HOKKAIDO UNIVERSITY

Title	A Conserved Noncoding Sequence Can Function as a Spermatocyte-Specific Enhancer and a Bidirectional Promoter for a Ubiquitously Expressed Gene and a Testis-Specific Long Noncoding RNA
Author(s)	Kurihara, Misuzu; Shiraishi, Akira; Satake, Honoo; Kimura, Atsushi P.
Citation	Journal of Molecular Biology, 426(17), 3069-3093 https://doi.org/10.1016/j.jmb.2014.06.018
Issue Date	2014-08-26
Doc URL	http://hdl.handle.net/2115/57322
Rights	(C) 2014 Elsevier Ltd. All rights reserved.
Type	article (author version)
File Information	JMB_426_3069-.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

A conserved noncoding sequence can function as a spermatocyte-specific enhancer and a bidirectional promoter for a ubiquitously expressed gene and a testis-specific long noncoding RNA

Misuzu Kurihara[1], Akira Shiraishi[2], Honoo Satake[2], and Atsushi P. Kimura[1,3,*]

[1]Graduate school of Life Science, Hokkaido University, Sapporo 060-0810, Japan

[2]Suntory Foundation for Life Sciences, Bioorganic Research Institute, Osaka 618-8503, Japan

[3]Department of Biological Sciences, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan

Running title: A dual promoter–enhancer in mouse spermatocytes

*Corresponding author: Atsushi P. Kimura, Department of Biological Sciences, Faculty of Science, Hokkaido. University, Sapporo 060-0810, Japan. Phone and Fax: 81-11-706-4452. E-mail: akimura@sci.hokudai.ac.jp

Abstract

Tissue-specific gene expression is tightly regulated by various elements such as promoters, enhancers, and long noncoding RNAs (lncRNAs). In the present study, we identified a conserved noncoding sequence (CNS1) as a novel enhancer for the spermatocyte-specific mouse *testicular cell adhesion molecule 1* (*Tcam1*) gene. CNS1 was located 3.4 kb upstream of the *Tcam1* gene and associated with histone H3K4 mono-methylation in testicular germ cells. By the *in vitro* reporter gene assay, CNS1 could enhance *Tcam1* promoter activity only in GC-2spd(ts) cells, which were derived from mouse spermatocytes. When we integrated the 6.9-kb 5'-flanking sequence of *Tcam1* with or without a deletion of CNS1 linked to the enhanced green fluorescent protein gene into the chromatin of GC-2spd(ts) cells, CNS1 significantly enhanced *Tcam1* promoter activity. These results indicate that CNS1 could function as a spermatocyte-specific enhancer. Interestingly, CNS1 also showed high bidirectional promoter activity in the reporter assay, and consistent with this, the *Smarcd2* gene and lncRNA, designated *lncRNA-Tcam1*, were transcribed from adjacent regions of CNS1. While *Smarcd2* was ubiquitously expressed, *lncRNA-Tcam1* expression was restricted to testicular germ cells, although this lncRNA did not participate in *Tcam1* activation. Ubiquitous *Smarcd2* expression was correlated to CpG hypo-methylation of CNS1 and partially controlled by Sp1. However, for *lncRNA-Tcam1* transcription, the strong association with histone acetylation and histone H3K4 tri-methylation also appeared to be required. The present data suggest that CNS1 is a spermatocyte-specific enhancer for the *Tcam1* gene and a bidirectional promoter of *Smarcd2* and *lncRNA-Tcam1*.

Keywords: *Tcam1*; GC-2spd(ts); enhancer; long noncoding RNA; conserved noncoding sequence

Abbreviations used: Aip, aryl-hydrocarbon receptor-interacting protein; ANOVA, analysis of variance; CGI, CpG island; ChIP, chromatin immunoprecipitation; ChIP-seq, chromatin immunoprecipitation sequencing; CNS, conserved noncoding sequence; DMEM, Dulbecco's modified Eagle's medium; EGFP, enhanced green fluorescent protein; ENCODE, Encyclopedia of DNA elements; eRNA, enhancer RNA; hGH, human growth hormone; HS, hypersensitive site; LCR, locus control region; lncRNA, long noncoding RNA; PIC, pre-initiation complex; RACE, rapid amplification of cDNA ends; Smarcd2, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2; Tcam1, testicular cell adhesion molecule 1; TSS, transcriptional start site

Introduction

Tissue-specific gene activation is controlled by complicated mechanisms that involve the activity of promoters and enhancers, epigenetic modification, and noncoding transcription.[1–3] In mammals, many tissue-specific genes require distal enhancers as well as a proximal promoter for full activation.[4,5] The enhancer is a sequence to which transcription factors bind for increasing the rate of target gene transcription.[4,5] Recent studies have demonstrated that enhancers physically interact with the target gene promoter by looping out the intervening sequences and are associated with many transcription factors.[4,5] Genome-wide analyses have also revealed that a large number of enhancers are occupied by RNA polymerase II,[6–11] and consistent with this, many enhancers are actually transcribed into long noncoding RNAs (lncRNAs) that are often essential for target gene activation.[11,12] However, the relationship between lncRNAs and enhancers seems to be diverse and is not completely understood.[11–13]

To understand the regulatory mechanism for tissue-specific gene activation, various genes have been investigated as model genes.[14–20] The human growth hormone (*hGH*) gene cluster is one such model. This cluster is located on chromosome 17q22-24 and encompasses five paralogous growth hormone genes. Although the primary structure of the five genes is well conserved, *hGH-N* is specifically expressed in the pituitary and the other four genes are placenta-specific.[21,22] The tissue-specific activation of the *hGH* cluster is dependent on the 5'-distal locus control region (LCR),[23,24] and epigenetic regulation and noncoding transcription are known to play crucial roles in activation by the *hGH* LCR.[25–29] Interestingly, the *hGH* cluster is linked to two other tissue-specific genes: the B cell-specific *CD79b* gene, which is located between the cluster and LCR, and the testis-specific testicular cell adhesion molecule (*TCAM1P*) gene in the 3'-region of the cluster. Therefore, this locus is an excellent model for tissue-specific gene activation; however, the *TCAM1P* gene regulation has not been studied in detail.

TCAM1P is a highly conserved gene among placental mammals such as the cow, rat, mouse, and rhesus monkey. Although the human *TCAM1P* gene does not seem to encode a protein, the orthologous

gene in other species is translated to a protein related to cell adhesion.[30] The testis-specific expression of this gene has been confirmed in the rat and mouse,[30,31] and the mouse *Tcam1* gene has been found to be expressed in the 17-day-old testis, when germ cell meiosis reaches the late pachytene spermatocyte stage.[32] Consistent with this, mouse *Tcam1* mRNA has been found to be localized in such spermatocytes by *in situ* hybridization.[30] With regard to the regulation of this gene, a DNase I hypersensitive site (HS) has been detected in the rat *Tcam1* promoter, and a high level of histone acetylation has been observed at DNase I HS in rat plasmacytoma-derived Y3-Ag1.2.3 cells.[33,34] However, no *cis*-elements have been identified for *Tcam1* regulation and the regulatory mechanism in native testicular germ cells has not been investigated.

In the present study, we focused on conserved noncoding sequences (CNSs) to examine the regulation of the mouse *Tcam1* gene. There were six CNSs at the *Tcam1* locus, among which CNS1 was identified as a potential spermatocyte-specific enhancer. Interestingly, CNS1 also contained bidirectional promoter activity, and the *SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2* (*Smarcd2*) gene and a novel testis-specific lncRNA (designated *lncRNA-Tcam1*) were actually expressed from the upstream and downstream of CNS1, respectively. The results indicated that CNS1 may work as a spermatocyte-specific enhancer for the *Tcam1* gene and a bidirectional promoter of the ubiquitously expressed *Smarcd2* gene and the testicular germ cell-specific *lncRNA-Tcam1*. This is the first indication of a dual promoter–enhancer in mammals.

Results

Tcam1 is a spermatocyte-specific gene

The mouse *Tcam1* gene has been reported to be specifically expressed in the testis, particularly in germ cells at stages from pachytene spermatocytes to secondary spermatocytes.[30] We first attempted to

confirm this expression pattern of *Tcam1* mRNA. By northern blot analysis using total RNAs from 15 mouse tissues, we detected a specific signal for *Tcam1* at the 3.0-kb position only in the testis (Fig. 1a). This was also confirmed by the transcriptomic data (GSE9954), which demonstrated testis-specific expression of *Tcam1* mRNA among 22 mouse tissues (Fig. 1b). We then examined *Tcam1* mRNA expression at different developmental stages of the mouse testis (7, 14, 21, 28, and 56 days after birth). *Tcam1* mRNA was detected as early as 14 days after birth, when pachytene spermatocytes appear for the first time during mouse development,[35,36] and the signal intensity became stronger in 21 and 28-day-old testes (Fig. 1c). Expression was slightly decreased 56 days after birth.

To further confirm the localization of *Tcam1* mRNA, we performed *in situ* hybridization with testes at 21 and 28 days after birth. We used these stages of testes because *Tcam1* mRNA was expressed at the highest level at these stages in our northern blot analysis (Fig. 1c). The results indicated that *Tcam1* mRNA was localized in spermatocytes in all seminiferous tubules (Fig. 1d, e). Hybridization with a sense probe resulted in no positive signals (data not shown). Taken together, these findings confirm the spermatocyte-specific expression of the mouse *Tcam1* gene.

CNS1, CNS2, and CNS3 are candidate regulatory sequences for *Tcam1* gene activation by histone modification patterns

We next searched for candidate sequences for *Tcam1* regulation. We attempted to identify CNSs at the mouse *Tcam1* locus because important gene regulatory sequences are generally well conserved beyond species.[37] Using the rVista program (http://genome.lbl.gov/vista/index.shtml), we compared a 31-kb sequence of the mouse genome that contained 12 kb of the *Tcam1* gene and 4 kb and 15 kb of its 5’ and 3’ adjacent regions, respectively, with the corresponding human genomic sequence. We found two CNSs upstream of the *Tcam1* gene and four CNSs within the gene body (Fig. 2a). No CNS was present in the 3’-region of the *Tcam1* gene. We then assessed which CNSs could actually be involved in spermatocyte-specific *Tcam1* expression by determining the histone modification patterns. For this

purpose, we investigated histone H3K9 acetylation (H3K9ac), histone H3K4 mono-methylation (H3K4me1), and histone H3K4 tri-methylation (H3K4me3). Histone acetylation is consistently associated with open chromatin, leading to gene activation, and H3K4me1 and H3K4me3 are generally observed at the enhancer and promoter, respectively.[38,39] We conducted chromatin immunoprecipitation (ChIP) assays with specific antibodies for these epigenetic markers using nuclei from testicular germ cells and liver cells. Germ cells were isolated from adult testes as described previously,[40,41] and this fraction was presumed to contain approximately 30% of spermatocytes, as judged from our observations (R. Yoneda, M.K., and A.P.K., unpublished observations). The liver was investigated as a tissue that did not express the *Tcam1* gene.

We first prepared ten amplicons at the *Tcam1* locus: six CNSs, intron 1 of *Smarcd2* (Smarcd2 intron1), a region between CNS2 and *Tcam1* (CNS2-Tcam1), *Tcam1* promoter, and a region between CNS3 and CNS4 (genebody). Because CNS3 and CNS6 overlapped with or were close to repeat sequences, which made the primer design very difficult, we prepared the amplicons just upstream and downstream of them, respectively. As a result of ChIP and PCR, we detected high background signals at CNS1 and *Tcam1* promoter (data not shown), possibly due to the high GC content of these sequences, so we excluded these data. Consequently, we obtained histone modification levels of eight regions at the locus (Fig. 2a). Considering that the resolution of our ChIP analysis was 500–1000 bp and CNS1 and CNS2 were only 196 bp apart from each other, histone modification levels of CNS1 would be reflected by those of CNS2. Similarly, the status of *Tcam1* promoter would be reflected by CNS3, whose amplicon was positioned only 268 bp downstream of the transcriptional start site (TSS) of the *Tcam1* gene. Therefore, we re-designated these amplicons CNS1,2 and Tcam1pro-CNS3 (Fig. 2a).

The histone modification levels were calculated as the ratio of DNA in the antibody-bound chromatin to that in the input fraction. Because the nucleosome content could vary at each region, all the data were normalized to total histone which was revealed by the ChIP analysis with the antibody against histone H3. To normalize the data in different tissues, we used the *aryl-hydrocarbon receptor-interacting protein* (*Aip*) gene promoter because this gene was expressed at similar levels in

germ and liver cells by our qRT-PCR analysis (data not shown). As a negative control, we performed the ChIP assay with normal mouse IgG instead of the antibodies for modified histones.

In germ cells, high H3K9ac levels were observed in the *Smarcd2-Tcam1* intergenic region and intron 1 of the *Tcam1* gene (Fig. 2b). Statistical analysis revealed that the acetylation peak in germ cells was at Tcam1pro-CNS3 among the eight regions examined. In liver cells, H3K9ac was at background levels at all the amplicons except for Smarcd2 intron 1 (Fig. 2b). For the H3K4me1 marker, a clear peak was observed at CNS1,2 along with a low level of the modification at CNS2-Tcam1 in germ cells (Fig. 2c). In liver cells, no regions were marked with H3K4me1, except that Smarcd2 intron 1 might be slightly modified (Fig. 2c). This suggests that CNS1,2 may be a testicular germ cell-specific enhancer for the *Tcam1* gene. High H3K4me3 level was observed at CNS1,2, CNS2-Tcam1, and Tcam1pro-CNS3 in germ cells (Fig. 3a). Statistical analysis revealed that the H3K4me3 peak was CNS1,2 and Tcam1pro-CNS3. In liver cells, CNS1,2 was marked with H3K4me3 at a similar level to the *Aip* promoter (Fig. 3a).

To further access the histone modification patterns, we analyzed the ChIP sequencing (ChIP-seq) data deposited by other studies. We found that data for H3K4me3 in mouse spermatocytes and round spermatids were available (SRA accession: SRA097278),[42] and the corresponding data in the liver were obtained from the Encyclopedia of DNA elements (ENCODE) project (GEO accession: GSM769014).[43] In spermatocytes, depth peaks for the precipitated DNA were observed around CNS1,2 and *Tcam1* promoter, although the input DNA showed no peaks (Fig. 3b, Supplementary Fig. 1a). Such peaks disappeared in round spermatids that did not express *Tcam1* mRNA. Instead, the moderate peak around entire intergenic region was observed specifically in the precipitated DNA (Fig. 3c, Supplementary Fig. 1b). A similar pattern for H3K4me3 in round spermatids was also observed in another ChIP-seq data (GEO accession: GSE42629).[44] Compared with spermatocytes and spermatids, a small peak was observed around the *Smarcd2* promoter specifically in the precipitated DNA in the liver (Fig. 3d, Supplementary Fig. 1c). These data are consistent with the ChIP results (Fig. 3a), given that high levels of the H3K4me3 marker were detected at CNS1,2 and Tcam1pro-CNS3. Collectively, our re-analysis

using previous ChIP-seq data (SRA097278) confirmed that CNS1,2 and Tcam1pro-CNS3 were marked with H3K4me3 in spermatocytes.

Taken together, in germ cell, CNS1,2 was marked with H3K9ac, H3K4me1, and H3K4me3, and Tcam1pro-CNS3 was marked with H3K9ac and H3K4me3. Therefore, these three CNSs could be important regulatory elements for the *Tcam1* gene, and we focused on CNS1, CNS2, and CNS3 in the following analyses.

CNS1 can function as an enhancer of the *Tcam1* promoter in the GC-2spd(ts) cell

We assessed the enhancer activity of the three CNSs by the *in vitro* reporter gene assay. Based on TSS that was previously determined,[32] we cloned a 1644-bp sequence just upstream of the *Tcam1* gene as a promoter and connected it to the luciferase gene. CNS1, CNS2, and CNS3 were amplified by PCR with mouse genomic DNA, and the CNS1 and CNS2 sequences were cloned into 5' of the *Tcam1* promoter (Fig. 4a). CNS3 was connected to 3' of the luciferase gene because it was located downstream of the *Tcam1* promoter. These constructs (CNS1-Pro-luc, CNS2-Pro-luc, and Pro-luc-CNS3) were transiently transfected into three cell lines, GC2-spd(ts), Hepa1-6, and NIH3T3-3-4, and luciferase activity was measured and compared with that of cells transfected with the construct without CNSs (Tcam1-Pro-luc) (Fig. 4a). GC-2spd(ts) cells are derived from mouse spermatocytes[45] while Hepa1-6 and NIH3T3-3-4 cells are derived from mouse hepatocytes and embryonic fibroblasts, respectively.

Compared with the Tcam1-Pro-luc construct, luciferase activities from CNS2-Pro-luc and Pro-luc-CNS3 were slightly increased or at similar levels in the three cell lines (Fig. 4a). In contrast, when CNS1 was linked to the *Tcam1* promoter, luciferase activity was fivefold higher than that of the Tcam1-Pro-luc construct in GC-2spd(ts) cells but not in Hepa1-6 and NIH3T3-3-4 cells (Fig. 4a). This GC-2spd(ts)-specific activity of CNS1 was orientation-independent because CNS1 enhanced *Tcam1* promoter activity to a comparable extent when cloned in the reverse orientation (Fig. 4b). The activity was also detected when CNS1 was connected to 3' of the luciferase gene in both directions (Fig. 4c) and

when the construct was linearized before transfection (Supplementary Fig. 2a). Although shorter *Tcam1* promoters showed higher activity than the 1644-bp promoter, CNS1 could increase their luciferase activity (Supplementary Fig. 2b, c). These results suggest that CNS1 is a spermatocyte-specific enhancer for the *Tcam1* gene. To further characterize enhancer activity, we divided the CNS1 sequence into two halves, CNS1-(1-257) and CNS1-(258-373), and cloned them into the upstream region of the *Tcam1* promoter of the Tcam1-Pro-luc construct. By transfecting these constructs into GC-2spd(ts) cells, luciferase activities were increased in comparison with that of the Tcam1-Pro-luc construct; however, fold increases were 2.2 and 2.1, respectively (Fig. 4b). This indicates that the entire sequence of CNS1 is required for the full activity of this GC-2spd(ts)-specific enhancer.

Because CNS1 was located just upstream of the *Smarcd2* gene (Fig. 2a), it was possible that CNS1 had promoter activity that affected luciferase gene expression in our reporter assay. Therefore, we assessed whether CNS1 could function as a promoter. We connected the CNS1 sequence directly to the luciferase gene in both directions and transfected the constructs into GC-2spd(ts) cells. The results indicated that both directions of CNS1 had very strong promoter activity (Fig. 5a). Compared with the *Tcam1* promoter, CNS1 showed approximately 220-fold higher promoter activity in both directions. We also investigated the CNS1 promoter activity in Hepa1-6 and NIH3T3-3-4 cells. CNS1 showed strong promoter activity in both cell lines, but fold increases relative to the *Tcam1* promoter were 23 in Hepa1-6 cells and approximately 10 in NIH3T3-3-4 cells (Fig. 5a).

Then, we assessed the possibility that the fivefold increase in luciferase activity observed between Tcam1-Pro-luc and CNS1-Pro-luc in GC-2spd(ts) cells may have resulted from the strong promoter activity of CNS1. Using the CNS1-Pro-luc construct, we introduced a polyadenylation signal between CNS1 and the *Tcam1* promoter (CNS1-polyA-Pro-luc). In this construct, transcription of the *Tcam1* promoter driven by CNS1 should be stopped by the poly(A) signal. In fact, a transcript from the promoter in GC-2spd(ts) cells transfected with the CNS1-polyA-Pro-luc construct was dramatically decreased by qRT-PCR (Fig. 5b, left). However, the luciferase mRNA levels were unchanged between the cells transfected with the two constructs (Fig. 5b, right), and the luciferase activity of the

CNS1-polyA-Pro-luc construct was similar to that of CNS1-Pro-luc (Fig. 5c). This indicates that in the CNS1-Pro-luc construct, transcription driven by the CNS1 promoter elongated through the *Tcam1* promoter but stopped in the middle of the luciferase gene. Therefore, we concluded that the luciferase activity of CNS1-Pro-luc did not reflect the strong promoter activity of CNS1 and that CNS1 truly possessed enhancer activity in GC-2spd(ts) cells.

CNS1 is a bidirectional promoter of the ubiquitously expressed *Smarcd2* gene and a testicular germ cell-specific *lncRNA-Tcam1*

Our *in vitro* reporter analysis showed that CNS1 could drive transcription bidirectionally (Fig. 5a). While the *Smarcd2* gene, which was considered to be expressed ubiquitously, was presumed to be regulated by the CNS1 promoter, there were no annotated genes opposite CNS1 (Fig. 2a, Fig. 6a). Thus, we investigated whether any transcripts were generated from the adjacent regions of CNS1. RT-PCR with eight mouse tissues showed that *Samarcd2* was expressed in all tissues, and in the opposite direction, a novel transcript was exclusively detected in the testis (Fig. 6b). This indicates that CNS1 is actually a bidirectional promoter for the ubiquitously expressed *Smarcd2* gene and a novel testis-specific transcript.

We further characterized this novel transcript. To determine the 5’ and 3’ ends of the transcript, we performed rapid amplification of cDNA ends (RACE) analysis. A single band was detected by 5’RACE, and 10 subclones that we sequenced contained one nucleotide as TSS. In addition, 3’RACE resulted in the amplification of a single band, and all subclones contained one nucleotide as the 3’ end. The results indicated that the full length of the transcript consisted of 2404 nucleotides (Fig. 6a). According to the coding potential calculator (CPC) tool version 0.9r2,[46] this transcript was classified as lncRNA, and we designated it *lncRNA-Tcam1*. *lncRNA-Tcam1* was presumed to be polyadenylated at its 3’ end because we observed poly(A) sequences longer than the oligo(dT) length in the subclones obtained by our 3’RACE.

To check whether *lncRNA-Tcam1* was really transcribed as a single transcript, we first performed RT-PCR with testis RNA using primers at its 5'and 3'ends; however, we failed to amplify a specific signal (data not shown). This may be due to a low expression level or the presence of repetitive sequences at both ends of *lncRNA-Tcam1*. We then generated testis cDNA by reverse transcription with the primer at the 3' end and performed PCR to amplify the 478-bp region of the 5' end. This resulted in successful amplification of a specific signal (data not shown). Therefore, we concluded that *lncRNA-Tcam1* was transcribed as a single transcript.

We next investigated the expression pattern of *lncRNA-Tcam1*. By RT-PCR analysis of testes during postnatal development, *lncRNA-Tcam1* was first detected 14 days after birth, and the transcript level was increased at 21 and 28 days and decreased thereafter (Fig. 6c). This expression pattern was correlated to *Tcam1* mRNA (Fig. 1c). We then fractionated the testis into germ, Sertoli, and Leydig cells and investigated *lncRNA-Tcam1* expression. RT-PCR detected *lncRNA-Tcam1* signal in germ cells but not in somatic cells (Fig. 6d). This was also correlated to *Tcam1* mRNA. Finally, we fractionated germ cells into nuclear and cytoplasmic subfractions and checked the subcellular localization of *lncRNA-Tcam1*. *Gapdh* signals amplified from the primer pairs at exons 5 and 6 and at intron 5 and exon 6 confirmed that our fractionation was successful (Fig. 6e). The *lncRNA-Tcam1* signal was exclusively detected in the nucleus (Fig. 6e), which strongly suggests that it was not translated to any peptides and functioned as an RNA molecule.

DNA methylation status in CNS1

Surprisingly, CNS1 was a promoter for a ubiquitously expressed gene (*Smarcd2*), while it was only slightly associated with the H3K4me3 marker in the liver and the modification level was much lower than that in testicular germ cells (Fig. 3). We attributed this to another epigenetic marker: CpG methylation. Because we found a CpG island (CGI) encompassing exon 1 of the *Smarcd2* gene, CNS1, and CNS2 (Supplementary Fig. 3a), CpG methylation status could be one factor affecting the promoter

activity of CNS1. To investigate the methylation of this CGI, bisulfite sequencing analysis was performed using genomic DNA from spermatocytes and adult liver cells. Spermatocytes were collected by sorting germ cells as described previously.[41] After the bisulfite reaction and PCR, we sequenced 10 subclones from each sample and checked which cytosine was converted to thymine. We examined 54 CpGs in CGI, and almost all the CpGs were demethylated in both spermatocytes and the liver; we did not observe any difference between these two tissues (Supplementary Fig. 3b). This result suggests that the hypo-methylation of CGI was associated with the ubiquitous promoter activity of CNS1.

Sp1 contributes to CNS1 promoter activity for the *Smarcd2* gene

To gain mechanistic insight into promoter and enhancer activity of CNS1, we attempted to identify evolutionarily conserved sequences that were important regulatory elements for many genes.[37,47] We collected CNS1 sequences from seven different mammalian species and compared them (Fig. 7a). As a result, we found that some sequences were well conserved among these species and three of the conserved sequences overlapped with the consensus Sp1 binding site. Sp1 is a well-known transcription factor which is expressed ubiquitously and therefore can be involved in the *Smarcd2* gene activation. During murine spermatogenesis, spermatogonia and early and mid-pachytene spermatocytes are the major source of Sp1,[48,49] but some testis-specific splice variants were reported to be present at later stages.[50,51] In addition, the DNA-binding affinity of Sp1 from male germ cells is greater than that from other tissues[50] and the interaction with other proteins and post-translational modifications can change its activity.[52–54] Therefore, it was also possible that Sp1 was involved in the enhancer activity of CNS1.

We first checked whether Sp1 was expressed in the three cell lines we used (GC-2spd(ts), Hepa1-6, NIH3T3-3-4). In this experiment, we used four primer sets, one of which detected a ubiquitously expressed Sp1 transcript (8.2 kb in full length) as well as some splice variants such as 8.8-kb and 2.4-kb transcripts. The other sets were applied to specifically detect three germ cell-specific Sp1 variants (4.1, 3.7, and 3.2 kb) as reported by Thomas *et al.*,[51] although it was pointed out that the primers for a 3.7-kb

variant might also amplify an 8.2-kb transcript.[49] As a result of qRT-PCR, Sp1 mRNAs were detected in all three cell lines, and GC-2spd(ts) and NIH3T3-3-4 cells expressed all the variants at higher levels than Hepa1-6 cells (Supplementary Fig. 4).

We then prepared the CNS1 sequence, in which all three Sp1 sites were mutated not to be recognized by Sp1, and investigated its promoter and enhancer activity. When mutated CNS1 was linked to the *Tcam1* promoter, enhancer activity was significantly increased in GC-2spd(ts) and NIH3T3-3-4 cells (Fig. 7b). In contrast, the mutation had no effect on enhancer activity in Hepa1-6 cells (Fig. 7b). When mutated CNS1 was directly connected to the luciferase gene in the forward orientation, CNS1 promoter activity was not changed (Fig. 7c). However, in the reverse orientation, the mutation significantly decreased CNS1 promoter activity in all the cell lines (Fig. 7c). These suggest that Sp1 is not a key factor for the CNS1 enhancer but contributes to the promoter activity for the *Smarcd2* gene.

CNS1 functions as a promoter of *lncRNA-Tcam1* and an enhancer for *Tcam1* in the chromatin context

We finally assessed whether CNS1 could drive *lncRNA-Tcam1* transcription and enhance *Tcam1* promoter activity in the chromatin context. It is also interesting to know whether *lncRNA-Tcam1* may be involved in *Tcam1* regulation. To test these possibilities, we prepared a sequence encompassing the *Tcam1* promoter, *lncRNA-Tcam1*, and CNS1. We obtained a BAC clone (B6Ng01-276I01) encompassing the entire mouse *Tcam1* locus and tried to subclone a 14.4-kb fragment by *Tth111*I digestion. However, in the process of subcloning, the fragment was shortened, and we could only obtain a 6.9-kb sequence. Because this 6.9-kb fragment still contained the intact sequence encompassing the *Tcam1* promoter, *lncRNA-Tcam1*, and CNS1, we decided to use this for further analysis. The 6.9-kb sequence was linked to the enhanced green fluorescent protein (EGFP) gene and the resulting construct was named lncRNA-6.9kb-EGFP (Fig. 8a). We also prepared the ΔCNS1-lncRNA-EGFP construct by

deleting a 861-bp region containing CNS1 with a restriction enzyme, *Ksp*I (Fig. 8a).

The two constructs were transfected into GC-2spd(ts) with a vector containing the puromycin resistance gene, and by selection with puromycin, we obtained GC-2spd(ts) cells that stably expressed the antibiotic resistance gene. To further select cell clones that contained the lncRNA-6.9kb-EGFP or ΔCNS1-lncRNA-EGFP construct stably integrated into genomic DNA, we performed cell cloning by the limited dilution method. We successfully established 11 GC-2spd(ts) clones for the lncRNA-6.9kb-EGFP construct and 10 clones for ΔCNS1-lncRNA-EGFP. Copy numbers of the transgene in these clones were calculated by real time PCR using the β-actin gene as a control of two copies. Copy number ranged from 0.9 to 27.8 for cell clones with lncRNA-6.9kb-EGFP and from 1.4 to 18.7 for clones with ΔCNS1-lncRNA-EGFP. We also checked whether or not the integrated transgene was intact by long PCR with genome DNA purified from each clone. We used a primer pair that could amplify most region of the transgenes (Supplementary Fig. 5). As a result of PCR, signals with expected sizes were observed for all the clones along with extra bands at various sizes (Supplementary Fig. 5). The unexpected signals might be non-specific or derived from truncated transgene constructs, but this result suggested that at least one intact transgene was integrated in all the clones.

Using the established cell clones, we first investigated whether the bidirectional transcription of *Smarcd2* and *lncRNA-Tcam1* was impaired by deleting CNS1. However, because the *Smarcd2* gene was endogenously expressed in GC-2spd(ts) cells, we only examined *lncRNA-Tcam1* expression in stable cell clones. The RNA level was measured by qRT-PCR and was normalized to *Gapdh* mRNA level and transgene copy number. The levels greatly differed between clones, probably due to a position effect, but the average expression level was 15.3-fold higher in clones with lncRNA-6.9kb-EGFP than in those with ΔCNS1-lncRNA-EGFP (Fig. 8b). This difference was statistically significant, and therefore, the data indicate that CNS1 possesses promoter activity for *lncRNA-Tcam1* in the chromatin context.

Following this, we measured EGFP mRNA levels per copy in the established clones to evaluate the enhancer activity of CNS1. The EGFP mRNA level was again varied; however, comparative analysis revealed that cell clones integrated with lncRNA-6.9kb-EGFP expressed significantly higher levels of

EGFP mRNA than cells with ΔCNS1-lncRNA-EGFP (Fig. 8c). The average EGFP mRNA level was 4.5-fold higher in the clones with intact CNS1. This indicates that CNS1 could function as an enhancer element for *Tcam1* gene expression even when it is integrated into chromatin.

We finally examined whether the expression of *lncRNA-Tcam1* was correlated to that of EGFP in GC-2spd(ts) cell clones. Using qRT-PCR, we measured the *lncRNA-Tcam1* level in each cell clone with lncRNA-6.9kb-EGFP and compared it with the EGFP mRNA level (Fig. 8d). There was no statistical correlation between the *lncRNA-Tcam1* and EGFP levels. Therefore, it was less possible that *lncRNA-Tcam1* was involved in the regulation of *Tcam1* transcription.

Discussion

CNS1 may be a spermatocyte-specific enhancer for the *Tcam1* gene

An enhancer is a sequence that increases the transcription rate of its target gene and is usually located in a remote upstream or downstream region.[4,5] It enhances gene transcription when it physically interacts with the target gene promoter in the nucleus.[4,5] In the present study, CNS1 increased *Tcam1* promoter activity *in vitro* in GC-2spd(ts) cells (Fig. 4a, Supplementary Fig. 2), which indicates that CNS1 is a potential enhancer in these cells. To assess the enhancer activity of CNS1 in the chromatin context, we established stable cell clones using two constructs (Fig. 8). One contained the 6.9-kb upstream sequence of the *Tcam1* gene, and in the other construct, we deleted the CNS1 region. EGFP mRNA was expressed at significantly higher levels in cell clones with the full 6.9-kb sequence than in clones without CNS1. Possibly, integrated transgenes might be under control of neighboring enhancers, and some transgenes might be truncated as suggested by our genome PCR (Supplementary Fig. 5). However, there is no reason to expect that one construct was more affected by enhancers or more frequently truncated than the other. Therefore, the most reasonable explanation for these results is that

CNS1 has enhancer activity for the *Tcam1* promoter in GC-2spd(ts) cells.

Because CNS1 was positioned between two *Ksp*I sites and no other recognition sites of this enzyme were present in the construct, we cut out the 861-bp *Ksp*I sequence to delete CNS1 from the 6.9-kb sequence. However, this also deleted sequences other than CNS1 because CNS1 was only 373 bp in length. It is possible that the significant decrease in EGFP expression in cells with ΔCNS1-lncRNA-EGFP resulted from deletion of a sequence other than CNS1. We cannot completely rule out this possibility; however, considering the *in vitro* enhancer activity and histone modification patterns, the significant decrease in EGFP expression was most likely due to the deletion of CNS1.

Interestingly, the enhancer activity of CNS1 was only observed in GC-2spd(ts) cells, which were derived from mouse spermatocytes (Fig. 4a). This indicates that some transcription factors, which were expressed in GC-2spd(ts) but not in Hepa1-6 and NIH3T3-3-4 cells, bound to CNS1 and increased *Tcam1* promoter activity. However, the results do not necessarily mean that CNS1 is a genuine enhancer in native spermatocytes. One important question is the extent to which GC-2spd(ts) cells maintain the characteristics of spermatocytes. Some studies reported that GC-2spd(ts) cells expressed spermatocyte-specific genes and contained gene regulatory mechanisms similar to those of native spermatocytes.[55–57] For example, a regulatory mechanism of the spermatocyte-specific histone *H1t* gene was reported to be conserved in GC-2spd(ts) cells.[57–59] This indicates that some properties of GC-2spd(ts) cells are similar to those of native spermatocytes. However, it is obvious that cells of this cell line are not the same as spermatocytes, as evidenced by the fact that *Tcam1* mRNA was expressed at a much lower level in GC-2spd(ts) cells than in the testis (data not shown). In this context, it is interesting to note that a promoter of the spermatocyte/spermatid-specific thioredoxin-3 gene could be appropriately activated in GC-2spd(ts) cells despite its low level of endogenous expression.[60,61] It is possible that GC-2spd(ts) cells contain transcription factors for *Tcam1* activation by CNS1 but can only bind to the CNS1 sequence when the locus is associated with active chromatin markers.

The histone modification state is very important for estimating the function of a sequence in the cell. For our ChIP analyses, we used a germ cell fraction that contained not only spermatocytes but also

spermatogonia, spermatids, and spermatozoa, and we estimated that 30% of this fraction consisted of spermatocytes. Because spermatogonia was estimated to be 4% in germ cells (R. Yoneda, M.K, and A.P.K., unpublished observations), the rest of the fraction contained spermatids and spermatozoa. Importantly, after meiosis, histones are replaced with transition proteins and eventually protamines, and this replacement begins in the middle of spermiogenesis.[62] Therefore, a substantial proportion of spermatids and spermatozoa did not contain histones, although mature spermatozoa was reported to retain 1% histones.[63] This indicates that the population of spermatocytes should be much higher than 30% in germ cells containing histones. Taken together with the observation that the H3K4me3 pattern in a germ cell fraction was similar to that in spermatocytes (compare Fig. 3a with Fig. 3b), the histone modification peaks detected in germ cells mostly reflected the pattern occurring in spermatocytes.

In general, H3K4me1 is associated with enhancers, and many studies considered the regions marked with this modification to be enhancers.[64–66] Some reports actually demonstrated that regions marked with H3K4me1 possessed enhancer activity.[67,68] In the present study, CNS1 was associated with H3K4me1 in germ cells but not in the liver, which suggests that it is a germ cell-specific enhancer for the *Tcam1* gene. Considering that *Tcam1* was exclusively expressed in spermatocytes, CNS1 could be a true enhancer for spermatocyte-specific *Tcam1* expression.

CNS1 is a bidirectional promoter

A promoter is the sequence to which RNA polymerase binds to begin transcription, and it usually comprises a core promoter and a proximal promoter.[69–71] The core promoter is the region approximately 35 bp upstream and/or downstream of TSS and is required for the recruitment of general transcription factors to form the pre-initiation complex (PIC).[69,70] The proximal promoter contains the 5'-adjacent sequences to the core promoter, and it usually extends approximately 200–300 bp upstream of TSS.[71] In the present study, the region between the *Smarcd2* gene and *lncRNA-Tcam1* was estimated to be 187 bp, based on the TSS of *lncRNA-Tcam1* that we determined and that of *Smarcd2* reported in the DBTSS

database (http://dbtss.hgc.jp/). Therefore, the *Smarcd2* gene and *lncRNA-Tcam1* should share the proximal promoter when both of them are transcribed, and this promoter overlaps with CNS1. Although the core promoter is probably different between these two transcriptional units, we conclude that CNS1 is necessary to create PIC for both *Smarcd2* and *lncRNA-Tcam1*.

Interestingly, CNS1 was a bidirectional promoter only in testicular germ cells and functioned unidirectionally in other tissues. Bidirectional promoters are reportedly associated with CGIs more often than unidirectional promoters, as shown by full genome computer analyses.[72,73] Consistent with this, CNS1 was included in CGI that encompassed exon 1 of *Smarcd2*, CNS1, and CNS2. In general, hyper- and hypo-methylation of CGI in a bidirectional promoter resulted in gene silencing and activation, respectively, of both transcripts, as reported for some cancer-related genes.[74] However, the expression patterns of the two transcripts driven by CNS1 were different, and its hypo-methylation was correlated to the ubiquitous expression of *Smarcd2* but not to *lncRNA-Tcam1*. According to our present data, high levels of histone acetylation and histone H3K4 methylation may be necessary for *lncRNA-Tcam1* activation. These results strongly suggest that the transcripts driven by CNS1 are not coordinately regulated, unlike many reported examples.[72–75]

Notably, the *in vivo* promoter activity of CNS1 was not as high as its *in vitro* activity. While we could detect *Tcam1* mRNA by 30 cycles of PCR in our RT-PCR analysis, both *Smarcd2* mRNA starting from exon 1 and the *lncRNA-Tcam1* transcript could be detected by 40 cycles. With regard to the *Smarcd2* gene, a major mRNA is transcribed from exon 2 and its level is much higher than that of the transcript from exon 1.[76] CpG hypo-methylation of CNS1 and its weak association with the H3K4me3 marker may be sufficient for the low level of *Smarcd2* expression; however, the activation of *lncRNA-Tcam1* may require high levels of active histone modifications.

Bidirectional transcription from an enhancer is reminiscent of enhancer RNAs (eRNAs). eRNAs were originally identified as noncoding transcripts induced at enhancers on membrane depolarization of neurons. [11] Several studies have reported eRNAs at enhancers responsive to androgen, estrogen, and p53, and some have been found to play important roles in enhancer functions.[7,11,77–79] eRNAs are

generally characterized as short (1-2 kb), bidirectionally transcribed products, nonpolyadenylated RNAs, and transcripts from H3K4me1-enriched enhancers. Transcription from CNS1 at the *Tcam1* locus was similar to that of eRNAs in that it occurred bidirectionally from the H3K4me1-enriched enhancer (CNS1) but was otherwise different. For example, bidirectional transcription from CNS1 was only observed in testicular germ cells; in other tissues, CNS1 activated the *Smarcd2* gene alone. In addition, both *Smarcd2* mRNA and *lncRNA-Tcam1* were longer than 2 kb and polyadenylated, and *Smarcd2* mRNA could be translated. Therefore, the transcription of *Smarcd2* and *lncRNA-Tcam1* is different from that of eRNAs.

A mechanism by which CNS1 functions as a dual promoter–enhancer

There are two examples a dual promoter-enhancer element in the chicken. One is the −1.9-kb element of the chicken lysozyme gene. This element was originally identified as a hormone-responsive enhancer element and was later found to function as a promoter for lncRNA, lipopolysaccharide inducible noncoding RNA (LINoCR), which was necessary for nucleosome repositioning and eviction of negative regulator proteins in response to lipopolysaccharide in macrophages.[80–82] The other example is an enhancer of the chicken *mim-1* gene, induced by the Myb protein. This enhancer also harbored Myb-inducible promoter activity for a noncoding RNA, and this noncoding transcription was necessary for nucleosomal remodeling at the enhancer.[83] In mammals, genome-wide analysis revealed that 70% of extragenic RNA polymerase II peaks were related to the chromatin signature of enhancers,[9] which suggested that many enhancers could be promoters of noncoding RNAs. However, there is no clear indication of a dual promoter–enhancer element in mammals, and this is the first report of such an element.

How can CNS1 function as both a promoter and an enhancer? The hypo-methylation of a CGI containing CNS1 was observed not only in spermatocytes but also in the liver, and this, together with a weak association with H3K4me3, seemed to be linked to ubiquitous *Smarcd2* activation. On the other

hand, for the activation of *lncRNA-Tcam1* in spermatocytes, the chromatin in CNS1 probably needs to be more strongly associated with H3K4me3 as well as with other active chromatin markers like H3K9ac. Moreover, CNS1 was marked with H3K4me1 in spermatocytes and could act as an enhancer for the *Tcam1* gene. Therefore, in spermatocytes, CNS1 can bidirectionally drive the transcription of *Smarcd2* and *lncRNA-Tcam1* and enhance *Tcam1* gene expression, while in the liver, it can only drive *Smarcd2* expression as a promoter (Fig. 9).

To identify functional elements in CNS1, we focused on three Sp1-binding sites. By mutating all of them, we investigated whether Sp1 contributed to bidirectional promoter and/or enhancer activity of CNS1. The results indicated that Sp1 played a role in the promoter activity for *Smarcd2* but not for *lncRNA-Tcam1* in all the cell lines (Fig. 7c). In GC-2spd(ts) and NIH3T3-3-4 cells, which expressed Sp1 mRNAs at higher levels than Hepa1-6, Sp1 repressed the CNS1 enhancer activity (Fig. 7b). These suggest that the promoter activity for *Smarcd2* is partially controlled by Sp1 but is not coordinated with that for *lncRNA-Tcam1*, and that Sp1 is not a factor to enhance the *Tcam1* promoter activity. Considering that halved CNS1 sequences showed lower enhancer activity than intact CNS1 in our reporter assay (Fig. 4b), the entire 373-bp sequence may be necessary for its full enhancer activity. Possibly, several transcription factors other than Sp1 may bind to various regions of CNS1 or a large protein complex may be formed at CNS1.

Recently, a BET family protein, Brdt, was reported to play crucial roles in the regulation of testicular germ cell-specific gene expression during meiosis.[84] To test the possibility that *Tcam1* is controlled by Brdt, we analyzed the transcriptomic and ChIP-seq data (GEO accession: GSE39909, GSE39910, GSE39908). However, we could not find the binding signal of Brdt at the *Tcam1* locus, and *Tcam1* expression was not changed in the testis from *Brdt*-deficient mice (data not shown). Similarly, *Smarcd2* expression was not affected by *Brdt*-deficiency (data not shown). Therefore, it is unlikely that Brdt controls the *Tcam1* and *Smarcd2* gene.

Function of *lncRNA-Tcam1*

In the two examples of a dual promoter–enhancer element in the chicken, noncoding transcription was necessary for nucleosome remodeling to activate the target gene of the enhancer.[80–83] However, our data indicate that *lncRNA-Tcam1* transcription is not correlated to *Tcam1* promoter activity (Fig. 8d). What is the function of this lncRNA? The nuclear localization of *lncRNA-Tcam1* (Fig. 6e) may provide some hints. Many nuclear lncRNAs reported till date have two main functions: gene regulation and the formation of nuclear structures.[85] In general, lncRNAs that constitute some nuclear structures are expressed at high levels.[86] For example, *NEAT1* is a component of the nuclear paraspeckle, and its expression is as abundant as that of *XIST* in the nucleus.[87–89] In contrast, *lncRNA-Tcam1* is expressed at a low level; therefore, it is more likely to be involved in the regulation of some genes.

To test gene regulatory activity, we overexpressed *lncRNA-Tcam1* in GC-2spd(ts), Hepa1-6, and NIH3T3-3-4 cells; however, endogenous *Tcam1* gene expression remained unchanged (data not shown). In addition, the expression levels of the endogenous *Smarcd2* and *Gh* genes, which were linked to *Tcam1*, were not correlated to *lncRNA-Tcam1* in cell clones with the lncRNA-6.9kb-EGFP construct (data not shown). Therefore, we think that *lncRNA-Tcam1* may contribute to gene regulation at other loci, as is the case for several lncRNAs that were reported to work *in trans*. For example, *lincRNA-p21*, located next to the *p21* gene, was not related to *p21* gene regulation but activated or repressed many genes in the canonical p53 pathway and played a role in triggering apoptosis.[90,91] Transcription of *lncRNA-Tcam1* was dramatically induced 14–21 days after birth during postnatal testis development, when many important protein-coding genes are upregulated.[62] Therefore, *lncRNA-Tcam1* may play roles in gene activation during this period. Further studies will be necessary to reveal the actual function of *lncRNA-Tcam1*.

Materials and Methods

Animals

The mice (C57/BL6) were maintained at 25°C with a photoperiod of 14:10 hours light: dark with free access to food and water. Experimental procedures used in this study were approved by the Institutional Animal Use and Care Committee at Hokkaido University.

RNA analyses

Northern blot, *in situ* hybridization, RT-PCR, and quantitative RT-PCR (qRT-PCR) were done as described previously.[92] qRT-PCR was also performed using the 7300 real-time PCR system and KOD SYBR qPCR Mix (Toyobo, Osaka, Japan) in a total volume of 10 μl per well. The amplification condition was 98 °C for 2 min and 40 cycles of 98 °C for 10 sec, 60 °C for 10 sec, and 68 °C for 1 min. Dissociation curves were obtained to confirm the specificity of the amplified DNA, and in some cases, the amplified product was checked by agarose gel electrophoresis. Probes for northern blot and *in situ* hybridization were obtained by RT-PCR using adult testis cDNA, and primer pairs are listed in Table 1. Primer sequences for RT-PCR and qRT-PCR are also shown in Table 1.

Fractionation of the testis into germ, Sertoli, and Leydig cells and isolation of spermatocytes

The testis from 8-week-old mice was sorted into Leydig cell, germ cell, and Sertoli cell fractions as described previously with slight modifications.[40] Briefly, the tunica albuginea was removed from each testis and the tissue was placed in Dulbecco's modified Eagle's medium (DMEM) containing 0.1 % collagenase (Wako Pure Chemicals, Osaka, Japan) for about 15 minutes at 32 °C in a water bath with occasional agitation. The tubules were separated from the dispersed interstitial cells by unit gravity sedimentation for 5 min, and the supernatant was used as a Leydig cell-rich fraction. The tubules were then dissociated with 0.1 % collagenase and 1.5 kU/ml DNase I (Wako Pure Chemicals) for 30 min at

32 °C. The resulting cell suspension was filtered through nylon gauze (50 μm) twice and placed on 5 % Nycodentz (Sigma, St. Louis, USA) in an equal volume of Krebs ringer that was underlayered by an equal volume of 15 % Nycodentz. After the centrifugation at 120g for 3 min, the cells recovered from the layer formed between 5 % Nycodentz and DMEM were used as a Sertoli cell-rich fraction, and the cells between 5 % and 15 % Nycodentz were used as a germ cell-rich fraction. To isolate spermatocytes, the germ cells were treated with Hoechst blue and red and sorted by a JSAN cell sorter as described previously.[41] The purity of each cell fraction was checked in every experiment by qRT-PCR for marker genes.[41]

ChIP assay

ChIP was conducted with germ cells purified from the 8-week-old testis and the liver as previously described [92] with monoclonal antibodies specific for histone H3, H3K9ac, H3K4me1, and H3K4me3, or with 30 μg of normal mouse IgG. The amplification efficiency was normalized by calculating the ratio of the signal in the bound chromatin to that in the input fraction. Modification levels of H3K9ac, H3K4me1, and H3K4me3 were also normalized versus total histone H3. The values were further normalized to comparable signals at the ubiquitously expressed *Aip* promoter (defined as 1.0). The antibodies were kindly gifted by Dr. Hiroshi Kimura at Osaka University.[93] The qPCR was performed using primer pairs listed in Table 1.

ChIP sequencing data

ChIP sequencing data were collected from Gene expression omnibus (GEO) database[94] and Sequence read archive (SRA) database[95] with sra format (Table 2).

Alignment of short read data and depth calculation

The short read sequences were extracted in fastq format from downloaded sra format data with sratoolkit (version 2.3.5-2).[95] For comparative analysis between the pair end read data and single end data, 2nd read (3' end) were eliminated from pair end data. The extracted sequences were mapped to the mouse reference genome (build mm9 for liver data or mm10 for spermatocyte and round spermatid data)[96] with bowtie (version 2.1.0).[97] The alignment parameter was set to allow single mutation in alignment and to ignore the mismatch penalty for low quality nucleotides with lower quality value than 20. Then, the read depth for each position of genome was calculated with depth command in sumtools (version 0.1.19).[98] The depth for each sample was normalized with total number of reads (RPM) or means of depth for *Aip* promoter region (4125000-4127500bp of chromatin 19) and plotted. For the sample with duplicate experiments, the plot displays the means of two experiments.

Reporter constructs

Sequences of all the primers described in this section are listed in Table 1. All the constructs were subject to the sequencing analysis prior to transfection studies.

A putative promoter region of the *Tcam1* gene was amplified by KOD FX (Toyobo) with mouse genomic DNA using primer pairs listed in Table 1. The 1644-bp promoter fragment was ligated upstream of the luciferase gene in a pGL3-Basic vector (Promega Corporation, Madison, WI) at the *Sma*I site, and we named the resulting construct Tcam1-Pro-luc. All the CNS sequences including a half part of CNS1 were also amplified by PCR with KOD FX Neo. By inserting them into the blunted *Mlu*I site of Tcam1-Pro-luc, we generated the constructs, CNS1-Pro-luc, CNS2- Pro-luc, reversed CNS1-Pro-luc, CNS1-(1-257)-Pro-luc, and CNS1-(258-373)-Pro-luc. To generate the Pro-luc-CNS3, Pro-luc-CNS1, and Pro-luc-reversed CNS1 construct, we cloned the CNS3 or CNS1 fragment into the blunted *BamH*I site of Tcam1-Pro-luc. The polyA-signal sequence was obtained by digesting a pGL3-Basic vector with *BamH*I and *Xba*I and the resulting 262-bp fragment was blunted and

phosphorylated before ligation. The poly(A) sequence was inserted into the blunted *Nhe*I site located between CNS1 and the *Tcam1* promoter of CNS1-Pro-luc. CNS1-luc and reversed CNS1-luc constructs were generated by inserting CNS1 into the *Sma*I site of a pGL3-Basic vector.

For generating mutated CNS1 constructs, a CNS1 fragment was isolated from the CNS1-luc construct by digestion with *Nhe*I and *Xho*I and employed as a PCR template. The first round PCR reactions were performed with CNS1 forward and mutagenesis 1 reverse primers and with mutagenesis 1 forward and CNS1 reverse primers (Table 1). The products were purified, combined, and used as a template for the second round PCR, in which 30 cycles of reaction was performed with KOD FX Neo using CNS1 forward and CNS1 reverse primers after two cycles without the primers. The resulting product, which contained two GC-boxes (103-110, 128-135) mutated from GCCCCGCC to AAAAAAAA, was subcloned into a pBluescript vector (Stratagene, La Jolla, CA) at *EcoR*V site. After sequencing, a single nucleotide at 3' end and four nucleotides at 5' end were somehow missing in all the subclones. To repair these deletions, we performed PCR with CNS1 forward primer and T7 promoter primer using one of the subclones as a template. The product was digested with *EcoR*I and subcloned into a pBluescript vector at *EcoR*V and *EcoR*I site. The resulting subclone was further used as a template of PCR to generate another mutation at a Sp1 site. We performed PCR with mutagenesis 2 forward primer and CNS1 reverse primer and with T3 promoter primer and mutagenesis 2 reverse primer. The products were purified, combined, and used as a template for the second round PCR, which was performed as above by using CNS1 reverse primer and T3 promoter primer. This resulted in generation of CNS1, in which a GC- box (186-193) was mutated from GACCCGCC to AAAAAAAA besides two mutated GC-boxes. The product was directly used for generation of luciferase constructs. For the mutCNS1-luc construct, the PCR product was digested with *Kpn*I and inserted into a pGL3 basic vector at *Sma*I and *Kpn*I sites. For the reversed mutCNS1-luc construct, the PCR product was digested with *Hind*III and inserted into a pGL3 basic vector at *Sma*I and *Hind*III sites. To generate the mutCNS1-Pro-luc construct, the PCR product was digested with *Kpn*I and inserted into the Tcam1-Pro-luc construct at *Mlu*I and *Kpn*I site. The *Mlu*I site was blunted before ligation.

To generate a construct which contained a long upstream sequence of the *Tcam1* gene linked to the luciferase gene, we digested a BAC clone, B6Ng01-276I01, obtained from RIKEN Bioresource center, with *Tth111*I and collected a 14,383-bp fragment. We then inserted the fragment into pGL3-Basic at the *Sma*I site, but both of the two subclones we obtained included only 6969-bp of the fragment and lost 7414-bp 5' sequence. Because this fragment still encompassed CNS1, CNS2, the entire *lncRNA-Tcam1*, and the *Tcam1* promoter, we decided to use this construct for our analyses. We named the construct lncRNA-6.9kb-luc. To delete the CNS1 sequence, we digested lncRNA-6.9kb-luc with *Ksp*I and self-ligated the larger fragment, and the resulting construct was ΔCNS1-lncRNA-luc. This construct lost a 861-bp region containing CNS1 (Fig. 8a). To generate lncRNA-6.9kb-EGFP, we first destroyed the *Xho*I site of the pEGFP-1 vector (Clontech laboratories Inc., Palo Alto, CA, USA) by the digestion with *Xho*I, the blunting with T4 DNA polymerase, and its self-ligation, and the resulting vector was pEGFP-1Δ*Xho*I. Next, we digested lncRNA-6.9kb-luc with *Nhe*I to obtain the 4476-bp 5' sequence of the 6.9-kb *Tcam1* upstream sequence. These *Nhe*I fragments were blunted and inserted into the blunted *Pst*I site of pEGFP-1Δ*Xho*I. The resulting construct was further digested with *Xho*I and *Sal*I, and the *Xho*I fragment of lncRNA-6.9kb-luc, which contained 3' sequences of the 6.9-kb *Tcam1* upstream sequence, was ligated. The middle four nucleotides of the *Xho*I and *Sal*I recognition sites were identical, so we could ligate these fragments. The resulting constructs were lncRNA-6.9kb-EGFP. ΔCNS1-lncRNA-EGFP was made by connecting the 3041-bp *Kpn*I-*Xho*I fragment of ΔCNS1-lncRNA-luc to the 7269-bp *Xho*I-*EcoR*I fragment of lncRNA-6.9kb-EGFP. The *Kpn*I and *Eco*RI sites were blunted before ligation.

Cell culture, reporter gene transfection, and luciferase activity assay

GC-2spd(ts) cells (CRL-2196) were obtained from American Type Culture Collection. NIH3T3-3-4 (RCB1862) and Hepa1-6 (RCB1638) cells were obtained from RIKEN Cell Bank (Tsukuba, Japan). All the cells were cultured in DMEM containing 10 % fetal bovine serum, 100 U/ml

penicillin, 100 μg/ml streptomycin, and 292 μg/ml L-glutamine (Invitrogen, Carlsbad, CA). For the reporter gene assay, constructs were transfected into these cells in 24-well dishes using GeneJuice (Novagen**,** Inc., Madison, WI) according to directions, and luciferase activity was measured as described previously.[92]

5'RACE and 3'RACE

For 5'RACE, cDNA was generated using gene specific primers and mouse testis RNA with Superscript III reverse transcriptase (Invitrogen). After purification of cDNA with QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), oligodeoxycytidine was added by terminal deoxynucleotidyl transferase (Takara). The first round PCR was performed with abridged anchor primer (AAP) and gene specific primer 1 (GSP1). For the second nested amplification, GSP2 and abridged universal amplification primer (AUAP) were used.

For 3'RACE, cDNA was prepared by reverse transcription with oligo(dT) connected to an adaptor sequence (AP). The first PCR amplification was conducted by using the primer with adaptor sequence and GSP3. The second nested amplification was carried out with the same adaptor primer and GSP4.

All the amplified products were subcloned into a pBluescript vector (Stratagene, La Jolla, CA) by the TA-cloning method, and 10 subclones for each sample were sequenced. All the primer sequences are listed in Table 1.

Preparation of subcellular fractions of germ cells

Germ cells were dissolved in NP-40 lysis buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 0.5 % NP-40, pH 7.5) and subcellular fractions were prepared as described previously.[99]

Establishment of stable cell lines with GC-2spd(ts) cells

Each of the lncRNA-6.9kb-EGFP and ΔCNS1-lncRNA-EGFP construct was co-transfected with a pKO SelectPuro V810 vector (Lexicon Genetics, The Woodlands, Texas, USA) into GC-2spd(ts) cells in 35-cm^2 dishes by GeneJuice (Novagen) as above. The pKO SelectPuro vector was used for conferring the puromycin-resistence to the transfected cell. Twenty-four hours after the transfection, we started selection with 2-3 μg/ml of puromycin (Wako Pure Chemicals) and continued it for 11days. After selection, we counted the cell numbers and spread 2 cells per well in 96-well plates. This resulted in the growth of a single colony in most wells. We picked the wells containing a single colony and maintained the cells. For each clone, we isolated genomic DNA and total RNA, and assessed the copy number of the transgene and the expression levels of *lncRNA-Tcam1* and EGFP by the method described previously.[100]

Statistical analysis

The results were expressed as means ± S.D. The ChIP data was assessed by one-way analysis of variance (ANOVA) followed by Dunnett's test using Microsoft Excel statistical analysis functions (Microsoft Corp., Redmond, WA). Student's *t* test was performed using Microsoft Excel statistical analysis functions to compare the luciferase activity of CNS1-Pro-luc with that of mutCNS1-Pro-luc or of reversed CNS1-luc with reversed mutCNS1-luc. Statistical significance of EGFP or *lncRMA-Tcam1* expression between GC-2spd(ts) cell clones with lncRNA-6.9kb-EGFP and ΔCNS1-lncRNA-EGFP was analyzed with a Mann-Whitney U test. The relationship between *lncRNA-Tcam1* and EGFP in 11 clones with the lncRNA-6.9kb-EGFP construct was analyzed by the correlation coefficient statistical analysis. $P<0.05$ was considered statistically significant.

Accession number

Nucleotide sequence data for *lncRNA-Tcam1* reported in this paper is available in the DDBJ/EMBL/GenBank databases under the accession number AB902906.

Acknowledgements

We thank Dr. Shin Matsubara for his valuable comments on this manuscript. This work was supported by Grants-in-aid for Scientific Research 19770048 and 21770068 to A.P.K. from the Ministry of Education, Culture, Sports, Science and Technology, Japan. M.K. was supported by a research fellowship of the Japan Society of the Promotion of Science (24-6230).

References

1. Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–32

2. Geisler, S. & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14**, 699–712

3. Visel, A., Rubin, E.M. & Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. *Nature* **461**, 199–205

4. Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A. & Bejerano, G. (2013). Enhancers: five essential questions. *Nat. Rev. Genet.* **14**, 288–95

5. De Laat, W. & Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506

6. Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T. & Kraus, W.L. (2011). A Rapid, Extensive, and Transient Transcriptional Response to Estrogen Signaling in Breast Cancer Cells. *Cell* **145**, 622–634

7. Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., Glass, C.K., Rosenfeld, M.G. & Fu, X.-D. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–4

8. Ren, B. (2010). Transcription: Enhancers make non-coding RNA. *Nature* **465**, 173–4

9. De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L. & Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384

10. Koch, F., Jourquin, F., Ferrier, P. & Andrau, J.-C. (2008). Genome-wide RNA polymerase II: not genes only! *Trends Biochem. Sci.* **33**, 265–273

11. Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G. & Greenberg, M.E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–7

12. Natoli, G. & Andrau, J.-C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19

13. Ørom, U.A. & Shiekhattar, R. (2013). Long Noncoding RNAs Usher In a New Era in the Biology of Enhancers. *Cell* **154**, 1190–1193

14. De Laat, W., Klous, P., Kooren, J., Noordermeer, D., Palstra, R., Simonis, M., Splinter, E. & Grosveld, F. (2008). Chapter 5 Three‐Dimensional Organization of Gene Expression in Erythroid Cells. *Curr. Top. Dev. Biol.* **82**, 117–139

15. Deschamps, J. (2007). Ancestral and recently recruited global control of the Hox genes in development. *Curr. Opin. Genet. Dev.* **17**, 422–7

16. Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development* **134**, 2549–60

17. Palstra, R., de Laat, W. & Grosveld, F. (2008). Chapter 4 β‐Globin Regulation and Long‐Range Interactions. *Adv. Genet.* **61**, 107–142

18. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803

19. Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D. & Weintraub, H. (1989). MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell* **58**, 823–831

20. Tontonoz, P., Hu, E., Graves, R.A., Budavari, A.I. & Spiegelman, B.M. (1994). mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. *Genes Dev.* **8**, 1224–1234

21. Chen, E.Y., Liao, Y.-C., Smith, D.H., Barrera-Saldaña, H.A., Gelinas, R.E. & Seeburg, P.H. (1989). The human growth hormone locus: Nucleotide sequence, biology, and evolution. *Genomics* **4**, 479–497

22. Ho, Y., Liebhaber, S.A. & Cooke, N.E. (2004). Activation of the human GH gene cluster: roles for targeted chromatin modification. *Trends Endocrinol. Metab.* **15**, 40–45

23. Jones, B.K., Monks, B.R., Liebhaber, S.A. & Cooke, N.E. (1995). The human growth hormone gene is regulated by a multicomponent locus control region. *Mol. Cell. Biol.* **15**, 7010–21

24. Su, Y. (2000). The Human Growth Hormone Gene Cluster Locus Control Region Supports Position-independent Pituitary- and Placenta-specific Expression in the Transgenic Mouse. *J. Biol. Chem.* **275**, 7902–7909

25. Elefant, F. (2000). Targeted Recruitment of Histone Acetyltransferase Activity to a Locus Control Region. *J. Biol. Chem.* **275**, 13827–13834

26. Ho, Y., Elefant, F., Cooke, N. & Liebhaber, S. (2002). A Defined Locus Control Region Determinant Links Chromatin Domain Acetylation with Long-Range Gene Activation. *Mol. Cell* **9**, 291–302

27. Ho, Y., Elefant, F., Liebhaber, S.A. & Cooke, N.E. (2006). Locus Control Region Transcription Plays an Active Role in Long-Range Gene Activation. *Mol. Cell* **23**, 365–375

28. Kimura, A.P., Liebhaber, S.A. & Cooke, N.E. (2004). Epigenetic modifications at the human growth hormone locus predict distinct roles for histone acetylation and methylation in placental gene activation. *Mol. Endocrinol.* **18**, 1018–32

29. Kimura, A.P., Sizova, D., Handwerger, S., Cooke, N.E. & Liebhaber, S.A. (2007). Epigenetic activation of the human growth hormone gene cluster during placental cytotrophoblast differentiation. *Mol. Cell. Biol.* **27**, 6555–68

30. Nalam, R.L., Lin, Y.-N. & Matzuk, M.M. (2010). Testicular cell adhesion molecule 1 (TCAM1) is not essential for fertility. *Mol. Cell. Endocrinol.* **315**, 246–253

31. Ono, M., Nomoto, K. & Nakazato, S. (1999). Gene structure of rat testicular cell adhesion molecule 1 (TCAM-1), and its physical linkage to genes coding for the growth hormone and BAF60b, a component of SWI/SNF complexes. *Gene* **226**, 95–102

32. Sakatani, S., Takahashi, R., Okuda, Y., Aizawa, A., Otsuka, A., Komatsu, A. & Ono, M. (2000). Structure, expression, and conserved physical linkage of mouse testicular cell adhesion molecule-1 (TCAM-1) gene. *Genome* **43**, 957–962

33. Komatsu, A., Otsuka, A. & Ono, M. (2002). Novel regulatory regions found downstream of the rat B29/Ig-β gene. *Eur. J. Biochem.* **269**, 1227–1236

34. Osano, K. & Ono, M. (2003). State of histone modification in the rat Ig-beta/growth hormone locus. *Eur. J. Biochem.* **270**, 2532–2539

35. Bvé, A.R., Cavicchia, J.C., Millette, C.F., O'Brien, D.A., Bhatnagar, Y.M. & Dym, M. (1977). Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J. Cell Biol.* **74**, 68–85

36. Goetz, P., Chandley, A.C. & Speed, R.M. (1984). Morphological and temporal sequence of meiotic prophase development at puberty in the male mouse. *J. Cell Sci.* **65**, 249–63

37. Nelson, A.C. & Wardle, F.C. (2013). Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* **140**, 1385–95

38. Berger, S.L. (2002). Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**, 142–148

39. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E. & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–8

40. Zhang, J., Eto, K., Honmyou, A., Nakao, K., Kiyonari, H. & Abé, S. (2011). Neuregulins are essential for spermatogonial proliferation and meiotic initiation in neonatal mouse testis. *Development* **138**, 3159–68

41. Yoneda, R., Takahashi, T., Matsui, H., Takano, N., Hasebe, Y., Ogiwara, K. & Kimura, A.P. (2013). Three testis-specific paralogous serine proteases play different roles in murine spermatogenesis and are involved in germ cell survival during meiosis. *Biol. Reprod.* **88**, 118

42. Lesch, B.J., Dokshin, G.A., Young, R.A., McCarrey, J.R. & Page, D.C. (2013). 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. U. S. A.* **110**, 16061–6

43. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C. & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74

44. Erkek, S., Hisano, M., Liang, C.-Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., van der Vlag, J., Stadler, M.B. & Peters, A.H.F.M. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mol. Biol.* **20**, 868–75

45. Hofmann, M.C., Hess, R.A., Goldberg, E. & Millán, J.L. (1994). Immortalized germ cells undergo meiosis in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5533–7

46. Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L. & Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–9

47. Pennacchio, L.A. & Rubin, E.M. (2001). Genomic strategies to identify mammalian regulatory sequences. *Nat. Rev. Genet.* **2**, 100–9

48. Saffer, J.D., Jackson, S.P. & Annarella, M.B. (1991). Developmental expression of Sp1 in the mouse. *Mol. Cell. Biol.* **11**, 2189–99

49. Ma, W., Horvath, G.C., Kistler, M.K. & Kistler, W.S. (2008). Expression patterns of SP1 and SP3 during mouse spermatogenesis: SP1 down-regulation correlates with two successive promoter changes and translationally compromised transcripts. *Biol. Reprod.* **79**, 289–300

50. Persengiev, S.P., Raval, P.J., Rabinovitch, S., Millette, C.F. & Kilpatrick, D.L. (1996). Transcription factor Sp1 is expressed by three different developmentally regulated messenger ribonucleic acids in mouse spermatogenic cells. *Endocrinology* **137**, 638–46

51. Thomas, K., Sung, D.-Y., Yang, J., Johnson, K., Thompson, W., Millette, C., McCarrey, J., Breitberg, A., Gibbs, R. & Walker, W. (2005). Identification, characterization, and functional analysis of sp1 transcript variants expressed in germ cells during mouse spermatogenesis. *Biol. Reprod.* **72**, 898–907

52. Wierstra, I. (2008). Sp1: emerging roles--beyond constitutive activation of TATA-less housekeeping genes. *Biochem. Biophys. Res. Commun.* **372**, 1–13

53. Safe, S. & Abdelrahim, M. (2005). Sp transcription factor family and its role in cancer. *Eur. J. Cancer* **41**, 2438–48

54. Thomas, K., Wu, J., Sung, D.Y., Thompson, W., Powell, M., McCarrey, J., Gibbs, R. & Walker, W. (2007). SP1 transcription factors in male germ cell development and differentiation. *Mol. Cell. Endocrinol.* **270**, 1–7

55. Onorato, T.M., Brown, P.W. & Morris, P.L. Mono-(2-ethylhexyl) phthalate increases spermatocyte mitochondrial peroxiredoxin 3 and cyclooxygenase 2. *J. Androl.* **29**, 293–303

56. Li, W., Wu, Z., Zhao, J., Guo, S., Li, Z., Feng, X., Ma, L., Zhang, J., Liu, X. & Zhang, Y. (2011). Transient protection from heat-stress induced apoptotic stimulation by metastasis-associated protein 1 in pachytene spermatocytes. *PLoS One* **6**, e26013

57. Wolfe, S.A., Wilkerson, D.C., Prado, S. & Grimes, S.R. (2004). Regulatory factor X2 (RFX2) binds to the H1t/TE1 promoter element and activates transcription of the testis-specific histone H1t gene. *J. Cell. Biochem.* **91**, 375–83

58. Grimes, S.R., Prado, S. & Wolfe, S.A. (2005). Transcriptional activation of the testis-specific histone H1t gene by RFX2 may require both proximal promoter X-box elements. *J. Cell. Biochem.* **94**, 317–26

59. vanWert, J.M., Wolfe, S.A. & Grimes, S.R. (1995). Testis-Specific Expression of the Rat Histone H1t Gene in Transgenic Mice. *Biochemistry* **34**, 8733–8743

60. Jiménez, A., Zu, W., Rawe, V.Y., Pelto-Huikko, M., Flickinger, C.J., Sutovsky, P., Gustafsson, J.-A., Oko, R. & Miranda-Vizuete, A. (2004). Spermatocyte/spermatid-specific thioredoxin-3, a novel Golgi apparatus-associated thioredoxin, is a specific marker of aberrant spermatogenesis. *J. Biol. Chem.* **279**, 34971–82

61. Jiménez, A., Prieto-Álamo, M.J., Fuentes-Almagro, C.A., Jurado, J., Gustafsson, J.-Å., Pueyo, C. & Miranda-Vizuete, A. (2005). Absolute mRNA levels and transcriptional regulation of the mouse testis-specific thioredoxins. *Biochem. Biophys. Res. Commun.* **330**, 65–74

62. Rathke, C., Baarends, W.M., Awe, S. & Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. *Biochim. Biophys. Acta* **1839**, 155–68

63. Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schübeler, D., Stadler, M.B. & Peters, A.H.F.M. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat. Struct. Mol. Biol.* **17**, 679–87

64. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X., Green, R.D., Lobanenkov, V. V, Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M. & Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–12

65. Ong, C.-T. & Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283–93

66. Mercer, E.M., Lin, Y.C., Benner, C., Jhunjhunwala, S., Dutkowski, J., Flores, M., Sigvardsson, M., Ideker, T., Glass, C.K. & Murre, C. (2011). Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors. *Immunity* **35**, 413–25

67. Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S. & Brown, M. (2008). FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. *Cell* **132**, 958–970

68. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. & Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589

69. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. & Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* **8**, 424–36

70. Smale, S.T. & Kadonaga, J.T. (2003). The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449–79

71. Werner, T. (1999). Models for prediction and recognition of eukaryotic promoters. *Mamm. Genome* **10**, 168–175

72. Engström, P.G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Tan, S.L., Yang, L., Kunarso, G., Ng, E.L.-C., Batalov, S., Wahlestedt, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Wells, C., Bajic, V.B., Orlando, V., Reid, J.F., Lenhard, B. & Lipovich, L. (2006). Complex Loci in human and mouse genomes. *PLoS Genet.* **2**, e47

73. Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P. & Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. *Genome Res.* **14**, 62–6

74. Shu, J., Jelinek, J., Chang, H., Shen, L., Qin, T., Chung, W., Oki, Y. & Issa, J.-P.J. (2006). Silencing of bidirectional promoters by DNA methylation in tumorigenesis. *Cancer Res.* **66**, 5077–84

75. Guarguaglini, G., Battistoni, A., Pittoggi, C., Di Matteo, G., Di Fiore, B. & Lavia, P. (1997). Expression of the murine RanBP1 and Htf9-c genes is regulated from a shared bidirectional promoter during cell cycle progression. *Biochem. J.* **325 (Pt 1**, 277–86

76. Calvert, I., Peng, Z., Kung, H. & Raziuddin (1991). Cloning and characterization of a novel sequence-specific DNA-binding protein recognizing the negative regulatory element (NRE) region of the HIV-1 long terminal repeat. *Gene* **101**, 171–176

77. Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Vrielink, J.A.F.O., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., de Laat, W. & Agami, R. (2013). eRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Mol. Cell* **49**, 524–535

78. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H.-S., Glass, C.K. & Rosenfeld, M.G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–20

79. Lam, M.T.Y., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., Lee, C.Y., Watt, A., Grossman, T.R., Rosenfeld, M.G., Evans, R.M. & Glass, C.K. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–5

80. Hecht, A., Berkenstam, A., Strömstedt, P.E., Gustafsson, J.A. & Sippel, A.E. (1988). A progesterone responsive element maps to the far upstream steroid dependent DNase hypersensitive site of chicken lysozyme chromatin. *EMBO J.* **7**, 2063–73

81. Lefevre, P., Witham, J., Lacroix, C.E., Cockerill, P.N. & Bonifer, C. (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. *Mol. Cell* **32**, 129–39

82. Witham, J., Ouboussad, L. & Lefevre, P.F. (2013). A NF-κB-dependent dual promoter-enhancer initiates the lipopolysaccharide-mediated transcriptional activation of the chicken lysozyme in macrophages. *PLoS One* **8**, e59389

83. Wilczek, C., Chayka, O., Plachetka, A. & Klempnauer, K.-H. (2009). Myb-induced chromatin remodeling at a dual enhancer/promoter element involves non-coding rna transcription and is disrupted by oncogenic mutations of v-myb. *J. Biol. Chem.* **284**, 35314–24

84. Gaucher, J., Boussouar, F., Montellier, E., Curtet, S., Buchou, T., Bertrand, S., Hery, P., Jounier, S., Depaux, A., Vitte, A.-L., Guardiola, P., Pernet, K., Debernardi, A., Lopez, F., Holota, H., Imbert, J., Wolgemuth, D.J., Gérard, M., Rousseaux, S. & Khochbin, S. (2012). Bromodomain-dependent stage-specific male genome programming by Brdt. *EMBO J.* **31**, 3809–20

85. Batista, P.J. & Chang, H.Y. (2013). Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* **152**, 1298–1307

86. Dundr, M. & Misteli, T. (2010). Biogenesis of nuclear bodies. *Cold Spring Harb. Perspect. Biol.* **2**, a000711

87. Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A. & Lawrence, J.B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* **33**, 717–726

88. Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T. & Hirose, T. (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2525–30

89. Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B. & Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39

90. Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., Attardi, L.D., Regev, A., Lander, E.S., Jacks, T. & Rinn, J.L. (2010). A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. *Cell* **142**, 409–419

91. Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G. & Gorospe, M. (2012). *LincRNA-p21 Suppresses Target mRNA Translation*. *Mol. Cell* **47**, 648–655

92. Matsubara, S., Takahashi, T. & Kimura, A.P. (2010). Epigenetic patterns at the mouse prolyl oligopeptidase gene locus suggest the CpG island in the gene body to be a novel regulator for gene expression. *Gene* **465**, 17–29

93. Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N. & Nozaki, N. (2008). The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struct. Funct.* **33**, 61–73

94. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Muertter, R.N., Holko, M., Ayanbule, O., Yefanov, A. & Soboleva, A. (2011). NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res.* **39**, D1005–10

95. Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S., Feolo, M., Fingerman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Krasnov, S., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Karsch-Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., Wilbur, W.J., Yaschenko, E. & Ye, J. (2012). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **40**, D13–25

96. Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T.R., Giardine, B.M., Harte, R.A., Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R.M., Learned, K., Li, C.H., Meyer, L.R., Pohl, A., Raney, B.J., Rosenbloom, K.R., Smith, K.E., Haussler, D. & Kent, W.J. (2011). The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–82

97. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25

98. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9

99. Matsubara, S., Kurihara, M. & Kimura, A.P. (2014). A long non-coding RNA transcribed from conserved non-coding sequences contributes to the mouse prolyl oligopeptidase gene activation. *J. Biochem.* **155**, 243–56

100. Matsubara, S., Maruyama, Y. & Kimura, A.P. (2013). A 914-bp promoter is sufficient to reproduce the endogenous prolyl oligopeptidase gene localization in the mouse placenta if not subject to position effect. *Gene* **524**, 114–123

Figure legends

Fig. 1. Tissue distribution and localization of *Tcam1* mRNA. (a) Northern blot analysis of *Tcam1* in various mouse tissues. Total RNAs were purified from the indicated tissues obtained from two- to three-month-old mice. Each lane contained 20 μg RNA. After the agarose gel electrophoresis with formaldehyde, RNA was transferred to a nylon membrane and hybridized with a radio-labeled *Tcam1* probe. The signal was detected by autoradiography. *β-actin* was used as a loading control. (b) Expression of *Tcam1* mRNA in various mouse tissues by a GEO dataset (GSE9954). (c) *Tcam1* mRNA expression during postnatal testicular development. Total RNAs were prepared from testes at the indicated developmental stages, and northern blot analysis was conducted as in (a). (d, e) *In situ* hybridization analysis of *Tcam1* in the mouse testis. Frozen sections (10 μm) were prepared from mouse testes 21 (d) and 28 (e) days after birth. Neighboring sections were hybridized with digoxigenin-labeled sense (SS) or antisense (AS) cRNA probes for *Tcam1*. The signal was detected by using nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate substrates. The sections were counter-stained by methylgreen. Only sections hybridized with the AS probe are shown and the results with the SS probe, which showed no specific signals, are not presented. *Tcam1* was specifically expressed in spermatocytes. The bar represents 100 μm.

Fig. 2. Histone modifications of CNSs at the mouse *Tcam1* locus. (a) A genomic structure of the mouse *Tcam1* locus is illustrated at the top. Exons are indicated by solid and open boxes, which represent the translated and untranslated regions, respectively. The paintings below the gene structure show the sequence homology to the human *TCAM1P* locus. Conserved sequences depicted with blue are exons, and the other conserved regions (pink) are CNSs. Positions of amplicons for ChIP are indicated by gray boxes. Because the resolution of our ChIP analysis was 500–1000 bp, the amplicons at CNS2 and a region between the *Tcam1* promoter and CNS3 were designated CNS1,2 and Tcam1pro-CNS3, respectively. (b) Histone H3K9 acetylation at the mouse *Tcam1* locus. ChIP was conducted with chromatin isolated from testicular germ cells and liver cells. Sheared chromatin was immunoprecipitated with a monoclonal antibody against H3K9ac. DNA purified from the precipitated

(bound) fraction was subjected to real time PCR amplification using the primer pairs shown in (a). The amplification efficiency was normalized by calculating the ratio of the signal in the bound chromatin to that in the input fraction. Because the nucleosome content could vary, the level was also normalized to total histone H3, which was determined by ChIP with anti-histone H3 antibody. The value was further normalized to the comparable signal of the constitutively active *Aip* gene promoter designated as 1.0. The red bar represents the acetylation level in germ cells, and the blue bar in the liver. The immunoprecipitation was also performed with normal mouse IgG instead of the antibody against H3K9ac, and the results are represented by purple bars for germ cells and yellowish green bars for the liver. (c) Histone H3K4 mono-methylation at the mouse *Tcam1* locus. ChIP was conducted as in (b) using a monoclonal antibody against H3K4me1. The methylation levels were calculated and normalized as in (b). All the data are presented as mean ± S.D. from four independent experiments with two sets of testicular germ cells and two adult livers. One-way ANOVA followed by Dunnett's test was used to determine statistical differences in H3K9ac and H3K4me1 between examined regions in germ cells (**$P < 0.01$).

Fig. 3. Histone H3K4 tri-methylation at the mouse *Tcam1* locus. (a) The H3K4me3 pattern based on the ChIP-PCR analysis. ChIP was conducted as in Fig. 2 with chromatin isolated from testicular germ cells and liver cells, using a monoclonal antibody against H3K4me3. The red and blue bars represent the methylation levels in germ cells and in the liver, respectively. The data with IgG were represented by purple bars for germ cells and yellowish green bars for the liver. The modification levels were calculated and normalized as in Fig.2. All the data are expressed as mean ± S.D. from four independent experiments with two sets of testicular germ cells and two adult livers. One-way ANOVA followed by Dunnett's test was used to determine statistical difference in H3K4me3 between the indicated regions in germ cells (* $P <0.05$, ** $P < 0.01$). (b-d) The H3K4me3 patterns based on the ChIP-seq analysis. The ChIP-seq data for H3K4me3 in spermatocytes (b), round spermatids(c) (SRA097278), and liver (d) (GSM769014) were analyzed as described in Materials and Methods. Gene structures of *Smarcd2* and

Tcam1 are depicted by light blue and purple lines and rectangles, respectively. Amplicon positions for ChIP-PCR (Fig. 2a) are indicated by small triangles.

Fig. 4. *In vitro* reporter gene analysis for transcriptional activity of CNS1, CNS2, and CNS3. Reporter gene constructs were generated as indicated at the left side of the graph. In the figure, CNS1, CNS2, and CNS3 are indicated as C1, C2, and C3, respectively (a-c), and the 5' (CNS1-(1-257)) and 3' halves (CNS1-(258-373)) of CNS1 are shown with A and P (b). The constructs were transfected into GC-2spd(ts) (red bar), Hepa1–6 (yellow bar), or NIH3T3-3-4 cells (blue bar) by GeneJuice transfection reagent, and luciferase activity was measured two days later. The construct without any promoter for the luciferase gene was used for a comparison and the luciferase activity of this construct was set to 1.0. The data are presented as mean ± S.D. from four independent experiments. n = 4. (a) CNS1 increased the *Tcam1* promoter activity only in GC-2spd(ts) cells. (b) Enhancer activity of CNS1 in reverse orientation and halved CNS1 in GC-2spd(ts) cells. (c) Enhancer activity of CNS1 at the downstream of the luciferase gene in GC-2spd(ts) cells.

Fig. 5. Promoter activity of CNS1. (a) Bidirectional promoter activity of CNS1 in GC-2spd(ts), Hepa1-6, and NIH3T3-3-4 cells. CNS1 was connected directly to the luciferase gene in both orientations and the reporter gene assay was conducted as in Fig. 4 using three cell lines indicated. (b, c) The effect of poly(A) signal insertion between CNS1 and the *Tcam1* promoter on enhancer activity of CNS1. Using the CNS1-Pro-luc construct, we inserted the poly(A) signal sequence between CNS1 and the *Tcam1* promoter and the resulting construct was CNS1-polyA-Pro-luc. The constructs were transiently transfected into GC-2spd(ts) cells, and transcription of the promoter sequence (left) and the luciferase gene (right) was investigated (b). qRT-PCR was performed with total RNAs isolated from the transfected cells by using the oligo(dT) primer for reverse transcription. The *Gapdh* signal was amplified as a control and the expression level was presented relative to *Gapdh*. While transcription of the *Tcam1* promoter was greatly reduced by inserting the poly(A) signal, the luciferase gene expression

was not changed. Similarly, the luciferase activity was not affected by poly(A) signal insertion, and CNS1 increased *Tcam1* promoter activity to a comparable level with the construct without insertion (c).

Fig. 6. Expression of *lncRNA-Tcam1* in mouse tissues and the testis. (a) Schematic drawing of a 5' upstream region of the mouse *Tcam1* gene. The region transcribed as *lncRNA-Tcam1* is depicted by a gray box with an arrow indicating the transcriptional direction. A full length of *lncRNA-Tcam1* was determined by RACE analysis and the 2404-bp transcript contained the whole CNS2 and a part of CNS1. Since we observed the poly(A) tail in the subclones obtained by 3'RACE, the *lncRNA-Tcam1* transcript was presumed to be polyadenylated. (b) Expression of *lncRNA-Tcam1* and *Smarcd2* in various mouse tissues. RT-PCR was conducted by using total RNAs prepared from 8 adult mouse tissues with the oligo(dT) primer. *Gapdh* was amplified as an internal control. For detecting the *lncRNA-Tcam1* signal, a primer set of lncRNA-Tcam1-a was used in this analysis (Table 1). The cycle numbers of PCR were 30 for *Gapdh* and 40 for *Smarcd2* and *lncRNA-Tcam1*. (c) *lncRNA-Tcam1* expression during postnatal testicular development. Testes were collected at the indicated ages and cDNAs were generated by reverse transcription using an antisense primer specific to *lncRNA-Tcam1* or *Gapdh*. PCR was conducted to amplify the *lncRNA-Tcam1* transcript by using a primer set of lncRNA-Tcam1-b (Table 1). (d) Expression of *lncRNA-Tcam1* in testicular germ and somatic cells. Adult testes were fractionated into germ, Sertoli, and Leydig cells according to the procedure described in Materials and Methods. The purity of each cell fraction was calculated by the marker genes expression, and the germ, Sertoli, and Leydig cell fractions was estimated to contain 80%, 70%, and 68% of each cell type, respectively. cDNAs were prepared by reverse transcription with the oligo(dT) primer and the *lncRNA-Tcam1* signal was amplified with the lncRNA-Tcam1-b primer pairs. *Gapdh* was used as an internal control. (e) Subcellular localization of *lncRNA-Tcam1* in testicular germ cells. Mouse germ cells were isolated from adult testes as in (d) and fractionated into nuclear and cytoplasmic subfractions. Total RNAs were purified from both subfractions, and RT-PCR was performed by using the oligo(dT) primer for reverse transcription. A primer pair of lncRNA-Tcam1-b was used to detect

lncRNA-Tcam1 expression. *Gapdh*(in5-ex6) was amplified using primers designed in intron 5 and exon 6 to detect immature mRNA which should be localized only in the nucleus. *Gapdh*(ex5-ex6) was amplified by using primers designed in exon 5 and exon 6 to detect mature mRNA which is thought to be mainly localized in the cytoplasm. *Gapdh*(ex6-ex6) was used to amplify both immature and mature RNAs.

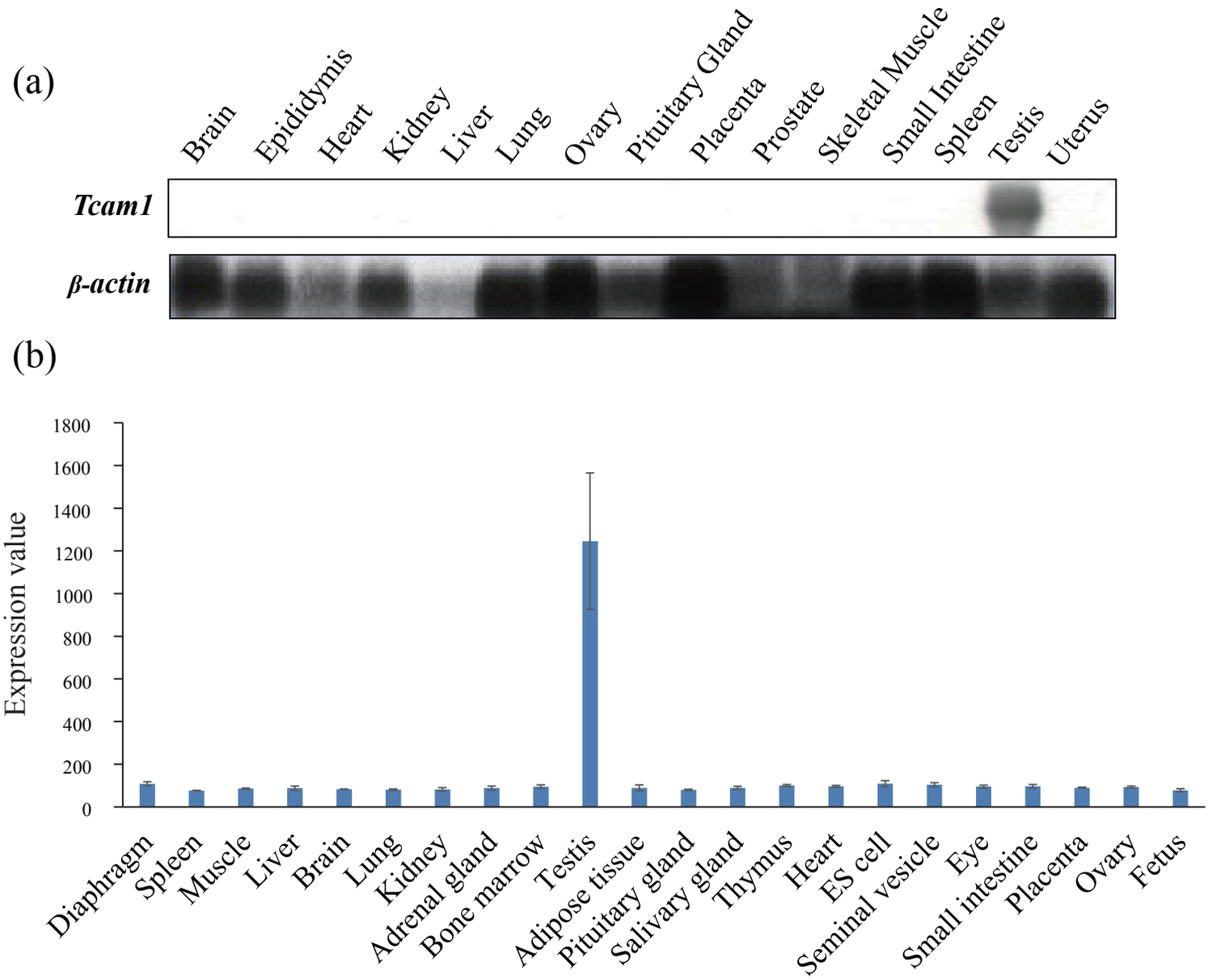
Fig. 7. Effects of mutation of three Sp1 binding sites on promoter and enhancer activity of CNS1. (a) A multiple alignment of CNS1 from seven mammalian species. CNS1 sequences were obtained for seven mammalian species, *Mus musculus*, *Rattus norvegicus*, *Homo sapiens*, *Macaca mulatta*, *Nomascus leucogenys*, *Bos Taurus*, and *Mustela putorius furo*, and compared using the DNASIS-Pro software (HITACHI Software Engineering, Yokohama, Japan). Conserved nucleotides among all seven species are indicated by asterisks. Three hexanucleotide elements for the Sp1 transcription factor binding site are highly conserved and marked by red boxes. (b) Enhancer activity of CNS1 with mutated Sp1 binding sites. We prepared CNS1 in which all three Sp1 sites were mutated not to be recognized by Sp1 and connected the mutated CNS1 at upstream of the *Tcam1* promoter. The construct was transfected into GC-2spd(ts), Hepa1-6, and NIH3T3-3-4 cells, and the luciferase activity was measured as in Fig. 4. The Tcam1-Pro-luc and CNS1-Pro-luc constructs were transfected for comparison. The data are presented as mean ± S.D. from four independent experiments. n = 4. Student's *t* test was performed to compare the luciferase activity of CNS1-Pro-luc with that of mutCNS1-Pro-luc. The luciferase activity of mutCNS1-Pro-luc was significantly higher than that of CNS1-Pro-luc in GC-2spd(ts) and NIH3T3-3-4 cells. $^{**}P < 0.01$. (c) Bidirectional promoter activity of CNS1 with mutated Sp1 binding sites. The mutated CNS1 was directly connected to the luciferase gene in both directions and the luciferase activity was measured as in (b). The activity of reversed mutCNS1-luc was significantly lower than that of reversed CNS1-luc in all the cell lines. $^{**}P < 0.01$, $^{*}P < 0.05$.

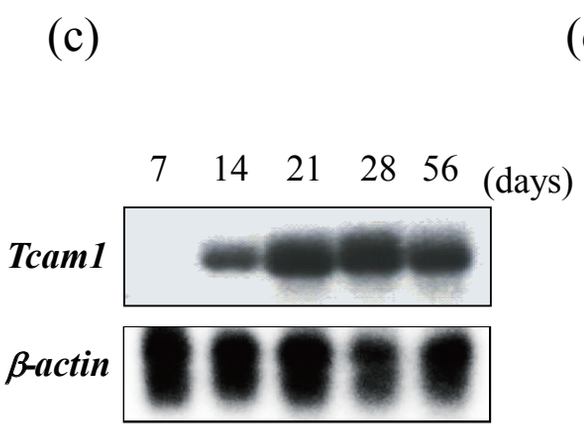
Fig. 8. Effects of CNS1 deletion on the *lncRNA-Tcam1* expression and *Tcam1* promoter activity and the

relationship between *lncRNA-Tcam1* and *Tcam1*. (a) A schematic drawing of transgene constructs. A 6.9-kb genomic fragment, which includes CNS1, *lncRNA-Tcam1*, and the *Tcam1* promoter, was obtained from a BAC clone. This fragment was linked to the EGFP gene and the resulting construct was lncRNA-6.9kb-EGFP. The ΔCNS1-lncRNA-EGFP construct was generated by deleting a 861-bp region including CNS1 from lncRNA-6.9kb-EGFP. (b, c) Expression of *lncRNA-Tcam1* and EGFP in GC-2spd(ts) cell clones. The two constructs were transfected into GC-2spd(ts) cells, and cell clones that were integrated with the transgene were selected. Eleven cell clones were obtained for the lncRNA-6.9kb-EGFP construct and ten clones were for ΔCNS1-lncRNA-EGFP. Using total RNAs purified from these cell clones, qRT-PCR was performed to determine the expression levels of *lncRNA-Tcam1* (b) and EGFP (c) per transgene copy number. Reverse transcription was conducted with the oligo(dT) primer, and an internal control was the *Gapdh* gene. Statistical evaluations for the comparison of the *lncRNA-Tcam1* and EGFP expression in cell clones with lncRNA-6.9kb-EGFP and ΔCNS1-lncRNA-EGFP were done by using the Mann-Whitney *U* test. The expression levels of both *lncRNA-Tcam1* (b) and EGFP mRNA (c) were significantly higher in clones with lncRNA-6.9kb-EGFP than in those with ΔCNS1-lncRNA-EGFP. ** $P < 0.01$, * $P < 0.05$. (d) Expression of EGFP and *lncRNA-Tcam1* in each cell clone with lncRNA-6.9kb-EGFP. The blue bar represents EGFP expression and the red bar shows the *lncRNA-Tcam1* level. The correlation coefficient statistical analysis implied no relationship between the expression levels of EGFP and *lncRNA-Tcam1*.

Fig. 9. A model of CNS1 function as an enhancer and a bidirectional promoter. Exon 1 of the *Smarcd2* and *Tcam1* gene is indicated by using solid and open boxes, which represent translated and untranslated region, respectively, and a transcribed region of *lncRNA-Tcam1* is drawn by a gray box. Transcriptional directions of these genes and lncRNA are shown by horizontal arrows. Positions of CNS1, CNS2, and CNS3 are indicated by yellow boxes (C1, C2, C3). In testicular germ cells, CNS1 functions as a bidirectional promoter of *lncRNA-Tcam1* (blue line) and *Smarcd2* mRNA (pink line). Sp1 may contribute to the CNS1 promoter activity for the *Smarcd2* gene. CGI is completely hypo-methylated,

and high levels of H3K9ac, H3K4me3, and H3K4me1 are observed at CNS1 and CNS2. The *Tcam1* promoter and CNS3 are also marked with H3K9ac and H3K4me3 in germ cells. In spermatocytes, CNS1 can work as an enhancer to increase *Tcam1* expression (green arrow). While *Smarcd2* and *Tcam1* mRNAs are probably translated to proteins, *lncRNA-Tcam1* probably functions as an RNA molecule, possibly contributing to gene regulation at other loci. In the liver, CNS1 only functions as a unidirectional promoter for *Smarcd2*, and Sp1 may contribute to this activity. In this tissue, CGI is hypo-methylated as in spermatocytes, but no histone modification markers for active chromatin are present except for weak association of H3K4me3.


(a)
Brain
Epididymis
Heart
Kidney
Liver
Lung
Ovary
Pituitary Gland
Placenta
Prostate
Skeletal Muscle
Small Intestine
Spleen
Testis
Uterus
Tcam1
β-actin
(b)
Expression value
1800
1600
1400
1200
1000
800
600
400
200
0
Diaphragm
Spleen
Muscle
Liver
Brain
Lung
Kidney
Adrenal gland
Bone marrow
Testis
Adipose tissue
Pituitary gland
Salivary gland
Thymus
Heart
ES cell
Seminal vesicle
Eye
Small intestine
Placenta
Ovary
Fetus


(c)
7
14
21
28
56
(days)
Tcam1
β-actin

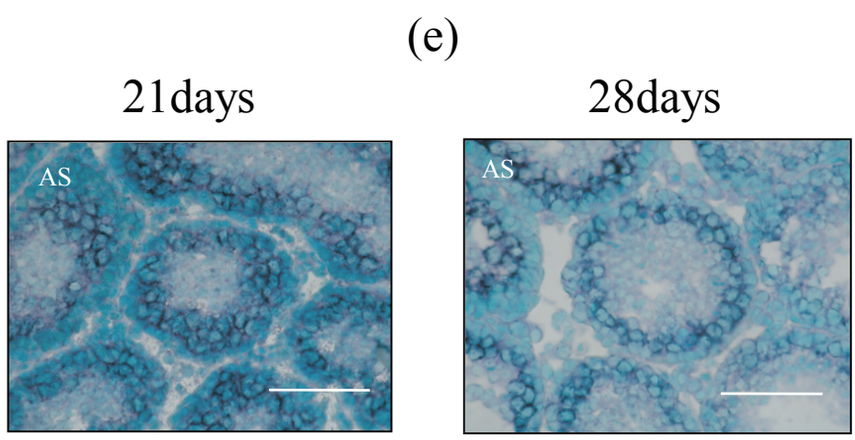

(d)
21days
AS
(e)
28days
AS

Fig. 1

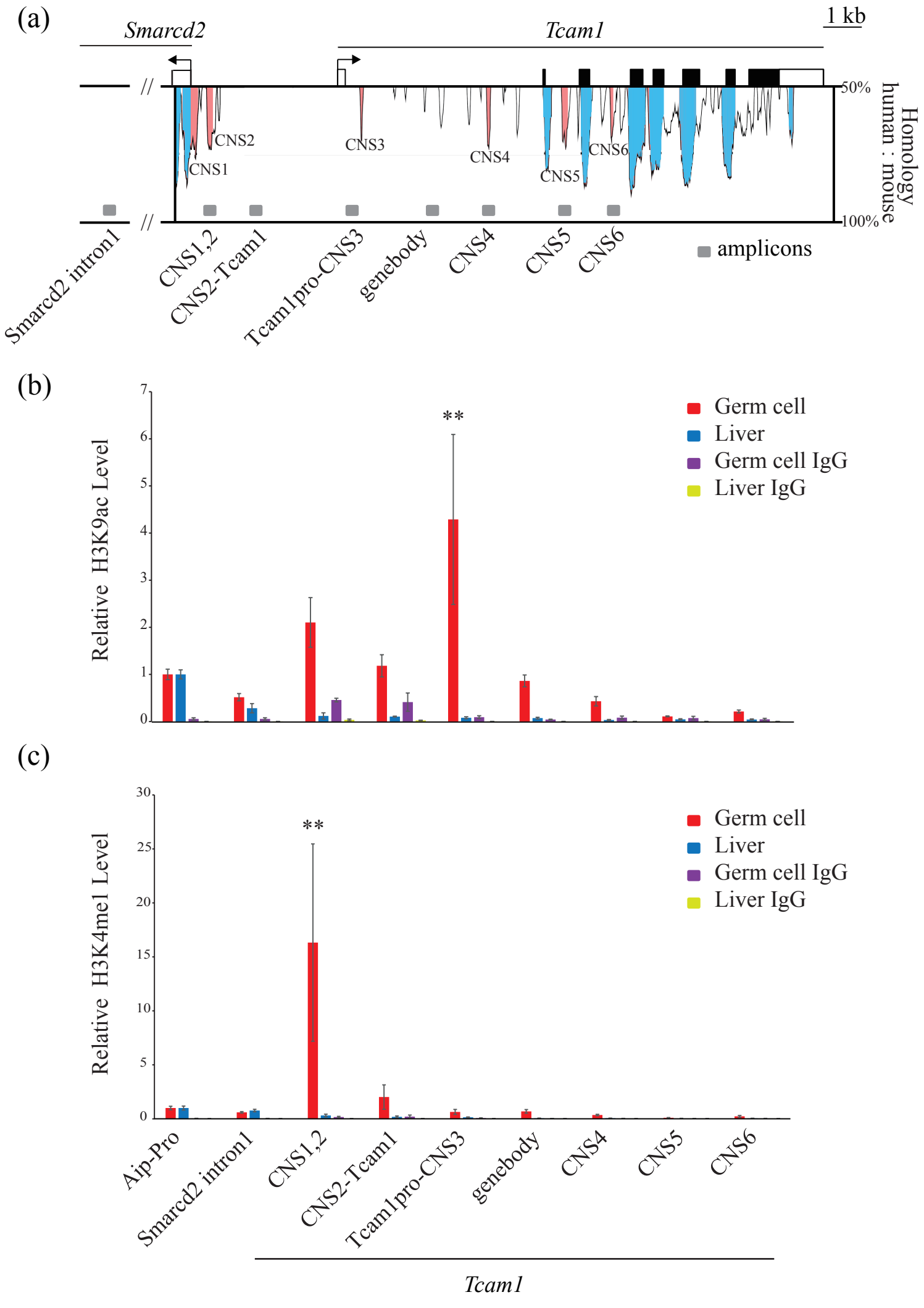

(a)
Smarcd2
Tcam1
1 kb
50%
100%
Homology human : mouse
CNS1
CNS2
CNS3
CNS4
CNS5
CNS6
amplicons
Smarcd2 intron1
CNS1,2
CNS2-Tcam1
Tcam1pro-CNS3
genebody
CNS4
CNS5
CNS6
(b)
Relative H3K9ac Level
Germ cell
Liver
Germ cell IgG
Liver IgG
**
(c)
Relative H3K4me1 Level
Germ cell
Liver
Germ cell IgG
Liver IgG
**
Aip-Pro
Smarcd2 intron1
CNS1,2
CNS2-Tcam1
Tcam1pro-CNS3
genebody
CNS4
CNS5
CNS6
Tcam1

Fig. 2

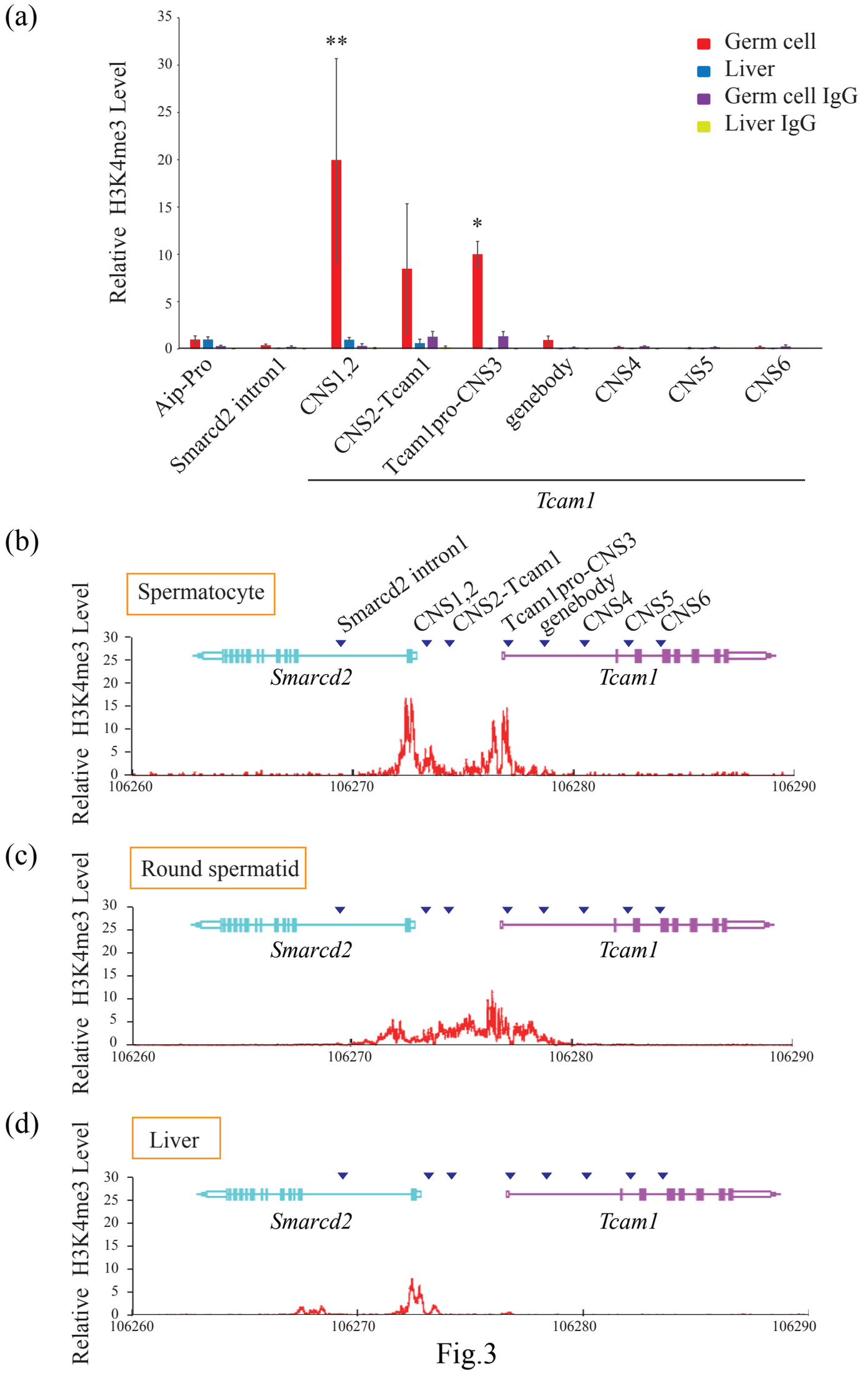

(a)
Relative H3K4me3 Level
35
30
25
20
15
10
5
0
**
*
Germ cell
Liver
Germ cell IgG
Liver IgG
Aip-Pro
Smarcd2 intron1
CNS1,2
CNS2-Tcam1
Tcam1pro-CNS3
genebody
CNS4
CNS5
CNS6
Tcam1
(b)
Spermatocyte
Smarcd2 intron1
CNS1,2
CNS2-Tcam1
Tcam1pro-CNS3
genebody
CNS4
CNS5
CNS6
Smarcd2
Tcam1
Relative H3K4me3 Level
30
25
20
15
10
5
0
106260
106270
106280
106290
(c)
Round spermatid
Smarcd2
Tcam1
Relative H3K4me3 Level
106260
106270
106280
106290
(d)
Liver
Smarcd2
Tcam1
Relative H3K4me3 Level
106260
106270
106280
106290

Fig.3

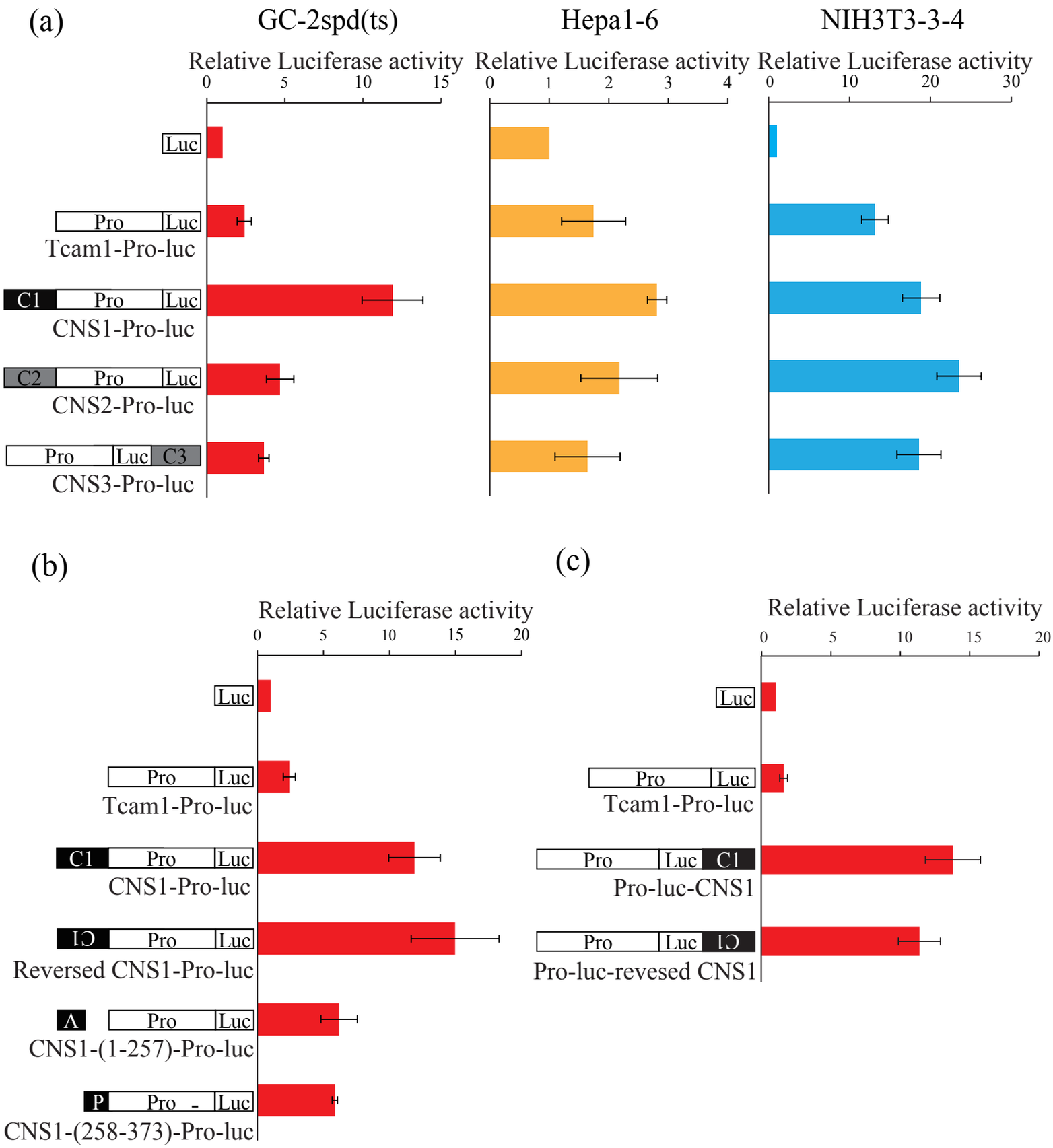

(a)
GC-2spd(ts)
Hepa1-6
NIH3T3-3-4
Relative Luciferase activity
0
5
10
15
1
2
3
4
20
30
Luc
Pro
Tcam1-Pro-luc
C1
CNS1-Pro-luc
C2
CNS2-Pro-luc
C3
CNS3-Pro-luc
(b)
(c)
Pro-luc-CNS1
Reversed CNS1-Pro-luc
A
CNS1-(1-257)-Pro-luc
P
CNS1-(258-373)-Pro-luc
Pro-luc-revesed CNS1

Fig. 4

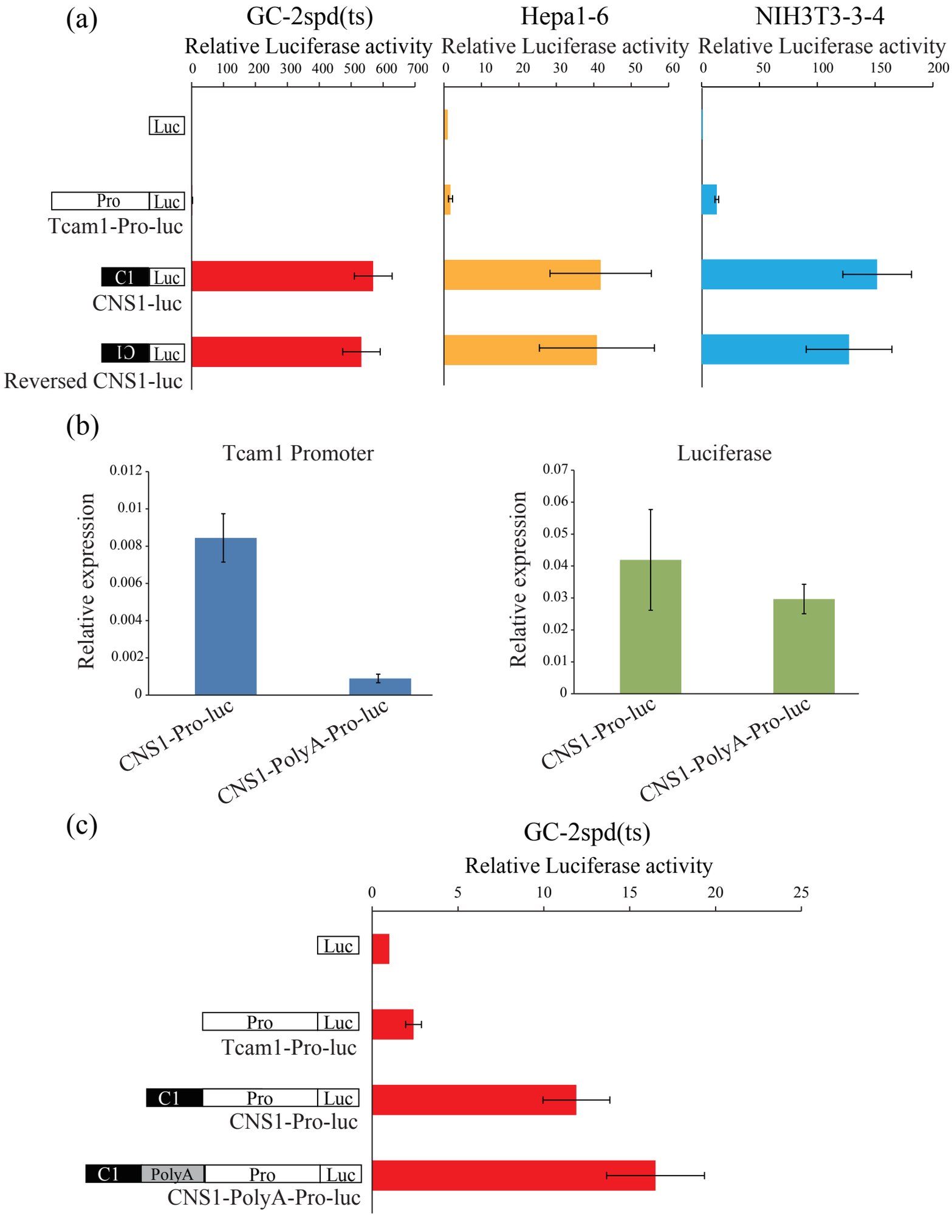

(a)
GC-2spd(ts)
Hepa1-6
NIH3T3-3-4
Relative Luciferase activity
Relative Luciferase activity
Relative Luciferase activity
0 100 200 300 400 500 600 700
0 10 20 30 40 50 60
0 50 100 150 200
Luc
Pro Luc
Tcam1-Pro-luc
C1 Luc
CNS1-luc
C1 Luc
Reversed CNS1-luc
(b)
Tcam1 Promoter
Luciferase
Relative expression
0.012
0.01
0.008
0.006
0.004
0.002
0
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
CNS1-Pro-luc
CNS1-PolyA-Pro-luc
(c)
GC-2spd(ts)
Relative Luciferase activity
0 5 10 15 20 25
Luc
Pro Luc
Tcam1-Pro-luc
C1 Pro Luc
CNS1-Pro-luc
C1 PolyA Pro Luc
CNS1-PolyA-Pro-luc

Fig. 5

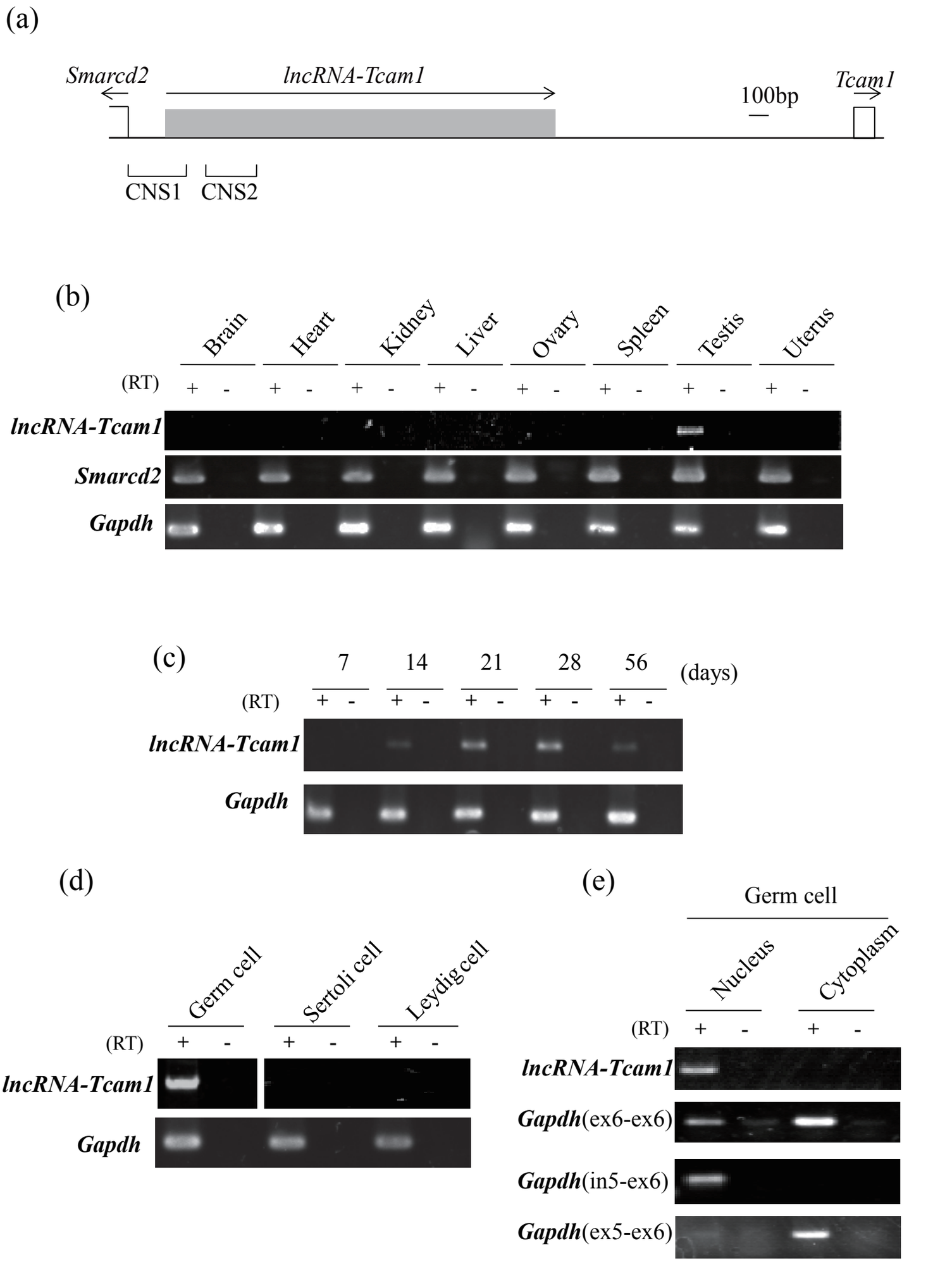
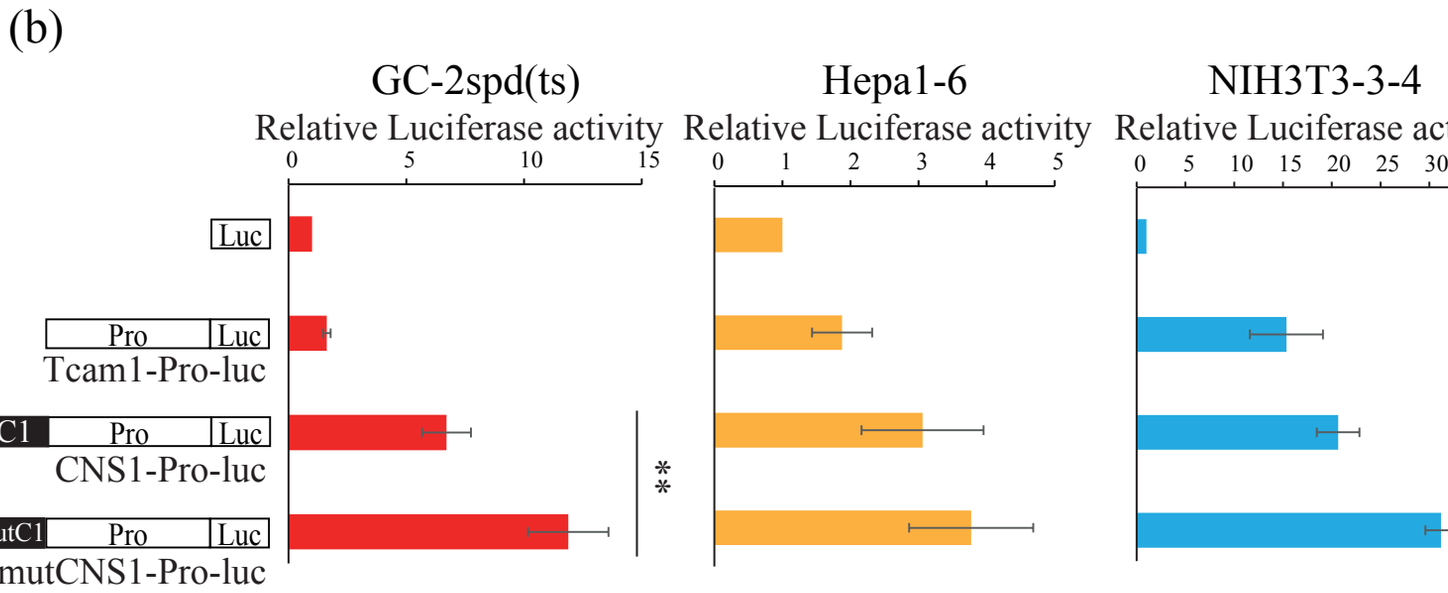

(a)
Smarcd2
lncRNA-Tcam1
100bp
Tcam1
CNS1
CNS2
(b)
Brain
Heart
Kidney
Liver
Ovary
Spleen
Testis
Uterus
(RT)
+ - + - + - + - + - + - + - + -
lncRNA-Tcam1
Smarcd2
Gapdh
(c)
7
14
21
28
56
(days)
(RT)
+ - + - + - + - + -
lncRNA-Tcam1
Gapdh
(d)
Germ cell
Sertoli cell
Leydig cell
(RT)
+ - + - + -
lncRNA-Tcam1
Gapdh
(e)
Germ cell
Nucleus
Cytoplasm
(RT)
+ - + -
lncRNA-Tcam1
Gapdh(ex6-ex6)
Gapdh(in5-ex6)
Gapdh(ex5-ex6)

Fig.6

(a)
Mus musculus
Rattus norvegicus
Homo sapiens
Macaca mulatta
Nomascus leucogenys
Bos taurus
Mustela putorius furo
Sp1
Sp1
Sp1


(b)
GC-2spd(ts)
Hepa1-6
NIH3T3-3-4
Relative Luciferase activity
Luc
Pro
Luc
Tcam1-Pro-luc
C1
Pro
Luc
CNS1-Pro-luc
mutC1
Pro
Luc
mutCNS1-Pro-luc
**

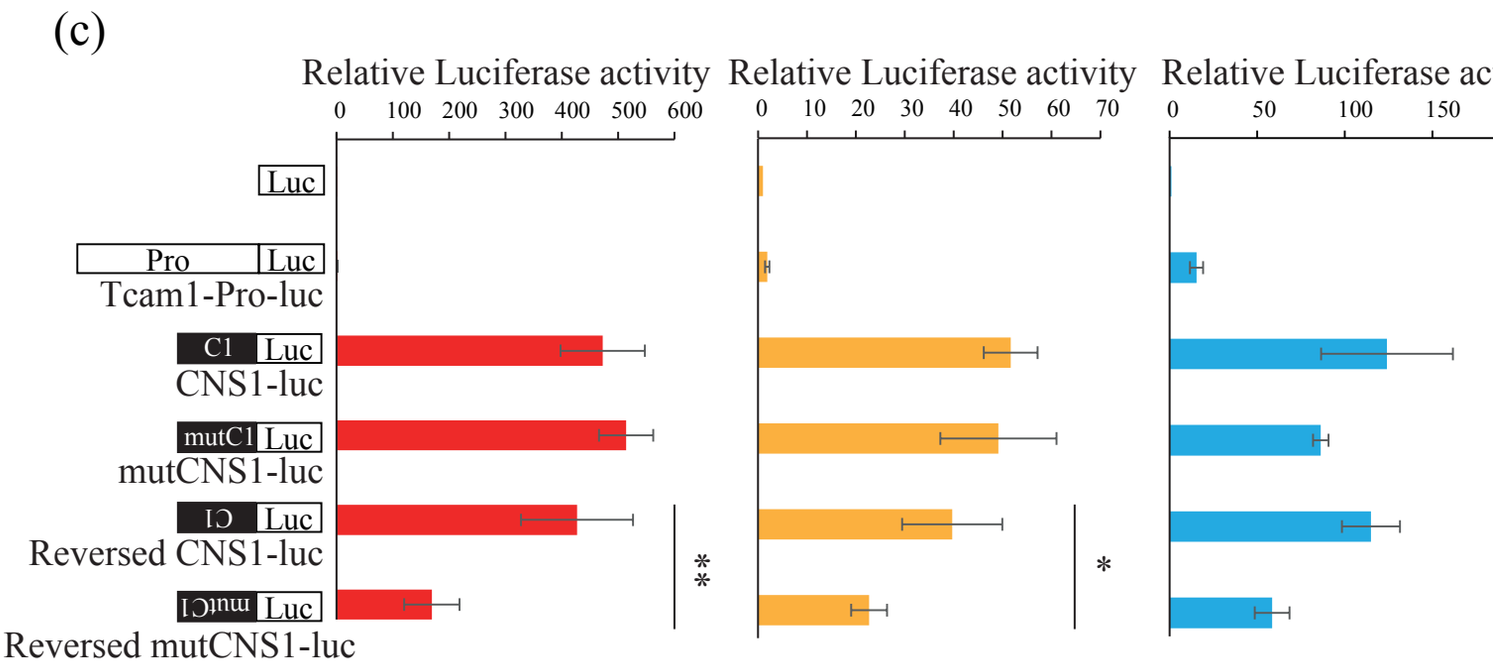

(c)
Relative Luciferase activity
Luc
Pro
Luc
Tcam1-Pro-luc
C1
Luc
CNS1-luc
mutC1
Luc
mutCNS1-luc
Luc
Reversed CNS1-luc
Luc
Reversed mutCNS1-luc
**
*
**

Fig. 7

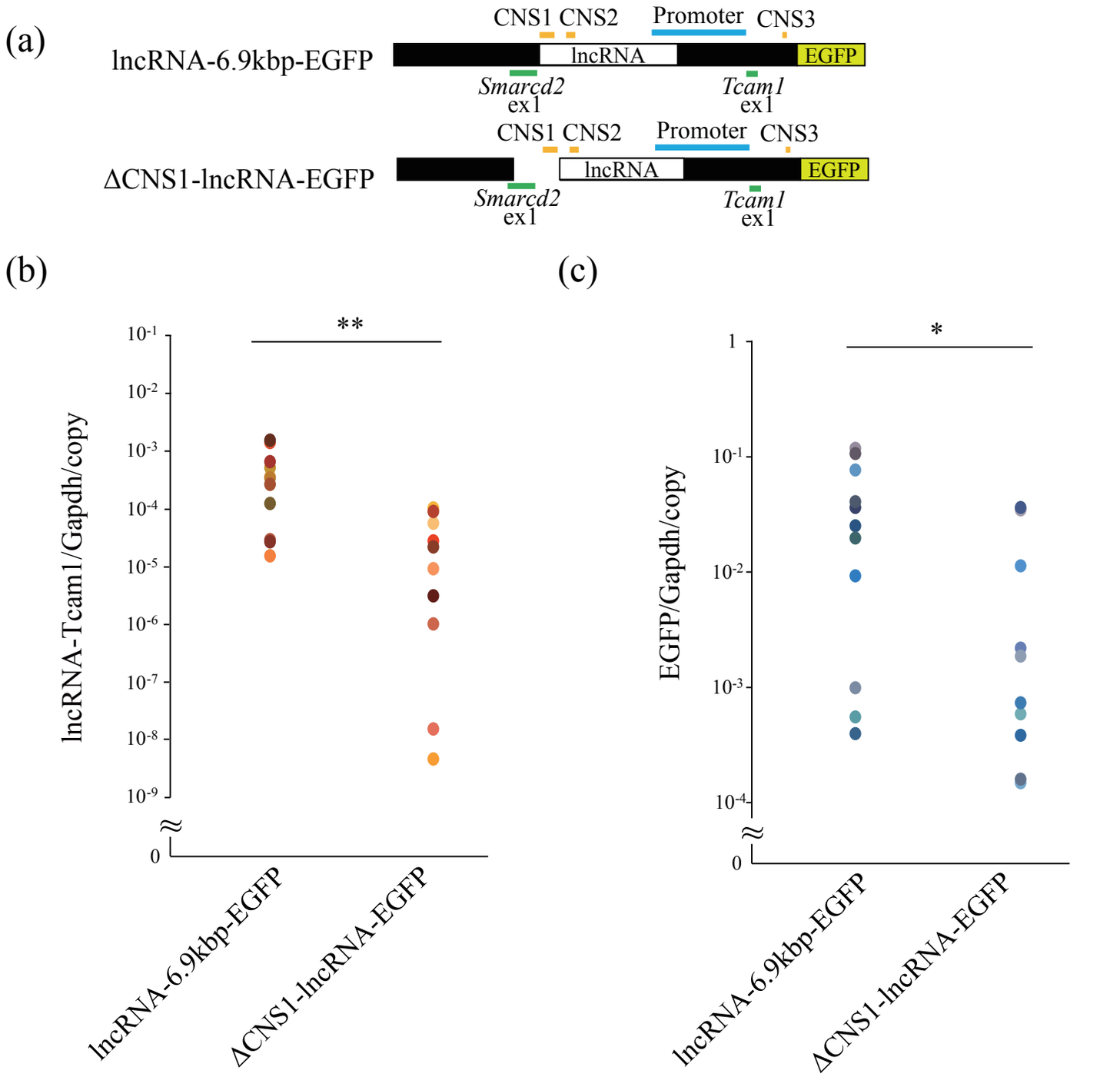

(a)
lncRNA-6.9kbp-EGFP
CNS1 CNS2
Promoter
CNS3
lncRNA
EGFP
Smarcd2 ex1
Tcam1 ex1
ΔCNS1-lncRNA-EGFP
(b)
lncRNA-Tcam1/Gapdh/copy
**
(c)
EGFP/Gapdh/copy
*
lncRNA-6.9kbp-EGFP
ΔCNS1-lncRNA-EGFP
(d)
Relative Expression per copy
EGFP
lncRNA-Tcam1
Clone #

Fig. 8

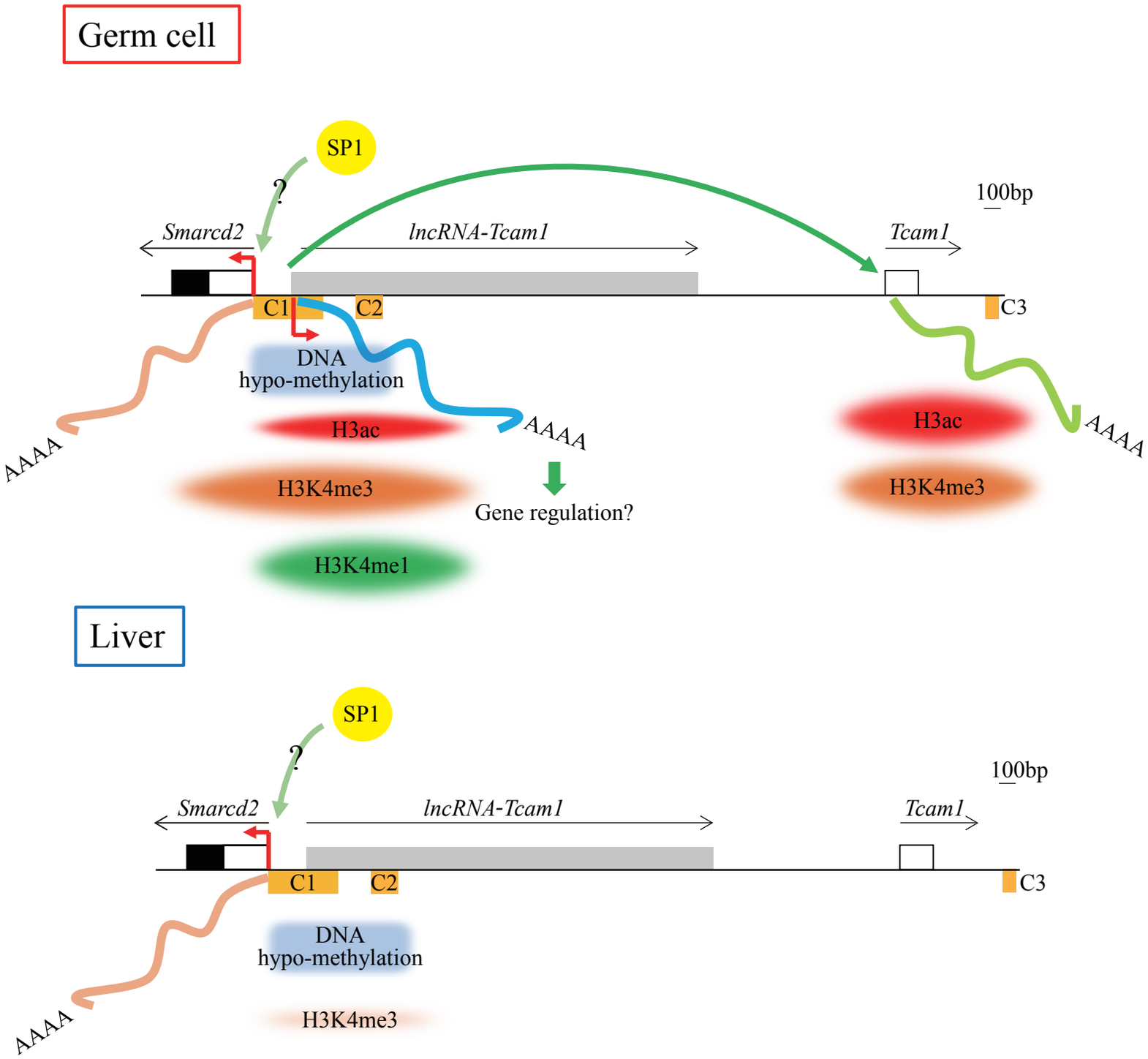
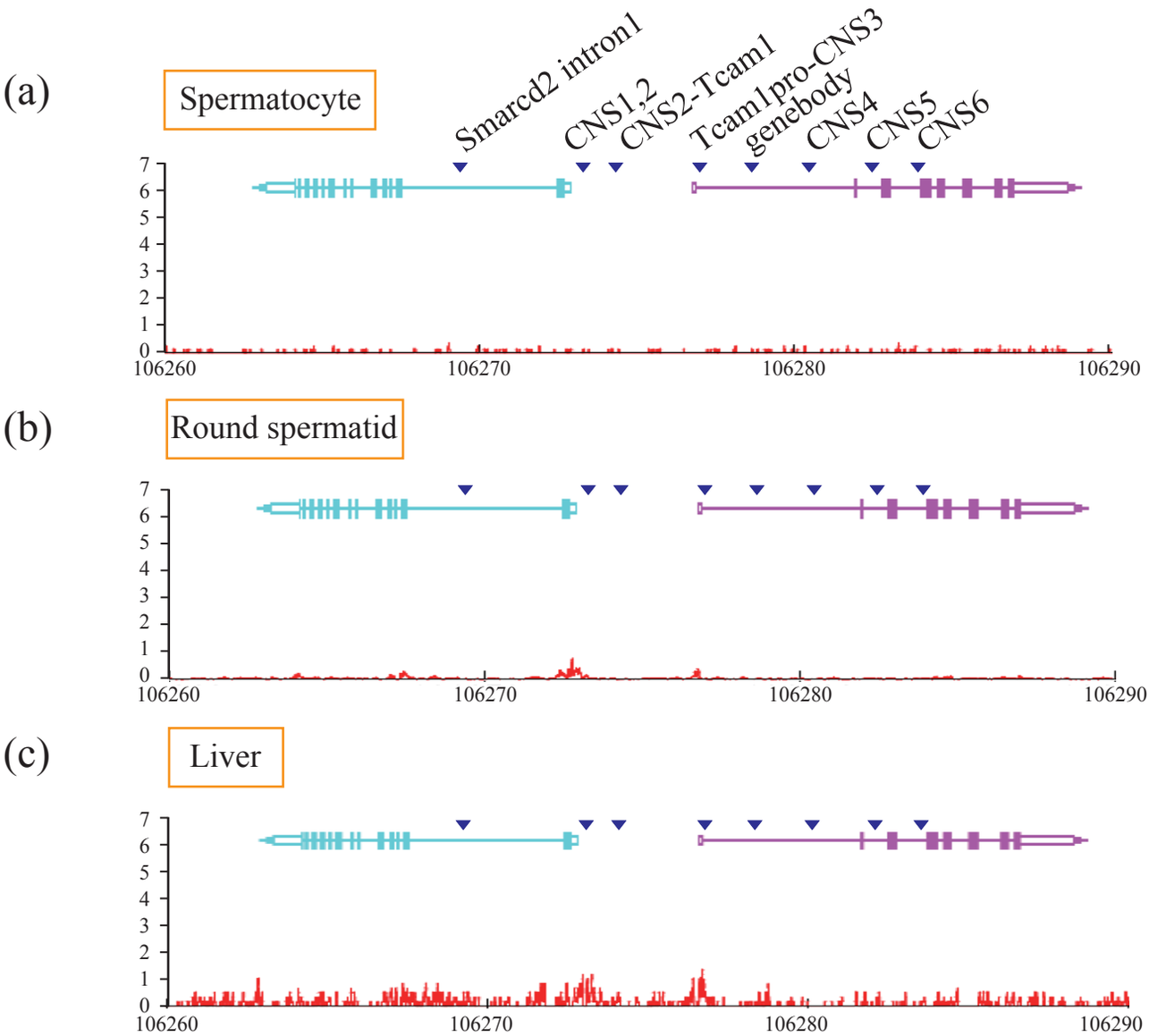
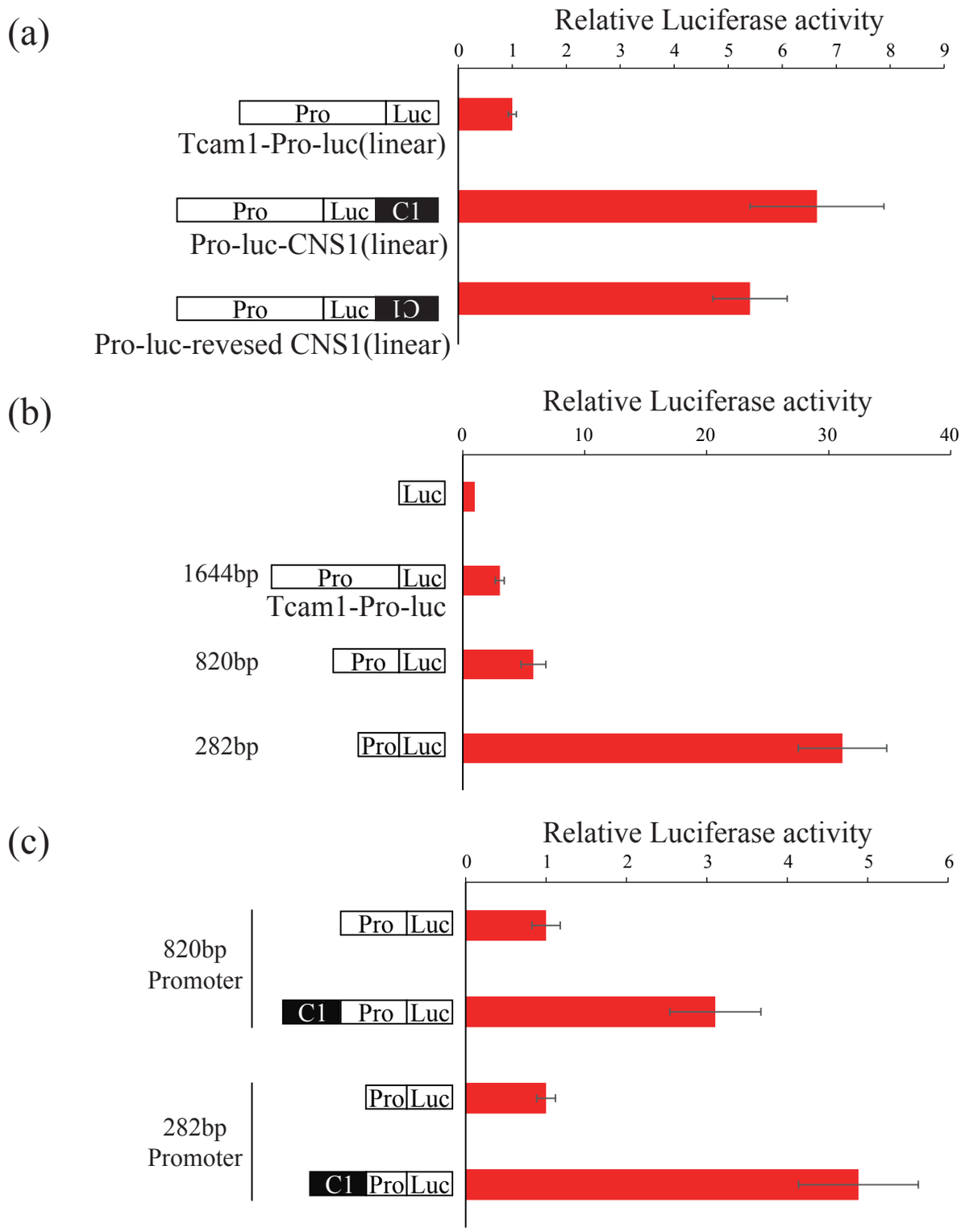

Germ cell
SP1
?
100bp
Smarcd2
lncRNA-Tcam1
Tcam1
C1
C2
C3
DNA
hypo-methylation
H3ac
H3K4me3
H3K4me1
AAAA
AAAA
AAAA
Gene regulation?
Liver
SP1
?
100bp
Smarcd2
lncRNA-Tcam1
Tcam1
C1
C2
C3
DNA
hypo-methylation
H3K4me3
AAAA

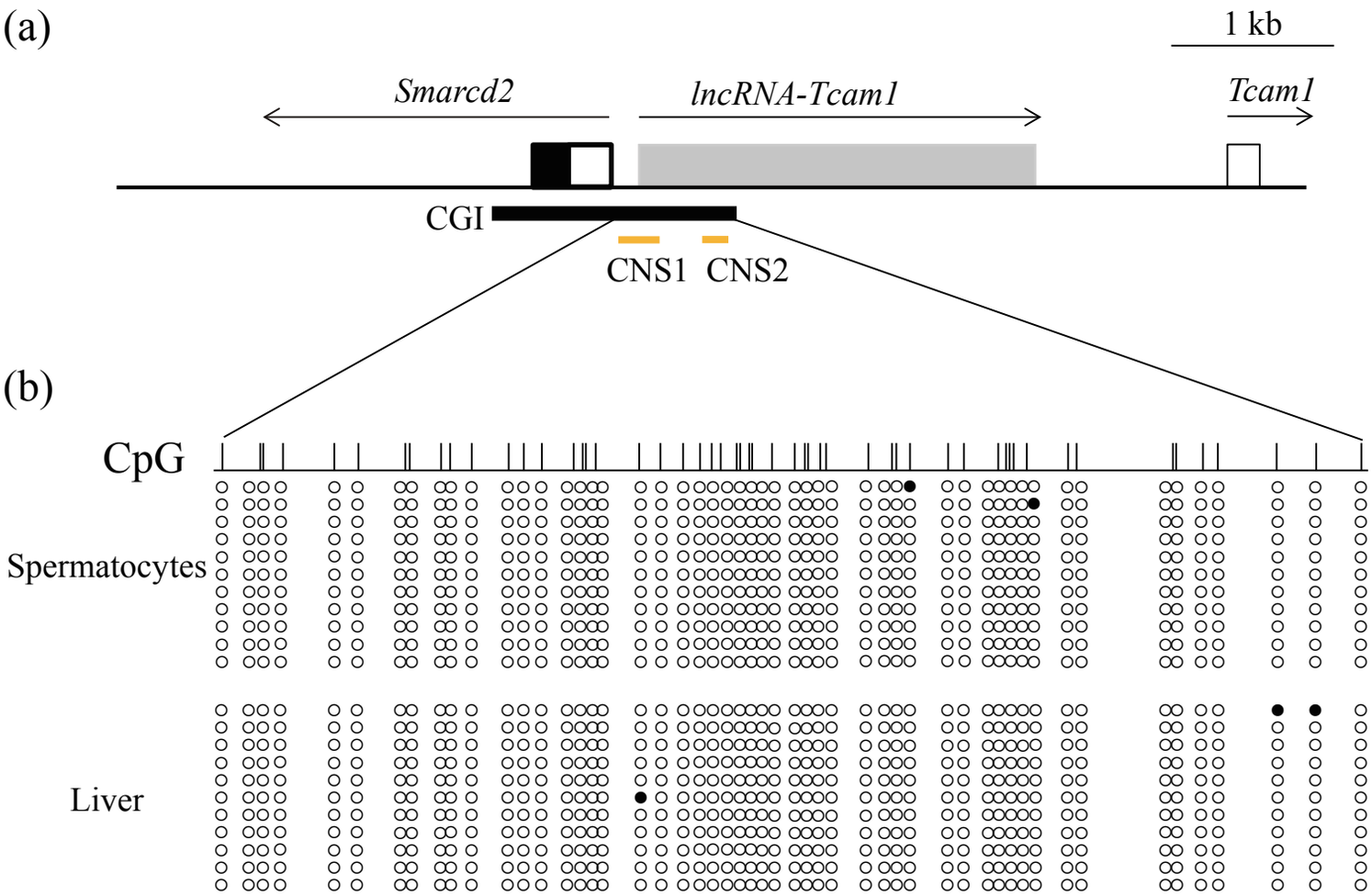
Fig. 9



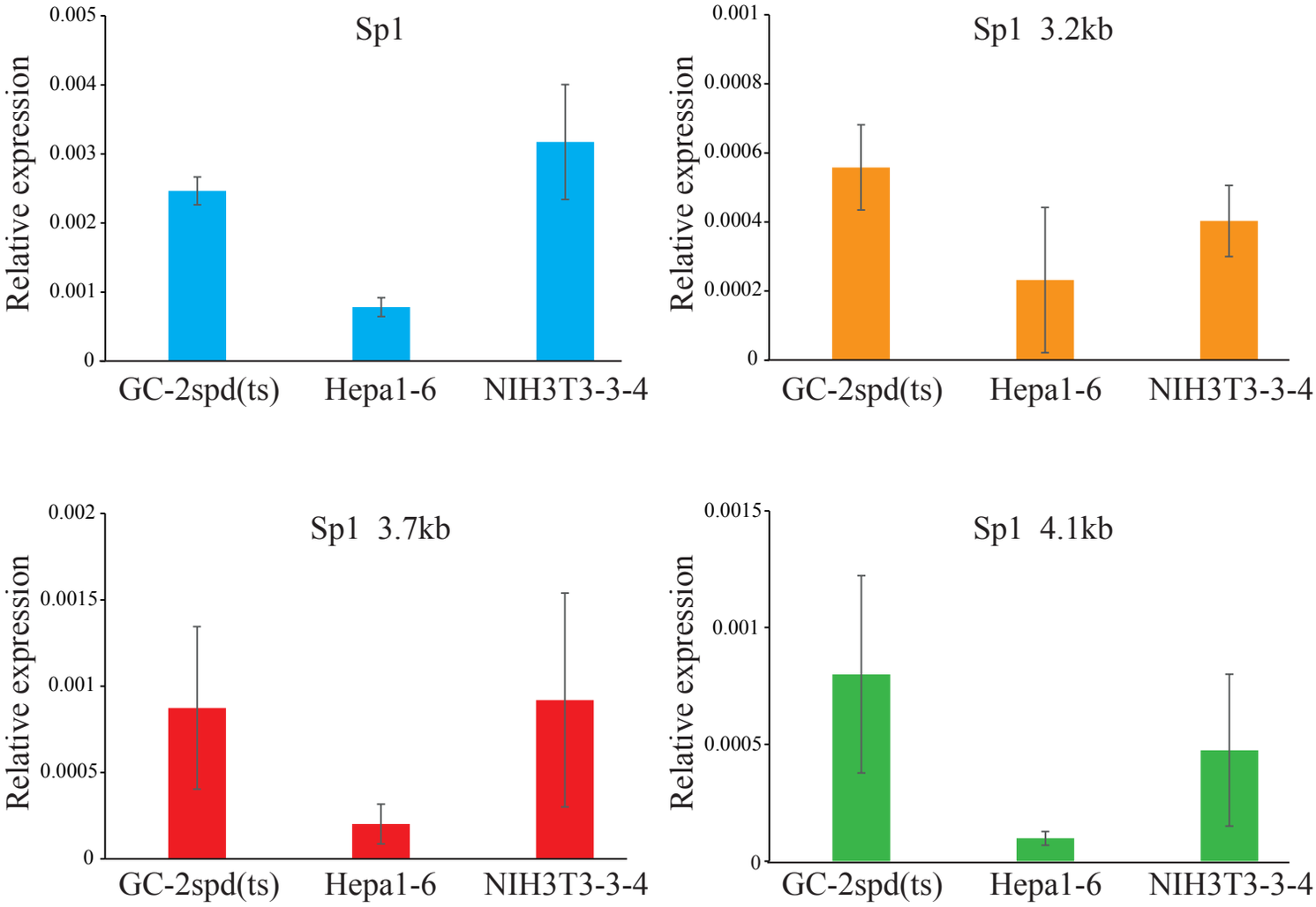
Supplementary Fig. 1. Data of input DNA for H3K4me3 in spermatocytes (a), round spermatids (b) (SRA097278), and liver (c) (GSM769014). Gene structures of *Smarcd2* and *Tcam1* are depicted by light blue and purple lines and rectangles, respectively. Amplicon positions for ChIP-PCR (Fig. 2a) are indicated by small triangles. No significant peaks were observed at the *Smarcd2-Tcam1* locus.



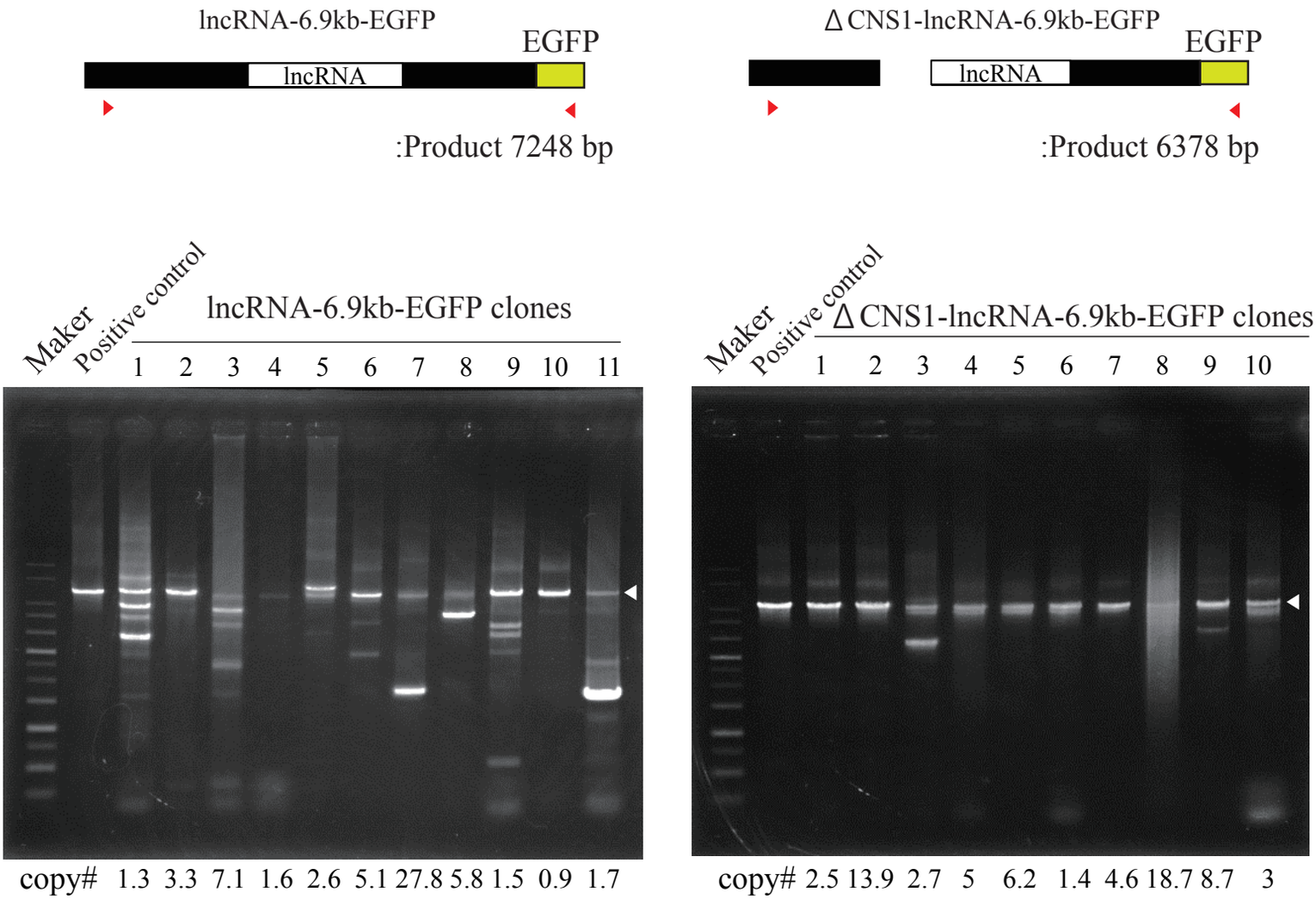
Supplementary Fig. 2. Transcriptional activity of Tcam1 promoter and CNS1 enhancer. (a) Transcriptional activity of CNS1 in linearized constructs. The Tcam1-Pro-luc, Pro-luc-CNS1, and Pro-luc-reversed CNS1 constructs were linearized by digestion with SalI before transfection. The constructs were transfected into GC-2spd(ts) cells by using GeneJuice transfection reagent, and luciferase activity was measured two days later. The value for the Tcam1-Pro-luc construct was set to 1.0. The data are presented as mean ± S.D. from four independent experiments. n = 4. CNS1 enhanced Tcam1 promoter activity in linearized constructs. (b) Transcriptional activity of various sizes of Tcam1 promoter. 820-bp and 282-bp Tcam1 promoter sequences were prepared by genome PCR and digestion with XbaI, respectively, and luciferase activity was measured as in (a). The construct without any promoter for the luciferase gene was used for a comparison and its activity was set to 1.0. Shorter sequence showed higher promoter activity. (c) Transcriptional activity of CNS1 to 820-bp and 282-bp Tcam1 promoters. CNS1 was connected to upstream of 820-bp and 282-bp Tcam1 promoters, and luciferase activity was measured as in (a). The value for constructs without CNS1 was set to 1.0. CNS1 significantly enhanced activity of both promoters.



Supplementary Fig. 3. DNA methylation patterns of CGI in spermatocytes and the liver. (a) Structure of the mouse *Tcam1* and *Smarcd2* locus. Exons of *Tcam1* and *Smarcd2* are depicted by open and filled boxes that represent untranslated and translated sequences, respectively. A region from which *lncRNA-Tcam1* is transcribed is depicted by a gray box. Transcriptional directions of the genes and lncRNA are shown by horizontal arrows. Positions of CGI, CNS1, and CNS2 are drawn with solid horizontal lines below the gene structure. To reveal DNA methylation states of the *Smarcd2* and *lncRNA-Tcam1* promoter, a 747-bp region of CGI, which included both CNS1 and CNS2, was investigated. (b) A detailed map of the 747-bp region in CGI and DNA methylation patterns. The 747-bp region contains 54 CpG dinucleotides which are shown with vertical lines. The methylation state of each CpG was determined by bisulfite sequencing. Genomic DNA was purified from 1×106 spermatocytes and a piece of liver and treated with sodium bisulfite. The 747-bp region was amplified by PCR and 10 subclones of the amplified product for each tissue were sequenced. Open and filled circles represent unmethylated and methylated cytosines, respectively.



Supplementary Fig. 4. Expression of *Sp1* mRNAs in GC-2spd(ts), Hepa1-6, and NIH3T3-3-4 cells. The qRT-PCR analysis was performed with total RNAs isolated from indicated cell lines. The oligo(dT) primer was used for reverse transcription. A primer pair for 'Sp1' was to detect a most common transcript which is ubiquitously expressed in various tissues. The other primers were for detection of testicular germ cell-specific splice variants. The *Gapdh* signal was amplified as an internal control and the relative expression of each *Sp1* mRNA was calculated as the ratio to *Gapdh*. The data are presented as mean ± S.D. from four independent experiments. n = 4.



Supplementary Fig. 5. Genomic PCR of stable GC-2spd(ts) cell clones. A schematic drawing of the transgene constructs and the primer positions are indicated with their product sizes at the top. PCR was performed using 3-200 ng genome DNAs from the clones with ExTaq polymerase (Takara) in a total volume of 10 μl. A part of the products were electrophoresed on a 0.7% agarose gel. As the positive control, PCR was also performed using 3 ng of the purified plasmids for lncRNA-6.9kb-EGFP and ΔCNS1-lncRNA-EGFP as templates. Each lane is marked with the clone number, and the copy number of each clone is also indicated. Positions of specific bands with expected sizes are indicated by arrowheads.

Table 1. Oligonucleotide primers used in this study

Designation	Forward	Reverse
Northern blot analysis and *in situ* hybridization		
Tcam1	5'-ATAGCCTGGCATGAGTTGCT-3'	5'-GCACCCTAAGACCGATTTCA-3'
β-actin	5'-ACATCCGTAAAGACCTCTATG-3'	5'-TAAAACGCAGCTCAGTAACAGT-3'
ChIP		
Aip-Promoter	5'-GGGCTTCAGCACAGAATCCA-3'	5'-TGAAAAATCCTGAGAGCCTCATT-3'
Smarcd2 intron1	5'-ACCCAGAGATGGCAGAATC-3'	5'-CAAGCACCAACCCCACATT-3'
CNS1,2	5'-CCAGAAGCCTGTATTGGTT-3'	5'-GGCAAGTTAGTGCAGTTAAG-3'
CNS2-Tcam1	5'-AGACCAAAGCCAGCATGAAT-3'	5'-CTCTCTGCCCAGGAGGTCTA-3'
Tcam1pro-CNS3	5'-CTGCCGTTAAATGCCTTCAG-3'	5'-GTGGGAAGGAACACTTGGAT-3'
gene body	5'-TCGGAGTGACCAGACAAGTG-3'	5'-CACACCCACAGCTCTAATCC-3'
CNS4	5'-TGCTGGCACTTAATGTGGTT-3'	5'-CACCCGGCTTGTTTGTTTTA-3'
CNS5	5'-TGAGAGAAATGCTGCTTTGG-3'	5'-TGAAAAGTCACATGCTGGAAA-3'
CNS6	5'-CTTAGCCATGGCCACCTTT-3'	5'-CCACTCACCTCCAGAAGGAA-3'
Reporter gene assay		
Tcam1 promoter	5'-ATAACGGCGTTGGCAGTGTG-3'	5'-TCCTCGATGCTTGGGGACCT-3'
CNS1	5'-ATACTCCAGATCCGGGATGT-3'	5'-ACGGAGCAAACAGCAAACAC-3
CNS1-(1-257)	5'-ATACTCCAGATCCGGGATGT-3'	5'-GAAAAGGCCGCCTCCCCCAA-3'
CNS1-(258-373)	5'-CCGGAGGAGCGGGAGCGGAA-3'	5'-ACGGAGCAAACAGCAAACAC-3'
CNS2	5'-TTTTAAGAGCCCATCTCGGG-3'	5'-GCATGCAAAATCCCTTCACC-3'
CNS3	5'-CCTTGGCTATCTTGGAACTC-3'	5'-TGCCTCTCTTCCCTGAACTA-3'
Mutagenesis 1	5'-AAAAAAAACAACCTAGACCCTGCAGAAAAAAAACC TGCCCGGCAACCCAATCG-3'	5'-TTTTTTTTCTGCAGGGTCTAGGTTGTTTTTTTTTGGG AGTTCGTAACCGCCTC-3'
Mutagenesis 2	5'-TCGCGAAAAAAAAATCAAACCAGCACCTCCCATA-3'	5'-GTTTGATTTTTTTTTCGCGAGGGCGGGATTTAAA-3'

RT-PCR		
lncRNA-Tcam1-a	5'-TTTTAAGAGCCCATCTCGGG-3'	5'-AGGCTTAGCTTTCCTGCTCT-3'
lncRNA-Tcam1-b	5'-TTTTAAGAGCCCATCTCGGG-3'	5'-TGGCACACAAGTGAGATCAA-3'
Gapdh	5'-CATGACCACAGTCCATGCCATC-3'	5'-TAGCCCAAGATGCCCTTCAGTG-3'
Gapdh (in5-ex6)	5'-CCTTCTTTGTAGGTGTCCCT-3'	5'-TAGCCCAAGATGCCCTTCAGTG-3'
Gapdh (ex5-ex6)	5'-TTGTGATGGGTGTGAACCAC-3'	5'-TAGCCCAAGATGCCCTTCAGTG-3'
lncRNA-Tcam1 for RT		5'-AGGCTTAGCTTTCCTGCTCT-3'
Gapdh for RT		5'-TAGCCCAAGATGCCCTTCAGTG-3'
5'RACE		
AAP	5'-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3'	
AUAP	5'-GGCCACGCGTCGACTAGTAC-3'	
lncRNA-Tcam1 for RT		5'-CTCAGAAATCCACCTGCCTC-3'
lncRNA-Tcam1 GSP1		5'-GCATGCAAAATCCCTTCACC-3'
lncRNA-Tcam1 GSP2		5'-TGCGAATTACCAGGCTTCCT-3'
3’RACE		
AP		5'-CTGATCTAGAGGTACCGGATCC-3'
lncRNA-Tcam1 GSP3	5'-ACAGCTTTGCTTGGGTTCTG-3'	
lncRNA-Tcam1 GSP4	5'-GGGCTCCTTTGTTCAAAGAG-3'	
qRT-PCR		
Tcam1 promoter	5'-CAGGAGATGGCTTCCCTACT-3'	5'-CCAGAAACTCGTGACGCTTA-3'
luciferase	5'-GGGACGAAGACGAACACTTC-3'	5'-GGTGTTGGAGCAAGATGGAT-3'
Gapdh	5'-TGCACCACCAACTGCTTAGC-3'	5'-GGCATGGACTGTGGTCATGAG-3'
EGFP	5-AGCAAAGACCCCAACGAGAA-3'	5’-GGCGGCGGTCACGAA-3'
lncRNA-Tcam1	5'-GACTGTCTGGGCAGAGTGAA-3'	5'-GAACCCAAGCAAAGCTGTAAAC-3'
β-actin	5'-CCATAGGCTTCACACCTTCCTG-3'	5'-GCACTAACACTACCTTCCTCAACCG-3'

Table 2 ChIP sequencing data

Cell/Organ	Antigen	Accession	source	Ref.
Liver	Input	GSM769034	GEO	43
Liver	H3K4me3	GSM769014	GEO	43
Round Spermatid	Input for H3K4me3	SRX336654	SRA	42
Round Spermatid	H3K4me3	SRX336652 SRX336653	SRA	42
Round Spermatid	Input for H3K4me3	GSM1046838 GSM1046839	GEO	44
Round Spermatid	H3K4me3	GSM1046840 GSM1046841	GEO	44
Round Spermatid	Input for Brdt	GSM984198	GEO	84
Round Spermatid	Brdt	GSM984200	GEO	84
Spermatocyte	Input for H3K4me3	SRX336651	SRA	42
Spermatocyte	H3K4me3	SRX336649 SRX336650	SRA	42
Spermatocyte	Input for Brdt	GSM984197	GEO	84
Spermatocyte	Brdt	GSM984199	GEO	84