Title

A rapidly evolved domain, the SCML2 DNA-binding repeats, contributes to chromatin binding of mouse SCML2†

Journal

Biology of Reproduction, 100(2)

ISSN

0006-3363

Authors

Maezawa, So

Alavattam, Kris G

Tatara, Mayu

et al.

Publication Date

2019-02-01

DOI

10.1093/biolre/ioy181

Peer reviewed

Biology of Reproduction, 2019, 100(2), 409–419
doi:10.1093/biolre/ioy181
Research Article
Advance Access Publication Date: 18 August 2018

OXFORD

Research Article

A rapidly evolved domain, the SCML2 DNA-binding repeats, contributes to chromatin binding of mouse SCML2†

So Maezawa[1,2,3,*], Kris G. Alavattam[1,2], Mayu Tatara[3], Rika Nagai[3], Artem Barski[2,4] and Satoshi H. Namekawa[1,2,*]

[1]Division of Reproductive Sciences, Division of Developmental Biology, Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; [2]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA; [3]Department of Animal Science and Biotechnology, School of Veterinary Medicine, Azabu University, Sagamihara, Kanagawa, Japan and [4]Division of Allergy and Immunology, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

*__Correspondence:__ Department of Animal Science and Biotechnology, School of Veterinary Medicine, Azabu University, Sagamihara, Kanagawa 252-5201, Japan. Tel: +81-42-850-2514; Fax: +81-42-754-7661; E-mail: s-maezawa@azabu-u.ac.jp; Division of Reproductive Sciences, Division of Developmental Biology, Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229. Tel: +1-513-803-1377; Fax: +1-513-803-1660;
E-mail: satoshi.namekawa@cchmc.org

†__Grant support:__ This work was supported by the Research Grant (FY13-510) from the March of Dimes Foundation to SHN, NIH Grants GM119134 and HL098691 to AB, and GM098605 and GM122776 to SHN. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Received 23 March 2018; Revised 20 July 2018; Accepted 16 August 2018

Abstract

Genes involved in sexual reproduction diverge rapidly as a result of reproductive fitness. Here, we identify a novel protein domain in the germline-specific Polycomb protein SCML2 that is required for the establishment of unique gene expression programs after the mitosis-to-meiosis transition in spermatogenesis. We term this novel domain, which is comprised of rapidly evolved, DNA-binding repeat units of 28 amino acids, the SCML2 DNA-binding (SDB) repeats. These repeats are acquired in a specific subgroup of the rodent lineage, having been subjected to positive selection in the course of evolution. Mouse SCML2 has two DNA-binding domains: one is the SDB repeats and the other is an RNA-binding region, which is conserved in human SCML2. For the recruitment of SCML2 to target loci, the SDB repeats cooperate with the other functional domains of SCML2 to bind chromatin. The cooperative action of these domains enables SCML2 to sense DNA hypomethylation in an in vivo chromatin environment, thereby enabling SCML2 to bind to hypomethylated chromatin. We propose that the rapid evolution of SCML2 is due to reproductive adaptation, which has promoted species-specific gene expression programs in spermatogenesis.

Summary Sentence

Here, we identify a novel protein domain in the germline-specific Polycomb protein SCML2; because the domain is comprised of rapidly evolved DNA-binding repeats, it has been termed the SCML2 DNA-binding (SDB) repeats.

Key words: SCML2, germline, spermatogenesis, meiosis, Polycomb.

Introduction

Genes involved in sexual reproduction diverge rapidly as a result of adaptive evolution [1], and the key genetic determinants of sexual reproduction code for rapidly evolved DNA-binding domains. Such factors include Y-linked *Sry*, which encodes the master regulator of male sex determination [2]; *Prdm9*, which encodes an essential protein that positions the sites of meiotic recombination [3, 4]; and *Rhox5* (also known as *Pem*), an X-linked homeobox gene that promotes the differentiation of male germ cells [5]. In this study, we identify a novel DNA-binding domain comprised of repeat units in the germline-specific Polycomb protein SCML2, which is required for the execution of a wholly unique gene expression program after the mitosis-to-meiosis transition of spermatogenesis [6]. We term this novel domain the SCML2 DNA-binding (SDB) repeats.

During the later stages of germ cell differentiation in the testes, germ cells go through epigenetic programming that is specific to males [7–9]. This includes the global suppression of a somatic gene expression program in late spermatogenesis after the mitosis-to-meiosis transition; at the same time, male germ cells undergo activation of spermatogenesis-specific genes [6, 10]. We recently identified mouse SCML2 as a suppressor of somatic/progenitor genes in late spermatogenesis, thereby ensuring the activation of spermatogenesis-specific genes [6]. Together with this global gene expression change, open chromatin is reorganized and SCML2 plays a critical role in this process [11]. SCML2 also accumulates on the sex chromosomes during meiosis and regulates gene expression of sex chromosome-linked genes [6, 12].

Polycomb proteins regulate heritable gene silencing and define cell type-specific gene expression [13–15]. Two major Polycomb complexes, Polycomb repressive complexes 1 and 2 (PRC1 and PRC2), are required for gene expression programs in spermatogenesis [6, 16, 17]. While SCML2 was initially identified as a component of PRC1 [6], our recent study demonstrated that SCML2 establishes PRC2-mediated H3K27me3 on the hypomethylated promoters of genes suppressed in spermatogenesis [18], uncovering the regulatory mechanisms of bivalent genomic domains (H3K27me3 and H3K4me2/3) in the male germline, which are proposed to have roles in the preparation of totipotency [10, 19, 20]. SCML2 also regulates PRC2-mediated H3K27me3 on pericentric heterochromatin, thereby promoting heterochromatin organization in late spermatogenesis [21].

In this study, by identifying the rapidly evolved SDB repeats as a DNA-binding domain in mouse SCML2, we elucidate molecular mechanisms by which SCML2 binds chromatin. Through utilization of the fluorescence recovery after photobleaching (FRAP) assay, we determine the mechanism by which SCML2 is enriched on hypomethylated chromatin. Since this mechanism underlies the establishment of germline-specific transcriptomes, we propose that the rapid evolution of SCML2 is due to a reproductive adaptation to gain species-specific gene expression programs in spermatogenesis.

Materials and methods

Evolutionary analysis

The phylogenetic tree of mammalian SCML2 was constructed through the Unweighted Pair Group Method with Arithmetic Mean algorithm [45] using CLUSTALW (http://www.genome.jp/tools-bin/clustalw). *Ka*/*Ks* values were calculated using the *Ka*/*Ks* Calculation tool (http://services.cbu.uib.no/tools/kaks).

Animals

Scml2-KO mice were previously reported [6]. All experimental work was approved by the Institutional Animal Care and Use Committee protocol no. IACUC2015-0032.

Cell culture

Mouse embryonic kidney (mK4) cells [46] (BioSample: SAMN02297925) were cultured in DMEM with GlutaMAX (Thermo Fisher Scientific), 10% fetal bovine serum, and antibiotics. To induce DNA hypomethylation, cells were treated with 0.5 μM 5-azacytidine (5azaC; Sigma) for 7 days or 100 μg/ml vitamin C (Sigma) for 24 h [34]. Transient transfections were performed in OptiMEM (Thermo Fisher Scientific) with Lipofectamine 3000 reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. For western blotting and ChIP-qPCR experiments, transfected cells were selectively collected via the MACSelect 4 System (Miltenyi Biotec) according to the manufacturer's instructions. Briefly, cells were co-transfected with pMACS 4.1 vector, which encodes a truncated CD4 marker. Transfected cells expressing truncated CD4 at the cell surface were labeled with MACS4 microbeads and collected by AutoMACS.

Western blotting

Whole nuclei were collected from transfected cells as described in [47]. Briefly, cells were resuspended in buffer A (10 mM HEPES-KOH pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT, protease inhibitors (Roche)). Triton X-100 (0.1%) was added, and the cells were incubated for 5 min on ice. Nuclei were collected in a pellet by low-speed centrifugation at 1300 × g for 6 min at 4°C. Nuclei were washed once in buffer A, lysed in Laemmli Buffer with Benzonase (NEB). Proteins were separated by SDS-PAGE, and blotted to an Immobilon-P PVDF Membrane (Millipore) by Trans-Blot Turbo Transfer System (BioRad). The membranes were blocked in StartingBlock T20 (TBS) Blocking Buffer (Thermo Fisher Scientific) for 30 min, probed with anti-H3K27me3 mouse monoclonal antibody (Active Motif) or anti-H3 rabbit polyclonal antibody (Millipore) in x1/20 diluted blocking buffer at 4°C overnight, and then incubated with an HRP-conjugated secondary antibody for 1 h at room temperature. Blots were developed using Immobilon Western Chemiluminescent HRP Substrate (Millipore) and exposed onto Super RX-N X-ray film (Fujifilm). To analyze SCML2 protein stability as shown in Supplementary Figure S3, whole cell lysates were analyzed by western blotting using an anti-SCML2 rabbit polyclonal antibody (homemade, described in [6]) or anti-Lamin B1 antibody (Abcam).

Fluorescence recovery after photobleaching

Full-length SCML2 and truncated forms of SCML2 were fused to EGFP and ectopically expressed in mK4 cells. FRAP experiments were performed using a confocal laser-scanning microscope (A1R, Nikon) equipped with a 60× (NA 1.2) water immersion objective and an incubation chamber (Tokai Hit), and were controlled via NIS-Elements (Nikon). Photobleaching and EGFP excitation were performed with a 488 nm laser. A small (1 μm) circular region of interest (ROI) was arbitrarily positioned in the nucleus. The EGFP bleaching-acquisition sequence consisted of prebleaching acquisition, bleaching of the ROI, followed by acquisition for 50 s. NIS-Elements was used to determine the mean ROI intensity values of FRAP signals (locations where the fluorescence was bleached away),

background signals (from regions outside cells), and reference signals (portions of cells that were not bleached). The background was subtracted from both the reference and FRAP ROIs. Relative intensity values for FRAP ROIs for each time point were normalized with reference signals and fit the FRAP recovery curves to the single exponential function. Each curve was obtained by plotting the mean relative intensities, calculated via 15-to-42 independent FRAP time courses from three independent experiments as a function of time after bleaching.

Purification of recombinant SCML2

Mouse *Scml2* cDNAs corresponding to full-length, the 10 SDB repeats, the RNA-binding region (RBR), and the Scm-like embedded domain (SLED) were PCR amplified with Q5 DNA polymerase (NEB), cloned into pGEX6P-1 (GE Healthcare) in frame with an N-terminal GST-tag encoded within the vector, and transformed into *E. coli* BL21-Gold (DE3). Expression of recombinant proteins was induced with 0.1 mM IPTG for 20 h at 22°C. The cells were washed with PBS and resuspended in Lysis Buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1 mM DTT, protease inhibitors (Roche)), incubated with lysozyme (2.5 μg/μl) for 30 min at 4°C, sonicated several times, and centrifuged at 21,000 × *g* for 20 min to clarify lysates. Lysates were incubated with 100 μl of Glutathione Sepharose 4B (GE Healthcare) for 2 h and washed four times with Lysis Buffer. For electrophoresis mobility shift assays (EMSA), the beads were washed twice with HRV 3C Buffer, and then proteins were eluted by cleavage of GST-tag with HRV 3C (Thermo Fisher Scientific) at 4°C overnight. For microplate-based protein DNA-binding assays, GST-tagged proteins were eluted with Elution Buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10% glycerol, 20 mM reduced glutathione, 1 mM DTT). Purified proteins were concentrated with Amicon Ultra Centrifugal Filters (Millipore). The protein solution was exchanged with 40% glycerol/PBS by Zeba Spin Desalting Columns (Pierce) and stored at –20°C.

EMSA assays

EMSA assays were performed on 5% polyacrylamide 0.5 × TBE gels. Biotinylated 88-bp dsDNA probes (5'-GTAAAACGACGGCCAGAGGAGCGGGAGAGGAGCGGGAG AGGAGCGGGAGAGGAGCGGGAGAGGAGCGGGAGGTCA TAGCTGTTTCCTG-3'), which contain the M13-forward sequence, 5 tandem repeats of human *ASCL-1* promoter region (AGGAGCGGGAG)—which was previously identified as a binding site for human SCML2 [32]—and the M13-reverse sequence, were PCR amplified using Phusion HF polymerase (NEB) and NTPs containing either unmodified, methylated, or hydroxymethylated dCTP residues (Zymo Research). The PCR primer sequences for probe construction were 5'-biotin-CAGGAAACAGCTATGAC-3' and 5'-biotin-GTAAAACGACGGCCAG-3'. Binding reactions (total 18 μl) were performed in Binding Buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 8% glycerol, 0.05% Triton X-100, 5 mM $MgCl_2$, 1 mM DTT) with 83 ng/μl BSA, 7 nM probe, and with or without purified SCML2 (0.1, 0.5, or 2.5 μM as indicated). After electrophoresis, samples were transferred to Hybond N + Membrane (GE Healthcare) and cross-linked by UV. The membrane was blocked with Bullet Blocking One (Nacalai) for 10 min, probed with streptavidin-HRP, and incubated with Immobilon Western Chemiluminescent HRP substrate (Millipore) to develop signals.

Microplate-based protein DNA-binding assay

For binding reactions, biotinylated dsDNA (40 ng) prepared as described above was mixed with purified GST-tagged proteins (2 μg), with or without unlabeled 88-bp dsDNA (2 μg) as a competitor, in 50 μl of Binding Buffer (25 mM Tris-HCl pH8.0, 50 mM NaCl, 5 mM $MgCl_2$, 0.05% Triton X-100, 1 mM DTT, 0.1% BSA). The protein-DNA mixture was added to a PIERCE Streptavidin Coated High Capacity White 96-Well Plate (Thermo Fisher Scientific) and incubated for 1 h at room temperature. The plate was washed three times with Wash Buffer (25 mM Tris-HCl pH8.0, 150 mM NaCl, 0.1% BSA, 0.05% Tween-20), low-salt Wash Buffer containing 75 mM NaCl, or high-salt Wash Buffer containing 300 mM NaCl as indicated. HRP-conjugated anti-GST mouse monoclonal antibody (Thermo Fisher Scientific) diluted with 50 μl of Wash Buffer was added to the plate and incubated for 1 h at room temperature. The plate was washed three times with Wash Buffer, low-salt Wash Buffer, or high-salt Wash Buffer as indicated. Chemiluminescence was developed with 100 μl Immobilon Western Chemiluminescent HRP Substrate (Millipore) and detected with a Synergy H1 Hybrid Multi-Mode Microplate Reader (BioTek). Four independent experiments were performed for statistical analyses.

Germ cell fractionation

Isolation of the Thy1$^+$ fraction was performed as described previously [48].

Co-Immunoprecipitation

For detection of phosphorylation of the SLED domain in SCML2, mK4 cells were seeded on a 60-mm dish and transfected with a vector expressing EGFP-SLED. Forty-eight hours after the transfection, cells were lysed with Buffer E (20 mM HEPES-KOH pH 7.9, 500 mM NaCl, 10% glycerol, 1 mM DTT, 0.5% TritonX-100, protease inhibitors (Roche), phosphatase inhibitors (Sigma)), and centrifuged at 8000 × *g* for 10 min at 4°C. Supernatant was transferred to a new tube, incubated with anti-GFP rabbit polyclonal antibody (Abcam) for 2 h at 4°C, and then incubated with Dynabeads Protein G (Thermo Fisher Scientific) for 1 h at 4°C. The beads were washed four times with Buffer E without protease inhibitors and phosphatase inhibitors, washed once with λ protein phosphatase reaction buffer (NEB), and incubated with or without λ protein phosphatase (NEB) for 30 min at 30°C. The bound proteins were eluted with Laemmli Buffer for 5 min at 95°C and analyzed by western blotting using anti-GFP or anti-phospho Ser/Thr/Tyr rabbit polyclonal antibody (Thermo Fisher Scientific).

ChIP-qPCR

Cells were suspended in Buffer A (described above). Triton X-100 (0.1%) was added, and the cells were incubated for 5 min on ice. Nuclei were collected in a pellet by low-speed centrifugation at 1300 × *g* for 6 min at 4°C. Nuclei were washed once in Buffer A, resuspended with Buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, protease inhibitors (Roche)), disrupted by sonication, and incubated for 30 min on ice. EGTA (final concentration of 10 mM), KCl (final concentration of 300 mM), and NP-40 (final concentration of 0.2%) were added to the lysates. The lysates were incubated for 1 h at 4°C with rotation at 20 rpm and centrifuged at 8000 × *g* for 10 min at 4°C. Using the supernatant, chromatin immunoprecipitation was carried out on an SX-8X IP-STAR compact automated system (Diagenode). Dynabeads Protein G (Thermo Fisher Scientific) were pre-incubated with 0.1% BSA for 2 h. Then, the chromatin fractions

were incubated with the beads conjugated with rabbit polyclonal antibodies against SCML2 (homemade, described in [6]), histone H3 (Millipore), or control IgG (Millipore) for 4 h at 4°C and washed four times with Wash Buffer (20 mM HEPES-KOH pH 7.9, 300 mM KCl, 0.05% NP-40, and 0.5 mM DTT). The beads were incubated with Elution Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM NaCl, 0.3% SDS, 0.1 μg/μl Proteinase K) for 30 min at 42°C. ChIPed DNA was purified by Agencourt AMPure XP (Beckman Coulter). qPCR was performed with FAST SYBR Green Master Mix (Thermo Fisher Scientific) on a StepOnePlus Real-time PCR system (Thermo Fisher Scientific). Primer sequences: *Hoxd11*-ChIP-Fw, CAC TCT TGT CCC TGG TGT CA; *Hoxd11*-ChIP-Rv, CTG GGA GCT TGT TGC TTC TT; *Cdkn2a*-ChIP-Fw, TAG ACG CCT GCG CAG AAC TT; *Cdkn2a*-ChIP-Rv, GCA AAA GCG CGA TTG ATG CC; rDNA-ChIP-Fw, CCT GTG AAT TCT CTG AAC TC; rDNA-ChIP-Rv, CCT AAA CTG CTG ACA GGG TG. Four independent experiments were performed for statistical analyses.

RNA-sequencing and data analysis

RNA-seq data (GSE89502) was analyzed as performed previously [48].

Results

Identification of a rapidly evolved domain, the SCML2 DNA-binding repeats, in mouse SCML2

There are two major splicing isoforms of mouse SCML2: an isoform that lacks the SPM (also known as SAM) domain, which interacts with Polycomb proteins [22, 23], and another isoform that contains the SPM domain and is a homolog of human SCML2A [24] (Figure 1A). The former is the predominant isoform expressed exclusively in spermatogenic cells, and the latter is not expressed in spermatogenesis (Supplementary Figure S1A and B). Unlike human SCML2, mouse SCML2 contains 10 repeats of 28-amino-acid units enriched with basic amino acids such as arginine and lysine (Figure 1B), and the theoretical isoelectric point (pI) of an individual repeat unit is strongly basic (pI = 9.53). This 10-repeat region has no homology to known functional domains or motifs. As described below in this study, we identify the DNA-binding function of this repeat region. Therefore, we termed this novel domain the SCML2 DNA-binding (SDB) repeats.

Our genomic search identified this domain of SCML2 as specific to a subset of rodents such as mice, rats, voles, and hamsters (Figure 1C), suggesting that the domain was acquired in rodent phylogenesis. Interestingly, the numbers of SDB repeats vary across species of rodents, ranging from 6 to 14 repeats (Figure 1C). Because other lineage-specific genomic domains on the X chromosome are associated with reproductive function [25–28], acquisition of this domain on the X-linked *Scml2* gene might be related to the reproductive fitness of rodents. Therefore, we sought to determine whether acquisition of this domain is associated with an evolutionary trait that reflects reproductive fitness. To this end, we calculated the ratio of nonsynonymous (*Ka*: associated with amino acid changes) to synonymous (*Ks*: without amino acids changes) nucleotide substitution rates, an indicator of selective pressures on genes. *Ka/Ks* values of SCML2 among rodents are relatively high (0.857 between hamster and rat, and 0.918 between the hamster/rat node and mouse: Figure 1D), suggesting that SCML2 undergoes rapid evolution. These values were compared to that of *Rnf2*, which encodes the main catalytic subunit of PRC1, mediating the monoubiquitination of histone H2A at Lysine 119; and *Ezh2*, which encodes the main catalytic subunit of PRC2, mediating H3K27me3. Consistent with the broad functions of PRC1 and PRC2, the *Rnf2* and *Ezh2* genes are subject to strong purifying selection that maintains amino acid sequences, so their respective *Ka/Ks* values are very low (less than 0.1 among rodents: Figure 1D). *Scml2* is a germline-specific X-linked gene, and X-linked genes related to spermatogenesis tend to have higher *Ka/Ks* values, presumably due to conferred advantages for reproductive fitness [29]. Therefore, our analyses suggest that SCML2 has undergone rapid evolution associated with reproductive fitness.

Notably, the *Ka/Ks* values for the region of *Scml2* encoding the 10 SDB repeats is higher than that of full-length *Scml2*, and the *Ka/Ks* value of the same region between hamster and rat is 1.22 (Figure 1D). Since *Ka/Ks* values above 1 are observed in only rare cases in which selection has acted to change a protein (i.e. a protein subject to positive selection) [30], these analyses suggest that the 10 SDB repeats is a rapidly evolved domain subject to positive selection, thereby conferring an advantage in reproductive fitness. We also found that the RBR [31] of SCML2, which we identify as a novel DNA-binding domain in this study (described below), has a relatively high *Ka/Ks* value (1.42 between human and rat: Supplementary Figure S2A). On the other hand, other domains of SCML2 have relatively low *Ka/Ks* values (Supplementary Figure S2A), suggesting that these domains are subject to purifying selection.

Another notable feature of the 10 SDB repeats is that each repeat unit is coded by a single exon. Therefore, the 10 SDB repeats is likely derived as the consequence of genomic duplications (Figure 1E). Below, we sought to determine the function of the 10 SDB repeats to infer the mechanisms behind this unique evolutionary trait.

SCML2 DNA-binding repeats and RNA-binding regions are two DNA-binding domains for mouse SCML2

To determine the molecular mechanisms by which SCML2 binds chromatin, we tested the ability of SCML2 to bind DNA. To this end, we expressed full-length SCML2 proteins in *Escherichia coli* (Figure 2A and B) for EMSA, and we found that full-length SCML2 directly bound DNA that contains unmethylated cytosine (Figure 2C). In cultured germline stem (GS) cells, which include a stem cell population, SCML2 preferentially binds transcription start sites of target genes where DNA is hypomethylated, based on ChIP-seq analysis [6]. To determine whether SCML2's affinity changes depending on the methylation status of DNA, we performed EMSA using DNA substrates that contain 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), an intermediate in active DNA demethylation. However, full-length SCML2 bound DNA substrates containing either 5mC or 5hmC (Figure 2D). Together, these results indicate that SCML2 binds to both methylated and unmethylated DNA in vitro.

Next, to determine the domains of SCML2 that function in DNA binding, we investigated the DNA-binding properties of the individual domains that comprise SCML2. Human SCML2 has two major functional domains: the DNA-binding SLED [32] and the RBR [31]. In mice, the 10-repeat domain and RBR directly bound DNA while, unexpectedly, the mouse SLED did not bind DNA (Figure 2E).

To independently confirm the DNA-binding properties of SCML2's RBR and 10 SDB repeats, we performed a microplate-based protein DNA-binding assay (Figure 3). Briefly, we incubated biotinylated dsDNA with a GST-tagged protein of interest, immobilized the dsDNA-protein complex on streptavidin-coated microplates, and then examined their binding properties with an

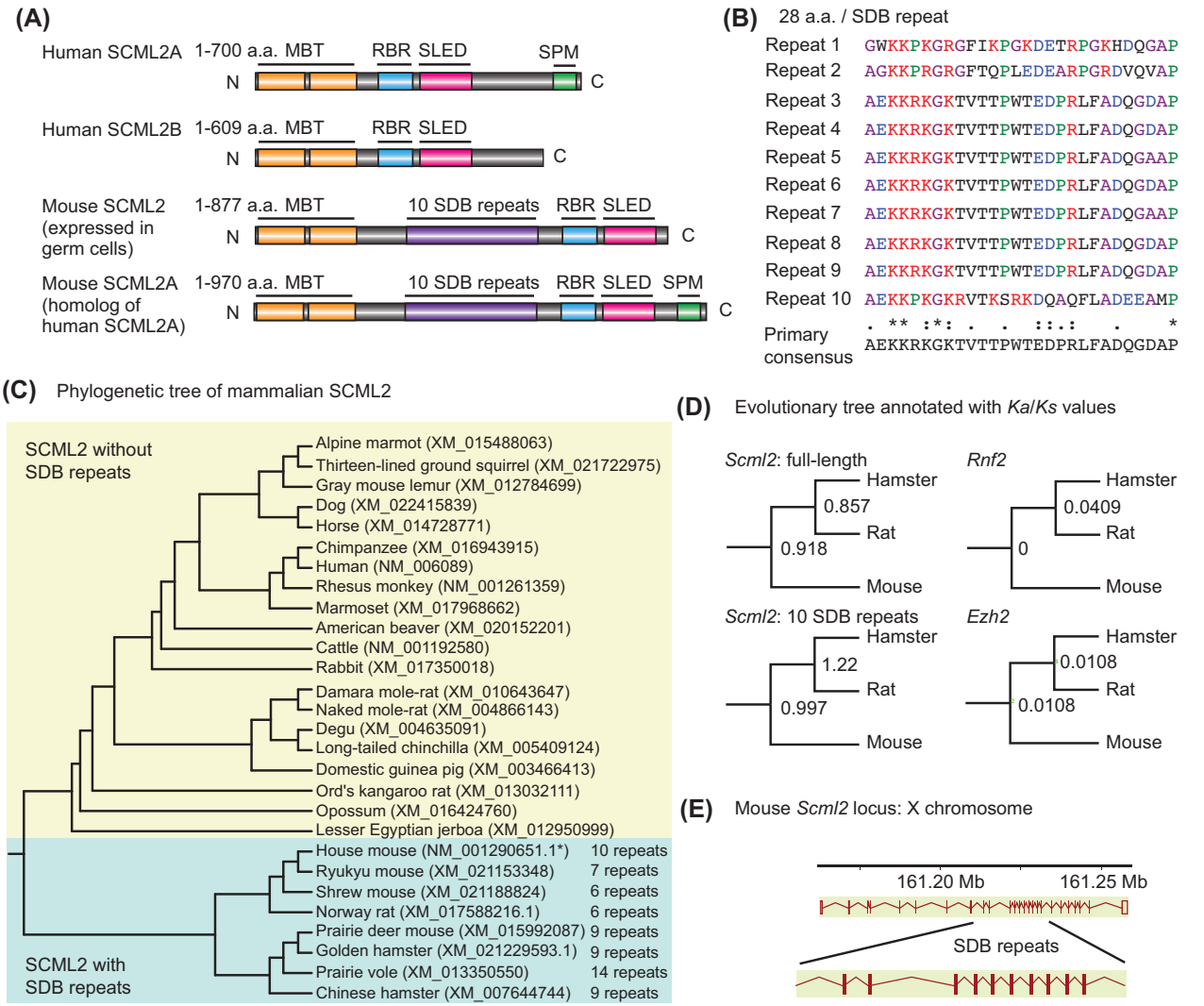

Figure 1. The SCML2 DNA-binding (SDB) repeats. (**A**) Schematic of human and mouse SCML2 proteins. (**B**) Amino acid sequence of the 10 SDB repeats. (**C**) Phylogenetic tree of mammalian SCML2. * For the sequence of house mouse SCML2, a cDNA clone similar to NM_001290651.1 was used. (**D**) *Ka*/*Ks* values and a *Ka*/*Ks* annotated evolutionary tree. (**E**) Genomic structure of the mouse *Scml2* locus.

HRP-labeled anti-GST antibody followed by chemiluminescence. The microplate-based protein DNA-binding assay confirmed the DNA-binding properties of the RBR and 10 SDB repeats (Figure 3). Their specific binding was confirmed with experiments using an excessive amount of competitor DNA. Furthermore, the DNA-binding properties of the RBR and 10 SDB repeats remained stable after wash steps with high-salt buffer (300 mM NaCl), and we also confirmed that SLED does not bind to DNA after wash steps with low-salt buffer (75 mM NaCl). Together, we conclude that mouse SCML2 has two independent DNA-binding domains: the RBR and the SDB.

Hypomethylated DNA promotes SCML2 binding to chromatin in an in vivo chromatin environment

When we identified the DNA-binding domains of SCML2, we noticed that recombinant SCML2 protein did not distinguish DNA methylation status in the EMSA. This raises the possibility that SCML2 facilitates the sensing of DNA hypomethylation status in an in vivo chromatin environment. To determine whether SCML2 responds to DNA methylation status in vivo, we utilized the FRAP method, a live-imaging assay to measure the mobility of chromatin-binding proteins. In this setup, we took advantage of somatic cells, where germline-specific SCML2 is not endogenously expressed, to monitor the mobility of full-length SCML2 fused to EGFP (EGFP-SCML2). We used the somatic cell line mK4, derived from embryonic kidney mesenchyme [33], to monitor EGFP-SCML2. The recovery kinetics after photobleaching were examined under hypomethylated conditions using two demethylation agents: 5-azacytidine, a DNMT1 inhibitor, and vitamin C, which promotes TET activity for DNA demethylation [34]. We confirmed that neither 5-azacytidine nor vitamin C affected the stability of SCML2 (Supplementary Figure S3A). Both 5-azacytidine and vitamin C induced slower recovery of EGFP-SCML2 after photobleaching (Figure 4A), suggesting that SCML2 tightly binds to chromatin under hypomethylated conditions, which slows the recovery of GFP signals to photobleached sites. This is indicative of two simultaneous events: (1) under hypomethylated conditions, bleached EGFP-SCML2 tightly binds to sites of photobleaching; and (2) under the same conditions,

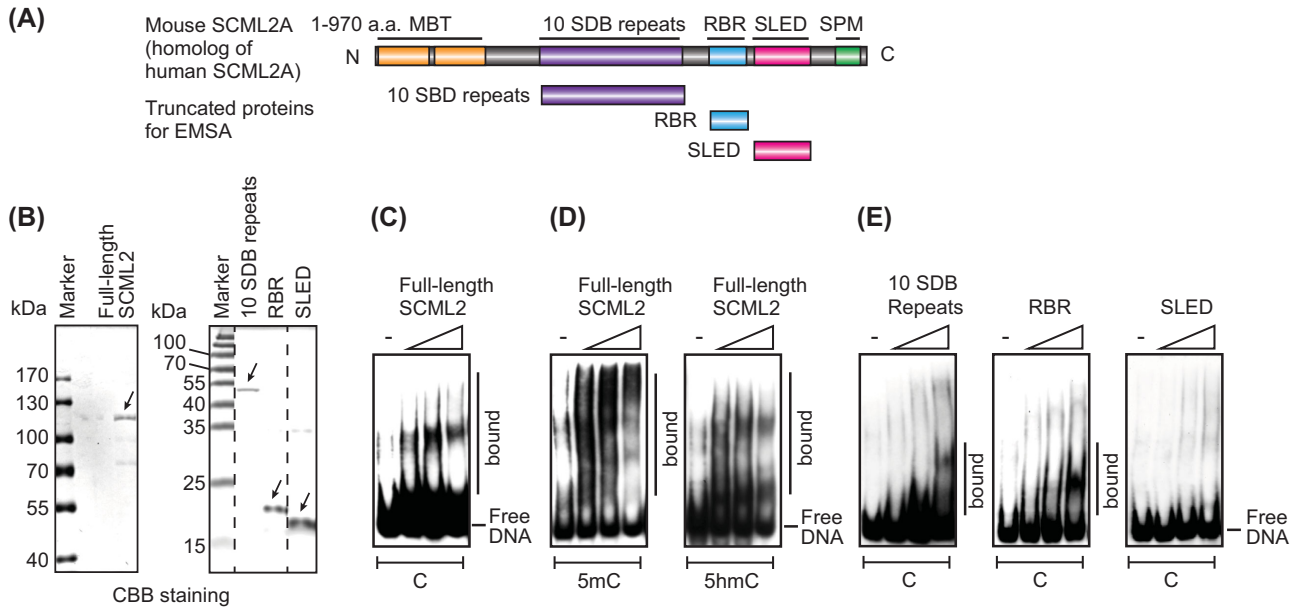

Figure 2. The SCML2 DNA-binding (SDB) repeats and RNA-binding regions (RBR) are two DNA-binding domains in mouse SCML2. (**A**) Schematic of mouse SCML2 protein and the truncated proteins used for EMSA. (**B**) CBB staining of full-length and truncated SCML2 proteins used for EMSA. (**C–E**) EMSA assays using probes with unmethylated cytosine (C), methylated cytosine (mC), or hydroxymethylated cytosine (hmC) residues. EMSA assays were performed with or without SCML2 (0.1, 0.5, or 2.5 μM).

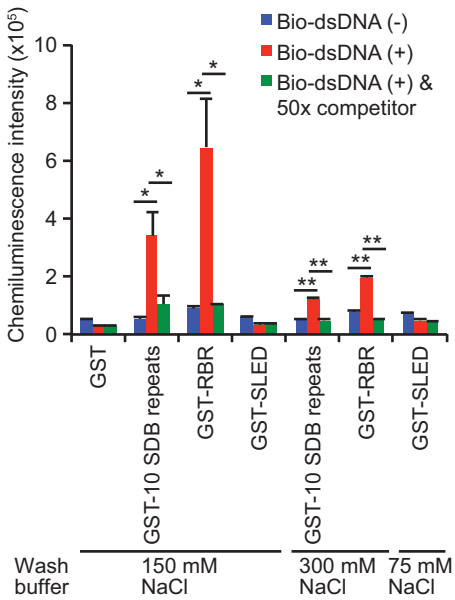

Figure 3. Microplate-based protein DNA-binding assays. Readouts from microplate-based protein DNA-binding assays. Bio-dsDNA: biotinylated ds-DNA, 50x: 50 times. As negative controls, DNA binding was tested without biotinylated double-strand DNA probes (Bio-dsDNA (–): shown by the blue bars). $^{*}P < 0.05$, $^{**}P < 0.005$, unpaired t-test. Mean ± SEM (error bars); n = 4.

nonbleached EGFP-SCML2 tightly binds to the area outside bleached sites. Together, these events cause a delay in the recovery of GFP signals at sites of photobleaching. The recovery rate is calculated based on the amount of time needed for 50% of plateaued intensity to be recovered ($t_{1/2}$). Under control conditions, when there is no treatment with demethylation agents, it took 8.56 s to recover 50% of plateaued intensity ($t_{1/2}$) at sites of photobleaching (Figure 4B). The mobile fraction is determined from the plateaued value of the relative intensity. Both 5-azacytidine and vitamin C induced delayed recovery ($t_{1/2}$) and a lower mobile fraction (Figure 4B). Although we could not eliminate the possible indirect effects of these demethylation reagents, these quantitative results suggest that hypomethylated DNA promotes SCML2 binding to chromatin in vivo.

Next, using a series of domain deletion mutants derived from the EGFP-SCML2 construct (Figure 4C), we examined the chromatin-binding functions of each SCML2 domain. SCML2 mutants lacking the malignant brain tumor domains (ΔMBT), which are chromatin readers of histone posttranslational modifications [35]; the 10 SDB repeats (Δ10-rep); or the RBR (ΔRBR) demonstrated faster recovery and higher mobility (Figure 4D), indicating a reduction in chromatin-binding ability. Given the functions of the 10 SDB repeats and the RBR in DNA binding (Figures 2E and 3), it is conceivable that SCML2 binds to chromatin through these domains. Interestingly, the SLED deletion mutant (ΔSLED) exhibited slower recovery and lower mobility than full-length SCML2 (Figure 4D), suggesting that ΔSLED has a higher affinity to chromatin in comparison to full-length SCML2. Thus, counterintuitively, SLED negatively regulates the chromatin binding of SCML2.

We sought to further define the specific chromatin-binding abilities of the SCML2 domains with SCML2 truncation proteins (Figure 5A). We found that MBT and SLED do not bind to chromatin (Supplementary Figure S3B), while the SDB repeats and RBR bind to chromatin (Figure 5B). Importantly, the 10 SDB repeats truncation protein (10-rep) has a higher affinity for chromatin than the 5 SDB repeats truncation protein (5-rep), suggesting that chromatin-binding ability increases as the number of repeats increases. Furthermore, 10-rep with RBR (10-rep + RBR) has a higher affinity than each single domain, suggesting that the chromatin-binding abilities of the SDB repeats and RBR are cooperative. Finally, we examined which domain is important for sensing hypomethylation

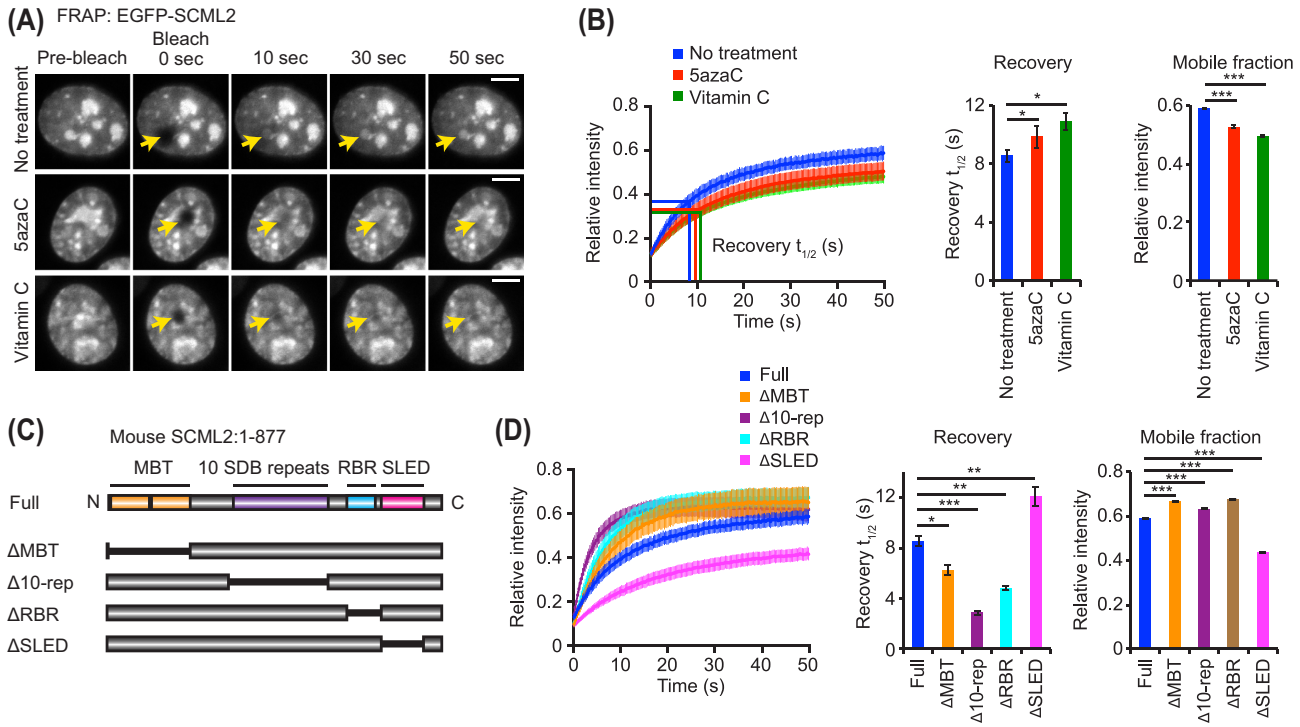

Figure 4. Cooperative action of SCML2 domains for chromatin-binding and sensing hypomethylated status. (**A**) Representative live images of FRAP assays in mK4 cells expressing EGFP-SCML2 with or without demethylation agents. 5azaC: 5-azacytidine. FRAP assays performed without demethylation agents, labeled no treatment, are shown in the panel. Bars: 5 μm. (**B**) Relative intensity, recovery, and mobile fraction of FRAP assays using full-length SCML2. Recovery represents $t_{1/2}$ (s). n $\geq$ 33. Error bars for FRAP curves and recovery $t_{1/2}$ represent 95% confidence interval of the mean. Error bars for mobile fractions represent SEM. (**C**) Schematics of mouse SCML2 proteins and truncated proteins. (**D**) FRAP assays using the truncated proteins shown in (**C**). n $\geq$ 18. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0001$, unpaired t-test.

status. Δ10-rep and ΔRBR lost the ability to sense hypomethylation status (Figure 5C and Supplementary Figure S3C). Since individual domains did not respond to methylation status (Figure 5D and Supplementary Figure S3D), we propose that the overall function of SCML2 for chromatin binding and hypomethylation sensing is defined by the collective action of the domains; specifically, the SDB repeats and the RBR promote chromatin binding through direct DNA binding, while the SLED antagonizes this process. The MBT also has a regulatory role in this process since ΔMBT reduced the chromatin-binding ability (Figure 4D).

To understand why the SLED negatively regulates the binding of SCML2 to chromatin, we investigated the physical properties of the SLED. We found that SLED is enriched with Ser, Thr, and Tyr residues, which are potential targets of phosphorylation (Figure 5E). To confirm whether the SLED is a target of phosphorylation, we expressed a GFP fusion protein of the SLED and tested the phosphorylation at Ser, Thr, and Tyr residues after immunoprecipitation using an anti-GFP antibody. We found that the SLED is phosphorylated at Ser, Thr, and/or Tyr residues, and confirmed the phosphorylation through dephosphorylation via λ protein phosphatase treatment (Figure 5F). These data suggest the possibility that the phosphorylation of SLED could drive its negative regulation of SCML2 chromatin binding.

SCML2 binds to hypomethylated chromatin to facilitate H3K27me3

Finally, we sought to determine the functional consequences of SCML2's chromatin-binding function. Since SCML2 is enriched at the hypomethylated promoters of target genes in cultured GS cells, as determined by ChIP-seq analysis [6], we examined how SCML2 functions after recruitment to hypomethylated chromatin. We validated SCML2 binding at target loci via ChIP followed by quantitative PCR (qPCR). SCML2 was enriched at the promoters of bivalent-domain genes such as *Cdkn2a* and *Hoxd11* in Thy1^{+} spermatogonia, as well as in mK4 cells with ectopically expressed tandem affinity purification (TAP)-tagged-SCML2 (Figure 6A). Because SCML2 regulates H3K27me3 on bivalent genomic domains [18] and pericentric heterochromatin in spermatogenesis [21], we tested the hypothesis that chromatin binding of SCML2 induces H3K27me3 using our mK4 cell model. While demethylation reagents did not change the level of H3K27me3 in the absence of SCML2, the level of H3K27me3 increased after treatment with demethylation agents in the presence of ectopically expressed SCML2 (Figure 6B). These data provide a functional link between DNA hypomethylation and the SCML2-mediated induction of H3K27me3. The massive, genome-wide increase of H3K27me3 is likely due to the increase on both target gene promoters and pericentric heterochromatin. Together with the data above, the rapidly evolved SDB repeats may alter the species-specific function of SCML2 to recognize hypomethylated promoters at target loci and thereby induce H3K27me3 to establish bivalent domains.

Discussion

Here, we identify the SDB repeats of mouse SCML2 as a rapidly evolved DNA-binding domain. Furthermore, by uncovering the molecular function of SCML2, our study reveals how SCML2 is

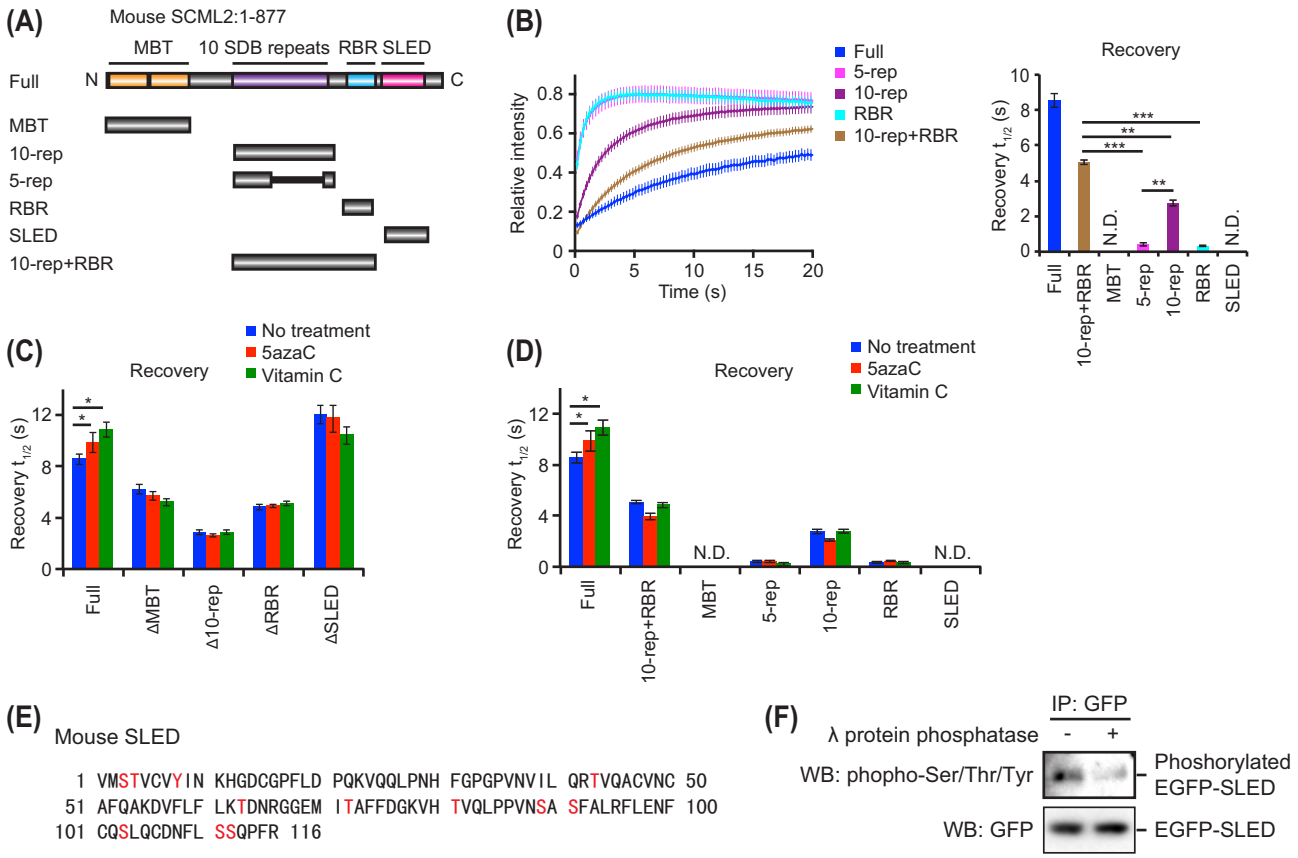

Figure 5. Function of each SCML2 domain for chromatin-binding and sensing hypomethylated status. (**A**) Schematics of mouse SCML2 proteins and truncated proteins. (**B**) FRAP assays using the truncated proteins shown in (**A**). n ≥ 18. (**C, D**) FRAP assays using the truncated proteins with or without demethylation agents. n ≥ 15. (**E**) Amino acid sequence of mouse SLED. Ser, Thr, and Tyr residues are shown in red. (**F**) Detection of Ser/Thr/Tyr phosphorylation after immunoprecipitation. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0001$, unpaired t-test.

recruited to chromatin. The chromatin-binding function of SCML2 results from the synergy of three domains: the MBT, 10 SDB repeats, and RBR (Figure 7). Furthermore, the SLED negatively regulates the chromatin-binding function of SCML2. Thus, rather than a simple mechanism mediated by a specific catalytic domain, it is possible that allosteric changes to distal domains within SCML2 regulate its chromatin-binding function according to the local chromatin environment. Although in vitro EMSAs did not reveal the direct response of SCML2 to hypomethylated DNA, the FRAP assays revealed that hypomethylated DNA promotes SCML2 binding to chromatin in an in vivo chromatin environment. While the mechanism by which Polycomb proteins are recruited to target loci remains elusive, one key Polycomb subunit is of particular interest: KDM2B, which recognizes DNA hypomethylation at CpG sites and recruits PRC1 [36–38]. Therefore, it is conceivable that SCML2 works with a direct sensor of DNA hypomethylation (such as KDM2B) in a complex, and the functions of each domain of SCML2 may modulate the function of the sensor protein. An alternative possibility is that SCML2 is recruited to sites of open chromatin, which are generally associated with DNA hypomethylation. In accord with this possibility, our recent studies demonstrated that SCML2 binds to hypomethylated promoters in undifferentiated spermatogonia [18], and these promoters possess features of open chromatin during spermatogenesis [11].

Here we suggest that the mouse SLED does not have a DNA-binding function, unlike the human SLED, which binds DNA [32]. One possible reason for this is that acquisition of the SDB domain in rodent SCML2 displaced the DNA-binding function of the SLED. This may be an example of evolutionary adaptation to gain reproductive fitness. However, the SLED is highly conserved between human and mouse (Supplementary Figure S2B), and the *Ka/Ks* value for the SLED is relatively low compared to the rapidly evolved SDB repeats and RBR (Supplementary Figure S2A), suggesting that the SLED is subject to purifying selection in the course of evolution. These evolutionary traits do not explain the reason why the functions of SLED differ between human and mouse. Further investigation is warranted to understand how functional differences between the human and mouse SLED arose.

Variations in the numbers of SDB repeats across rodent species is particularly intriguing because copy number variation of DNA-binding domains potentially determines the species-specific function of proteins required for reproduction. One example of such a repeat domain is the zinc fingers of PRDM9. In rodents, sequences and copy numbers of the zinc fingers of PRDM9 are rapidly evolved and subject to positive selection [4]. PRDM9 is involved in speciation [39, 40], and the rapid evolution of PRDM9 zinc fingers was proposed to influence the turnover of recombination hotspots, impede recombination in hybrids, and contribute to speciation in mice [41]. It is

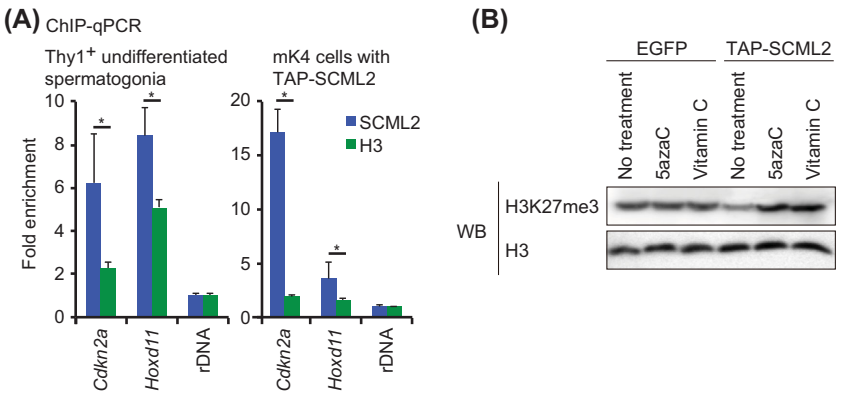

Figure 6. SCML2 induces H3K27me3 in response to DNA hypomethylation. (**A**) SCML2 enrichment was detected with ChIP-qPCR using Thy1$^+$ undifferentiated spermatogonia (left) and mK4 cells expressing TAP-SCML2 (right). IgG data was used for normalization; fold enrichment was calculated based on the normalized data. The relative values of ChIP-qPCR data to input samples were normalized against rDNA. To determine the enrichment of SCML2 on target loci, ChIP data using anti-SCML2 antibody (blue bars) were compared with that of anti-panH3 antibody (green bars). Mean ± SEM (error bars); n = 4. $^*P < 0.05$, unpaired t-test. (**B**) Western blot detecting the induction of H3K27me3 in mK4 cells expressing TAP-SCML2 or EGFP control with and without treatment with demethylation agents. Bars: 5 μm. Two independent experiments were performed.

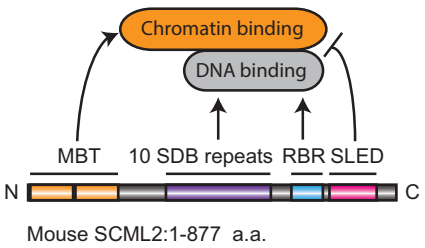

Figure 7. Model of the functions of each domain of SCML2. The 10 SDB repeats and RBR are DNA-binding domains for mouse SCML2. The chromatin-binding function of SCML2 results from the synergy of three domains: the MBT, 10 SDB repeats, and RBR, while the SLED negatively regulates the chromatin-binding function of SCML2.

tempting to speculate that SCML2 establishes a species-specific gene expression program in spermatogenesis due to the species-specific characteristic of the SDB repeats. A promising research direction is to determine the function of the SDB repeats in the establishment of such a species-specific gene expression program. Given the function of SCML2 in the regulation of H3K27me3 and bivalent domains, it is also intriguing to test whether there is a species-specific mechanism involved in the recovery of totipotency after fertilization.

Curiously, a rapidly evolved hydrophilic repeat domain was recently reported in GCNA proteins, which are also specific to the germline [42]. While SCML2's SDB repeats are strongly basic (theoretical pI = 9.53), GCNA's hydrophilic repeats are strongly acidic (theoretical pI = 4.17). Due to this strongly acidic feature, it is unlikely that the hydrophilic repeats of GCNA have the ability to bind DNA. However, the hydrophilic repeats of both SCML2 and GCNA were likely derived in the course of evolution to increase functional diversity and reproductive fitness. A notable common feature between SCML2 and GCNA is their X-linkage. In 1984, Rice hypothesized that X chromosomes accumulate male-biased genes [43]. This is because the X chromosome is hemizygous in males, so mutations on the X chromosome can readily present a phenotype without masking compensation from another allele. Thus, if mutations are beneficial to reproductive fitness, the X chromosome is an ideal site for preferential preservation. With respect to Rice's hypothesis, the unique repeats of both SCML2 and GCNA, with their putative evolutionary advantages, might have been acquired due to their X-linkage. It is noteworthy that other rapidly evolved proteins, such as SRY and RHOX5, are located on the sex chromosomes too; furthermore, X-linked genes expressed in spermatids tend to be newly and rapidly evolved [25–29, 44]. Therefore, the unique evolutionary trait of the SCML2 SDB repeats could be, in part, attributed to its location on the X chromosome. It is interesting to speculate that acquisition of the SDB repeats may contribute to robust reproduction (for example, many siblings, short reproductive cycles) in a specific subset of rodents.

Supplementary data

Supplementary data are available at *BIOLRE* online.

Supplementary Figure S1. SCML2 lacks the SPM domain in mouse spermatogenesis. (**A**) RNA-seq data during spermatogenesis. *Scml2* transcripts expressed during spermatogenesis lack the sequence encoding the SPM domain. (**B**) DNA sequences of two types of *Scml2* cDNA, with and without the SPM domain, are shown.
Supplementary Figure S2. Evolutionary aspects of SCML2. (**A**) *Ka/Ks* values and a *Ka/Ks* annotated evolutionary tree for each domain of SCML2 or full-length SCML2 without 10 SDB repeats of SCML2. (**B**) Amino acid sequence of human and mouse SLED. Ser, Thr, and Tyr residues are shown in red. Nine out of 12 Ser, Thr, and Tyr residues of mice SCML2 are conserved in human SCML2.
Supplementary Figure S3. Functions of SCML2 domains for chromatin binding. (**A**) Stability of SCML2 protein after treatment with 5-azacytidine and vitamin C. (**B**) Relative intensities of FRAP assays using MBT or SLED truncated proteins. The MBT and SLED do not bind chromatin. n ≥ 18. (**C, D**) Relative intensities of FRAP assays using the labeled truncated proteins with or without demethylation agents. n ≥ 15. Error bars for FRAP curves represent 95% confidence interval of the mean.

Acknowledgments

We thank Kazuteru Hasegawa, Yueh-Chiang Hu, Jérôme Déjardin, Eda Yildrim, and members of the Namekawa lab for discussion and helpful comments regarding the manuscript.

References

1. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. *Nat Rev Genet* 2002; **3**(2):137–144.
2. Whitfield LS, Lovell-Badge R, Goodfellow PN. Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature* 1993; **364**(6439):713–715.
3. Ponting CP. What are the genomic drivers of the rapid evolution of PRDM9? *Trends Genet* 2011; **27**(5):165–171.
4. Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genet* 2009; **5**(12):e1000753.
5. Maiti S, Doskow J, Sutton K, Nhim RP, Lawlor DA, Levan K, Lindsey JS, Wilkinson MF. ThePemHomeobox gene: rapid evolution of the homeodomain, X chromosomal localization, and expression in reproductive tissue. *Genomics* 1996; **34**(3):304–316.
6. Hasegawa K, Sin HS, Maezawa S, Broering TJ, Kartashov AV, Alavattam KG, Ichijima Y, Zhang F, Bacon WC, Greis KD, Andreassen PR, Barski A et al. SCML2 establishes the male germline epigenome through regulation of histone H2A ubiquitination. *Dev Cell* 2015; **32**(5):574–588.
7. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 2008; **9**(2):129–140.
8. Kota SK, Feil R. Epigenetic transitions in germ cell development and meiosis. *Dev Cell* 2010; **19**(5):675–686.
9. Kimmins S, Sassone-Corsi P. Chromatin remodelling and epigenetic features of germ cells. *Nature* 2005; **434**(7033):583–589.
10. Sin HS, Kartashov AV, Hasegawa K, Barski A, Namekawa SH. Poised chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biol* 2015; **13**(1):53.
11. Maezawa S, Yukawa M, Alavattam KG, Barski A, Namekawa SH. Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res* 2018; **46**(2): 593–608.
12. Adams RA, Maezawa S, Alavattam KG, Abe K, Sakashita A, Shroder M, Broering TJ, Sroga Rios J, Thomas MA, Lin X, Price CM, Barski A et al. RNF8 and SCML2 cooperate to regulate ubiquitination and H3K27 acetylation for escape gene activation on the sex chromosomes. *PLoS Genet* 2018; **14**(2):e1007233 in press.
13. Aranda S, Mas G. Regulation of gene transcription by Polycomb proteins. *Sci Adv* 2015; **1**(11):e1500737–e1500737.
14. Geisler SJ, Paro R. Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. *Development* 2015; **142**(17):2876–2887.
15. Laugesen A, Helin K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* 2014; **14**(6):735–751.
16. Maezawa S, Hasegawa K, Yukawa M, Sakashita A, Alavattam KG, Andreassen PR, Vidal M, Koseki H, Barski A, Namekawa SH. Polycomb directs timely activation of germline genes in spermatogenesis. *Genes Dev* 2017; **31**(16):1693–1703.
17. Mu W, Starmer J, Fedoriw AM, Yee D, Magnuson T. Repression of the soma-specific transcriptome by Polycomb-repressive complex 2 promotes male germ cell development. *Genes Dev* 2014; **28**(18):2056–2069.
18. Maezawa S, Hasegawa K, Yukawa M, Kubo N, Sakashita A, Alavattam KG, Sin HS, Kartashov AV, Sasaki H, Barski A, Namekawa SH. Polycomb protein SCML2 facilitates H3K27me3 to establish bivalent domains in the male germline. *Proc Natl Acad Sci USA* 2018; **115**(19):4957–4962.
19. Lesch BJ, Dokshin GA, Young RA, McCarrey JR, Page DC. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proc Natl Acad Sci USA* 2013; **110**(40):16061–16066.
20. Lesch BJ, Page DC. Poised chromatin in the mammalian germ line. *Development* 2014; **141**(19):3619–3626.
21. Maezawa S, Hasegawa K, Alavattam KG, Funakoshi M, Sato T, Barski A, Namekawa SH. SCML2 promotes heterochromatin organization in late spermatogenesis. *J Cell Sci* 2018; **131**(17):jcs217125.
22. Peterson AJ, Kyba M, Bornemann D, Morgan K, Brock HW, Simon J. A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Mol Cell Biol* 1997; **17**(11):6683–6692.
23. Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, Ishikura T, Toyoda T, Bernstein BE, Koseki H. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* 2013; **26**(6):565–577.
24. Lecona E, Rojas LA, Bonasio R, Johnston A, Fernandez-Capetillo O, Reinberg D. Polycomb protein SCML2 regulates the cell cycle by binding and modulating CDK/CYCLIN/p21 complexes. *PLoS Biol* 2013; **11**(12):e1001737.
25. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res* 2004; **14**(10a):1861–1869.
26. Ellis PJ, Bacon J, Affara NA. Association of Sly with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? *Hum Mol Genet* 2011; **20**(15):3010–3021.
27. Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK, Page DC. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* 2013; **45**(9):1083–1087.
28. Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BA, Smith AD, Dean MD, Good JM. Contrasting levels of molecular evolution on the mouse X chromosome. *Genetics* 2016; **203**(4):1841–1857.
29. Sin HS, Ichijima Y, Koh E, Namiki M, Namekawa SH. Human postmeiotic sex chromatin and its impact on sex chromosome evolution. *Genome Res* 2012; **22**(5):827–836.
30. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* 2002; **18**(9):486–487.
31. Bonasio R, Lecona E, Narendra V, Voigt P, Parisi F, Kluger Y, Reinberg D. Interactions with RNA direct the Polycomb group protein SCML2 to chromatin where it represses target genes. *Elife* 2014; **3**:e02637.
32. Bezsonova I. Solution NMR structure of the DNA-binding domain from Scml2 (sex comb on midleg-like 2). *J Biol Chem* 2014; **289**(22):15739–15749.
33. Valerius MT, Patterson LT, Witte DP, Potter SS. Microarray analysis of novel cell lines representing two stages of metanephric mesenchyme differentiation. *Mech Dev* 2002; **112**(1-2):219–232.
34. Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, Mahapatra S, Tam A, Laird DJ, Hirst M, Rao A, Lorincz MC, Ramalho-Santos M. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 2013; **500**(7461):222–226.
35. Bonasio R, Lecona E, Reinberg D. MBT domain proteins in development and disease. *Seminars in Cell & Developmental Biology* 2010; **21**(2):221–230.
36. Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW, Koseki H, Brockdorff N et al. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* 2012; **1**:e00205.
37. He J, Shen L, Wan M, Taranova O, Wu H, Zhang Y. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol* 2013; **15**(4):373–384.
38. Wu X, Johansen JV, Helin K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol Cell* 2013; **49**(6):1134–1146.
39. Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 2009; **323**(5912):373–375.
40. Davies B, Hatton E, Altemose N, Hussin JG, Pratto F, Zhang G, Hinch AG, Moralli D, Biggs D, Diaz R, Preece C, Li R et al. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* 2016; **530**(7589):171–176.
41. Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes Dev* 2016; **30**(3):266–280.
42. Carmell MA, Dokshin GA, Skaletsky H, Hu YC, von Wolfswinkel JC, Igarashi KJ, Bellott DW, Nefedov M, Reddien PW. A widely employed

germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. *Elife* 2016; **5**:e19993.
43. Rice WR. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 1984; **38**(4):735–742.
44. Zhang YE, Vibranovski MD, Landback P, Marais GA, Long M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol* 2010; **8**(10):e1000494.
45. Sneath PHA, Sokal RR. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. W. H. Freeman; 1973.
46. Teneng I, Stribinskis V, Ramos KS. Context-specific regulation of LINE-1. *Genes Cells* 2007; **12**(10):1101–1110.
47. Mendez J, Stillman B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol Cell Biol* 2000; **20**(22):8602–8612.
48. Maezawa S, Yukawa M, Alavattam KG, Barski A, Namekawa SH. Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res* 2018; **46**(2):593–608.