC6 qPCR	GCTGTGAACCAACATCAGCTC
F-AKT1 qPCR	AGCGACGTGGCTATTGTGAAG
R-AKT1 qPCR	GCCATCATTCTTGAGGAGGAAGT

- ChIP qPCR (Related to Figures 3, 6, 7, S2 and S3) -

F-hCDKN1A Target	GCTCAGCGTGACGTGTT
R-hCDKN1A Target	GGGAAAGCCTGTGTGCTATT
F-hGADD45a Target	GCAGCGGCCCAATTAGT
R-hGADD45a Target	GCCTGCTTTCTGCACTCA
F-hNOTCH2NL Target	AAGGTGGAGGCAGGAGAA
R-hNOTCH2NL Target	CCGCGCCTCAGAAAGAATAA
F-hATF3 Target	AGGTGTTTCTGCCCTTCAC
R-hATF3 Target	CTGCCCGTGGATGACTATTT
F-hGFI1 Target	CCGAGTTTCGTGGCTGTTTA
R-hGFI1 Target	TCCTAGGGCTAGAGCTTTGT
F-hEGR1 Target	AACCCTTTCGCCACCATC
R-hEGR1 Target	GCATCCCTCTAACACATGACTC
F-hSRC Target	TGGGTGACCGTAAGCAATTC
R-hSRC Target	AGGTGTTCAGTAGATGTTGGC
F-hGFI1B promoter	CCAAGTGTCGAGGGTTAGAAAT
R-hGFI1B promoter	GTGAACAAGCAGCCAATGAAG
F-hATF4 promoter	CGGAAGGATGCGTCTGTG
R-hATF4 promoter	GGGCGGGCAAAGTAGAAA

- 3'RACE PCR (Related to Figure S5) -

Adapter oligo-dT	GCTCGCGAGCGCGTTTAAACGCGCACGCGTTTTTTTTTTTTTTTTTTVN
SON E1 (F1)	CAGAAGCTCGCCTACAGTTT
SON E2 (F2)	CAGCCTGCCACTGTCAATA
Adapter Reverse part1 (R1)	GCTCGCGAGCGCGTTTAAAC
Adapter Reverse part2 (R2)	GCGTTTAAACGCGCACGCGT

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell Lines and Cell Culture

K562, MV4;11 and ML-2 cell lines were cultured in RPMI 1640 medium with 10% fetal bovine serum. Primary human $CD34^+$ bone marrow cells were purchased from Lonza and were cultured in STEMSPAN SFEM (StemCell Technologies) supplemented with 1% penicillin and streptomycin, 100 ng/ml of recombinant human SCF, 100 ng/ml of recombinant human FLT3L and 100 ng/ml of recombinant human TPO for 1 to 2 days before transfection. Primary mouse bone marrow cells were cultured in STEMSPAN supplemented with 15% FBS, 1% penicillin and streptomycin, 1% Glutamate, 10 ng/ml recombinant mouse IL-3 50ng/ml recombinant mouse SCF and 50 ng/ml recombinant mouse FLT3L for 1 day after lentivirus infection. All the cytokines were purchased from PeproTech.

Plasmid Construction

Plasmids containing siRNA-resistant full-length SON cDNA (siRR-SON F) and SON SR+RB (serine/arginine-rich region and RNA-binding domain) were created as previously described (Ahn et al., 2011). The plasmid containing SON E, short splice variant of SON, was constructed as follows. The region from internal HpaI site to the 3' end of SON E cDNA was amplified using the forward primer (SON HpaI-F: 5'-GAATCTTCAATTACGTTAACA-3') and the reverse primer (Flag-SON E NotI-R: 5'-TTTGCGGCCGCTATTTGTCATCGTCATCCTTGTAGTCTGGCCGGCC AAACTCAGTTTAGTTCTTCTATAGTAGCTCCTCCTG-3'). The sequence for the Flag tag was added as indicated in the reverse primer. Then, the C-terminus of SON F in pcDNA3 HA-siRR SON F- Flag was replaced by the amplified SON E C-terminal fragment at the HpaI and NotI sites to create pcDNA3 HA-siRR SON E-Flag construct. To construct V5-tagged plasmids, the C-terminus region was amplified from cDNA using specific primers (SON BmgBI-F: 5'-GCAAGTGATGTTGGACGTGACAGATC-3' and V5-SON F NotI-R: 5'-TTTGCGGCCGCTA

CGTAGAATCGAGACCGAGGAGAGGGTTAGGGATAGGCTTACCTGGCCGGCCATACCTATT CAAGAAAAACATACAATT-3') and subcloned into plasmids. Full-length cDNA clones of human menin (pcDNA3 Flag-menin, #32079) and MLL-ENL (pMSCV Flag-MLL-pl-ENL, #20873) were purchased from Addgene. Myc-tagged menin truncation mutant constructs were kindly provided by Dr. Eun-Jung Cho (Sungkyunkwan University, Korea).

Reverse Transcription and Quantitative PCR (RT-qPCR)

Human leukemic cell lines from one 6-well, human $CD34^+$ BM cells from one 24-well, and patient samples were lysed and total RNA was isolated using the RNeasy Mini Kit (Qiagen). Total RNA from each sample was treated with RNase-free DNase I and used as a template to produce cDNA with oligo-dT using the SuperScript III First Strand Synthesis Kit (Life Technologies). Quantitative real time PCR (RT-qPCR) was performed in triplicate reactions on the iQ5 Real Time PCR Detection System (Bio-Rad) using the Fast Start Universal SYBR Green Master (Roche) and standard deviations were calculated. All PCR reactions were finished under the following program: initial denaturation step was 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 seconds, and annealing at 60°C for 60 seconds. PCR primers were listed in Table S3 in the supplemental material. Gene expression levels were normalized to *GAPDH*.

Antibodies

The antibodies used for ChIP, IP, and Western blotting were the following; SON antibody recognizing the N-terminus of SON (SON-N Ab) was generated against amino acids 74-88 of human SON, and was provided by Dr. Dong-Er Zhang (University of California San Diego). SON-N antibody was used for both ChIP and Western blot. Anti-SON (SON-C Ab, ab121759, for ChIP), anti-H3K27ac (ab4729), anti-H3K4me3 (ab8580), anti-SUZ12 (ab12073), anti-WDR5

(ab56919), anti-H3 (ab1791), and anti-H3K4me1 (ab8895), anti-H3K79me2 (ab3594) were purchased from Abcam. Anti-TRX2/MLL2 (A300-113A), anti-ASH2L (A300-489A), anti-menin (A300-105A), anti-MLL-C (MLL1, C-terminus, A300-374A), anti-LEDGF (A300-848A), anti-SET1A (A300-289A), anti-SET1B (A302-281A), anti-CFP1 (A303-161A), and anti-MLL-N for ChIP (MLL1, N-terminus, A300-086A) were purchased from Bethyl Laboratories. ASC2 antibody was a gift from Dr. Jae W Lee (Oregon Health & Science University). Anti-H3K27me3 (07-4490, Millipore), anti-MLL-N for IP and WB (MLL1 N-terminus, 39829, Active motif), anti-V5 (R960-25, Invitrogen), anti-HA (#2367, Cell Signaling Technology), anti-Flag M2 (F3165, Sigma), and anti-Actin (A5441, Sigma) were purchased from the indicated companies.

Chromatin Immunoprecpitation (ChIP)

K562 cells were incubated with 1% formaldehyde in 5 ml growth medium for 10 min at room temperature and cross-linking reaction was terminated by incubation with 125 mM glycine for 10 min. Subsequently cells were incubated for 15 min at 4°C with lysis buffer (5 mM PIPES pH 8.0 / 85 mM KCl / 0.5% NP-40 / 1x Complete Protease Inhibitor Cocktail (Roche)), collected by centrifugation for 5 min at 3,000g and resuspended in RIPA buffer (150 mM NaCl / 50 mM Tris-HCl, pH 8.0/ 1 mM EDTA / 1% sodium deoxycholate / 0.1% SDS / 1% Triton X-100 / 1x Complete Protease Inhibitor Cocktail). To shear chromatin to lengths ranging between 200–500 base pairs, crude nuclei were sonicated with the Ultrasonic disintegrator Sonicator S-4000 (Misonix). Sonicated DNA from each sample were incubated at 4°C overnight with 1–5 μg of specific antibodies or normal immunoglobulin G (IgG) as controls and magnetic bead protein A or G (Dynabeads Protein A or Protein G, Life Technologies). The magnetic beads were washed 5 times for 10min at 4°C on a rotating platform with 1ml wash buffer (100 mM Tris pH 7.5 / 500 mM LiCl / 1% NP-40 / 1% Sodium deoxycholate) and washed once with TE (10 mM Tris pH 7.5 / 0.1 mM EDTA). After washing, the washed beads were eluted by heating for 2hr at 65°C in

elution buffer (1% SDS / 0.1 M $NaHCO_3$) with proteinase K. ChIP DNA were purified and concentrated using the QIAquick PCR Purification Kit (Qiagen).

siRNA and Plasmid Transfection

Total SON siRNAs directed against human SON (siSON #1: GCAUUUGGCCCAUCUGAGAtt, Ahn et al, 2011; siSON #2: UGAGCGCUCUAUGAUGUCAtt, Lu et al, 2013), human SON E-specific siRNA (siSON E: CACCGGAGCUUGGAAAUUAtt), and negative control siRNA (UAACGACGCGACGACGUAAtt) were custom synthesis products by Life Technologies (Silencer Select siRNA). For K562 and ML2 cells, 0.5×10^6 cells were nucleofected with 100 - 200 pmol of siRNA or 5 µg of plasmid using Cell Line Nucleofector Kit V (Lonza) according to the manufacturer's instructions. For MV4;11 cells, 0.5×10^6 cells were nucleofected with 150 - 200 pmol of siRNA using Cell Line Nucleofector Kit L (Lonza) according to the manufacturer's instructions. For human $CD34^+$ bone marrow cells, 0.4×10^6 cells were nucleofected with 80 - 150 pmol of siRNA or 4 µg of plasmid using Human $CD34^+$ Cell Nucleofector Kit (Lonza) according to the manufacturer's instructions. Human embryonic kidney (HEK) 293 cells were transiently transfected with 5 µg of plasmids using PEI.

ChIP-Seq Analysis

ChIP-sequencing data files were aligned to the hg19 human reference genome using Bowtie (version 0.12.9) and standard parameters. Peak calling was performed with MACS v1.4.2 software using default parameters. To identify high confidence SON binding peaks, the MACS peak calling output from two different experimental samples were used. BigWig files were generated by first extending the 5' ends of uniquely aligned, non-duplicate ChIP-seq reads by the average DNA fragment length (150bp for histone marks, 250bp for transcription factors) in the 3' direction using BEDtools. Identified peaks were then annotated to the nearest

transcription start site. Peaks that were identified in both experimental samples (overlapping peaks) with a false discovery rate (FDR) of 0.001 and a tag density (TD) of 12 in at least one of the experimental samples were identified as significant high confidence peaks. When determining the peak overlaps from each analysis, an in-house script was used to determine the percentage of the region of the smaller peak that overlapped with the larger peak. Overlap cut-off threshold was set to 50%, such that 50% of the smaller peak in one replicate was required to overlap with the peak in the other replicate to be considered an overlapping peak.

The genomic distributions of binding sites (Figures 1B and 1C) were analyzed using the cis-regulatory element annotation system (CEAS v1.0.2). The genes closest to the binding site on both strands were classified into functional categories such as promoter (from -1kb to +100bp), 5' UTR, first exon, first intron, exon, intron, 3' UTR, and intergenic region. The genes were also divided into defined groups according to the enrichment of the SON across TSS regions (5 kb window surrounding the TSS). Tag density heatmap (Figure 1D) was generated using the R package pheatmap v0.7.7 (peak extensions 5 kb upstream and 5 kb downstream of the peak summit and bin size 10 bp). Motif analyses (Figures 1E and S1C) were performed using HOMER, the methods of which are freely available at http://biowhat.ucsd.edu/homer/. Gene-associated region annotations (Figure 1F) were obtained with Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7. ChIP density plots (Figures 2B and S2D) were made using HOMER by calculating average tag densities across ±5kb regions surrounding indicated peaks. ChIP-seq read density files (Figures 2C, 2D and S2B) were generated using IGV tools and were viewed in Integrative Genomics Viewer (IGV) (http://www.broadinstitute.org/igv/). Previously published ChIP-seq data for H3K27ac, H3K4me3, and H2A.Z from K562 (downloaded from ENCODE/Broad Institute) and MNase-seq data for nucleosome position from K562 (downloaded from ENCODE/Stanford/BYU) were analyzed. Enriched ChIP-seq regions at promoters (5 kb window surrounding the TSS) for SON and

histone marks were combined together to generate a unified track consisting of all merged enriched regions.

Preparation of Nuclear Extraction and Immunoprecipitation (IP)

Nuclear extracts were prepared from control, siRNA- or plasmid-transfected K562 or MV4;11 cells using the Dignam protocol (Dignam et al., 1983). In brief, harvested cells were resuspended in three packed cell volumes of buffer A (10 mM HEPES pH 7.9 / 1.5 mM $MgCl_2$ / 10 mM KCl / 1 mM DTT / 0.1% NP-40 / Protease Inhibitor Cocktail) and homogenized using needle and syringe with 25 to 30 gentle strokes. Lysed cells were centrifuged at 13,000×*g* for 10 min and the nuclei pellet was resuspended in two packed volumes of buffer C (20 mm HEPES pH 7.9 / 420 mM KCl / 1.5 mM $MgCl_2$ / 1 mM DTT / 25% glycerol / Protease Inhibitor Cocktail). The nuclei suspension was gently stirred for 30 min at 4°C and centrifuged 15 min at 13,000 *g* to remove debris. Sufficient volume of buffer D (20 mM HEPES pH 7.9 / 0.5 mM DTT / 25% glycerol / Protease Inhibitor Cocktail) was added to the nuclei extract. For IP with K562 cells, nuclear extracts were pre-cleared with protein A-sepharose beads (Life Technologies) for 1 hour and incubated either with rabbit IgG or SON antibody at 4°C overnight on a rotator. Beads were washed four times with wash buffer (20 mM HEPES pH 7.9 / 150 mM NaCl / 0.05% (v/v) NP-40) and eluted by boiling in SDS buffer and analyzed by SDS–PAGE.

For Co-IP with exogenously expressed proteins, HEK293 cells were co-transfected with V5 and HA-tagged plasmids encoding wild type SON F or its alternative spliced variant SON E and either Flag-tagged menin plasmid or empty vector. Cells were lysed with lysis buffer (50 mM Tris-HCl, pH 8.0 / 150 mM NaCl / 0.5% NP-40 / 10% glycerol / Protease Inhibitor Cocktail / 50 U/mL of Benzonase nuclease (Sigma)) for 1 hour. Whole cell extracts were pre-cleared and incubated with the V5 or HA antibody and protein G-sepharose beads (Life technologies)

overnight. Co-IP complexes precipitated by bead were eluted by boiling in SDS buffer and subjected to Western blot analyses.

Size Fractionation and Analysis

Gel filtration chromatography was carried out at 4 °C using AKTA system (GE Healthcare). Nuclear extract from control or SON siRNA transfected K562 cells were prepared fresh, passed through a 0.22-μm pore size MILLEX-GS filter (Millipore) and size-fractionated by fast-protein liquid chromatography (FPLC). 5mg of nuclear protein was applied to a Superose 6 10/300 GL column (GE Healthcare) equilibrated in FPLC buffer (20 mM Tris-HCl / 0.2 mM EDTA / 5 mM $MgCl_2$ / 0.1 M KCl / 10% glycerol / 0.5 mM DTT / 1 mM benzamidine / 0.2 mM PMSF / pH 7.9) at 0.3 mL/min. 0.5 mL of elutes were collected and prepared for Western blot.

Analysis of Minigene Splicing

The minigene construct containing *TUBG1 e*xon 7- intron 7- exon 8 were used to examine SON-mediated splicing as described previously (Ahn et al., 2011). Briefly, HeLa cells in 6 well plates were transfected with 100 pmol of control siRNA or SON siRNA. The next day, minigene alone or minigene plus various amounts of SON F or SON E constructs (siRNA-resistant form) were transfected as indicated in each experiment. RNAs were isolated 30h after minigene transfection, and RT-PCR was performed using forward primer targeting *TUBG1* exon 7 and the reverse primer targeting the bovine growth hormone (BGH) terminator sequence present in the pcDNA vector.

3’ RACE PCR

The transcription terminating site of the primary transcript of SON E was determined by 3' RACE PCR. For the 3’ RACE PCR, K562 total RNA was reverse transcribed by SuperScript III First Strand Synthesis Kit (Life Technologies) with the adapter oligo-dT primer. The first PCR was

performed using the first adapter primer (Adapter Reverse Part1) and SON E primer (SON E1), which specifically bind with SON exon 5a region. A second PCR was achieved using the second adapter primer (Adapter Reverse Part2) and SON E primer (SON E2). PCR products were visualized on a 1% agarose gel and cDNA fragments were cloned and sequenced. The primer sets for RACE PCR are listed in Table S3.

Primary Patient Samples

The bone marrow mononuclear cells and/or peripheral blood mononuclear cells from AML patients (FAB subtype M2, P1 - P11) as well as bone marrow mononuclear cells from healthy donors (N1 – N4) were obtained from the Stem Cell and Xenotransplantation Core Facility of the University of Pennsylvania. All samples were obtained after written informed consent according to the University of Pennsylvania IRB approved protocols. Peripheral blood samples from additional 5 patients diagnosed with AML or MDS (P12 – P16) and healthy donors (N5 – N11) were obtained from the BioBank of Mitchell Cancer Institute, University of South Alabama. These patient samples were collected according to University of South Alabama IRB approved protocols. Cells were purified by Ficoll-Paque (GE Healthcare) density-gradient centrifugation and frozen as viable cells. Details of the patient samples were listed in Supplemental Information Table S2.

Preparation of Leukemic Blasts and Lin^- / $c\text{-}Kit^+$ Bone Marrow Cells from Mice

Mouse leukemic blasts (a kind gift from Dr. Dong-Er Zhang) were obtained from C57BL/6 mice with AML1-ETO9a-induced leukemia generated by retroviral transduction-transplantation (Yan et al., 2006). MLL-AF9 leukemic blasts were obtained from the spleen and bone marrow of MLL-AF9a knock-in mice (Corral et al., 1996). Normal mouse Lin^-/$c\text{-}Kit^+$ bone marrow cells were prepared by Lineage Cell Depletion Kit and CD117 MicroBeads (Miltenyl Biotec).

Lentiviral Vector Construction and Lentivirus Production

Lentiviral vector for SON E overexpression was prepared by subcloning SON E cDNAs into the pCDH-MCS-T2A-copGFP-MSCV vector (System Bioscience). Lentivirus was produced by cotransfection of HEK 293T cells with expression plasmid, pMDLg/pRRE, pRSV-REV, and pVSVG. Viral supernatants were collected after 48 h and clarified by filtration before use. Ultracentrifugation was performed for lentivirus concentration with the Optima L-100 XP centrifuge (Beckman) using an SW55TI rotor (Beckman) at 19,400 rpm for 2 hr at 20°C. Supernatant was completely removed and virus pellets resuspended in PBS.

Colony Forming Unit Assay and Serial Replating

Mouse total bone marrow cells were transduced with recombinant lentivirus in the 4 ug/ml polybrene. Infected cells were incubated overnight in the above mouse BM culture media. After 3 days of culture, 2 x 10^4 cells were plated in methylcellulose medium (Methocult GF M3434, StemCell Technologies). Colony number counting and re-plating were repeated every 7-10 days. The colony-forming units (CFUs) were quantified as the average and standard deviation of at least triplicate determinations.

Oncomine Database Analysis

Analysis of SON isoform expression in leukemia and normal cells were conducted using the Oncomine database (www.oncomine.org). The Oncomine database was used as previously described (Rhodes et al., 2004).

Statistical Analysis

In all graphs data are expressed as mean ± SD of three independent experiments, except when otherwise indicated. Differences were analyzed by Student's t test. *P*-values < 0.05 were considered significant.

SUPPLEMENTAL REFERENCES

Corral, J., Lavenir, I., Impey, H., Warren, A. J., Forster, A., Larson, T. A., Bell, S., McKenzie, A. N., King, G., and Rabbitts, T. H. (1996). An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. Cell *85*, 853-861.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Maia, S., Haining, W. N., Ansen, S., Xia, Z., Armstrong, S. A., Seth, N. P., Ghia, P., den Boer, M. L., Pieters, R., Sallan, S. E.*, et al.* (2005). Gene expression profiling identifies BAX-delta as a novel tumor antigen in acute lymphoblastic leukemia. Cancer Res *65*, 10050-10058.

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., and Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia *6*, 1-6.

Stegmaier, K., Ross, K. N., Colavito, S. A., O'Malley, S., Stockwell, B. R., and Golub, T. R. (2004). Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. Nat Genet *36*, 257-263.

Yan, M., Kanbe, E., Peterson, L. F., Boyapati, A., Miao, Y., Wang, Y., Chen, I. M., Chen, Z., Rowley, J. D., Willman, C. L., and Zhang, D. E. (2006). A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. Nat Med *12*, 945-949.