© 2014. Published by The Company of Biologists Ltd | Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

RESEARCH ARTICLE

STEM CELLS AND REGENERATION

Identification of *Ssm1b*, a novel modifier of DNA methylation, and its expression during mouse embryogenesis

Sarayu Ratnam[1], Peter Engler[1], Grazyna Bozek[1], Liqun Mao[1,*], Andrej Podlutsky[2,‡], Steve Austad[2,§], Terence Martin[1] and Ursula Storb[1,¶]

ABSTRACT

The strain-specific modifier Ssm1 is responsible for the strain-dependent methylation of particular *E. coli gpt*-containing transgenic sequences. Here, we identify *Ssm1* as the KRAB-zinc finger (ZF) gene 2610305D13Rik located on distal chromosome 4. *Ssm1b* is a member of a gene family with an unusual array of three ZFs. Ssm1 family members in C57BL/6 (B6) and DBA/2 (D2) mice have various amino acid changes in their ZF domain and in the linker between the KRAB and ZF domains. *Ssm1b* is expressed up to E8.5; its target transgene gains partial methylation by this stage as well. At E9.5, *Ssm1b* mRNA is no longer expressed but by then its target has become completely methylated. By contrast, in D2 embryos the transgene is essentially unmethylated. Methylation during B6 embryonic development depends on *Dnmt3b* but not *Mecp2*. In differentiating B6 embryonic stem cells methylation spreads from *gpt* to a co-integrated *neo* gene that has a similarly high CpG content as *gpt*, but *neo* alone is not methylated. In adult B6 mice, *Ssm1b* is expressed in ovaries, but in other organs only other members of the *Ssm1* family are expressed. Interestingly, the transgene becomes methylated when crossed into some, but not other, wild mice that were kept outbred in the laboratory. Thus, polymorphisms for the methylation patterns seen among laboratory inbred strains are also found in a free-living population. This may imply that mice that do not have the *Ssm1b* gene may use another member of the *Ssm1* family to control the potentially harmful expression of certain endogenous or exogenous genes.

KEY WORDS: *Ssm1*, KRAB-zinc finger genes, DNA methylation, Gene inactivation, ES cells, Mouse development

INTRODUCTION

The role of epigenetic modification in the control of gene expression has been abundantly demonstrated, including the need for silencing of most of the genome while allowing tissue-specific expression of a subset of sequences. DNA methylation is involved in gene silencing. The known essential *de novo* DNA methyltransferases do not have DNA sequence specificity, so other factors are presumed to direct the methylases to specific targets. In the case of the maintenance methyltransferase Dnmt1 hemimethylated DNA may suffice. However, for *de novo* methylation, a targeting mechanism must exist. The strain-specific modifier Ssm1 is a candidate for a novel targeting factor as it causes specific gene silencing via DNA methylation and chromatin compaction (Padjen et al., 2005).

An extensive analysis has been carried out with a target of Ssm1, the HRD transgene (supplementary material Fig. S1), which is a complex construct designed to study V(D)J recombination (Engler et al., 1991). When HRD is carried in certain inbred strains of mice (*Mus musculus*), such as C57BL/6 (B6), it is highly methylated at CpG nucleotides. HRD is, however, unmethylated in other strains, such as DBA/2 (D2) (Engler et al., 1991; Weng et al., 1995; Padjen et al., 2005). Unmethylated HRD transgenes are transcribed and undergo V(D)J recombination (Engler and Storb, 1999). When an unmethylated HRD is crossed into B6 or any one of six other methylating strains that we examined (Engler et al., 1991) it becomes methylated within one generation, leading to the conclusion that *Ssm1b* is dominant. Both HRD-methylating and non-methylating strains are spread throughout the phylogeny of inbred laboratory mice (Tsang et al., 2005), suggesting either that it represents an ancient polymorphism found in the wild ancestors of laboratory mice or that a mutation occurred very early in the history of mouse domestication which then assorted itself among inbred strains. As we show below, the former hypothesis appears to be correct. Since the B6 phenotype is dominant [(B6×D2)F1 mice methylate HRD], in this paper we use the designation *Ssm1b* when a B6 allele is homo- or heterozygous in the embryonic stem cells (ESCs)/mice under investigation and *Ssm1d* when the ESCs/mice are homozygous D2. As discussed in detail below, we do not know whether *Ssm1b* and *Ssm1d* are allelic variants or come from different loci because the region of distal chromosome 4 where *Ssm1b* resides is not available for other strains besides B6. We postulate that *Ssm1d* mice express a related *Ssm1* gene that is responsible for the suppression of related targets.

Within the original HRD transgene, we have identified a discrete segment, derived from the *gpt* gene of *E. coli*, that is the major determinant for Ssm1-mediated methylation (Engler et al., 1998). Methylation spreads into the surrounding chromosome in a strain-dependent fashion, and the methylation status is independent of the transgene integration site and transgene copy number (Engler et al., 1998), suggesting that the level of Ssm1 modifier is not limiting within this range.

A detailed analysis of transgenic embryos has shown that methylation occurs around the time of implantation, coincident with global methylation changes of endogenous loci (Weng et al., 1995). Analysis of post-implantation embryos revealed that strain-specific methylation is initiated prior to embryonic day (E) 6.5 in *Ssm1b* mice (Weng et al., 1995). A strain-independent pattern of partial methylation occurs in the trophectoderm (Weng et al., 1995). To address earlier stages, ESCs were derived from E3.5 blastocysts of *Ssm1b* and *Ssm1d* mice carrying HRD transgenes (Weng et al., 1995). Some methylation of the HRD transgene was found in undifferentiated ESCs of both mouse strains. Upon differentiation, HRD became more methylated in *Ssm1b* but less methylated in *Ssm1d* cells.

[1]Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637, USA. [2]University of Texas Health Science Center San Antonio, San Antonio, TX 78229, USA.
*Present address: Department of Medicine, University of Chicago, Chicago, IL 60637, USA. ‡Present address: University of Alaska, Fairbanks, AK 99775, USA. §Present address: Department of Biology, University of Alabama at Birmingham, AL 35294, USA.

¶Author for correspondence (stor@uchicago.edu)

Received 12 November 2013; Accepted 24 March 2014

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

The HRD transgene is in active chromatin in *Ssm1d* adults, but heterochromatic in adult *Ssm1b* mice (Padjen et al., 2005). In undifferentiated ESCs of both strains, the transgene is in a chromatin state intermediate between active and inactive. This intermediate state is still observed in both B6 and D2 ESCs 1 week after removal of leukemia inhibitory factor (LIF) and feeder fibroblasts, except that in B6 the HRD transgene becomes associated with the methylated DNA-binding protein Mecp2. After differentiation in culture, in B6 ESCs HRD is heterochromatic, whereas in D2 the HRD transgene assumes an active chromatin state (Padjen et al., 2005). HRD transgenic RNA is expressed in D2 in all stages, but in B6 only in undifferentiated ESCs and during the first 3.5 days after differentiation in culture. HRD RNA in B6 is already reduced at 7 days of differentiation, and is not expressed in E10 embryos. The first increase in HRD DNA methylation in B6 precedes the loss of expression of the HRD transgene. However, complete methylation coincides with chromatin compaction, at which time HRD transcription ceases.

The *Ssm1*/HRD system is unique in that a genetically defined modifier directs methylation to a defined target sequence. Understanding the molecular details of this process should help determine how methylation and chromatin patterns are established, and perhaps to what extent silencing is controlled as opposed to occurring by default (lack of activating signals).

Here we describe the identification of the *Ssm1* gene and the other *Ssm1* family members, their expression during development and in adult mice, the gain of HRD methylation during development, the role of Dnmt3b and Mecp2, the methylation of HRD in wild mice, and finally a model concerning the function of *Ssm1* and the *Ssm1b* versus *Ssm1d* strain dichotomy.

RESULTS

Mapping and identification of the *Ssm1* gene

Originally, *Ssm1* was localized to a ∼10 cM region on distal chromosome 4 using a BXD recombinant inbred panel (see figure 5 in Engler et al., 1991; Taylor, 1989). Map position was refined by a 500 mouse backcross (Engler and Storb, 2000) and further with a 2000 mouse intercross.

One of the backcross (bx) offspring defines the centromeric border of the *Ssm1* candidate region (Fig. 1A) (Engler and Storb, 2000). Its genome is D2 at *Nppa* (148.0 Mb on chromosome 4; NCBI Annotation Release 103) and D2 at several more centromeric loci until a single nucleotide polymorphism (SNP) at 147.4 Mb. This mouse does not methylate HRD (Me–), thus placing *Ssm1* telomeric of 147.4 Mb. All other analyses were consistent with this location (not shown).

In order to define the telomeric end of the interval containing *Ssm1*, a 2000 mouse intercross (4000 meioses) was analyzed. The informative recombinant from the intercross (ix) is shown at the top of Fig. 1A. The recombinant chromosome is B6 from 147.4 to 147.9 Mb, but D2 at *Mfn2* (147.9) and *Nppa* (148.0) as well as at distal loci. When crossed with a D2 tester HRD transgenic mouse, offspring inheriting the recombinant chromosome methylated the HRD transgene (Me+), showing that *Ssm1* is centromeric of *Nppa*. Again, all other mapping data were consistent with this assignment (not shown). Thus, *Ssm1* resides in a 0.5 Mb interval between 147.4 and 147.9 Mb on mouse chromosome 4 (Fig. 1A).

About a dozen genes are found in this interval, ranging from well characterized to only predicted. Of these, six genes were plausible candidates; all are C2-H2 zinc finger (ZF) genes containing KRAB domains N-terminal of the ZFs (Fig. 1A). One attractive model for Ssm1 action is that the ZFs bind the DNA target and the KRAB domain recruits repressive factors.

To functionally assess the KRAB-ZF (KZF) genes, BACs from a B6 library carrying at least one of the *Ssm1b* candidates were introduced into *Ssm1d* fertilized eggs carrying the HRD target. A BAC containing *Ssm1b* should cause methylation of HRD in the normally non-methylating strain. Five overlapping BACs contain all of the candidate genes (Fig. 1A). *Ssm1d* C3H/HeJ females were mated with *Ssm1d* D2 males carrying hemizygous HRD. Batches of (C3H×D2)F1 eggs were injected with one of the five B6 BACs. Three of the BACs had no effect on HRD transgene methylation but mice with either of two overlapping BACs (RP23-469B8 and RP23-282C23) showed a dramatic increase in methylation of HRD (Fig. 1B-D). These BACs share the complete KZF 'c' gene (147.6 Mb; boxed). HRD methylation was initially analyzed by Southern blots of tail DNA (Fig. 1B) and in various organs (Fig. 1C) and was further confirmed by bisulfite analysis (Fig. 1D).

Although it was possible that *Ssm1* was an miRNA or other regulatory RNA (see below), we first investigated the candidate KZF gene at 147.6 Mb that caused HRD methylation. A cDNA transgene was made in the vector pCXN2 (Fig. 2A) (Niwa et al., 1991), which is known to be functional in undifferentiated and differentiating ESCs (Alexopoulou et al., 2008). Two 147.6 cDNA transgenic lines (founders 1 and 2) showed methylation of HRD by Southern blots and bisulfite sequencing (Fig. 2B). Since some copies of HRD are unmethylated, we assume that only a subset of cells in the early embryo express the KZF transgenic cDNA; mosaic integration of transgenes has been observed frequently (Wilkie et al., 1986; Chandler et al., 2007). Some transgenic lines did not methylate HRD but we have no evidence to suggest that these lines expressed the 147.6 cDNA transgenes at the appropriate stage of development.

Further generations from founder 1 that carried the 147.6 cDNA were also analyzed and they showed almost complete HRD methylation (Fig. 2C). Thus, the presence of the 147.6 KZF gene leads to methylation of HRD and we therefore conclude that the KZF gene at position 147.6 Mb is *Ssm1*. The HRD-methylating *Ssm1* gene will henceforth be referred to as *Ssm1b*.

To confirm that the 147.6 cDNA causes HRD methylation also in ESCs, the same 147.6 cDNA transgene in the pCXN2 vector was furnished with a FLAG tag just before the STOP codon and introduced into D2 ESC lines (supplementary material Fig. S3A). Two independent D2 ESC lines carrying the 147.6-FLAG cDNA transgene showed expression of Ssm1b-FLAG (supplementary material Fig. S3B) and increased HRD methylation (by bisulfite analysis) already in undifferentiated ESCs (supplementary material Fig. S3C), with a further rise in HRD DNA methylation upon differentiation for 7 days after removal of LIF and feeder cells (supplementary material Fig. S3C). Similar to the bisulfite analyses of the *Ssm1b* transgenic mice (Fig. 2B), the HRD methylation levels varied considerably between sequences of HRD DNA clones (supplementary material Fig. S3C).

Further, we conclude that *Ssm1b* function is based on the KZF protein; the 147.6 cDNA contains no known miRNA and has no similarity with known miRNAs (miRBase, http://www.mirbase.org).

NCBI designated the *Ssm1b* gene as 2610305D13Rik and the mRNA as NM_145078. A difference found in the *Ssm1b* mRNA that we amplified in comparison to the published sequence NM_145078 is the addition of 29 nt at the end of exon 1 (supplementary material Fig. S2). These 29 nt correspond to a sequence in the *Ssm1b* gene and appear to be due to differential RNA splicing (lack of excision of one short intron). The *Ssm1b* gene has five exons, with one very large intron (∼22 kb) between exon 2 and exon 3 (Fig. 2D). In addition, there is an initiator motif (TCATTCT) and a downstream promoter element (GGTCA) (supplementary material Fig. S2), which together comprise

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

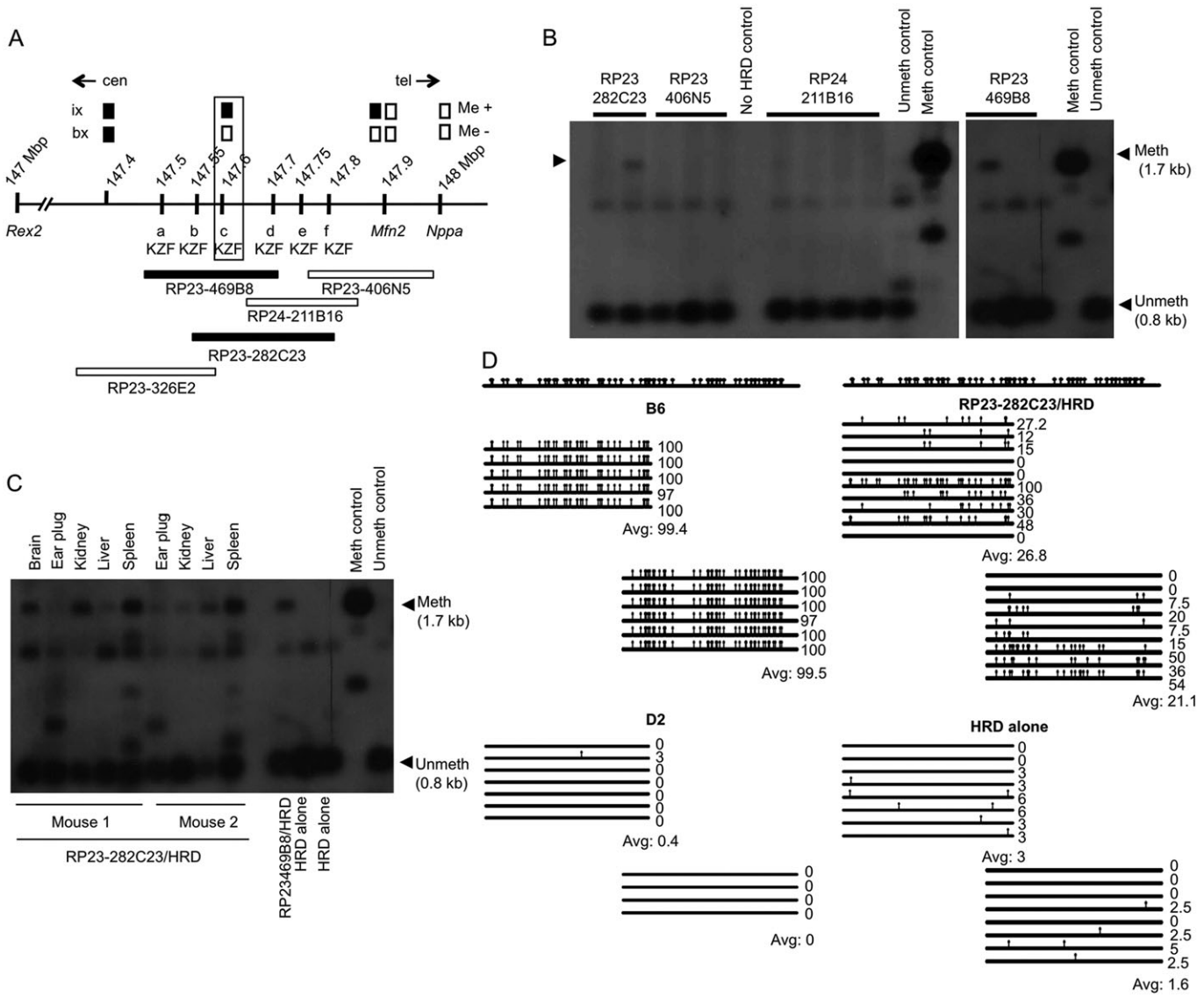

Fig. 1. Mapping and identification of the *Ssm1* gene. (A) Map of distal chromosome 4. Me+ and Me– indicate methylation present or absent of HRD in recombinants (ix, intercross; bx, backcross). Boxes indicate genotype: black, B6 genotype; white, D2 genotype. a, b, c, d, e, f are six candidate KZF genes. Position (Mb) on chromosome 4 is indicated (NCBI Annotation, Release 103); at 147.6 Mb *Ssm1b* is outlined. Beneath are five overlapping BACs. Centromeric/telomeric (cen/tel) orientation is indicated. (B) Southern blot showing HRD (see map in supplementary material Fig. S1) methylation in BAC RP23-282C23 or BAC RP23-469B8 transgenic mice. Meth, methylated; unmeth, unmethylated band. Different lanes for the same BAC represent littermates. (C) Southern blot showing HRD methylation in two mice carrying independent integrations of BAC RP23-282C23. (D) Bisulfite analysis: increased HRD/*gpt* methylation in tail DNA from mouse 1 of C. HRD alone, littermates without BAC. Each horizontal line represents the sequence from an individual bacterial colony. Vertical bars, meCpGs. The percentage of CpGs methylated is indicated. Above is shown all the CpGs in this sequence. First part of *gpt* analyzed, 530 nt; second part of *gpt* analyzed, 540 nt; overlap between the two parts of *gpt*, 162 nt.

a core promoter (Burke and Kadonaga, 1996, 1997; Yang et al., 2007). That this is the promoter region for *Ssm1b* has yet to be verified experimentally. *Ssm1b* encodes a KZF protein of 407 amino acids with one KRAB-A box (Urrutia, 2003; Vissing et al., 1995), a 218 amino acid linker, and three functional C2-H2 ZFs (Fig. 2E; supplementary material Fig. S2). A variant appears to be expressed in *Ssm1d* mice (as discussed below). *Ssm1b* and *Ssm1d* might be allelic variants (although a contig of distal chromosome 4 does not seem to exist for any of the *Ssm1d* strains) with related targets and repressive functions (see below).

Expression of *Ssm1* in ESCs

Using primers 18F and 2515R (supplementary material Table S1) a 2.5 kb product was amplified and sequenced from the cDNAs obtained from B6 and D2 ESCs (Fig. 3A). Interestingly, the B6 ESCs expressed not only *Ssm1b* mRNA, but also several related mRNAs that have SNPs resulting in altered amino acids in the ZF and linker regions, but not in the KRAB domain (Fig. 3B). Since none of the other mRNAs maps to the *Ssm1* candidate region in distal chromosome 4 (Fig. 1A), they are not involved in the methylation of HRD. One of these other mRNAs has identical ZFs to *Ssm1b* (c, Fig. 3B) but has changes in the linker. Thus, the linker might be involved in the target specificity as well.

Another surprising observation is that none of the *Ssm1b* family related genes was identified in the mouse genome using the BLAST tool from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The chromosome that contains the *Ssm1b* gene still has unsequenced gaps and it is possible that some or all of the *Ssm1*-like genes reside in these gaps.

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

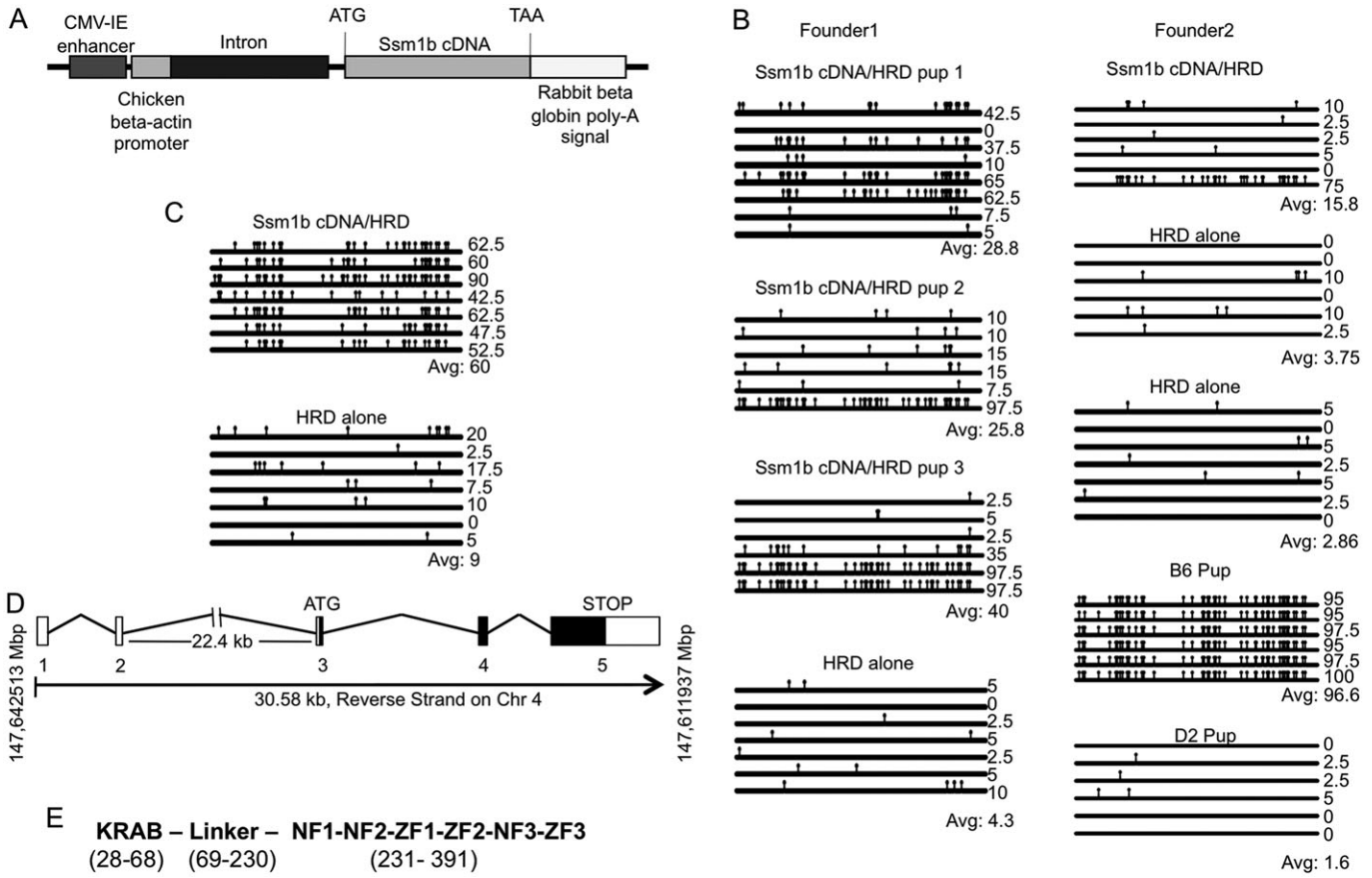

Fig. 2. HRD gains methylation upon introduction of *Ssm1b* cDNA into mice. (A) The *Ssm1b* cDNA construct that was introduced into a non-methylating mouse strain (DBA/C3H). (B) Bisulfite analysis: gain of HRD methylation in newborn pups from two independent founder mice with the *Ssm1b* cDNA transgene. B6 and D2 pups provide controls for complete and low methylation, respectively. (C) Bisulfite analysis showing gain and spread of methylation in the next generation of pups [(founder 1×DBA)F1] carrying the *Ssm1b* cDNA transgene. (D) *Ssm1b* gene structure showing exons (1-5) and introns on chromosome 4. Black box, translated sequence; white box, untranslated sequence. (E) Ssm1b protein domains (amino acid boundaries are numbered). NF1-NF3 are non-functional ZFs, whereas ZF1-ZF3 are functional ZFs.

Several *Ssm1* family mRNAs are also expressed in D2 ESCs (Fig. 3C). D2 ESCs do not, however, express *Ssm1b* mRNA. *Ssm1b* mRNA and all of the *Ssm1* family mRNAs share unusual ZFs: two C-C/H-H fingers (ZF1, ZF2) are followed by an inactive Y-C/H-H and a third functional C-C/H-H finger (ZF3) (Fig. 2E; supplementary material Fig. S2). The other five KZF genes mapping near *Ssm1b* on distal chromosome 4 (Fig. 1A) do not have the unusual ZFs; all contain larger numbers of functional ZFs. The levels of *Ssm1b* and *Ssm1* family mRNAs are one-third the levels in ESCs after 4 days differentiation of ESCs, and halved again by 7 days (supplementary material Fig. S4A).

Clearly, the *Ssm1b* sequence is not seen in *Ssm1d* ESCs among 34 sequenced *Ssm1* cDNAs, whereas in B6 six of the 23 sequences are *Ssm1b* (Fig. 3B,C). Thus, *Ssm1d* cells either do not express *Ssm1b* or, if they do, then *Ssm1b* mRNA would be, at the most, one-ninth the level in *Ssm1b*. However, transfected Ssm1b in *Ssm1d* ESCs causes higher methylation of HRD DNA (supplementary material Fig. S3). These findings suggest that it is indeed the protein sequence of Ssm1b that causes the methylation, rather than higher expression of Ssm1b.

Expression of *Ssm1* in early mouse development

Ssm1b is expressed at the blastocyst stage as shown above (ESCs are derived from the inner cell mass of blastocysts). To analyze the expression of *Ssm1b* at further stages of mouse embryonic development, mRNA was collected from E6.5, E7.5, E8.5 and E9.5 B6 embryos, and *Ssm1b* cDNA was amplified using primers 18F and 2515R (supplementary material Table S1). *Ssm1b* mRNA was expressed in E6.5, E7.5 and E8.5 (early and late) embryos in addition to the other *Ssm1*-like mRNAs. But in the E9.5 embryos only other *Ssm1* family mRNAs were found (Fig. 4). Presumably, the Ssm1b protein binds to its genomic targets before E9.5.

Ssm1 expression was also analyzed in various organs in an adult B6 mouse (Fig. 4). Certain organs, such as the heart, showed high expression levels of *Ssm1* family mRNA but none (except for the ovaries, not shown) expressed *Ssm1b*. Thus, *Ssm1b* expression seems to be tightly regulated and only active up to E8.5.

Analysis of HRD methylation during development

HRD methylation in extra-embryonic tissue (EET) had been analyzed by Southern blot of E6.5 to E12.5 post-implantation embryos (Weng et al., 1995). In contrast to the embryo proper, there was no differential methylation between D2 and (B6×D2)F1 EET (supplementary material Table S2); instead, both showed intermediate bands suggesting that *Ssm1b* might not be expressed in EET.

In the current study, embryos at different stages of development (E6.5-9.5) from (B6×D2)F1 crosses were analyzed for HRD (*gpt*) methylation. At E6.5, HRD was not completely methylated but was significantly more methylated than in later stage D2 embryos (Fig. 5; supplementary material Fig. S5). In E7.5 and E8.5 (B6×D2) F1 embryos, HRD methylation was about the same on average,

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

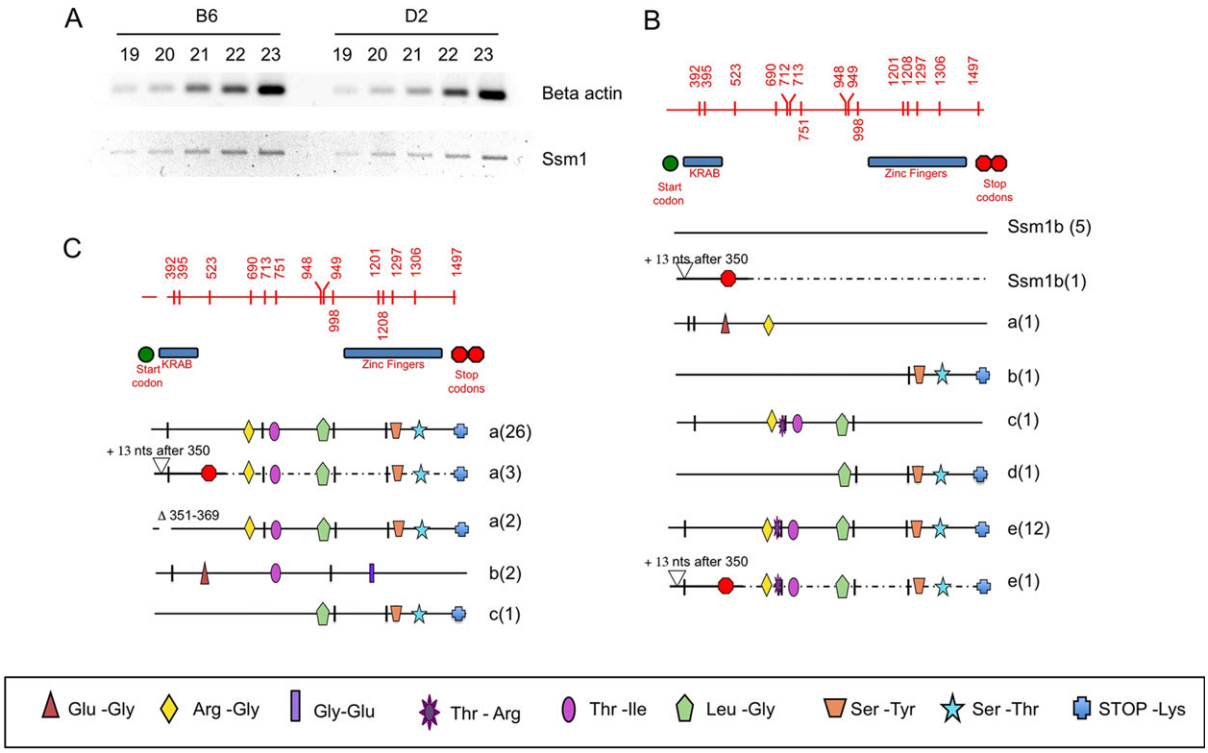

Fig. 3. Expression of *Ssm1* in B6 and D2 ESCs. (A) RT-PCR of *beta actin* and *Ssm1* in B6 and D2 undifferentiated ESCs. The number of PCR cycles is indicated. (B) *Ssm1* mRNAs in B6 ESCs. Each horizontal line represents the mRNA sequence of each member of the Ssm1 family amplified. Numbers in parentheses indicate the number of bacterial clones for each sequence. The region shown is coding sequence of *Ssm1* from nt 276 to nt 1499 (supplementary material Fig. S2). Lines labeled *Ssm1b* indicate sequences matching the genomic sequence of *Ssm1b* and hence considered the *Ssm1b* cDNA. Lines marked a, b, c, d and e represent the other *Ssm1* family sequences. Alternative splice forms of the same sequence with either addition (+13 or +39) or deletion (Δ) of nucleotides are labeled with the same letter as the original sequence. These changes lie within the translated RNA. In some splice forms the addition of 13 nt leads to a STOP codon 153 nt further 3′. Vertical marks indicate silent SNPs differing from the *Ssm1b* sequence. SNPs causing amino acid changes are marked with a symbol with the actual amino acid changes depicted beneath. Red top line displays the *Ssm1* region with vertical marks for all SNPs with nucleotide positions indicated. The positions of the KRAB (nt 357-477, supplementary material Fig. S2) and ZF (nt 966-1448) domains and the start (nt 276-278) and STOP (nt 1497-1499) codons are also depicted. (C) Expression of *Ssm1* in D2 ESCs. There is no expression of *Ssm1b* in the D2 ESCs.

although at E8.5 two sequences were almost completely methylated (supplementary material Fig. S5). In contrast to the low level of methylation in D2 at E9.5, B6 showed essentially complete methylation of HRD by E9.5 ($P<0.0002$). This steep increase of the spreading of methylation between E8.5 and E9.5 is unlikely to be due to Ssm1b protein since its mRNA is not seen at E9.5 (Fig. 4). Instead, it might be related to the switch from Dnmt3b to Dnmt3a that occurs around this stage (Watanabe et al., 2002). This would mean that spreading of DNA methylation requires *de novo* methyltransferase activity.

Association of *Ssm1b* with Dnmt3b

Because the earliest sign of HRD inactivation in *Ssm1b* mice is the increased DNA methylation in (B6×D2)F1 ESCs compared with D2 ESCs (Weng et al., 1995; Padjen et al., 2005), Ssm1b might operate by *de novo* DNA methylation. The *Ssm1b* gene product lacks a methyltransferase domain and hence it is likely that it works in concert with one of the *de novo* methyltransferases, Dnmt3a and/or Dnmt3b (Okano et al., 1999). To determine the role of Dnmt3b in Ssm1b function, the HRD transgene was crossed into a *Dnmt3b* knockout background. Multiple crosses were required to obtain E10.5 embryos that have a homozygous deletion of *Dnmt3b*, possess one *Ssm1b* allele and contain the HRD transgene that was unmethylated in the *Ssm1d* parent (supplementary material Fig. S6).

In the $Dnmt3b^{-/-}$ embryos in an *Ssm1b* background and their placentas, the HRD transgene was mostly unmethylated (by Southern blot analysis), whereas in littermates that were either $Dnmt3b^{+/+}$ or $Dnmt3b^{+/-}$ the HRD transgene was completely methylated (Fig. 6B). This result was confirmed by bisulfite analysis, where all the clones from the $Dnmt3b^{-/-}$ embryos show greatly decreased levels of methylation (Fig. 6A), suggesting that Ssm1b activity is linked to *de novo* DNA methylation. The few HRD transgene copies with higher methylation might have been caused by Dnmt3a activity (Borgel et al., 2010). Thus, HRD methylation in *Ssm1b* mice depends mainly on *de novo* DNA methylation by Dnmt3b. *Dnmt3b* mutations are frequent in the human ICF syndrome (Owen and Bowie, 1972). Since 30-40% of patients have no *Dnmt3b* mutations, a co-factor such as *Ssm1b* might be mutated.

Role of the methylated DNA-binding protein Mecp2 in Ssm1b function

Since it is possible that Ssm1 plays a role in targeting of the methyl-CpG (meCpG) DNA-binding domain (MBD) protein Mecp2 to HRD (*gpt*), HRD was crossed into either of two types of *Mecp2*-deficient mice, one with a truncation of *Mecp2* within exon 4 (Shahbazian et al., 2002) and the other with complete inactivation of *Mecp2* (Guy et al., 2001). On the *Ssm1b* background, the mice lacking Mecp2 activity still methylate HRD (supplementary material Fig. S8). This suggests that the readout of the *Ssm1b*-associated DNA methylation does not require Mecp2. It has not been determined whether Ssm1 interacts with other MBD proteins.

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

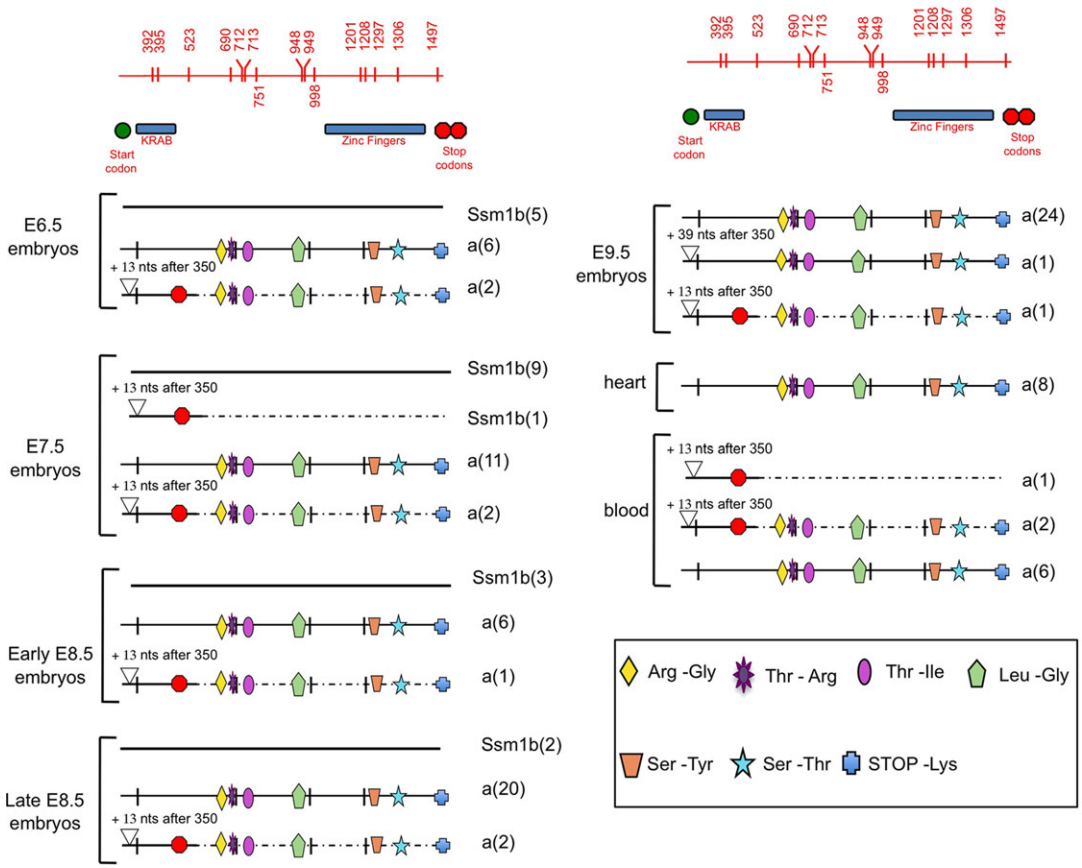

Fig. 4. Expression of *Ssm1* in early mouse development. B6 females were crossed to B6 males and embryos from various stages of development were analyzed for the expression of *Ssm1b* (see Fig. 3B legend for details). *Ssm1b* was expressed up to late E8.5 together with the other *Ssm1* family sequences. At E9.5 and in heart and blood, *Ssm1b* expression ceased while the other *Ssm1* family sequences continued to be expressed.

HRD methylation and *Ssm1b* expression in wild-derived mice

To determine whether *Ssm1b* might be an artifact of laboratory mice, we analyzed HRD methylation in mice recently descended from wild ancestors. D2 mice with an unmethylated HRD transgene were crossed with outbred wild-derived mice (Miller et al., 2002) and methylation of HRD was determined in the offspring. We found that they are a mix of HRD-methylating and non-methylating mice (Fig. 7A). Offspring of some wild mice showed complete methylation of HRD, whereas others showed no HRD methylation. Thus, polymorphisms for the methylation patterns seen among laboratory inbred strains are also found in a free-living population.

To determine whether there were any *Ssm1b* or *Ssm1*-like genes in wild mice we obtained tail DNA from the mothers of both the HRD-methylating and non-methylating wild mice and amplified the *Ssm1*

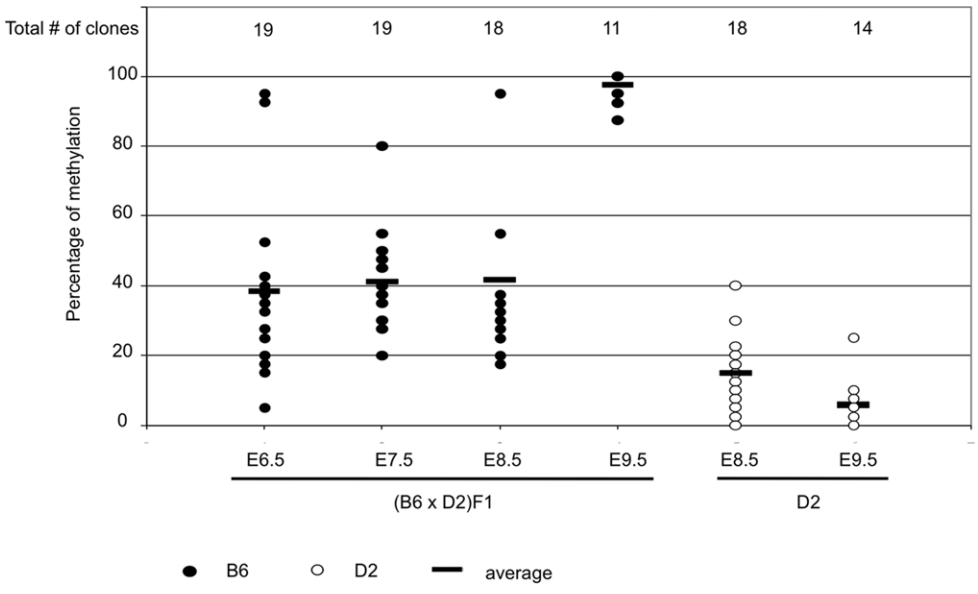

Fig. 5. HRD-*gpt* methylation during early mouse development. Gain in HRD-*gpt* methylation (bisulfite analysis) from E6.5 to E9.5 (B6×D2)F1 mouse embryos (see supplementary material Fig. S4). Circles represent percentages of CpGs methylated per *gpt* sequence from individual bacterial colonies. E8.5 and E9.5 embryos from D2 mice are shown as controls. To compare two sets of methylated sequences, a Mann-Whitney two-sample test (a non-parametric rank test) was performed on the percentages of methylated sites per clone (Mann and Whitney, 1947).

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

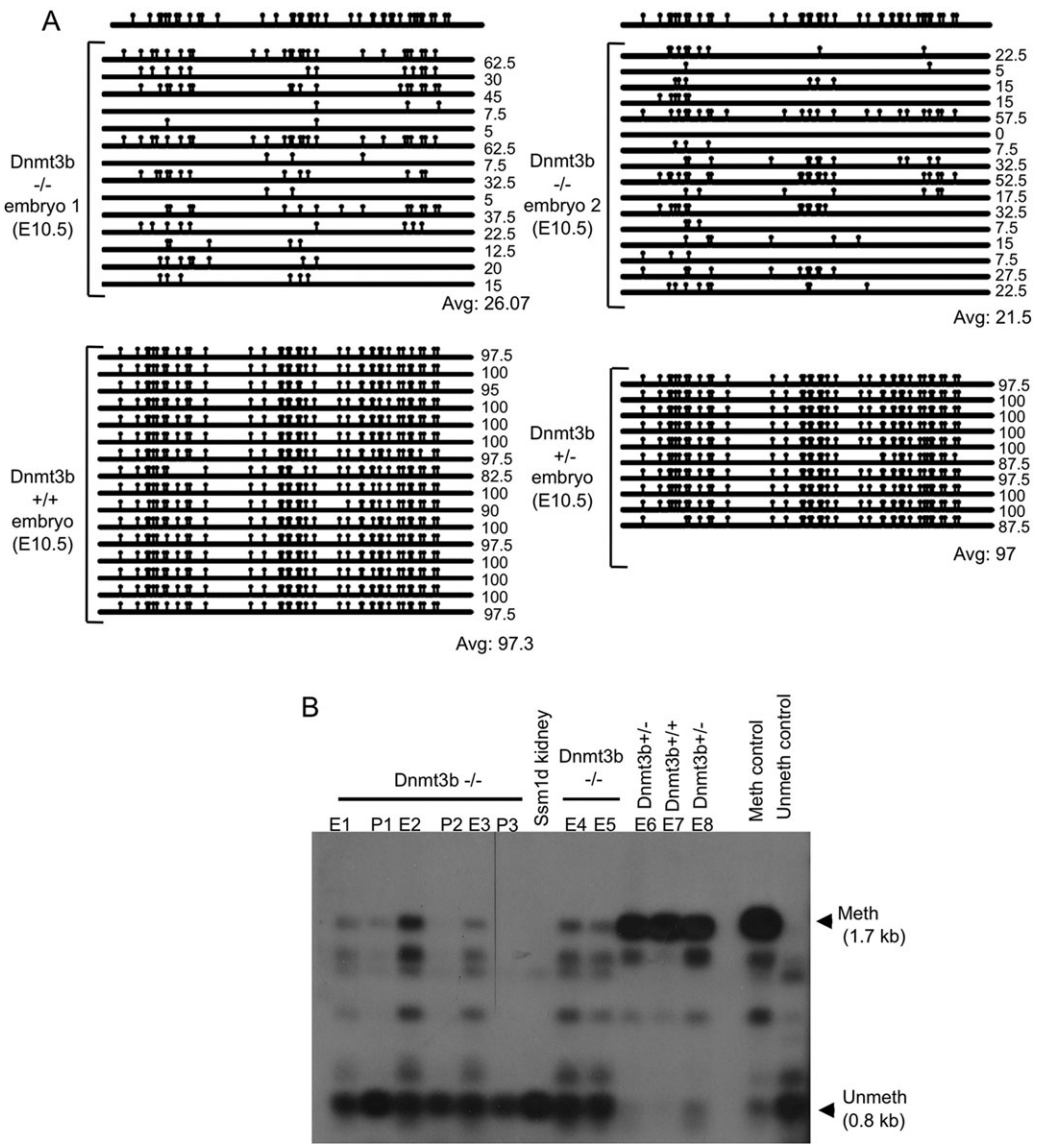

Fig. 6. Role of Dnmt3b in HRD methylation. (A) Bisulfite analysis (see Fig. 1 legend). (B) Southern blot (see Fig. 1 legend). There is low HRD methylation in B6 E10.5 embryos lacking Dnmt3b compared with Dnmt3b^{+} littermates. E1-E8: E10.5 embryos; P1-P3: E10.5 placentas.

region using primers 495F and 2515R (supplementary material Table S1), which amplify the last exon of the *Ssm1* gene family. HRD-methylating wild mice #1 and #4 have a gene that is almost identical to *Ssm1b* (Fig. 7B, 'b'), but with one amino acid change (R→H) at a non-conserved position in the last functional ZF (Thomas and Schneider, 2011) and T→I and H→Y in the linker between the KRAB and ZF domains (Fig. 7B). Non-methylating mouse #6 does not have an identical gene to *Ssm1b* but a similar one that is also present in the methylating mice. The R→H change in the ZF of Me+ mice #1 and #4 might interfere with zinc coordination by the His-His domain; therefore, it is uncertain whether this gene product is responsible for HRD methylation. This uncertainty is supported by the vastly different ratios of 'b' to 'a+c' between the two Me+ mice. These ratios are low in B6 mRNA (Fig. 3B) and genomic DNA (not shown). Possibly, the primers that amplify the *Ssm1* family genes in laboratory mice do not match the related gene in wild mice. In fact, for Me– wild mouse #2 the 495F and 2515R primers did not amplify any *Ssm1*-like gene, nor did any other B6 primer pairs in the *Ssm1b* gene (Fig. 7C). These primers were originally designed to amplify the linker, the ZFs and the 3′UTR of the *Ssm1b* gene (supplementary material Fig. S2). The forward primers, including 495F, are located in the linker; thus, even silent SNPs might have been incompatible. Therefore, whether the gene responsible for HRD methylation in wild mice is *Ssm1b* remains to be ascertained.

HRD is repressed when it is transfected into established B6 ESCs and the repression spreads to adjacent sequences

ESCs derived from HRD transgenic (B6×D2)F1 mice show partial HRD methylation that is greater than the methylation level in D2 ESCs (Padjen et al., 2005). This might indicate that *Ssm1b* begins to function at the blastocyst stage or that Ssm1b needs to interact with its target at an earlier pre-blastocyst stage. To distinguish between these possibilities an HRD-*neo* transgene (supplementary material Fig. S7C) was transfected into established B6 ESCs.

Similar to the findings with ESCs derived from HRD transgenic (B6×D2)F1 mice, the *gpt* region of the transfected HRD transgene shows partial methylation in the undifferentiated ESCs, i.e. ESCs grown on fetal fibroblasts and LIF, but gains complete methylation upon differentiation, as seen in all three clones with independent transfection of HRD-*neo* (supplementary material Fig. S7B). Thus,

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

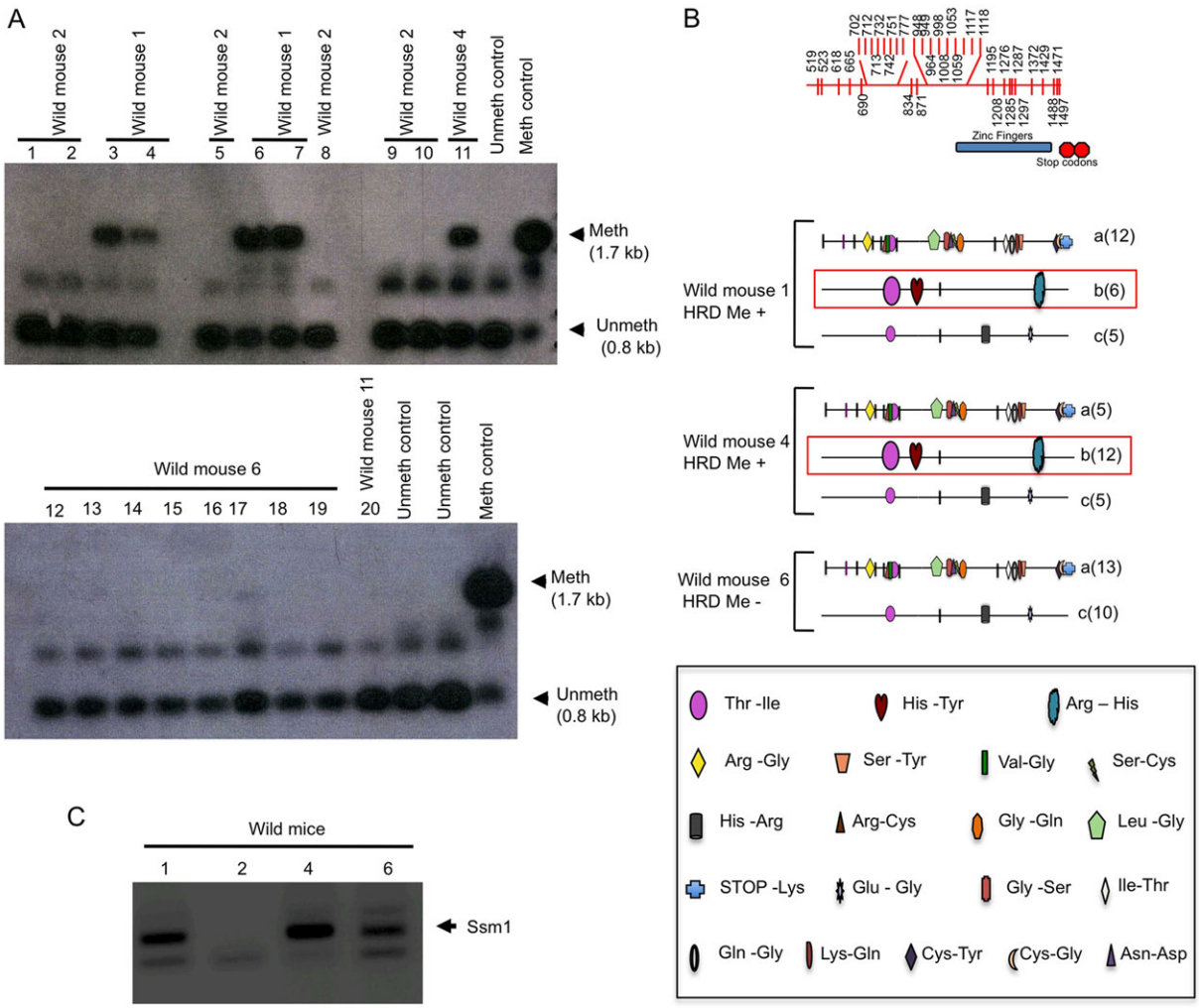

Fig. 7. An *Ssm1b*-like gene in HRD/*gpt*-methylating wild-derived mice. (A) Southern blot of liver DNA from offspring of wild female mice mated to D2-HRD males. Methylation of HRD is shown by offspring 3, 4, 6 and 7 from wild mouse 1 and offspring 11 from wild mouse 4; no methylation of HRD is apparent in offspring 1, 2, 5, 9 and 10 from wild mouse 2, offspring 12-19 from wild mouse 6, and offspring 20 from wild mouse 11. (B) Tail DNA from wild-derived mothers was analyzed for the presence of part of the *Ssm1* gene by PCR amplification with primers 495F and 2515R. The red box indicates sequence very similar to *Ssm1b* that is seen in females with HRD-methylating offspring (HRD Me+) but is absent in the female whose offspring did not methylate HRD (HRD Me–). (C) PCR amplification of *Ssm1* from tail DNA of wild-derived mice 1, 2, 4 and 6 using primers 495F and 2515R. Various primers complementary to this region were tested but were unsuccessful in amplifying *Ssm1* in wild-derived mouse 2.

the presence of *Ssm1b* at the ESC stage is sufficient for the later methylation of *gpt*.

The *neo* part of the transfected HRD-*neo* transgene is unmethylated in the undifferentiated B6 ESCs but becomes methylated upon differentiation (supplementary material Fig. S7A). By contrast, the control transgene *neo* alone (without *gpt*) stayed unmethylated even after differentiation (supplementary material Fig. S7A). One explanation for the unexpected partial methylation of the *neo* gene in the B6-*neo*-10 cell line at D28 (supplementary material Fig. S7A) could be its integration into a methylated region of the B6 genome. Thus, the complete methylation of HRD and its spreading to adjacent DNA sequences occurs during ESC differentiation and does not require the presence of the HRD target before the blastocyst/ESC stage, suggesting that *Ssm1b* begins to act in undifferentiated and early differentiating ESCs. Also, this analysis shows that HRD (and presumably its *gpt* portion) and not *neo* (which has a similarly high CpG/CpX ratio to *gpt*: 9.4% in *neo* versus 8.4% in *gpt*) is a specific target for DNA methylation by Ssm1b and it is the spreading of methylation from HRD to the adjacent *neo* gene that leads to the complete methylation of *neo*.

DISCUSSION

The *Ssm1b* gene is located on mouse distal chromosome 4 and is a member of a novel family of KZF genes. As our study shows, Ssm1b is responsible for the specific methylation of the HRD transgene in mice. The *Ssm1b* gene by itself does not encode a methyltransferase, since it lacks the conserved ‘transferase’ sequence present in all DNA methyltransferases from bacteria to man (Bestor, 2000), but Ssm1b causes HRD methylation via the *de novo* methyltransferase Dnmt3b. Since Dnmt3b does not have any sequence specificity beyond CpG dinucleotides (Okano et al., 1998; Yoder et al., 1997), the study of HRD methylation by Ssm1 might provide insights into the mechanisms that determine how specific loci are targeted for methylation and heterochromatin formation early in development.

Ssm1b belongs to a family of related KZFs genes with only three functional C2-H2 ZFs interspersed with three non-functional ones (Fig. 2E; supplementary material Fig. S2). We broadly label these unusual KZF genes as the *Ssm1* family. The *Ssm1* family is thus unique in terms of the organization of its ZFs. It has two non-functional ZFs, followed by two functional ones, a non-functional ZF,

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

and finally a third functional ZF (Fig. 2E). In addition, the *Ssm1* genes possess a fairly long and unique linker region between the KRAB and ZF domains. The role of the linker in KZF proteins is not known but it has been found that the location of KZFs in the nucleus varies depending on the linker sequence (Fleischer et al., 2006). Thus, the unique nature of the ZFs and linker is likely to be responsible for the targeting of Ssm1b protein to specific regions of the genome.

Polymorphisms in *Ssm1* genes are not an artifact of domestication as we also observed them in wild-derived mice, leading to the conclusion that this is indeed a conserved, and presumably important, gene family. This raises the question of what maintains such polymorphisms in nature. A possible answer is that certain environmental conditions favor one set of genes and that other conditions favor the alternative set. What such hypothetical conditions might be is of course unknown. However, mice, as the most widespread mammalian species on earth other than humans (Berry and Bronson, 1992), are ideally suited to address this question, but trying to answer it would require a broad census of multiple wild populations.

With regards to the expression and function of *Ssm1* family genes during development, the Ssm1b protein seems to be functional in undifferentiated ESCs. There is already more methylation of HRD/*gpt* in undifferentiated B6 ESCs than in undifferentiated D2 ESCs (Padjen et al., 2005). When HRD is transfected into established B6 ESCs (supplementary material Fig. S7B) *gpt* becomes methylated to the same percentage as in undifferentiated ESCs from HRD transgenic B6 mice (Padjen et al., 2005), suggesting that, indeed, Ssm1b most likely acts in blastocysts, if not earlier in development.

The early expression of Ssm1b in blastocysts is supported by a transcriptome analysis comparing ESCs grown in fetal calf serum (as our ESCs) and cells grown without serum in 2i medium (Marks et al., 2012). The latter are more undifferentiated and presumably more like cells of the inner cell mass. One of the genes expressed in the Marks et al. study was *Ssm1b* (2610305D13Rik) on chromosome 4. Three ESC lines grown in 2i medium had an average rpkm (reads per kb transcript per million mapped sequence reads) of 194.1, whereas the ESC lines grown in serum medium had an average rpkm of only 61.3. Thus, if these numbers are accepted as linear relationships, Ssm1b expression increases 3.2-fold in ESCs grown in 2i medium compared with the cells grown in serum.

If cells grown in 2i medium are indeed more undifferentiated, the finding that Ssm1b expression is higher in these cells is in line with our observation that Ssm1b expression is reduced upon differentiation: Ssm1b expression drops by almost half upon differentiating the ESCs for 4 days (supplementary material Fig. S4B), as compared with 'undifferentiated' ESCs grown with LIF on feeder cells (Fig. 3B). Thus, our data and the data from Marks et al. confirm that Ssm1b is expressed in the inner cell mass cells during mouse development. Identifying a possibly earlier stage of Ssm1b expression and function during pre-implantation embryonic development is an important goal that we are addressing in a separate study.

At the undifferentiated stage, both *Ssm1b* and *Ssm1d* ESCs show a biphasic pattern of active and inactive chromatin (Padjen et al., 2005). This agrees with recent reports that multiple genes in ESCs (Giadrossi et al., 2007; Bernstein et al., 2007), as well as stem cells in general (Azuara et al., 2006), have a mixed chromatin pattern. These findings were suggested to indicate that genes in stem cells are poised for a choice of either activation or inactivation. Thus, HRD behaves like an endogenous gene. However, we cannot rule out the possibility that the 'biphasic' pattern might be the result of heterogeneity in the cell population. An unusual observation about the *Ssm1* system is that, once differentiation is fully underway, different mouse strains treat HRD chromatin in opposite ways. Divergence between the strains appears to be initiated ~7 days after differentiation of ESCs in culture: at this stage, the only sign of the initiation of HRD inactivation in *Ssm1b* cells is a significant increase in Mecp2 binding to the *gpt* core of HRD and a slight reduction of HRD mRNA (Padjen et al., 2005). In the corresponding *Ssm1b* E7.5 embryos, *Ssm1b* is still expressed, but expression ceases after E8.5 and HRD has become completely methylated at E9.5. Therefore, *Ssm1b* must have initiated methylation at a stage earlier than E9.5, and this methylation then spread across the transgene during further development.

The active demethylation of meCpGs has recently been attributed to the TET proteins that convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), and further to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC), leading to replacement of meCpG by cytosine and thus DNA demethylation (reviewed by Bhutani et al., 2011). It appears that tissue-specific genes become actively demethylated in preparation for their activation. By contrast, the 5meC form of HRD/*gpt* induced during early embryonic development through *Ssm1b* does not undergo these dynamic changes; instead, HRD DNA remains methylated in fetal and adult life. This finding lends support to our preliminary ChIP-Seq analysis of endogenous Ssm1b target genes (not shown), which has revealed repetitive DNA sequences, such as SINEs, as the potential targets of *Ssm1b*. If unique tissue-specific genes were the targets, in an adult tissue such as liver in which over 50% of the cells are hepatocytes that express the metallothionein (MT) gene, which is, like the HRD gene, driven by the MT promoter, a significant proportion of the *gpt* sequences should be unmethylated. Although the bisulfite assays that we performed would not distinguish 5hmC from 5mC, it is unlikely that HRD would remain 100% methylated if it underwent the first step of demethylation. In fact, in mice that were treated with cadmium to induce MT expression, the HRD transgene was only expressed in *Ssm1d* mice, in which its DNA was not methylated (Weng et al., 1995).

So far, preliminary ChIP-Seq data (unpublished observations) seem to indicate that the endogenous targets of *Ssm1b* are repetitive DNA sequences. The repression of another group of repetitive sequences, intracisternal A particles (IAPs), was found to be due to a different KZF gene (Rowe et al., 2010). We propose that Ssm1 proteins are required for the repression early in development of repetitive DNA sequences, the expression of which would swamp the function of unique genes (Li et al., 1992). Many repetitive elements are suppressed by DNA methylation (Walsh et al., 1998). It is thus likely that the multiple variant mRNAs of the *Ssm1* family that we have found (and perhaps extended family members that were not RT-PCR amplified using primers specific for *Ssm1b*) are specific for different, but related, repetitive sequences. The Ssm1 family proteins have many amino acid polymorphisms in the linker and the ZF region that could be responsible for differential target specificities. For instance, it appears that *gpt* is a target only for Ssm1b. Presumably, the Ssm1b protein binds to a specific motif in *gpt* via its functional ZFs, whereas other Ssm1 family proteins do not bind *gpt* owing to amino acid changes in their ZFs that prevent their recognition of *gpt* as a target. Although *gpt* is not a target of Ssm1-like proteins, presumably various interrelated repetitive sequences will be targets. The *Ssm1* gene family might be just one of several families of multi-loci genes that are responsible for the essential repression of different repetitive genomic elements. One reason why a repressor such as Ssm1 was not discovered earlier might be that the target recognition sequences are short and, because of the redundancy of the system, were not found. We do not

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

know whether, in analogy with Zfp57 (Takikawa et al., 2013), Ssm1b targets imprinted genes.

The finding of *gpt* as a discrete target for Ssm1b, together with the finding that *neo* with a similarly high density of CpGs is not a target, makes it unlikely that Ssm1 recognizes all foreign, potentially parasitic DNA sequences. There are no obvious repeats in *gpt* as potential methylation targets (Reinhart et al., 2002). Thus, it is possible that Ssm1b recognizes a specific sequence within *gpt* that it binds to and initiates methylation, and endogenous targets might share some common features with *gpt*.

All *Ssm1*-like genes that we have found encode the identical KRAB protein domain. KRAB is a repressive domain that interacts with KRAB-associated protein (KAP1; also known as Trim28). Ssm1b may bind to its targets and recruit methyltransferases either directly or indirectly by interaction through the KRAB domain with proteins that are, or bind to, *de novo* methyltransferases. Spreading of methylation beyond the target sequences might be due to propagation of chromatin structure (Groner et al., 2010). Our previous experiments showed that DNA methylation of HRD precedes certain DNA chromatin modifications (Padjen et al., 2005). This raises the question of whether Ssm1b protein induces DNA methylation by interacting directly with Dnmt3b and/or other *de novo* DNA methyltransferases and steering them to potential endogenous targets, or whether it interacts with MBD proteins or other suppressive factors to direct them to partly methylated or unmethylated DNA. As we have shown, *Ssm1*-associated DNA methylation does not require Mecp2, although it is still possible that Ssm1b interacts with other MBD proteins. It is also possible that Ssm1b recruits a chromatin-modifying factor that we have not yet examined.

Thus, *Ssm1b* is a novel gene that points to a new family of KZF genes initiating specific DNA methylation and chromatin modification. The *Ssm1* family is likely to be involved in the repression of repetitive DNA sequences in the epigenetic control of early development, and discovering the details of its function provides an exciting challenge.

MATERIALS AND METHODS

Transgenic mice

Experiments with mice were performed in compliance with USDA animal welfare and PHS humane animal care guidelines.

Backcrosses and intercrosses to map *Ssm1*

The 500 mouse backcross was described by Engler and Storb (Engler and Storb, 2000). For intercrosses, (D2×B6)F1 mice were mated with each other and 2000 progeny were analyzed for HRD methylation.

ESCs

ESCs from HRD transgenic mice of both (B6×D2)F1 and D2 strains were isolated by A. Weng (Weng et al., 1995). B6 ESCs were obtained from the University of Chicago Transgenic Mouse Facility, D2 ESCs from Teruhiko Wakayama (Wakayama et al., 2001) and mouse embryonic fibroblasts (DR4) from Stanford Transgenic Research Center. The HRD-*neo* construct consists of the *Eco*RI-*Hin*dIII HRD transgene (Engler and Storb, 1987) linked to *pko-neo* (Sambrook et al., 1989). The *neo* gene alone was obtained from the *pMC1 Neo-PolyA* vector (Agilent Technologies).

BACs and cDNA transgenes

Bacterial artificial chromosomes (BACs) were obtained from Children's Hospital Oakland Research Institute (CHORI). *Ssm1b* cDNA transgenes were prepared from *Ssm1b* mRNA from undifferentiated B6 ESCs (Fig. 3) and cloned into the cDNA expression vector pCXN2 (Niwa et al., 1991). All BAC and *Ssm1b* cDNA transgenic mice were made at the University of Chicago Transgenic Mouse Facility. For ESCs, the FLAG tag was introduced immediately before the STOP codon of the *Ssm1b* cDNA (supplementary material Fig. S3A), cloned into pCXN2 and transfected into the D2 ESCs (Tompers and Labosky, 2004).

RT-PCR for *Ssm1* transcripts

RNA was isolated using RNA STAT60 (CS-110, AMSBIO) or the RNAqueous-4PCR kit (AM1914, Ambion). cDNA was made using the Superscript III first-strand synthesis system (18080-051, Invitrogen). *Ssm1b* cDNA was amplified using primers 18F and 2515R (supplementary material Table S1), gel-purified using the GeneJet gel extraction kit (K0691, Fermentas) and cloned into a PCR cloning vector (240205, Stratagene). Plasmid DNA was isolated using the GeneJet plasmid mini prep kit (K0502, Fermentas).

PCR for analyzing the *Ssm1* gene in wild mice

DNA was isolated from tails of outbred descendants of wild mice trapped in Idaho (Miller et al., 2002). The *Ssm1* region was amplified using primers 495F and 2515R (supplementary material Table S1). Female wild mice were mated with HRD transgenic male D2 mice. Liver DNA of the offspring was analyzed for HRD methylation by Southern blot (Padjen et al., 2005).

Bisulfite sequencing

Bisulfite sequencing of *gpt* and *neo* was carried out as described previously (Padjen et al., 2005).

Dnmt3b-deficient mice

Dnmt3b$^{+/-}$ mice on the B6 background were obtained from Novartis (Ueda et al., 2006). The breeding scheme is detailed in supplementary material Fig. S6.

Mecp2-deficient mice

Two lines of *Mecp2*-deficient mice were obtained from Jackson Laboratory: B6.129P2(C)-Mecp2$^{tm1.Bird}$/J (Guy et al., 2001) and B6.129S-Mecp2^{itm1Hzo}/J (Shahbazian et al., 2002). HRD transgenic male offspring lacking *Mecp2* provided the necessary experimental combination (*Ssm1b/d*, HRD, *Mecp2*$^{-/-}$).

Acknowledgements

We thank L. Degenstein of the University of Chicago Transgenic Mouse Facility for production of transgenic mice and valuable discussions, D. Nicolae for statistical analysis, I. Swanson for *Ssm1* PCR assays of wild mice, L. Godley for *Dnmt3b* primer design, R. Chaillet and L. Doglio for thoughtful comments, W. Buikema of the University of Chicago DNA Sequencing Facility for DNA sequencing, T. Wakayama for D2 ESCs and J. Miyazaki for the pCXN2 vector.

Competing interests

The authors declare no competing financial interests.

Author contributions

S.R. supervised the production and analysis of the BAC and *Ssm1b* cDNA transgenic mice and the *Dnmt3b*$^{-/-}$ embryos, co-designed with U.S. and carried out most of the experiments reported in this paper, and wrote the paper together with U.S. P.E. designed and produced the HRD transgene, made HRD transgenic mice and discovered the strain-specific methylation, designed and carried out the mapping of *Ssm1b*, created variants of HRD and discovered that *gpt* is the Ssm1b target. G.B. performed all the mouse breedings and assisted in all the molecular biology experiments. L.M. carried out the experiments in supplementary material Fig. S4A, checked the BAC sequences conferring the Ssm1 effect for the presence of miRNAs, and cloned and sequenced *Ssm1* cDNA that was then used to make transgenic mice. A.P. crossed HRD transgenic DBA males with female wild-derived mice. S.A. established the colony of wild-derived mice and kept them outbred. T.M. provided ideas and discussion throughout this study and edited the paper. U.S. directed the study and wrote the paper together with S.R.

Funding

The studies were supported by National Institutes of Health (NIH) grants [RO3HD05827 and AI047380] and a Pilot Project grant from the University of Chicago Cancer Research Center to U.S.; support for the S.A. laboratory was by an NIH Nathan Shock Center grant [P30 AG13319]. S.R. was supported by a postdoctoral fellowship from the Cancer Research Institute. Deposited in PMC for release after 12 months.

Supplementary material

Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.105726/-/DC1

Development (2014) 141, 2024-2034 doi:10.1242/dev.105726

References

Alexopoulou, A. N., Couchman, J. R. and Whiteford, J. R. (2008). The CMV early enhancer/chicken beta actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors. *BMC Cell Biol.* **9**, 2.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M. et al. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532-538.

Bernstein, B. E., Meissner, A. and Lander, E. S. (2007). The mammalian epigenome. *Cell* **128**, 669-681.

Berry, R. J. and Bronson, F. H. (1992). Life history and bioeconomy of the house mouse. *Biol. Rev. Camb. Philos. Soc.* **67**, 519-550.

Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9**, 2395-2402.

Bhutani, N., Burns, D. M. and Blau, H. M. (2011). DNA demethylation dynamics. *Cell* **146**, 866-872.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T. and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **42**, 1093-1100.

Burke, T. W. and Kadonaga, J. T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10**, 711-724.

Burke, T. W. and Kadonaga, J. T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes Dev.* **11**, 3020-3031.

Chandler, K. J., Chandler, R. L., Broeckelmann, E. M., Hou, Y., Southard-Smith, E. M. and Mortlock, D. P. (2007). Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression. *Mamm. Genome* **18**, 693-708.

Engler, P. and Storb, U. (1987). High-frequency deletional rearrangement of immunoglobulin kappa gene segments introduced into a pre-B-cell line. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4949-4953.

Engler, P. and Storb, U. (1999). Hypomethylation is necessary but not sufficient for V(D)J recombination within a transgenic substrate. *Mol. Immunol.* **36**, 1169-1173.

Engler, P. and Storb, U. (2000). A linkage map of distal mouse chromosome 4 in the vicinity of Ssm1, a strain-specific modifier of methylation. *Mamm. Genome* **11**, 694-695.

Engler, P., Haasch, D., Pinkert, C. A., Doglio, L., Glymour, M., Brinster, R. and Storb, U. (1991). A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. *Cell* **65**, 939-947.

Engler, P., Doglio, L. T., Bozek, G. and Storb, U. (1998). A cis-acting element that directs the activity of the murine methylation modifier locus Ssm1. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10763-10768.

Fleischer, S., Wiemann, S., Will, H. and Hofmann, T. G. (2006). PML-associated repressor of transcription (PAROT), a novel KRAB-zinc finger repressor, is regulated through association with PML nuclear bodies. *Exp. Cell Res.* **312**, 901-912.

Giadrossi, S., Dvorkina, M. and Fisher, A. G. (2007). Chromatin organization and differentiation in embryonic stem cell models. *Curr. Opin. Genet. Dev.* **17**, 132-138.

Groner, A. C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P. and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet.* **6**, e1000869.

Guy, J., Hendrich, B., Holmes, M., Martin, J. E. and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27**, 322-326.

Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18**, 1-164.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A. F., Smith, A. et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590-604.

Miller, R. A., Harper, J. M., Dysko, R. C., Durkee, S. J. and Austad, S. N. (2002). Longer life spans and delayed maturation in wild-derived mice. *Exp. Biol. Med.* **227**, 500-508.

Niwa, H., Yamamura, K. and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193-199.

Okano, M., Xie, S. and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219-220.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257.

Owen, C. A., Jr and Bowie, E. J. (1972). Intravascular coagulation-fibrinolysis (ICF) syndrome. *Ann. Clin. Lab. Sci.* **2**, 26-30.

Padjen, K., Ratnam, S. and Storb, U. (2005). DNA methylation precedes chromatin modifications under the influence of the strain-specific modifier Ssm1. *Mol. Cell. Biol.* **25**, 4782-4791.

Reinhart, B., Eljanne, M. and Chaillet, J. R. (2002). Shared role for differentially methylated domains of imprinted genes. *Mol. Cell. Biol.* **22**, 2089-2098.

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J. et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Shahbazian, M. D., Young, J. I., Yuva-Paylor, L. A., Spencer, C. M., Antalffy, B. A., Noebels, J. L., Armstrong, D. L., Paylor, R. and Zoghbi, H. Y. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* **35**, 243-254.

Takikawa, S., Wang, X., Ray, C., Vakulenko, M., Bell, F. T. and Li, X. (2013). Human and mouse ZFP57 proteins are functionally interchangeable in maintaining genomic imprinting at multiple imprinted regions in mouse ES cells. *Epigenetics* **8**, 1268-1279.

Taylor, B. A. (1989). Recombinant inbred strains. In *Genetic Variation in the Laboratory Mouse* (ed. M. L. Lyon and A. G. Searle), pp. 773-796. Oxford: Oxford University Press.

Thomas, J. H. and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. *Genome Res.* **21**, 1800-1812.

Tompers, D. M. and Labosky, P. A. (2004). Electroporation of murine embryonic stem cells: a step-by-step guide. *Stem Cells* **22**, 243-249.

Tsang, S., Sun, Z., Luke, B., Stewart, C., Lum, N., Gregory, M., Wu, X., Subleski, M., Jenkins, N. A., Copeland, N. G. et al. (2005). A comprehensive SNP-based genetic analysis of inbred mouse strains. *Mamm. Genome* **16**, 476-480.

Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K. and Li, E. (2006). Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* **133**, 1183-1192.

Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. *Genome Biol.* **4**, 231.

Vissing, H., Meyer, W. K.-H., Aagaard, L., Tommerup, N. and Thiesen, H.-J. (1995). Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Lett.* **369**, 153-157.

Wakayama, T., Tabar, V., Rodriguez, I., Perry, A. C. F., Studer, L. and Mombaerts, P. (2001). Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer. *Science* **292**, 740-743.

Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116-117.

Watanabe, D., Suetake, I., Tada, T. and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech. Dev.* **118**, 187-190.

Weng, A., Magnuson, T. and Storb, U. (1995). Strain-specific transgene methylation occurs early in mouse development and can be recapitulated in embryonic stem cells. *Development* **121**, 2853-2859.

Wilkie, T. M., Brinster, R. L. and Palmiter, R. D. (1986). Germline and somatic mosaicism in transgenic mice. *Dev. Biol.* **118**, 9-18.

Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. and Martinez, E. (2007). Prevalence of the Initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**, 52-65.

Yoder, J. A., Soman, N. S., Verdine, G. L. and Bestor, T. H. (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues: studies with a mechanism-based probe. *J. Mol. Biol.* **270**, 385-395.