Studies on regulatory mechanisms at the testis-specific *Tcam1*/*TCAM1P* gene locus in mouse and human

(マウスとヒトにおける精巣特異的遺伝子 *Tcam1*/*TCAM1P* の発現調節機構に関する研究)

A DISSERATION

Submitted to the Graduate School of Life Science, Hokkaido University

In partial fulfillment of the requirements for the degree

Doctor of Life Science

by Misuzu KURIHARA

March, 2015

CONTENTS

ABBREVIATIONS..... 2

GENERAL INTRODUCTION 5

Chapter 1..... 12

Chapter 2..... 70

GENERAL DISCUSSION..... 93

ACKNOWLEDGEMENTS..... 102

REFERENCES..... 103

ABBREVIATIONS

AAP, abridged anchor primer

Aip, aryl-hydrocarbon receptor-interacting protein

ANOVA, analysis of variance

AP, adaptor sequence

AUAP, abridged universal amplification primer

BAC, bacterial artificial chromosome

CGI, CpG island

ChIP, chromatin immunoprecipitation

ChIP-seq, chromatin immunoprecipitation-sequencing

CNS, conserved noncoding sequence

DMEM, Dulbecco’s modified Eagle’s medium

DPE, dual promoter–enhancer

EGFP, enhanced green fluorescent protein

ENCODE, Encyclopedia of DNA elements

eRNA, enhancer RNA

FBS, fetal bovine serum

Gapdh, glyceraldehyde 3-phosphate dehydrogenase

GSP, gene specific primer

hCNS1, human CNS1

hGH, human growth hormone

H3K4me1, H3 Lysine 4 mono-methylation

H3K4me3, H3 Lysine 4 tri-methylation

H3K9ac, H3 Lysine 9 acetylation

HS, hypersensitive site

IGV, Integrative Genomics Viewer

LCR, locus control region

LINoCR, lipopolysaccharide inducible noncoding RNA

lncRNA, long noncoding RNA

mCNS, mouse CNS

mGh, mouse growth hormone

ORF, open reading frame

PIC, pre-initiation complex

qRT-PCR, quantitative RT-PCR

RACE, rapid amplification of cDNA ends

RT-PCR, reverse transcription-polymerase chain reaction

Smarcd2/SMARCD2, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2

Tcam1/TCAM1P, testicular cell adhesion molecule

TSS, transcriptional start site

TTS, transcriptional termination site

GENERAL INTRODUCTION

Our body is made up of about sixty trillion cells and each of them contains a common genome. Nonetheless, various types of organs and tissues exist in our body, which are established by accurate spatiotemporal patterns of gene expression. The gene expression is controlled by mechanisms operating at various levels such as transcription, RNA splicing, RNA metabolism, translation, and posttranslational modification. Among them, a fundamental process of gene expression is achieved at the transcriptional level, and therefore, the investigation of the transcriptional regulation is crucial for understanding how this complicated body is established.

Accurate spatiotemporal patterns of eukaryotic gene transcription are strictly controlled by the intricate process, which is orchestrated by promoter and enhancer elements. Promoters, which are usually located at immediate upstream of transcriptional start sites (TSSs), initiate gene transcription by recruiting RNA polymerase and general transcription factors[1–3]. On the other hand, enhancers are sequences containing multiple binding sites for a variety of transcription factors and increase the rate of target gene transcription[4,5]. They generally function independently of their orientation, and are scattered across the genomic sequence at various distances from their target genes[6,7]. Because the activity of enhancers is restricted to a particular tissue and cell type or to some stimulation, they can define when, where, and at what level each gene is transcribed.

To find out candidate enhancers, the comparison of genomic sequences between human and mouse has been one useful approach, because regulatory mechanisms of gene expression have been thought to be evolutionarily conserved. Indeed, regulatory sequences of *β-globin, IL4/IL13/IL5,* and *SCL* genes were identified by the comparative genomics[8]. However, while DNase I hypersensitive sites (HSs) are highly conserved in the *β-globin* locus control region (LCR) between human and mouse and they activate the tissue- and developmental stage-specific

expression of the cluster genes[9], the molecular mechanism is not necessarily conserved. Although common erythroid specific-transcription factors bind to DNase I HSs in both human and mouse *β-globin* LCR, which results in the establishment of highly active chromatin structures, it occurs at different timing during development between the two species[9]. In addition, recent comprehensive analyses have provided several lines of evidence that gene expression and their regulatory mechanisms have largely diverged between human and mouse lineages[10,11]. Since genomic sequences have been dynamically changed in the process of evolution after human-mouse divergence[12], a fine comparative analysis of orthologous genes, whose regulatory mechanisms are diverged between the two species, is essential for understanding the evolution of gene regulation.

The *human growth hormone (hGH)* gene cluster is one of the well investigated models for tissue-specific gene activation, and the regulatory mechanism is not conserved compared to *mouse growth hormone (mGh)* gene (Fig. I). The *hGH* cluster is located on chromosome 17q22-24 and consists of five paralogous genes that were generated by duplication of a single ancestral *GH* gene[13,14]. The *hGH-N* gene is specifically expressed in the pituitary, and the other four genes are placenta-specific. Activation of the *hGH* cluster depends on the multifunctional LCR which is located at 15-32 kb upstream of the cluster (Fig. I)[15,16]. In the pituitary, a transcription factor, Pit-1, recruits the histone acetyltransferase to DNase I HSI and establishes a broad histone acetylation domain, which encompasses the entire LCR and the *hGH-N* promoter, resulting in the *hGH-N* activation[17,18]. Furthermore, noncoding transcription at the LCR and the *CD79b* gene, which is linked to the cluster, enhanced the *hGH-N* expression[19,20]. In the placenta, the LCR functions as the insulator to establish the broad histone acetylation domain, leading to the activation of the four placental paralogues[21–23].

On the other hand, the *mGh* gene is located on chromosome 11, and no placental

paralogues are present (Fig. I). In the mouse pituitary, the *mGh* expression is regulated by transcription of a SINE B2 element between the two genes at upstream of *mGh*, *Cd79b* and *Scn4a* [24]. By noncoding transcription of the SINE B2 element, the chromatin status at the *mGh* locus shifts from heterochromatic to permissive euchromatic environment, which results in the *mGh* activation. Importantly, while the pituitary-specfic expresssion is conserved between *hGH-N* and *mGh*, there is no known LCR at the *mGh* locus. This suggests that duplication of *GH* is followed by alteration of the regulatory mechanism.

Interestingly, the *hGH/mGh* gene is linked to three other tissue-specific genes: the B lymphocytes-specific *CD79b/Cd79b* and striated muscle-specific *SCN4A/Scn4a* genes at its 5' region, and the testis-specific testicular cell adhesion molecule (*TCAM1P/Tcam1*) gene at downstream of the cluster (Fig. I). The regulatory mechanism of the *CD79b/Cd79b* gene was analyzed in some species. The human *CD79b* gene was reported to be sufficiently activated by its proximal promoter in the B cell[25]. For rat *Cd79b*, three DNase I HSs at the *Cd79b-Gh* intergenic region could be potential enhancers.[26] In chicken, some DNaseI HSs are scattered in and around the *Cd79b* gene and functioned as transcriptional activators for B cell-specific *Cd79b* expression[27,28]. However, regulatory mechanisms of *SCN4A/Scn4a* and *TCAM1P/Tcam1* genes remain to be revealed.

Notably, the genomic structure of the *TCAM1P/Tcam1* gene is not preserved between human and mouse. The mouse *Tcam1* gene encodes a cell adhesion protein and the protein coding potential is retained in other eutherian species such as rat, dog, cow and reshus moneky[29]. However, as a result of pseudogenization that occurred in a short period during evolution[30], the human *TCAM1P* gene is a pseudogene and unlikely to encode such a protein. Therefore, it is very interesting to know how its expression and regulatory mechanism are changed in parallel with pseudogenization. Validating the relationship of the change in the *TCAM1P* genomic structure with its expression and regulation will help us understand the

evolution of gene regulation. To figure out these issues, I excecuted comparative analysis of the expression and regulatory mechnism of the human and mouse *TCAM1P/Tcam1* gene.

In Chapter 1, I attempted to identify and characterize regulatory sequences for the mouse *Tcam1* gene, and found that a mouse conserved noncoding sequence 1 (mCNS1) was a novel potential spermatocyte-specific enhancer. Surprisingly, mCNS1 also functioned as a bidirectional promoter of the *SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2* (*Smarcd2*) gene, which is located at 5' upstream of *Tcam1*, and a novel testis-specific long noncoding RNA (designated *lncRNA-Tcam1*). From these results, I concluded that mCNS1 can be a dual promoter-enhancer (DPE) element.

DPE is a multifunctional genomic element which was found in mammals for the first time by this study. There are several kinds of multifunctional genomic elements such as transcriptional units overlapping with gene regulatory elements[31–33] or enhancers for the expression of multiple genes[34,35], but only a small number of them was reported. Nonetheless, many researchers assume the massive existence of multifunctional genomic elements, and therefore, this study provides a new and important model for gene regulatory mechanisms by such genomic elements.

In Chapter 2, I examined whether the expression pattern and the regulatory mechanism of *TCAM1P* are preserved after pseudogenization or not. Especially, I focused on human CNS1 (hCNS1) and attempted to examine whether hCNS1can function as a DPE element at the human *TCAM1P* locus, too. The *TCAM1P* expression was testis-specific, but its localization was slightly different from mouse *Tcam1*. hCNS1 could enhance *TCAM1P* promoter activity and functioned as a promoter of the human *SMARCD2* gene, which suggests that it can be a DPE in human genome.

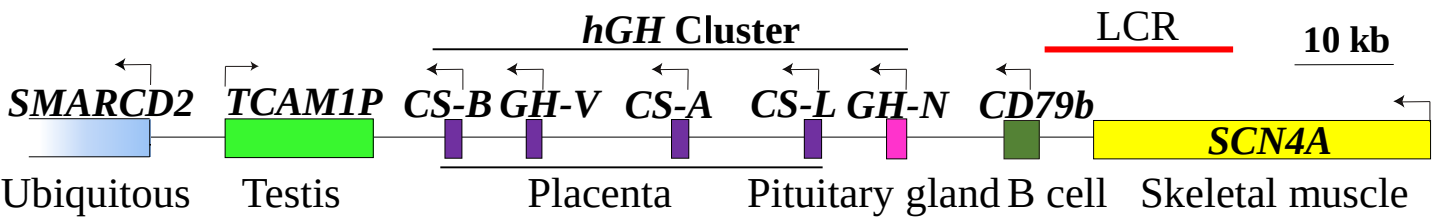
Collectively, this study highlights two major issues: the importance of multifunctional

genomic elements in transcriptional regulation and the alteration of gene regulation in parallel with pseudogenization. Both issues are extremely important but have not been well studied. Therefore, I believe that this study could be of great significance to comprehend the complexity of the genomic function.

Chapter 1:

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

(A) **Human chromosome 17**



(B) **Mouse chromosome 11**

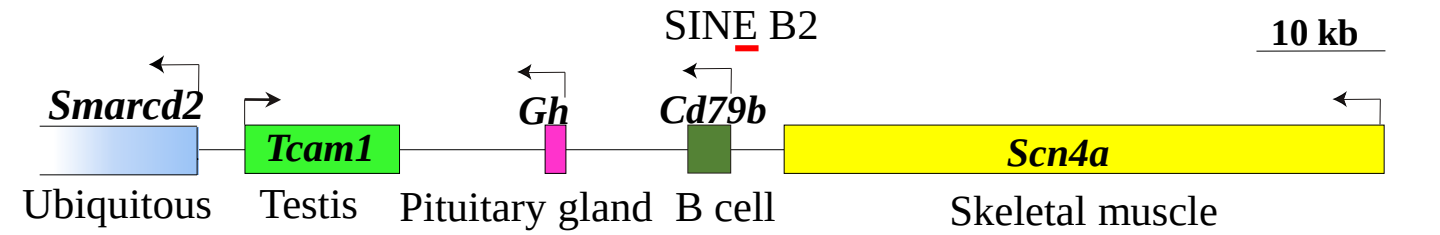


Fig. I. The schematic diagram of the human *TCAM1P* and mouse *Tcam1* locus.

(A) The *TCAM1P* gene locus on human chromosome 17 is drawn. Seven tissue-specific genes and one ubiquitous gene are present at downstream and upstream to *TCAM1P*, respectively. The tissue expressing each gene is indicated below the gene structure, and bent arrows indicate the transcriptional direction of each gene. The position of LCR, which controls the *hGH* cluster, is shown by a red line. (B) The mouse *Tcam1* locus on chromosome 11 is shown. Three tissue-specific genes and one ubiquitous gene are located at downstream and upstream to *Tcam1*, respectively. The tissue expressing each gene is indicated below the structure, and bent arrows indicate the transcriptional direction of each gene. The SINE B2 element drawn with a red line plays an important role in the mouse *Gh* gene activation.

Abstract

Tissue-specific gene expression is tightly regulated by various elements such as promoters, enhancers, and long noncoding RNAs. In the present study, I identified a conserved noncoding sequence, mCNS1, as a novel enhancer for the spermatocyte-specific mouse *Tcam1* gene. mCNS1 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, mCNS1 could enhance *Tcam1* promoter activity only in GC-2spd(ts) cells, which were derived from mouse spermatocytes. When I integrated the 6.9-kb 5'-flanking sequence of *Tcam1* with or without a deletion of mCNS1 linked to the enhanced green fluorescent protein gene into the chromatin of GC-2spd(ts) cells, mCNS1 significantly enhanced *Tcam1* promoter activity. These results indicate that mCNS1 could function as a spermatocyte-specific enhancer. Interestingly, mCNS1 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 mCNS1. 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 mCNS1 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 mCNS1 is a spermatocyte-specific enhancer for the *Tcam1* gene and a bidirectional promoter of *Smarcd2* and *lncRNA-Tcam1*.

Introduction

Tissue-specific gene activation is controlled by complicated mechanisms that involve the activity of promoters and enhancers, epigenetic modification, and noncoding transcription[36–38]. 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[33,39–43], and consistent with this, many enhancers are actually transcribed into long noncoding RNAs (lncRNAs) that are often essential for target gene activation[43,44]. However, the relationship between lncRNAs and enhancers seems to be diverse and is not completely understood[43–45].

To understand the regulatory mechanism for tissue-specific gene activation, various genes have been investigated as model genes[46–52]. The *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[14,53]. The tissue-specific activation of the *hGH* cluster is dependent on the 5'-distal LCR[15,16], and epigenetic regulation and noncoding transcription are known to play crucial roles in activation by the *hGH* LCR[17–19,21,22]. 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 *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[29]. The testis-specific expression of this gene has been confirmed in the rat and mouse[29,54], 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[55]. Consistent with this, mouse *Tcam1* mRNA has been found to be localized in such spermatocytes by *in situ* hybridization[29]. 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[26,56]. 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, I 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 mCNS1 was identified as a potential spermatocyte-specific enhancer. Interestingly, mCNS1 also contained bidirectional promoter activity, and the *Smarcd2* gene and a novel testis-specific lncRNA (designated *lncRNA-Tcam1*) were actually expressed from the upstream and downstream of mCNS1, respectively. The results indicated that mCNS1 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 DPE 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[29]. I first attempted to confirm this expression pattern of *Tcam1* mRNA. By northern blot analysis using total RNAs from 15 mouse tissues, I detected a specific signal for *Tcam1* at the 3.0-kb position only in the testis (Fig. 1-1A). This was also confirmed by the transcriptomic data (GSE9954), which demonstrated testis-specific expression of *Tcam1* mRNA among 22 mouse tissues (Fig. 1-1B). I 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[57,58], and the signal intensity became stronger in 21 and 28-day-old testes (Fig. 1-1C). Expression was slightly decreased 56 days after birth.

To further confirm the localization of *Tcam1* mRNA, I performed *in situ* hybridization with testes at 21 and 28 days after birth. I used these stages of testes because *Tcam1* mRNA was expressed at the highest level at these stages in my northern blot analysis (Fig. 1-1C). The results indicated that *Tcam1* mRNA was localized in spermatocytes in all seminiferous tubules (Fig. 1-1D, 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.

mCNS1, mCNS2, and mCNS3 are candidate regulatory sequences for *Tcam1* gene activation by histone modification patterns

I next searched for candidate sequences for *Tcam1* regulation. I attempted to identify CNSs at the mouse *Tcam1* locus because important gene regulatory sequences are generally well conserved beyond species[59]. Using the rVista program (http://genome.lbl.gov/vista/index.shtml), I 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. I found two CNSs upstream of the *Tcam1* gene and four CNSs within the gene body (Fig. 1-2A). No CNS was present in the 3'-region of the *Tcam1* gene. I then assessed which CNSs could actually be involved in spermatocyte-specific *Tcam1* expression by determining the histone modification patterns. For this purpose, I 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[60,61]. I 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[62,63], 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.

I first prepared ten amplicons at the *Tcam1* locus: six mCNSs, intron 1 of *Smarcd2* (Smarcd2 intron1), a region between mCNS2 and *Tcam1* (CNS2-Tcam1), *Tcam1* promoter, and a region between mCNS3 and mCNS4 (genebody). Because mCNS3 and mCNS6 overlapped

with or were close to repeat sequences, which made the primer design very difficult, I prepared the amplicons just upstream and downstream of them, respectively. As a result of ChIP and PCR, I detected high background signals at mCNS1 and *Tcam1* promoter (data not shown), possibly due to the high GC content of these sequences, so I excluded these data. Consequently, I obtained histone modification levels of eight regions at the locus (Fig. 1-2A). Considering that the resolution of my ChIP analysis was 500–1000 bp and mCNS1 and mCNS2 were only 196 bp apart from each other, histone modification levels of mCNS1 would be reflected by those of mCNS2. Similarly, the status of *Tcam1* promoter would be reflected by mCNS3, whose amplicon was positioned only 268 bp downstream of the TSS of the *Tcam1* gene. Therefore, I re-designated these amplicons CNS1,2 and Tcam1pro-CNS3 (Fig. 1-2A).

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, I used the *aryl-hydrocarbon receptor-interacting protein* (*Aip*) gene promoter because this gene was expressed at similar levels in germ and liver cells by my quantitative RT-PCR (qRT-PCR) analysis (data not shown). As a negative control, I 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. 1-2B). 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. 1-2B). 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. 1-2C). In liver cells, no regions were marked with H3K4me1, except that Smarcd2 intron 1 might be slightly modified (Fig. 1-2C). This

suggests that CNS1,2 is 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. 1-3A). 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. 1-3A).

To further access the histone modification patterns, I analyzed the ChIP-sequencing (ChIP-seq) data deposited by other studies. I found that data for H3K4me3 in mouse spermatocytes and round spermatids were available (SRA accession: SRA097278)[64], and the corresponding data in the liver were obtained from the Encyclopedia of DNA elements (ENCODE) project (GEO accession: GSM769014)[65]. In spermatocytes, depth peaks for the precipitated DNA were observed around CNS1,2 and *Tcam1* promoter, although the input DNA showed no peaks (Figs. 1-3B, 1-4A). 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 (Figs. 1-3C, 1-4B). A similar pattern for H3K4me3 in round spermatids was also observed in another ChIP-seq data (GEO accession: GSE42629)[66]. Compared with spermatocytes and spermatids, a small peak was observed around the *Smarcd2* promoter specifically in the precipitated DNA in the liver (Figs. 1-3D, 1-4C). These data are consistent with the ChIP results (Fig. 1-3A), given that high levels of the H3K4me3 marker were detected at CNS1,2 and Tcam1pro-CNS3. Collectively, my 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 mCNSs could be important regulatory elements for the *Tcam1* gene, and I focused on mCNS1, mCNS2, and mCNS3 in the following analyses.

mCNS1 can function as an enhancer of the *Tcam1* promoter in the GC-2spd(ts) cell

I assessed the enhancer activity of the three CNSs by the *in vitro* reporter gene assay. Based on TSS that was previously determined[55], I cloned a 1644-bp sequence just upstream of the *Tcam1* gene as a promoter and connected it to the luciferase gene. mCNS1, mCNS2, and mCNS3 were amplified by PCR with mouse genomic DNA, and the mCNS1 and mCNS2 sequences were cloned into 5' of the *Tcam1* promoter (Fig. 1-5A). mCNS3 was connected to 3' of the luciferase gene because it was located downstream of the *Tcam1* promoter. These constructs (mCNS1-Pro-luc, mCNS2-Pro-luc, and Pro-luc-mCNS3) 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. 1-5A). GC-2spd(ts) cells are derived from mouse spermatocytes[67] 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 mCNS2-Pro-luc and Pro-luc-mCNS3 were slightly increased or at similar levels in the three cell lines (Fig. 1-5A). In contrast, when mCNS1 was linked to the *Tcam1* promoter, luciferase activity was 5-fold 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. 1-5A). This GC-2spd(ts)-specific activity of mCNS1 was orientation-independent because mCNS1 enhanced *Tcam1* promoter activity to a comparable extent when cloned in the reverse orientation (Fig. 1-5B). The activity was also detected when mCNS1 was connected to 3' of the luciferase gene in both directions (Fig. 1-5C) and when the construct was linearized before transfection (Fig. 1-6A). Although shorter *Tcam1* promoters showed higher activity than the 1644-bp promoter, mCNS1 could increase their luciferase activity (Fig. 1-6B, C). These

results suggest that mCNS1 is a spermatocyte-specific enhancer for the *Tcam1* gene. To further characterize enhancer activity, I divided the mCNS1 sequence into two halves, mCNS1-(1-257) and mCNS1-(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. 1-5B). This indicates that the entire sequence of mCNS1 is required for the full activity of this GC-2spd(ts)-specific enhancer.

Because mCNS1 was located just upstream of the *Smarcd2* gene (Fig. 1-2A), it was possible that mCNS1 had promoter activity that affected luciferase gene expression in my reporter assay. Therefore, I assessed whether mCNS1 could function as a promoter. I connected the mCNS1 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 mCNS1 had very strong promoter activity (Fig. 1-7A). Compared with the *Tcam1* promoter, mCNS1 showed approximately 220-fold higher promoter activity in both directions. I also investigated the mCNS1 promoter activity in Hepa1-6 and NIH3T3-3-4 cells. mCNS1 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. 1-7A).

Then, I assessed the possibility that the fivefold increase in luciferase activity observed between Tcam1-Pro-luc and mCNS1-Pro-luc in GC-2spd(ts) cells has resulted from the strong promoter activity of mCNS1. Using the mCNS1-Pro-luc construct, I introduced a polyadenylation signal between mCNS1 and the *Tcam1* promoter (mCNS1-polyA-Pro-luc). In this construct, transcription of the *Tcam1* promoter driven by mCNS1 should be stopped by the poly(A) signal. In fact, a transcript from the promoter in GC-2spd(ts) cells transfected with the mCNS1-polyA-Pro-luc construct was dramatically decreased by qRT-PCR (Fig. 1-7B, left). However, the luciferase mRNA levels were unchanged between the cells transfected with the

two constructs (Fig. 1-7B, right), and the luciferase activity of the mCNS1-polyA-Pro-luc construct was similar to that of mCNS1-Pro-luc (Fig. 1-7C). This indicates that in the mCNS1-Pro-luc construct, transcription driven by the mCNS1 promoter elongated through the *Tcam1* promoter but stopped in the middle of the luciferase gene. Therefore, I concluded that the luciferase activity of mCNS1-Pro-luc did not reflect the strong promoter activity of mCNS1 and that mCNS1 truly possessed enhancer activity in GC-2spd(ts) cells.

mCNS1 is a bidirectional promoter of the ubiquitously expressed *Smarcd2* gene and a testicular germ cell-specific *lncRNA-Tcam1*

My *in vitro* reporter analysis showed that mCNS1 could drive transcription bidirectionally (Fig. 1-7A). While the *Smarcd2* gene, which was considered to be expressed ubiquitously, was presumed to be regulated by the mCNS1 promoter, there were no annotated genes opposite mCNS1 (Fig. 1-2A, Fig. 1-8A). Thus, I investigated whether any transcripts were generated from the adjacent regions of mCNS1. Reverse transcription-polymerase chain reaction (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. 1-8B). This indicates that mCNS1 is actually a bidirectional promoter for the ubiquitously expressed *Smarcd2* gene and a novel testis-specific transcript.

I further characterized this novel transcript. To determine the 5’ and 3’ ends of the transcript, I performed rapid amplification of cDNA ends (RACE) analysis. A single band was detected by 5’RACE, and 10 subclones that I 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. 1-8A). According to the coding potential calculator (CPC) tool version 0.9r2[68], this transcript was classified as lncRNA, and I designated it *lncRNA-Tcam1*. *lncRNA-Tcam1* was presumed to be polyadenylated at its 3' end because I observed poly(A) sequences longer than the oligo(dT) length in the subclones obtained by my 3'RACE.

To check whether *lncRNA-Tcam1* was really transcribed as a single transcript, I first performed RT-PCR with testis RNA using primers at its 5'and 3'ends; however, I 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*. I 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, I concluded that *lncRNA-Tcam1* was transcribed as a single transcript.

I 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. 1-8C). This expression pattern was correlated to *Tcam1* mRNA (Fig. 1-1C). I 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. 1-8D). This was also correlated to *Tcam1* mRNA. Finally, I fractionated germ cells into nuclear and cytoplasmic subfractions and checked the subcellular localization of *lncRNA-Tcam1*. *Glyceraldehyde 3-phosphate dehydrogenase* (*Gapdh*) signals amplified from the primer pairs at exons 5 and 6 and at intron 5 and exon 6 confirmed that my fractionation was successful (Fig. 1-8E). The *lncRNA-Tcam1* signal was exclusively detected in the nucleus (Fig. 1-8E), which strongly suggests that it was not translated to any peptides and functioned as an RNA molecule.

DNA methylation status in mCNS1

Surprisingly, mCNS1 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. 1-3). I attributed this to another epigenetic marker, CpG methylation. Because I found a CpG island (CGI) encompassing exon 1 of the *Smarcd2* gene, mCNS1, and mCNS2 (Fig. 1-9A), CpG methylation status could be one factor affecting the promoter activity of mCNS1. 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[63]. After the bisulfite reaction and PCR, I sequenced 10 subclones from each sample and checked which cytosine was converted to thymine. I examined 54 CpGs in CGI, and almost all the CpGs were demethylated in both spermatocytes and the liver; I did not observe any difference between these two tissues (Fig. 1-9B). This result suggests that the hypo-methylation of CGI was associated with the ubiquitous promoter activity of mCNS1.

Sp1 contributes to mCNS1 promoter activity for the *Smarcd2* gene

To gain mechanistic insight into promoter and enhancer activity of mCNS1, I attempted to identify evolutionarily conserved sequences that were important regulatory elements for many genes[8,59]. I collected mCNS1 sequences from seven different mammalian species and compared them (Fig. 1-11A). As a result, I 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[69,70], but some testis-specific splice variants were reported to be present at later stages[71,72]. In addition, the DNA-binding affinity of Sp1 from male germ cells is greater than that from other tissues[71] and the interaction with other proteins and post-translational modifications can change its activity[73–75]. Therefore, it was also possible that Sp1 was involved in the enhancer activity of mCNS1.

I first checked whether Sp1 was expressed in the three cell lines I used (GC-2spd(ts), Hepa1-6, NIH3T3-3-4). In this experiment, I 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.*[72], although it was pointed out that the primers for a 3.7-kb variant might also amplify an 8.2-kb transcript[70]. 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 (Fig. 1-10).

I then prepared the mCNS1 sequence, in which all three Sp1 sites were mutated not to be recognized by Sp1, and investigated its promoter and enhancer activity. When mutated mCNS1 was linked to the *Tcam1* promoter, enhancer activity was significantly increased in GC-2spd(ts) and NIH3T3-3-4 cells (Fig. 1-11B). In contrast, the mutation had no effect on enhancer activity in Hepa1-6 cells (Fig. 1-11B). When mutated mCNS1 was directly connected to the luciferase gene in the forward orientation, mCNS1 promoter activity was not changed (Fig. 1-11C). However, in the reverse orientation, the mutation significantly decreased mCNS1 promoter activity in all the cell lines (Fig. 1-11C). These suggest that Sp1 is not a key factor for the mCNS1 enhancer but contributes to the promoter activity for the *Smarcd2* gene.

mCNS1 functions as a promoter of *lncRNA-Tcam1* and an enhancer for *Tcam1* in the chromatin context

I finally assessed whether mCNS1 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, I prepared a sequence encompassing the *Tcam1* promoter, *lncRNA-Tcam1*, and mCNS1. I obtained a bacterial artificial chromosome (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 I 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 mCNS1, I 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. 1-13A). I also prepared the ΔmCNS1-lncRNA-EGFP construct by deleting a 861-bp region containing mCNS1 with a restriction enzyme, *Ksp*I (Fig. 1-13A).

The two constructs were transfected into GC-2spd(ts) with a vector containing the puromycin resistance gene, and by selection with puromycin, I 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 ΔmCNS1-lncRNA-EGFP construct stably integrated into genomic DNA, I performed cell cloning by the limited dilution method. I successfully established 11 GC-2spd(ts) clones for the lncRNA-6.9kb-EGFP construct and 10 clones for ΔmCNS1-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 ΔmCNS1-lncRNA-EGFP. I also checked whether or not the integrated transgene was intact by long PCR with genome DNA purified from each clone. I used a primer pair that could amplify most region of the transgenes (Fig. 1-12). As a result of PCR, signals with expected sizes were observed for all the clones along with extra bands at various sizes (Fig. 1-12). 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, I first investigated whether the bidirectional transcription of *Smarcd2* and *lncRNA-Tcam1* was impaired by deleting mCNS1. However, because the *Smarcd2* gene was endogenously expressed in GC-2spd(ts) cells but *lncRNA-Tcam1* was not, I 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 ΔmCNS1-lncRNA-EGFP (Fig. 1-13B). This difference was statistically significant, and therefore, the data indicate that mCNS1 possesses promoter activity for *lncRNA-Tcam1* in the chromatin context.

Following this, I measured EGFP mRNA levels per copy in the established clones to evaluate the enhancer activity of mCNS1. 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 ΔmCNS1-lncRNA-EGFP (Fig. 1-13C). The average EGFP mRNA level was 4.5-fold higher in the clones with intact mCNS1. This indicates that mCNS1 could function as an enhancer element for *Tcam1* gene expression even when it is integrated into chromatin.

I finally examined whether the expression of *lncRNA-Tcam1* was correlated to that of

EGFP in GC-2spd(ts) cell clones. Using qRT-PCR, I measured the *lncRNA-Tcam1* level in each cell clone with lncRNA-6.9kb-EGFP and compared it with the EGFP mRNA level (Fig. 1-13D). 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

mCNS1 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 this chapter, mCNS1 increased *Tcam1* promoter activity *in vitro* in GC-2spd(ts) cells (Figs. 1-5A, 1-6A, C), which indicates that mCNS1 is a potential enhancer in these cells. To assess the enhancer activity of mCNS1 in the chromatin context, I established stable cell clones using two constructs (Fig. 1-13). One contained the 6.9-kb upstream sequence of the *Tcam1* gene, and in the other construct, I deleted the mCNS1 region. EGFP mRNA was expressed at significantly higher levels in cell clones with the full 6.9-kb sequence than in clones without mCNS1. Possibly, integrated transgenes might be under control of neighboring enhancers, and some transgenes might be truncated as suggested by my genome PCR (Fig. 1-12). 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 mCNS1 has enhancer

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

Because mCNS1 was positioned between two *Ksp*I sites and no other recognition sites of this enzyme were present in the construct, I cut out the 861-bp *Ksp*I sequence to delete mCNS1 from the 6.9-kb sequence. However, this also deleted sequences other than mCNS1 because mCNS1 was only 373 bp in length. It is possible that the significant decrease in EGFP expression in cells with ΔmCNS1-lncRNA-EGFP resulted from deletion of a sequence other than mCNS1. I 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 mCNS1.

Interestingly, the enhancer activity of mCNS1 was only observed in GC-2spd(ts) cells, which were derived from mouse spermatocytes (Fig. 1-5A). 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 mCNS1 and increased *Tcam1* promoter activity. However, the results do not necessarily mean that mCNS1 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[76–78]. For example, a regulatory mechanism of the spermatocyte-specific histone *H1t* gene was reported to be conserved in GC-2spd(ts) cells[78–80]. 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[81,82]. It is possible that GC-2spd(ts) cells contain transcription factors for *Tcam1*

activation by mCNS1 but can only bind to the mCNS1 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 my ChIP analyses, I used a germ cell fraction that contained not only spermatocytes but also spermatogonia, spermatids, and spermatozoa, and I 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[83]. Therefore, a substantial proportion of spermatids and spermatozoa did not contain histones, although mature spermatozoa was reported to retain 1% histones[84]. 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. 1-3A with Fig. 1-3B), 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[85–87]. Some reports actually demonstrated that regions marked with H3K4me1 possessed enhancer activity[88,89]. In the present study, mCNS1 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, mCNS1 could be a true enhancer for spermatocyte-specific *Tcam1* expression.

mCNS1 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[1–3]. 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)[2,3]. The proximal promoter contains the 5'-adjacent sequences to the core promoter, and it usually extends approximately 200–300 bp upstream of TSS[1]. 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 I 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 mCNS1. Although the core promoter is probably different between these two transcriptional units, I conclude that mCNS1 is necessary to create PIC for both *Smarcd2* and *lncRNA-Tcam1*.

Interestingly, mCNS1 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[90,91]. Consistent with this, mCNS1 was included in CGI that encompassed exon 1 of *Smarcd2*, mCNS1, and mCNS2. 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[92]. However, the expression patterns of the two transcripts driven by mCNS1 were different, and its hypo-methylation was correlated to the ubiquitous expression of *Smarcd2* but not to *lncRNA-Tcam1*. According to my 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 mCNS1 are not coordinately regulated, unlike many reported examples[90–93].

Notably, the *in vivo* promoter activity of mCNS1 was not as high as its *in vitro* activity. While I could detect *Tcam1* mRNA by 30 cycles of PCR in my 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[94]. CpG hypo-methylation of mCNS1 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[43]. 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[40,43,95–97]. eRNAs are generally characterized as short (1-2 kb), bidirectionally transcribed products, nonpolyadenylated RNAs, and transcripts from H3K4me1-enriched enhancers. Transcription from mCNS1 at the *Tcam1* locus was similar to that of eRNAs in that it occurred bidirectionally from the H3K4me1-enriched enhancer (mCNS1) but was otherwise different. For example, bidirectional transcription from mCNS1 was only observed in testicular germ cells; in other tissues, mCNS1 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 mCNS1 functions as a DPE

There are two examples of a DPE 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[98–100]. 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[101]. In mammals, genome-wide analysis revealed that 70% of extragenic RNA polymerase II peaks were related to the chromatin signature of enhancers[33], which suggested that many enhancers could be promoters of noncoding RNAs. However, there is no clear indication of a DPE element in mammals, and this is the first report of such an element.

How can mCNS1 function as both a promoter and an enhancer? The hypo-methylation of a CGI containing mCNS1 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 mCNS1 probably needs to be more strongly associated with H3K4me3 as well as with other active chromatin markers like H3K9ac. Moreover, mCNS1 was marked with H3K4me1 in spermatocytes and could act as an enhancer for the *Tcam1* gene. Therefore, in spermatocytes, mCNS1 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. 1-14).

To identify functional elements in mCNS1, I focused on three Sp1-binding sites. By mutating all of them, I investigated whether Sp1 contributed to bidirectional promoter and/or enhancer activity of mCNS1. 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. 1-11C). In GC-2spd(ts) and NIH3T3-3-4 cells, which expressed Sp1 mRNAs at higher levels than Hepa1-6, Sp1 repressed the mCNS1 enhancer activity (Fig. 1-11B). 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 mCNS1 sequences showed lower enhancer activity than intact mCNS1 in my reporter assay (Fig. 1-5B), the entire 373-bp sequence may be necessary for its full enhancer activity. Possibly, several transcription factors other than Sp1 bind to various regions of mCNS1 or a large protein complex is formed at mCNS1.

Recently, a BET family protein, Brdt, was reported to play crucial roles in the regulation of testicular germ cell-specific gene expression during meiosis[102]. To test the possibility that *Tcam1* is controlled by Brdt, I analyzed the transcriptomic and ChIP-seq data (GEO accession: GSE39909, GSE39910, GSE39908). However, I 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 DPE element in the chicken, noncoding transcription was necessary for nucleosome remodeling to activate the target gene of the enhancer[98–101]. However,

my data indicate that *lncRNA-Tcam1* transcription is not correlated to *Tcam1* promoter activity (Fig. 1-13D). What is the function of this lncRNA? The nuclear localization of *lncRNA-Tcam1* (Fig. 1-8E) may provide some hints. Many nuclear lncRNAs reported till date have two main functions: gene regulation and the formation of nuclear structures[103]. In general, lncRNAs that constitute some nuclear structures are expressed at high levels[104]. For example, *NEAT1* is a component of the nuclear paraspeckle, and its expression is as abundant as that of *XIST* in the nucleus[105–107]. 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, I 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 *mGh* 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, I 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[108,109]. Transcription of *lncRNA-Tcam1* was dramatically induced 14–21 days after birth during postnatal testis development, when many important protein-coding genes are upregulated[83]. 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 chapter were approved by the Institutional Animal Use and Care Committee at Hokkaido University.

RNA analyses

Northern blot, *in situ* hybridization, RT-PCR, and qRT-PCR were done as described previously[110]. 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 1-1. Primer sequences for RT-PCR and qRT-PCR are also shown in Table 1-1.

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[62]. 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 min 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 120*g* 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[63]. The purity of each cell fraction was checked in every experiment by qRT-PCR for marker genes[63].

ChIP assay

ChIP was conducted with germ cells purified from the 8-week-old testis and the liver as

previously described [110] 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 Tokyo Institute of Technology[111]. The qPCR was performed using primer pairs listed in Table 1-1.

ChIP-seq data

ChIP-seq data were collected from Gene expression omnibus (GEO) database[112] and Sequence read archive (SRA) database[113] with sra format (Table 1-2).

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)[113]. 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)[114] with bowtie (version 2.1.0)[115]. 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)[116]. The depth for each sample was normalized with total number of reads (RPM) or means of depth for *Aip* promoter region (4125000-4127500 bp 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 1-1. 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 1-1. 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 I named the resulting construct Tcam1-Pro-luc. All the mCNS sequences including a half part of mCNS1 were also amplified by PCR with KOD FX Neo (Toyobo). By inserting them into the blunted *Mlu*I site of Tcam1-Pro-luc, I generated the constructs, mCNS1-Pro-luc, mCNS2- Pro-luc, reversed mCNS1-Pro-luc, CNS1-(1-257)-Pro-luc, and mCNS1-(258-373)-Pro-luc. To generate the Pro-luc-mCNS3, Pro-luc-mCNS1, and Pro-luc-reversed mCNS1 construct, I cloned the mCNS3 or mCNS1 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 mCNS1 and the *Tcam1* promoter of mCNS1-Pro-luc. mCNS1-luc and reversed mCNS1-luc constructs were generated by inserting mCNS1 into the *Sma*I site of a pGL3-Basic vector.

For generating mutated mCNS1 constructs, a mCNS1 fragment was isolated from the mCNS1-luc construct by digestion with *Nhe*I and *Xho*I and employed as a PCR template. The first round PCR reactions were performed with mCNS1 forward and mutagenesis 1 reverse primers and with mutagenesis 1 forward and mCNS1 reverse primers (Table 1-1). 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 mCNS1 forward and mCNS1 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, I performed PCR with mCNS1 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. I performed PCR with mutagenesis 2 forward primer and mCNS1 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 mCNS1 reverse primer and T3 promoter primer. This resulted in generation of mCNS1, 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 mut-mCNS1-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 mut-mCNS1-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 mut-mCNS1-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 that contained a long upstream sequence of the *Tcam1* gene linked to the luciferase gene, I digested a BAC clone, B6Ng01-276I01, obtained from RIKEN Bioresource center, with *Tth111*I and collected a 14,383-bp fragment. I then inserted the fragment into pGL3-Basic at the *Sma*I site, but both of the two subclones I obtained included only 6969 bp of the fragment and lost 7414-bp 5' sequence. Because this fragment still encompassed mCNS1, mCNS2, the entire *lncRNA-Tcam1*, and the *Tcam1* promoter, I decided to use this construct for my analyses. I named the construct lncRNA-6.9kb-luc. To delete the mCNS1 sequence, I digested lncRNA-6.9kb-luc with *Ksp*I and self-ligated the larger fragment, and the resulting construct was ΔmCNS1-lncRNA-luc. This construct lost a 861-bp region containing mCNS1 (Fig. 1-13A). To generate lncRNA-6.9kb-EGFP, I 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, I 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, thus I could ligate these fragments. The resulting constructs were lncRNA-6.9kb-EGFP. ΔmCNS1-lncRNA-EGFP was made by connecting the 3041-bp *Kpn*I-*Xho*I fragment of ΔmCNS1-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 (FBS), 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[110].

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 AP 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) by the

TA-cloning method, and 10 subclones for each sample were sequenced. All the primer sequences are listed in Table 1-1.

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[117].

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

Each of the lncRNA-6.9kb-EGFP and ΔmCNS1-lncRNA-EGFP construct was co-transfected with a pKO SelectPuro V810 vector (Lexicon Genetics, The Woodlands, Texas, USA) into GC-2spd(ts) cells in 35mm 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, I started selection with 2-3 μg/ml of puromycin (Wako Pure Chemicals) and continued it for 11 days. After selection, I 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. I picked the wells containing a single colony and maintained the cells. For each clone, I 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[118].

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 mCNS1-Pro-luc with that of mut-mCNS1-Pro-luc or of reversed mCNS1-luc with reversed mut-mCNS1-luc. Statistical significance of EGFP or *lncRMA-Tcam1* expression between GC-2spd(ts) cell clones with lncRNA-6.9kb-EGFP and ΔmCNS1-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 chapter is available in the DDBJ/EMBL/GenBank databases under the accession number AB902906.

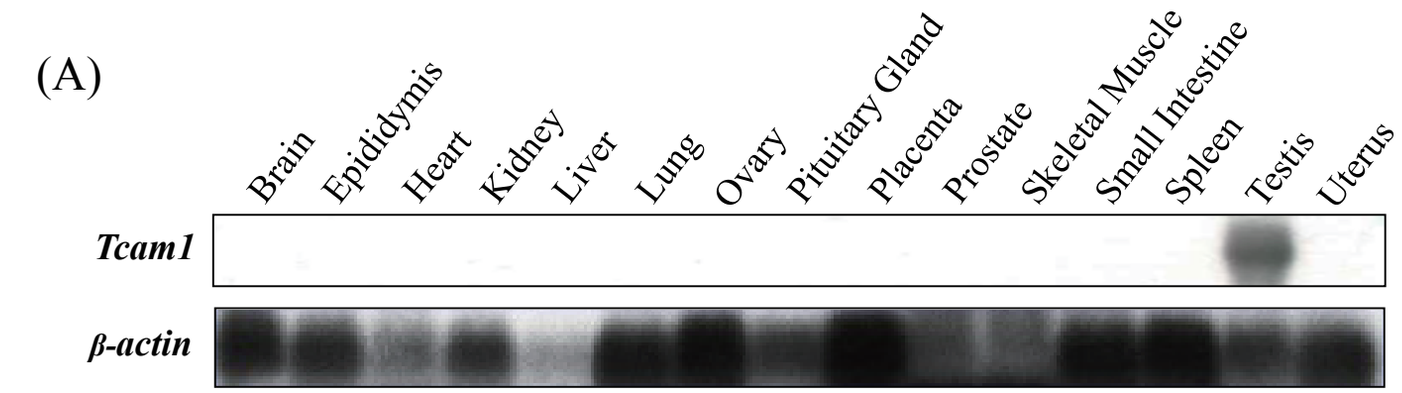
Table 1-1. Oligonucleotide primers used in Chapter 1

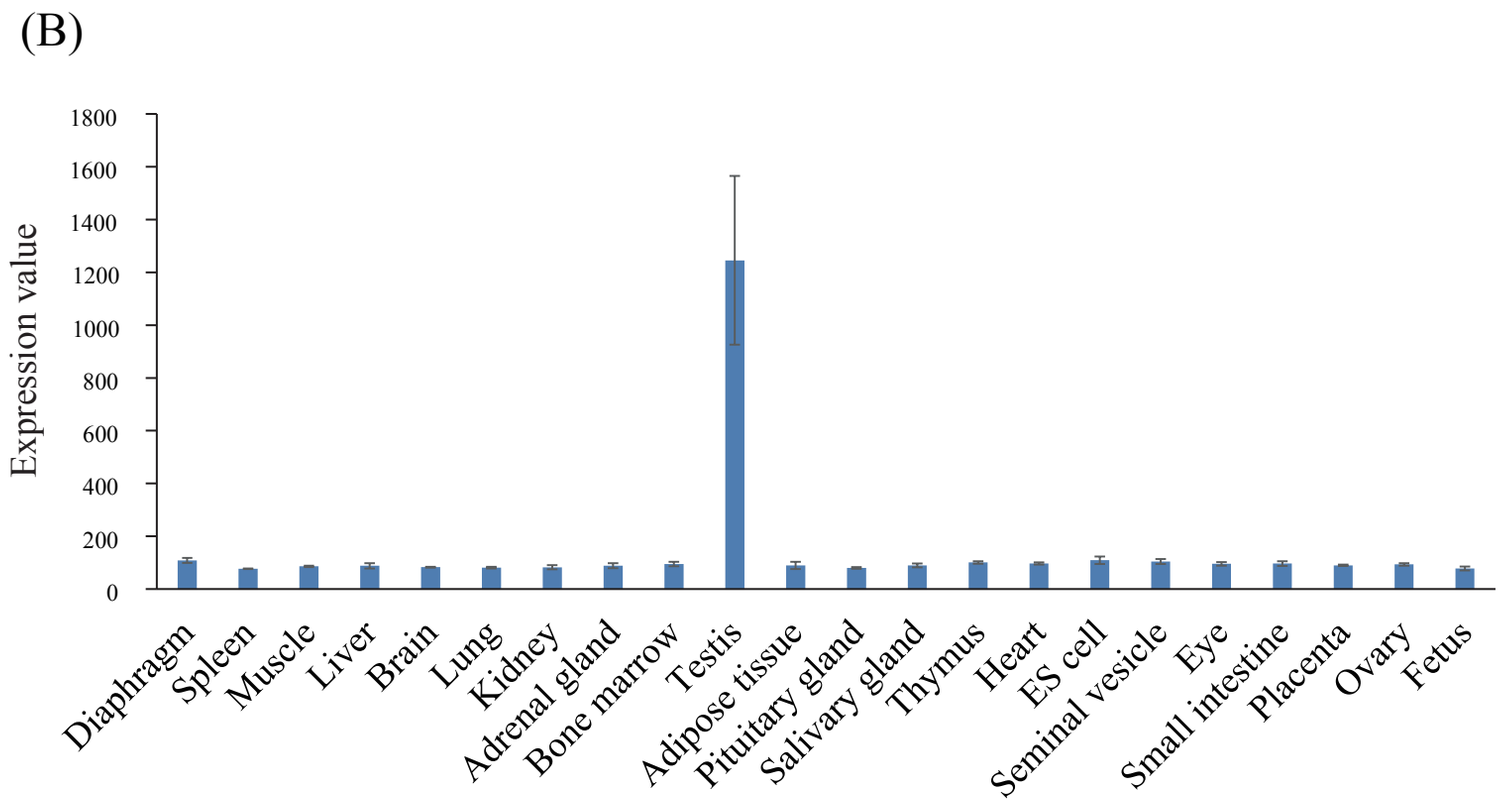
Designation	Forward	Reverse
Northern blot analysis and *in situ* hybridization		
Tcam1	5'-ATAGCCTGGCATGAGTTGCT-3'	5'-GCACCCTAAGACCGATTTCA-3'
β-actin	5'-ACATCCGTAAAGACCTCTATG-3'	5'-TAAAACGCAGCTCAGTAACAGT-3'
ChIP		
Aip-Promoter	5'-GGGCTTCAGCACAGAATCCA-3'	5'-TGAAAAATCCTGAGAGCCTCATT-3'
Smarcd2 intron1	5'-ACCCAGAGATGGCAGAATC-3'	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'

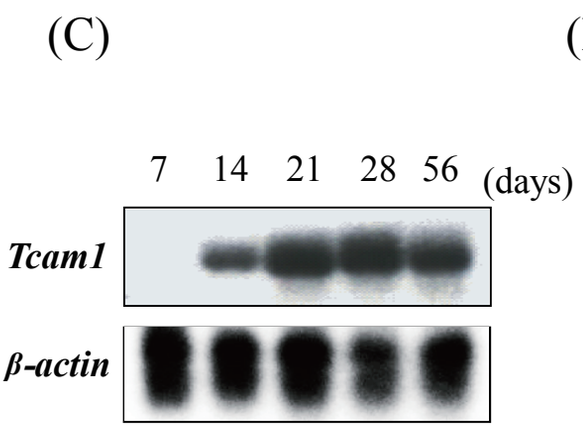
RT-PCR		
lncRNA-Tcam1-a	5'-TTTTAAGAGCCCATCTCGGG-3'	5'-AGGCTTAGCTTTCCTGCTCT-3'
lncRNA-Tcam1-b	5'-TTTTAAGAGCCCATCTCGGG-3'	5'-TGGCACACAAGTGAGATCAA-3'
Gapdh	5'-CATGACCACAGTCCATGCCATC-3'	5'-TAGCCCAAGATGCCCTTCAGTG-3'
Gapdh (in5-ex6)	5'-CCTTCTTTGTAGGTGTCCCT-3'	5'-TAGCCCAAGATGCCCTTCAGTG-3'
Gapdh (ex5-ex6)	5'-TTGTGATGGGTGTGAACCAC-3'	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 1-2. ChIP sequencing data

Cell/Organ	Antigen	Accession	source	Ref.
Liver	Input	GSM769034	GEO	65
Liver	H3K4me3	GSM769014	GEO	65
Round Spermatid	Input for H3K4me3	SRX336654	SRA	64
Round Spermatid	H3K4me3	SRX336652 SRX336653	SRA	64
Round Spermatid	Input for H3K4me3	GSM1046838 GSM1046839	GEO	66
Round Spermatid	H3K4me3	GSM1046840 GSM1046841	GEO	66
Round Spermatid	Input for Brdt	GSM984198	GEO	102
Round Spermatid	Brdt	GSM984200	GEO	102
Spermatocyte	Input for H3K4me3	SRX336651	SRA	64
Spermatocyte	H3K4me3	SRX336649 SRX336650	SRA	64
Spermatocyte	Input for Brdt	GSM984197	GEO	102
Spermatocyte	Brdt	GSM984199	GEO	102


(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

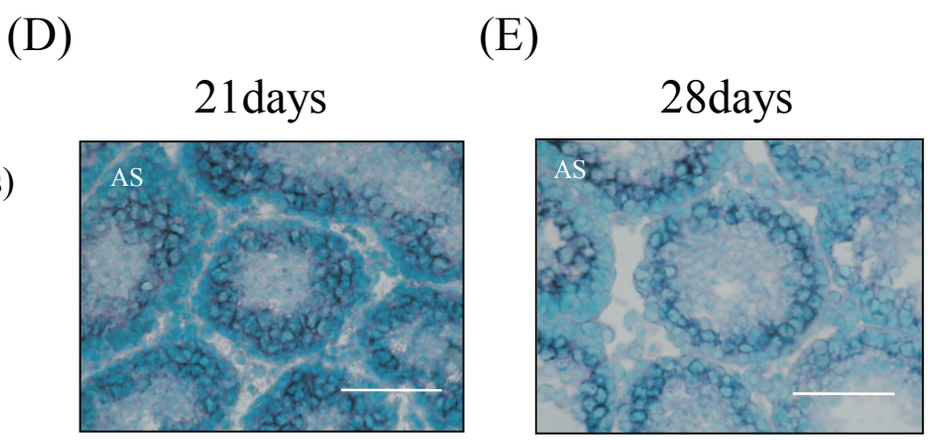

(D)
21days
AS
(E)
28days
AS

Fig. 1-1. 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.

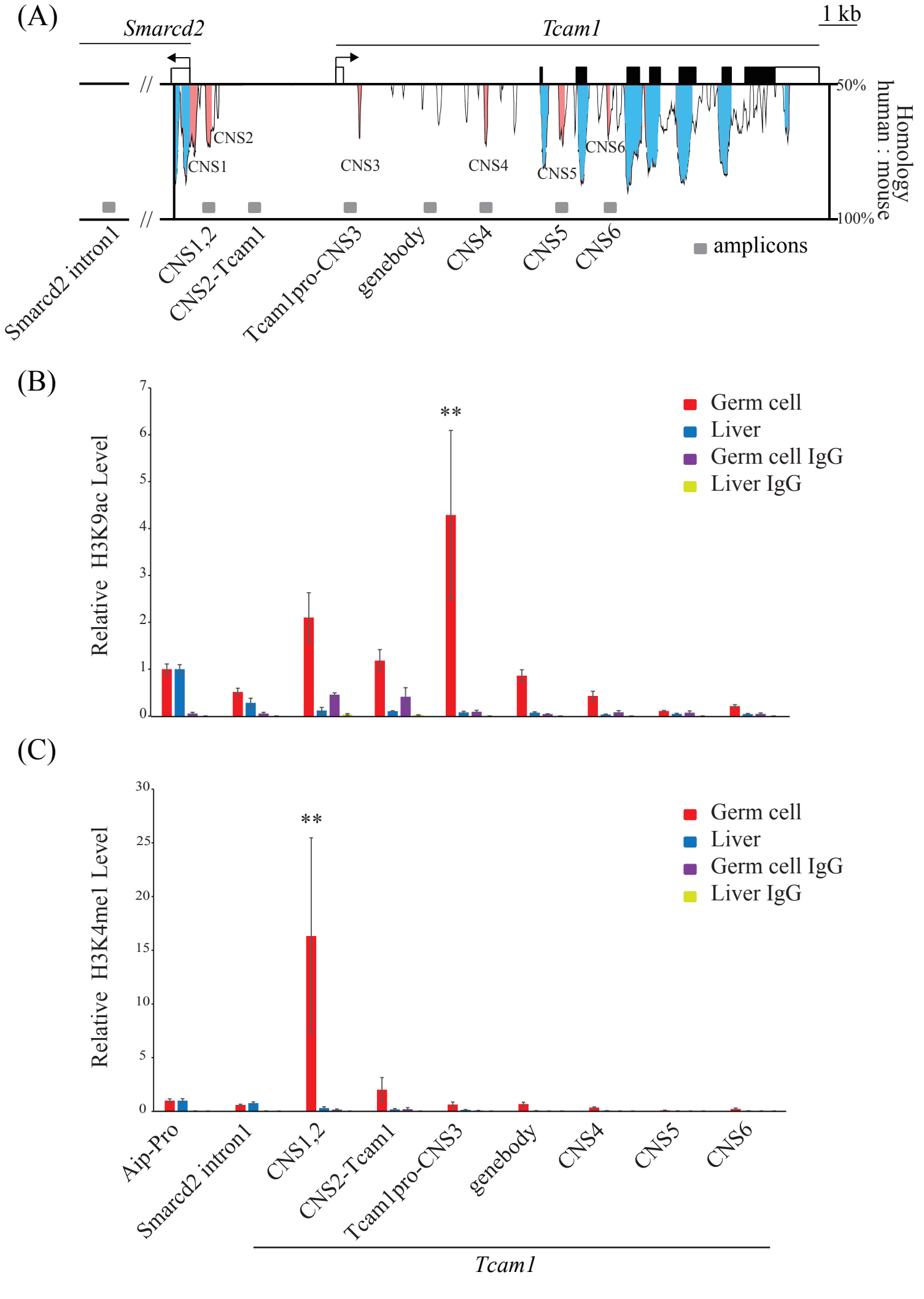

(A)
Smarcd2
Tcam1
1 kb
50%
100%
Homology human : mouse
CNS1
CNS2
CNS3
CNS4
CNS5
CNS6
amplicons
Smarcd2 intron1
CNS1,2
CNS2-Tcam1
Tcam1pro-CNS3
genebody
CNS4
CNS5
CNS6
(B)
Relative H3K9ac Level
Germ cell
Liver
Germ cell IgG
Liver IgG
**
(C)
Relative H3K4me1 Level
Germ cell
Liver
Germ cell IgG
Liver IgG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

Fig. 1-2. 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 mCNSs. Positions of amplicons for ChIP are indicated by gray boxes. Because the resolution of our ChIP analysis was 500–1000 bp, the amplicons at mCNS2 and a region between the *Tcam1* promoter and mCNS3 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 (**$P < 0.01$).

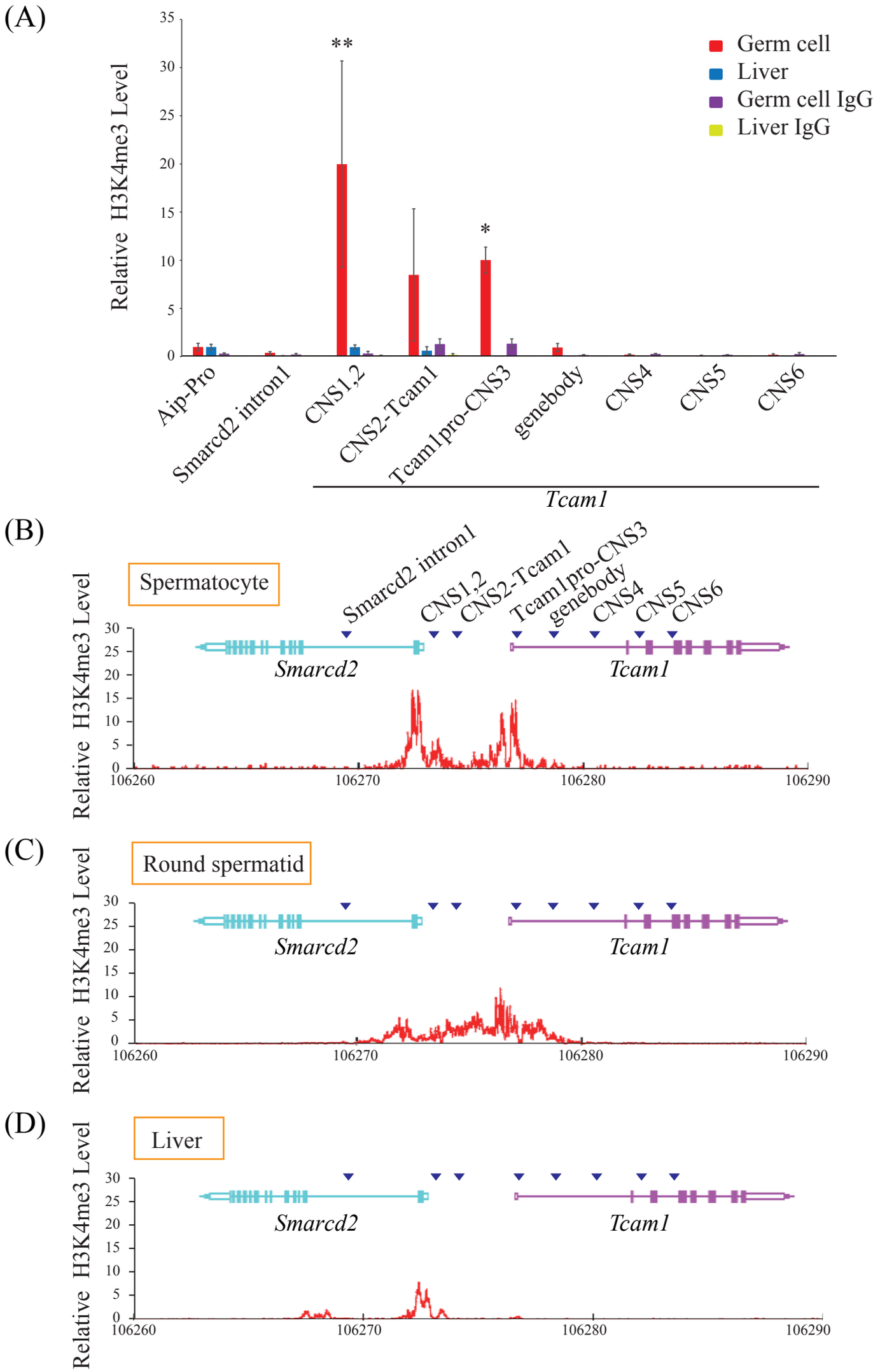

(A)
Relative H3K4me3 Level
Germ cell
Liver
Germ cell IgG
Liver IgG
**
*
Aip-Pro
Smarcd2 intron1
CNS1,2
CNS2-Tcam1
Tcam1pro-CNS3
genebody
CNS4
CNS5
CNS6
Tcam1
(B)
Spermatocyte
Smarcd2
Tcam1
106260
106270
106280
106290
(C)
Round spermatid
(D)
Liver

Fig. 1-3. 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. 1-2 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. 1-2. 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. 1-2A) are indicated by small triangles.

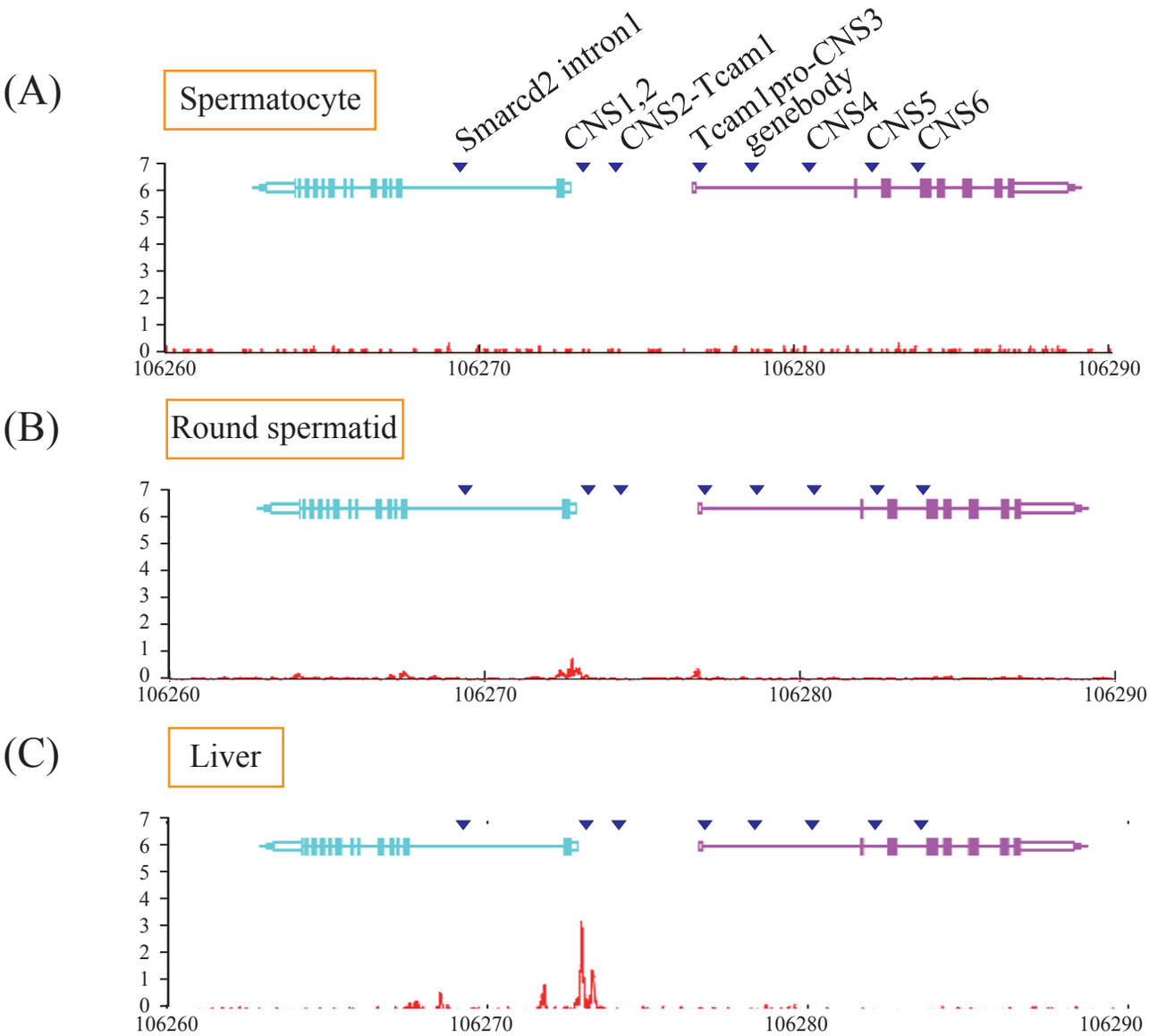


Fig. 1-4. Data of input DNA for H3K4me3 in spermatocytes (A), round spermatids (B)(SRA097278), and liver (C)(GSM769014). Gene structures of *Smarcd2*and *Tcam1* are depicted by light blue and purple lines and rectangles, respectively. Amplicon positions for ChIP-PCR (Fig. 1-2A) are indicated by small triangles. No significant peaks were observed at the *Smarcd2-Tcam1* locus.

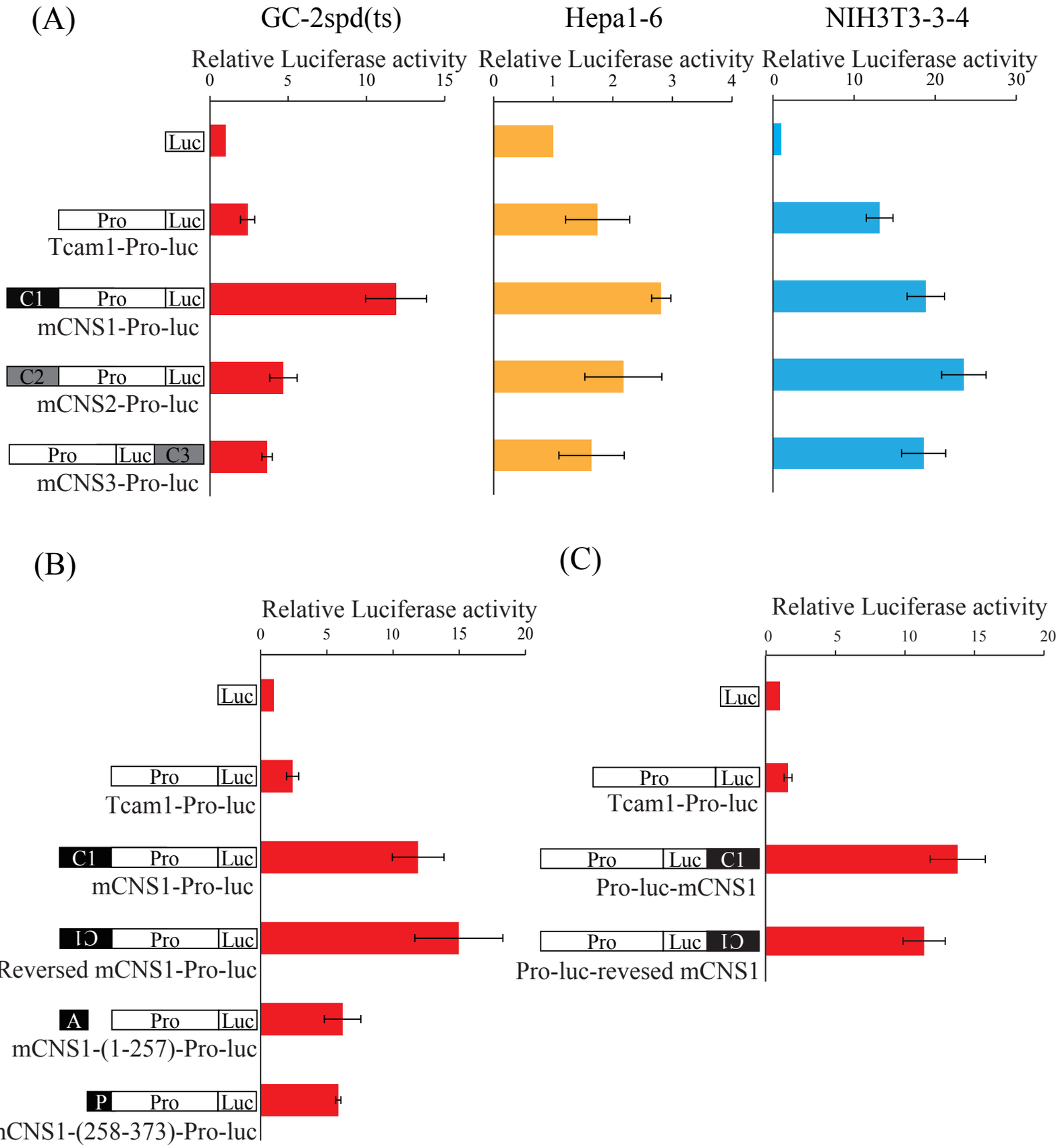

(A)
GC-2spd(ts)
Hepa1-6
NIH3T3-3-4
Relative Luciferase activity
0 5 10 15
0 1 2 3 4
0 10 20 30
Luc
Pro Luc
Tcam1-Pro-luc
C1 Pro Luc
mCNS1-Pro-luc
C2 Pro Luc
mCNS2-Pro-luc
Pro Luc C3
mCNS3-Pro-luc
(B)
Relative Luciferase activity
0 5 10 15 20
Luc
Pro Luc
Tcam1-Pro-luc
C1 Pro Luc
mCNS1-Pro-luc
C1 Pro Luc
Reversed mCNS1-Pro-luc
A Pro Luc
mCNS1-(1-257)-Pro-luc
P Pro Luc
mCNS1-(258-373)-Pro-luc
(C)
Relative Luciferase activity
0 5 10 15 20
Luc
Pro Luc
Tcam1-Pro-luc
Pro Luc C1
Pro-luc-mCNS1
Pro Luc C1
Pro-luc-revesed mCNS1

Fig. 1-5. *In vitro* reporter gene analysis for transcriptional activity of mCNS1, mCNS2, and mCNS3. Reporter gene constructs were generated as indicated at the left side of the graph. In the figure, mCNS1, mCNS2, and mCNS3 are indicated as C1, C2, and C3, respectively (A-C), and the 5' (mCNS1-(1-257)) and 3' halves (mCNS1-(258-373)) of mCNS1 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) mCNS1 increased the *Tcam1* promoter activity only in GC-2spd(ts) cells. (B) Enhancer activity of mCNS1 in reverse orientation and halved mCNS1 in GC-2spd(ts) cells. (C) Enhancer activity of mCNS1 at the downstream of the luciferase gene in GC-2spd(ts) cells.

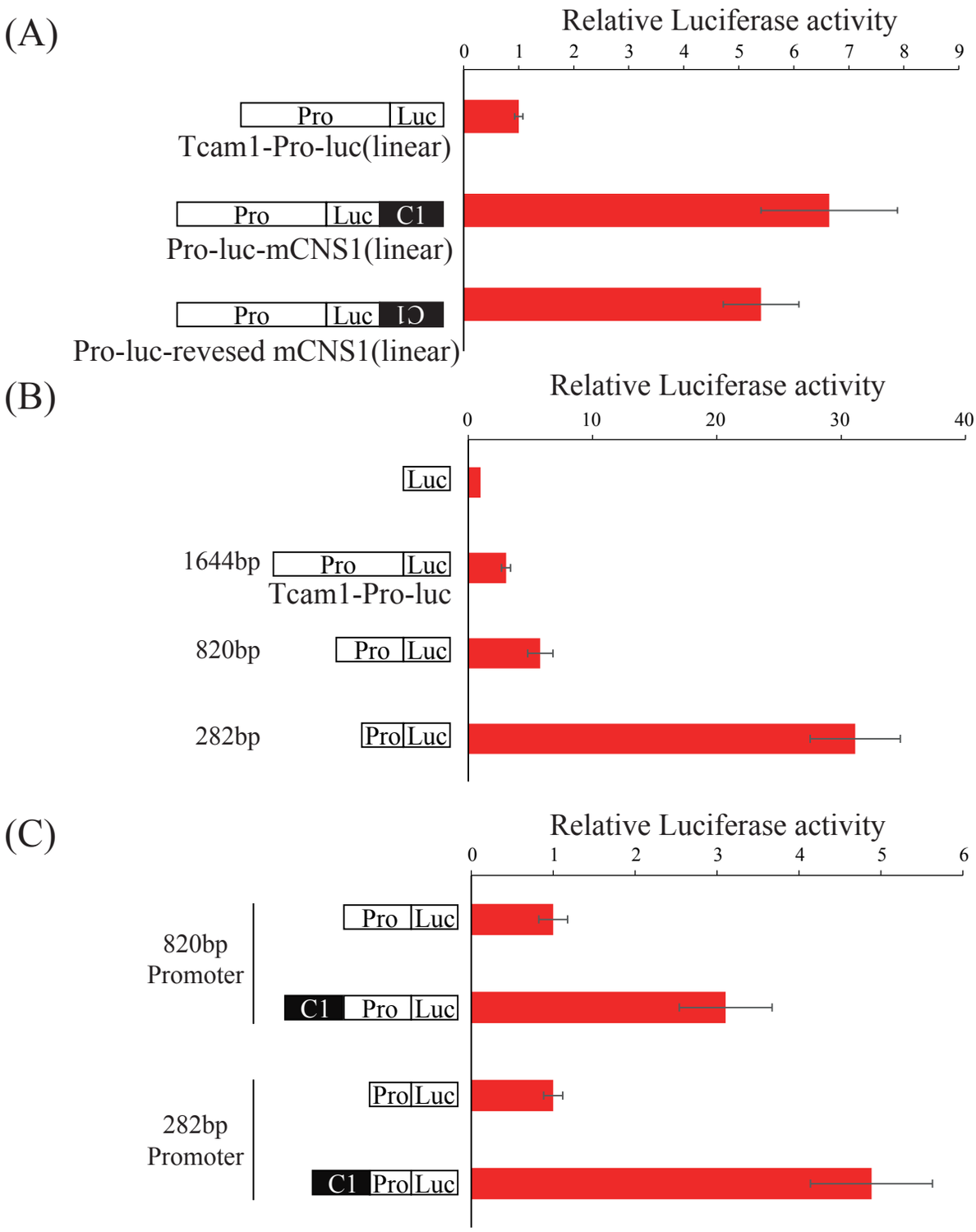
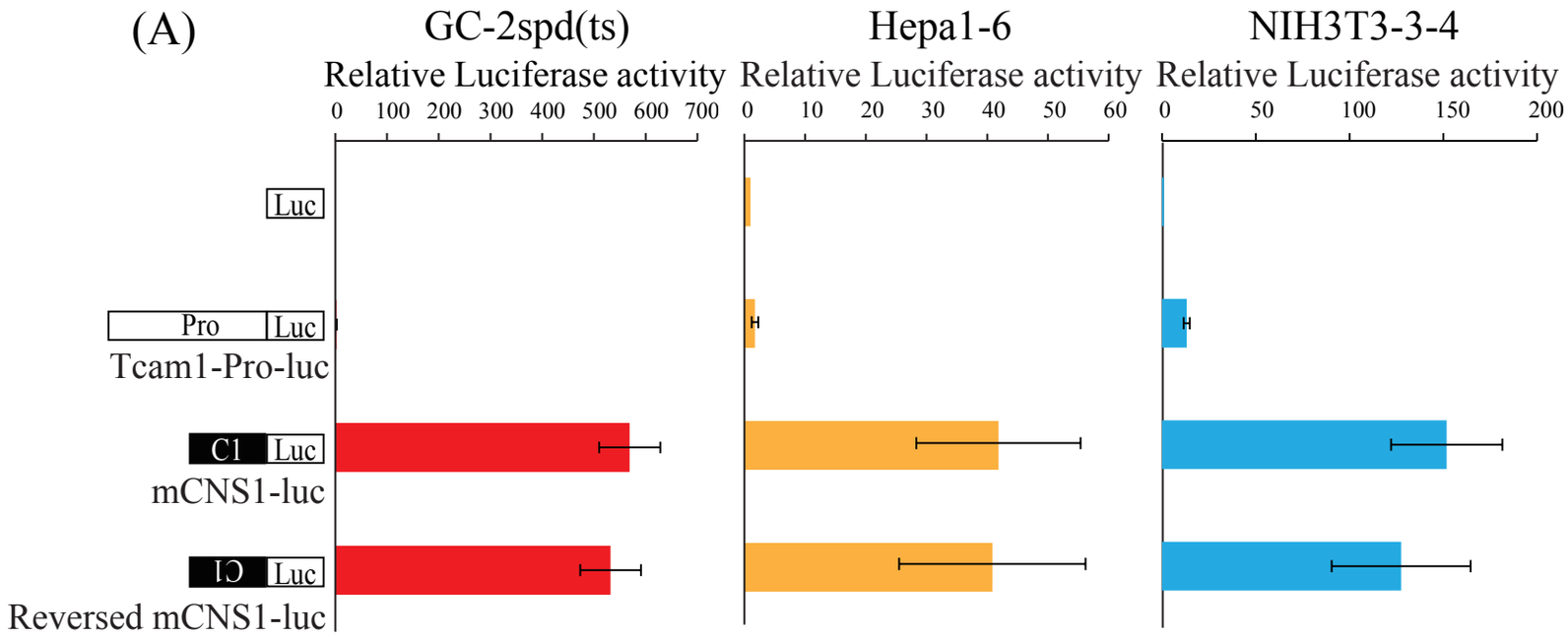
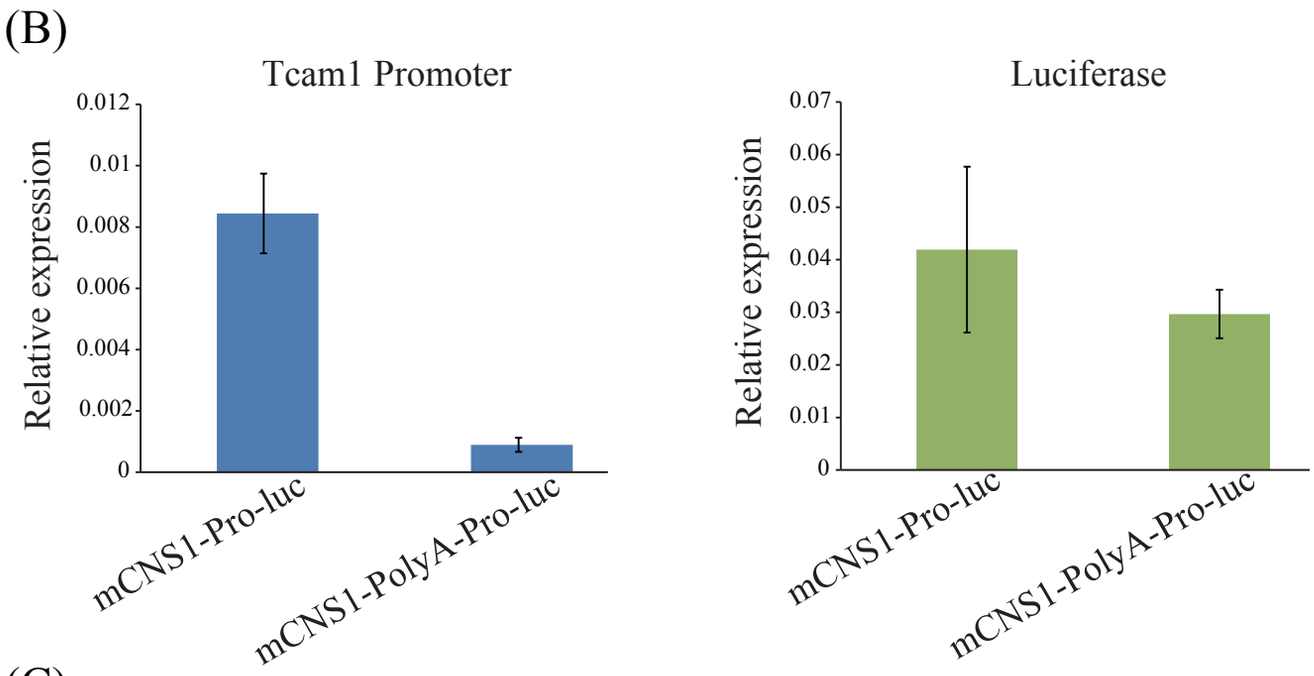


Fig. 1-6. Transcriptional activity of Tcam1 promoter and mCNS1 enhancer. (A) Transcriptional activity of mCNS1 in linearized constructs. The Tcam1-Pro-luc, Pro-luc-mCNS1, and Pro-luc-reversed mCNS1 constructs were linearized by digestion with SalI before transfection. The constructs were transfected 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. mCNS1 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 mCNS1 to 820-bp and 282-bp Tcam1 promoters. mCNS1 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 mCNS1 was set to 1.0. mCNS1 significantly enhanced activity of both promoters.


(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
mCNS1-luc
C1 Luc
Reversed mCNS1-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
mCNS1-Pro-luc
mCNS1-PolyA-Pro-luc
mCNS1-Pro-luc
mCNS1-PolyA-Pro-luc

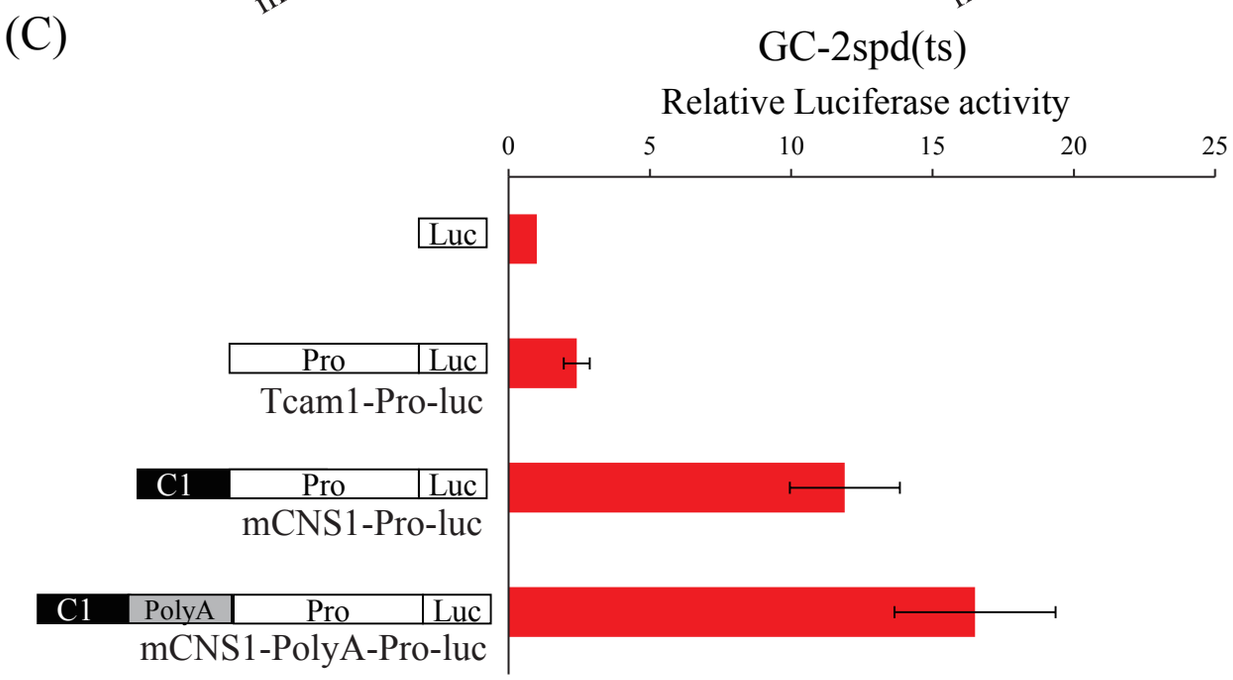

(C)
GC-2spd(ts)
Relative Luciferase activity
0 5 10 15 20 25
Luc
Pro Luc
Tcam1-Pro-luc
C1 Pro Luc
mCNS1-Pro-luc
C1 PolyA Pro Luc
mCNS1-PolyA-Pro-luc

Fig. 1-7. Promoter activity of mCNS1. (A) Bidirectional promoter activity of mCNS1 in GC-2spd(ts), Hepa1-6, and NIH3T3-3-4 cells. mCNS1 was connected directly to the luciferase gene in both orientations and the reporter gene assay was conducted as in Fig. 1-5 using three cell lines indicated. (B, C) The effect of poly(A) signal insertion between mCNS1 and the *Tcam1* promoter on enhancer activity of mCNS1. Using the mCNS1-Pro-luc construct, we inserted the poly(A) signal sequence between mCNS1 and the *Tcam1* promoter and the resulting construct was mCNS1-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 mCNS1 increased *Tcam1* promoter activity to a comparable level with the construct without insertion (C).

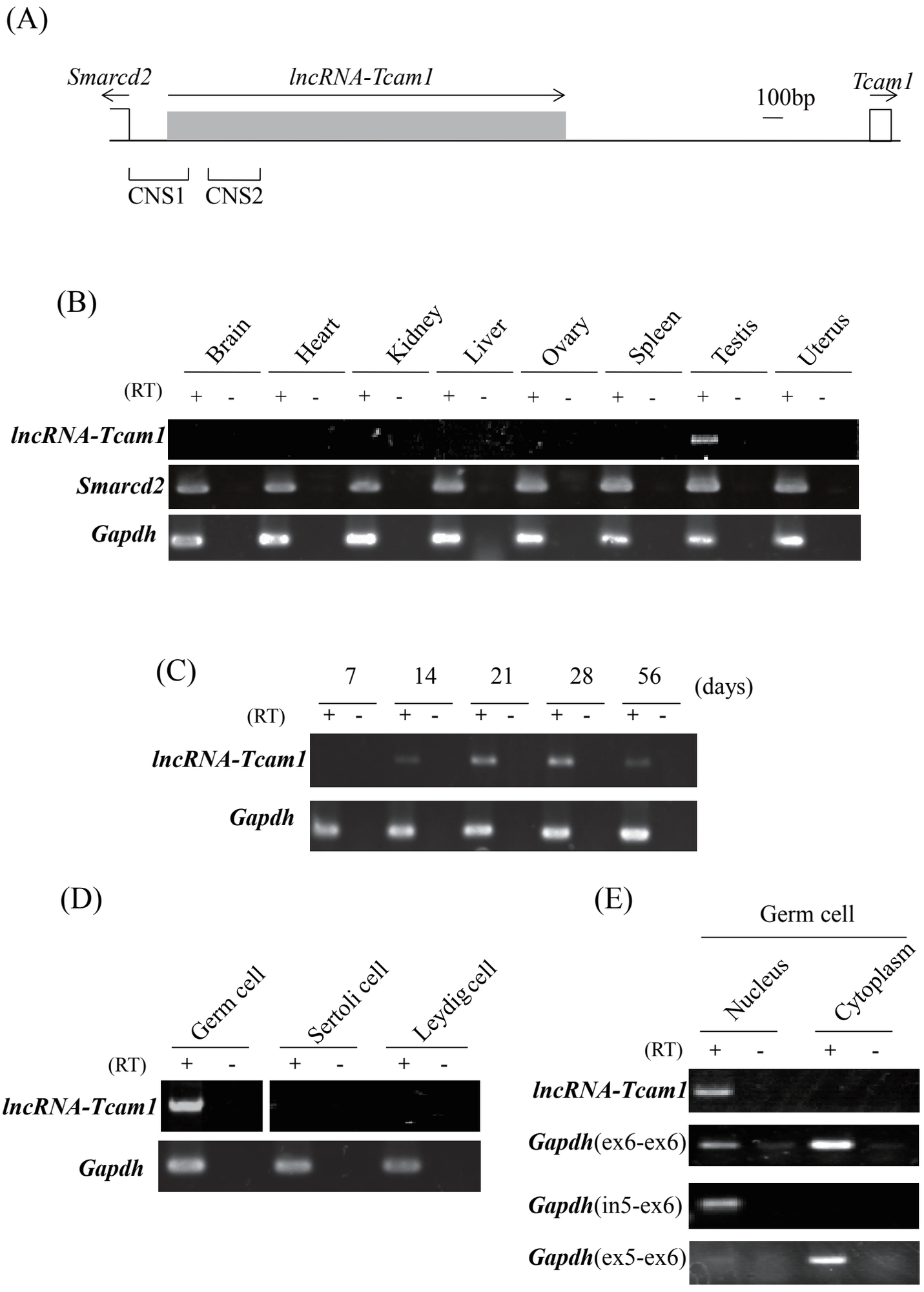

(A)
Smarcd2
lncRNA-Tcam1
100bp
Tcam1
CNS1
CNS2
(B)
Brain
Heart
Kidney
Liver
Ovary
Spleen
Testis
Uterus
(RT)
+ - + - + - + - + - + - + - + -
lncRNA-Tcam1
Smarcd2
Gapdh
(C)
7
14
21
28
56
(days)
(RT)
+ - + - + - + - + -
lncRNA-Tcam1
Gapdh
(D)
Germ cell
Sertoli cell
Leydig cell
(RT)
+ - + - + -
lncRNA-Tcam1
Gapdh
(E)
Germ cell
Nucleus
Cytoplasm
(RT)
+ - + -
lncRNA-Tcam1
Gapdh(ex6-ex6)
Gapdh(in5-ex6)
Gapdh(ex5-ex6)

Fig. 1-8. 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 mCNS2 and a part of mCNS1. 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.

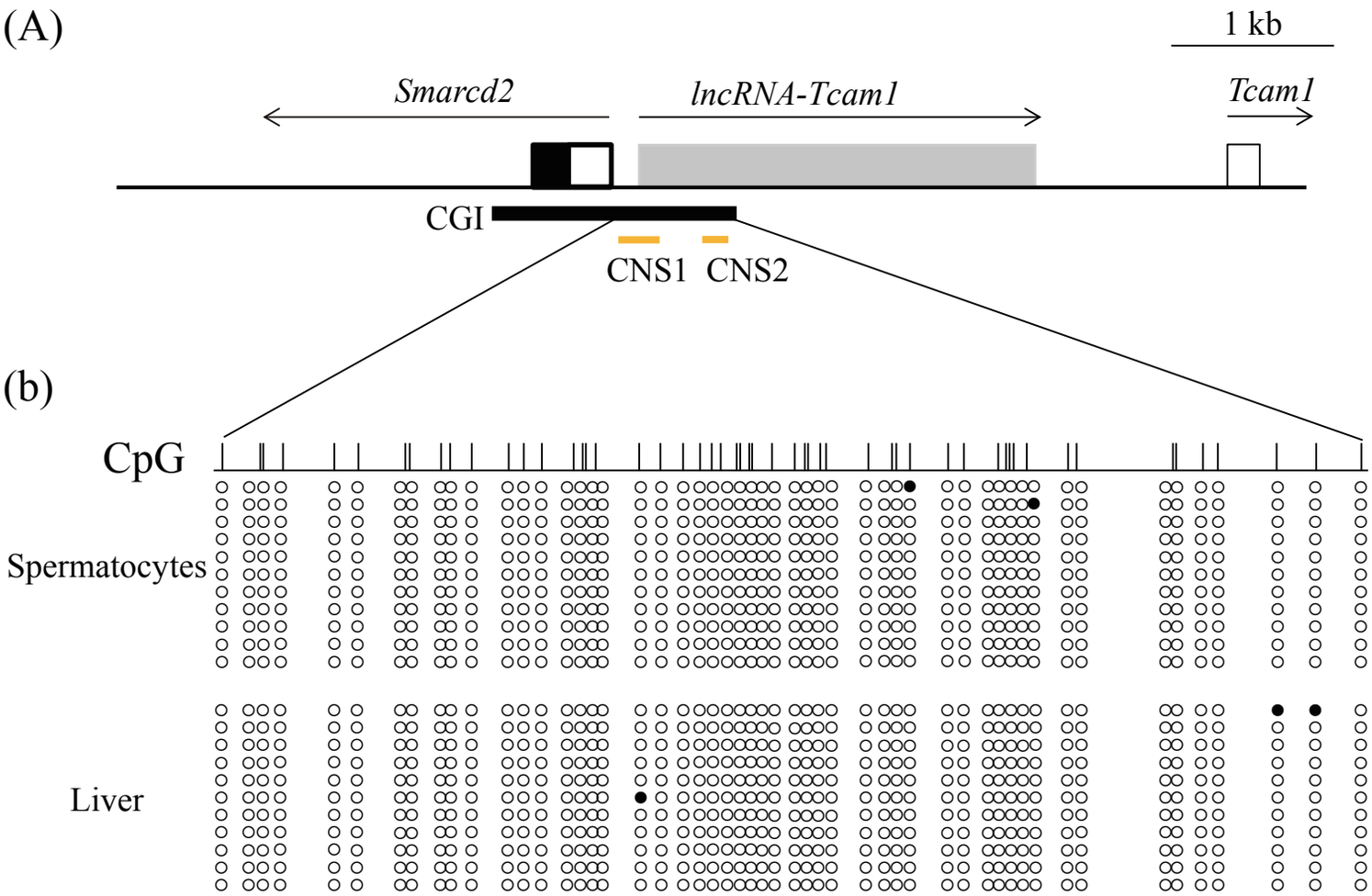


Fig. 1-9. DNA methylation patterns of CGI in spermatocytes and the liver. (A) Structure of the mouse *Tcam1* and *Smarcd2* locus. Exons of *Tcam1* and *Smarcd2* are depicted by open and filled boxes that represent untranslated and translated sequences, respectively. A region from which *lncRNA-Tcam1* is transcribed is depicted by a gray box. Transcriptional directions of the genes and lncRNA are shown by horizontal arrows. Positions of CGI, mCNS1, and mCNS2 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 mCNS1 and mCNS2, 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 $1^6 \times 10$ 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.

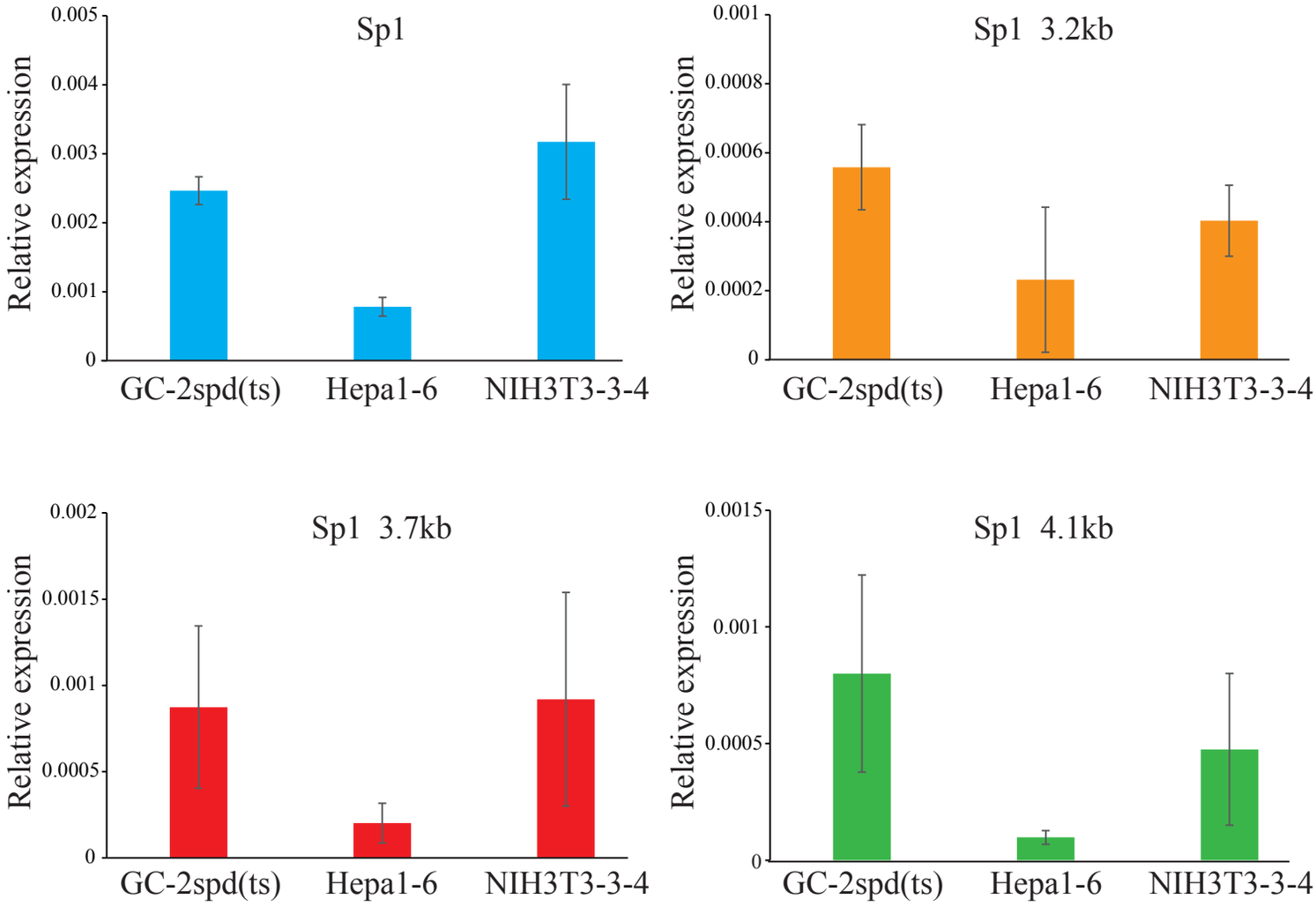
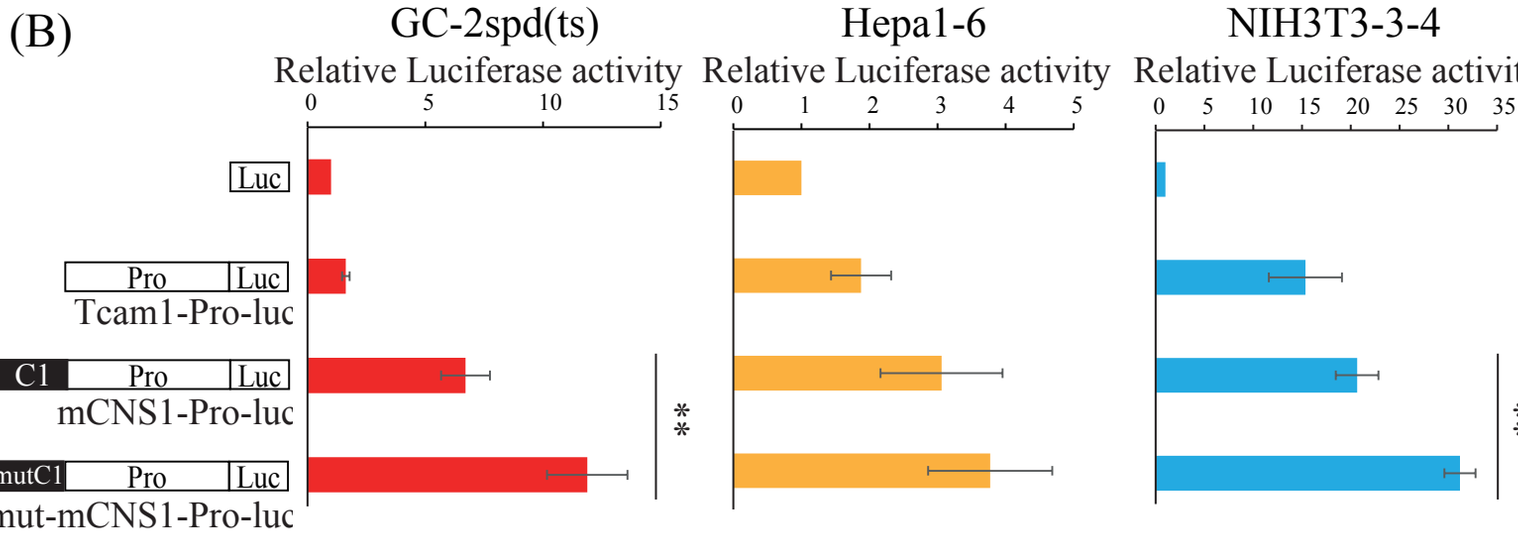


Fig. 1-10. Expression of *Sp1* mRNAs in GC-2spd(ts), Hepa1-6, and NIH3T3-3-4 cells. The qRT-PCR analysis was performed with total RNAs isolated from indicated cell lines. The oligo(dT) primer was used for reverse transcription. A primer pair for 'Sp1' was to detect a most common transcript which is ubiquitously 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.


(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
mCNS1-Pro-luc
mutC1
Pro
Luc
mut-mCNS1-Pro-luc
**

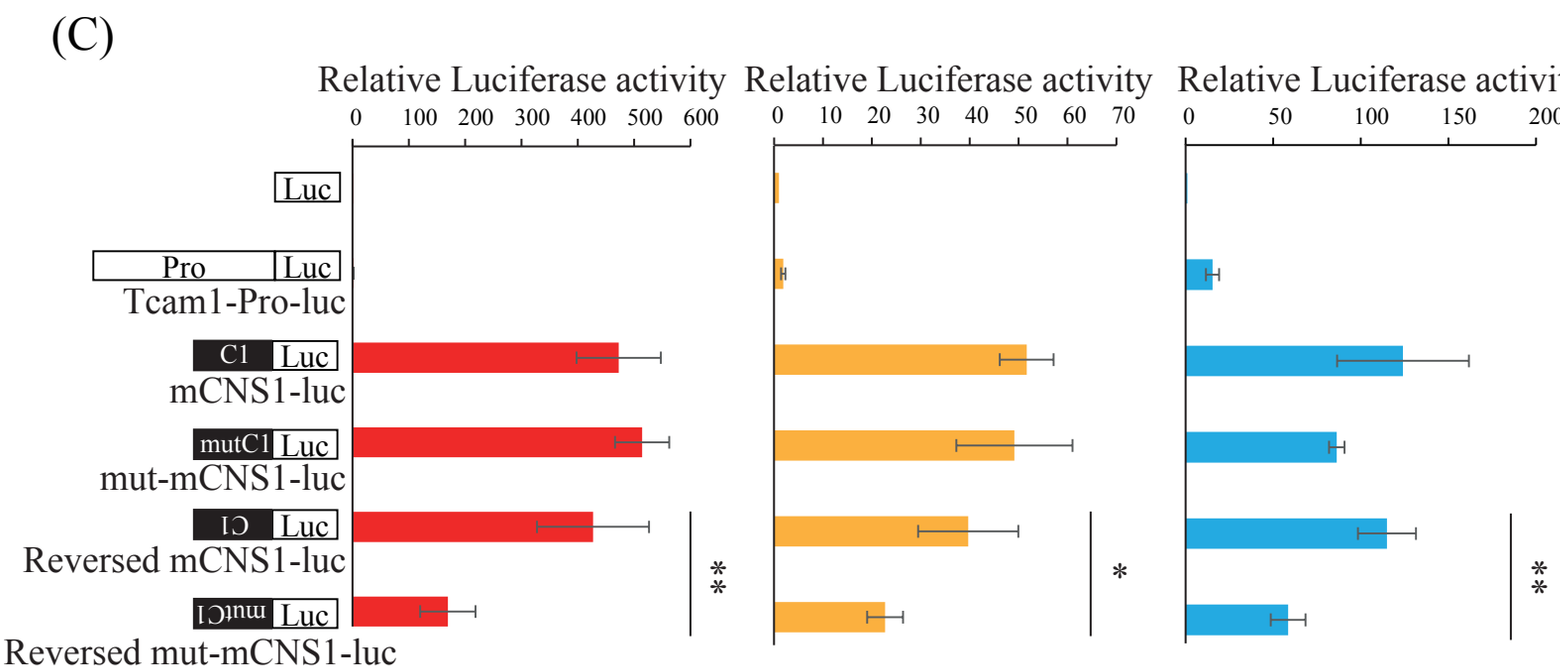

(C)
Relative Luciferase activity
Luc
Pro
Luc
Tcam1-Pro-luc
C1
Luc
mCNS1-luc
mutC1
Luc
mut-mCNS1-luc
Reversed mCNS1-luc
Reversed mut-mCNS1-luc
**
*
**

Fig. 1-11. Effects of mutation of three Sp1 binding sites on promoter and enhancer activity of mCNS1. (A) A multiple alignment of mCNS1 from seven mammalian species. mCNS1 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 octanucleotide elements for the Sp1 transcription factor binding site are highly conserved and marked by red boxes. (B) Enhancer activity of mCNS1 with mutated Sp1 binding sites. We prepared mCNS1 in which all three Sp1 sites were mutated not to be recognized by Sp1 and connected the mutated mCNS1 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. 1-5. The Tcam1-Pro-luc and mCNS1-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 mCNS1-Pro-luc with that of mut-mCNS1-Pro-luc. The luciferase activity of mut-mCNS1-Pro-luc was significantly higher than that of mCNS1-Pro-luc in GC-2spd(ts) and NIH3T3-3-4 cells. $^{**}P < 0.01$. (C) Bidirectional promoter activity of mCNS1 with mutated Sp1 binding sites. The mutated mCNS1 was directly connected to the luciferase gene in both directions and the luciferase activity was measured as in (B). The activity of reversed mut-mCNS1-luc was significantly lower than that of reversed mCNS1-luc in all the cell lines. $^{**}P < 0.01$, $^{*}P < 0.05$.

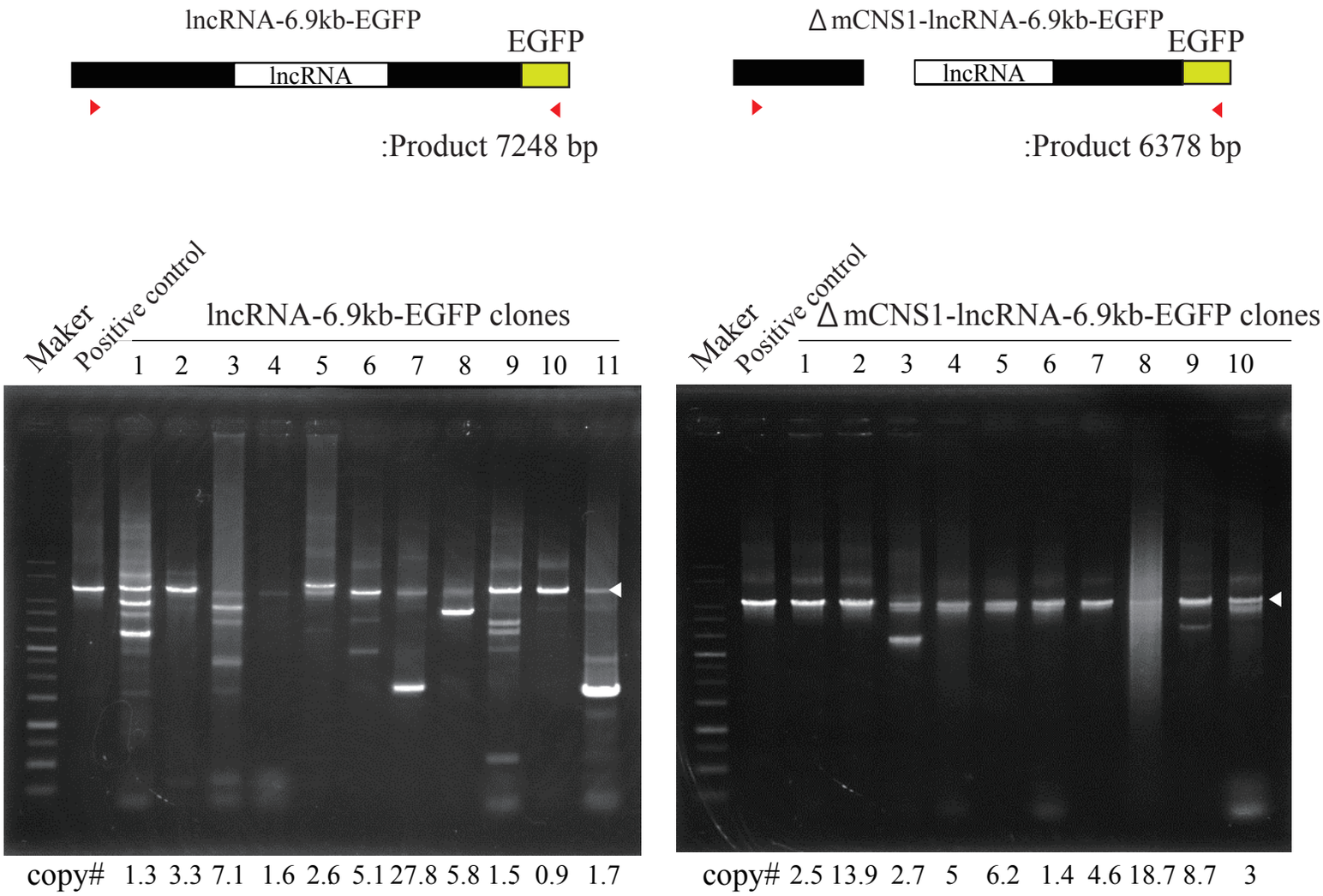


Fig. 1-12. Genomic PCR of stable GC-2spd(ts) cell clones. A schematic drawing of the transgene constructs and the primer positions are indicated with their product sizes at the top. PCR was performed using 3-200 ng genome DNAs from the clones with ExTaq polymerase (Takara) in a total volume of 10μl. 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 ΔmCNS1-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.

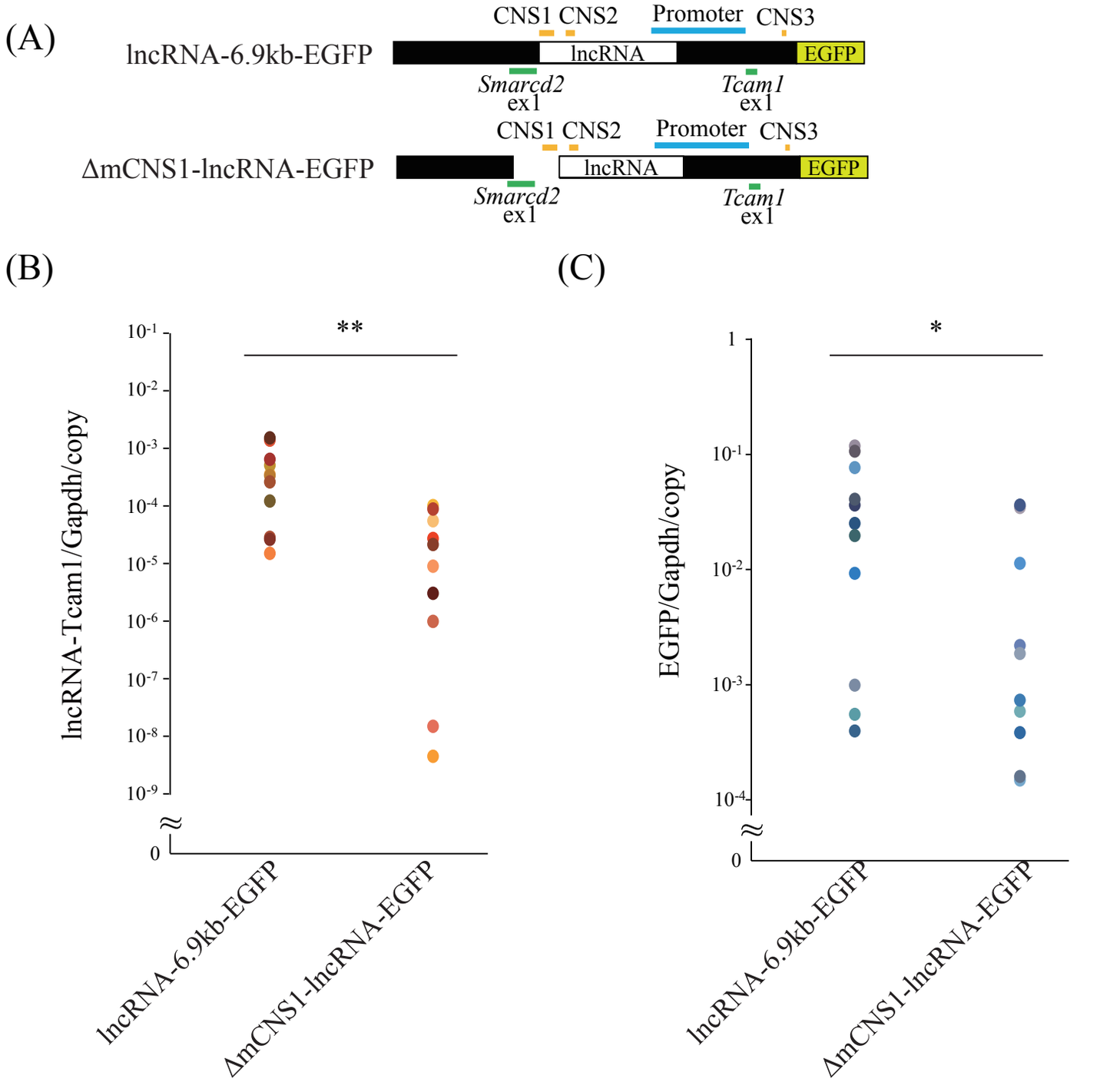

(A)
lncRNA-6.9kb-EGFP
CNS1 CNS2
Promoter
CNS3
lncRNA
EGFP
Smarcd2 ex1
Tcam1 ex1
ΔmCNS1-lncRNA-EGFP
(B)
lncRNA-Tcam1/Gapdh/copy
**
lncRNA-6.9kb-EGFP
ΔmCNS1-lncRNA-EGFP
(C)
EGFP/Gapdh/copy
*
lncRNA-6.9kb-EGFP
ΔmCNS1-lncRNA-EGFP
(D)
Relative Expression per copy
EGFP
lncRNA-Tcam1
Clone #
1 2 3 4 5 6 7 8 9 10 11

Fig. 1-13 Effects of mCNS1 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 mCNS1, *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 ΔmCNS1-lncRNA-EGFP construct was generated by deleting a 861-bp region including mCNS1 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 ΔmCNS1-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 ΔmCNS1-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 ΔmCNS1-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*.

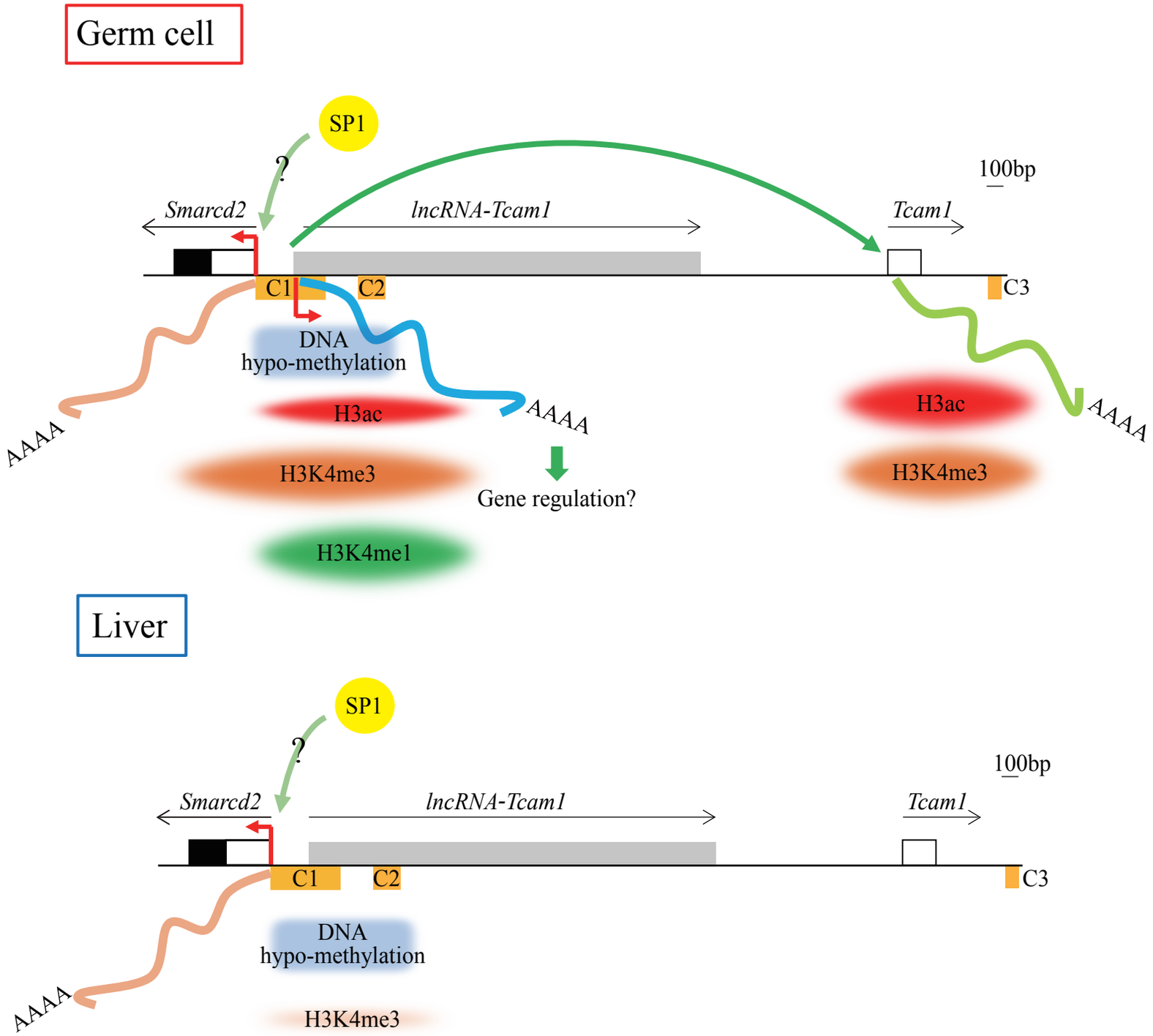


Fig. 1-14. A model of mCNS1 function as an enhancer and a bidirectional promoter. Exon 1 of the *Smarcd2* and *Tcam1* gene is indicated by using solid and open boxes, which represent translated and untranslated region, respectively, and a transcribed region of *lncRNA-Tcam1* is drawn by a gray box. Transcriptional directions of these genes and lncRNA are shown by horizontal arrows. Positions of mCNS1, mCNS2, and mCNS3 are indicated by orange boxes (C1, C2, C3). In testicular germ cells, mCNS1 functions as a bidirectional promoter of *lncRNA-Tcam1* (blue line) and *Smarcd2* mRNA (pink line). Sp1 may contribute to the mCNS1 promoter activity for the *Smarcd2* gene. CGI is completely hypo-methylated, and high levels of H3K9ac, H3K4me3, and H3K4me1 are observed at mCNS1 and mCNS2. The *Tcam1* promoter and mCNS3 are also marked with H3K9ac and H3K4me3 in germ cells. In spermatocytes, mCNS1 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, mCNS1 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.

Chapter 2:

Characterization of the human *TCAM1P* pseudogene and its activation by a potential dual promoter–enhancer: comparison with a protein-coding mouse orthologue

Abstract

TCAM1P is a pseudogene, which was disabled after human-mouse divergence. Here I found that *TCAM1P* was specifically expressed in the human testis, with different cell type-specificity from that of mouse *Tcam1*, and characterized its transcripts. At the mouse locus, a multifunctional DPE controls *Tcam1* gene expression. In this study, the corresponding human sequence was found to potentially function as DPE, although the molecular mechanism was different between the two species. The data suggest the presence of DPE in human genome for the first time, and provide an important model of evolutionary changes in a regulatory mechanism of a pseudogene.

Introduction

Functions and expression patterns of orthologous genes are generally conserved between closely related species[119,120], but it is not the case for pseudogenes. Pseudogenes have lost their original functions in the process of evolution, and are silenced or expressed as noncoding RNAs, sometimes with different patterns from their ancestral genes[121]. Recently, some of the transcribed pseudogenes were revealed to be functional and have novel roles[122–124]. Thus, it is of great importance to have a better understanding of pseudogenes, but many issues remain to be resolved. Especially, regulatory mechanisms of pseudogenes are largely unknown.

In chapter 1, I investigated a regulatory mechanism at the mouse *Tcam1* gene locus, and identified mCNS1 as a multifunctional DPE. mCNS1 could function as a spermatocyte-specific enhancer for *Tcam1* and a bidirectional promoter of the *Smarcd2* gene and a testis-specific lncRNA, *lncRNA-Tcam1* (Fig. 1-14). This was the first indication of DPE in mammals, although two examples had been reported in chicken[99,101].

Interestingly, while the synteny of *Smarcd2* and *Tcam1* genes is conserved among eutherians, a human orthologue of *Tcam1*, *TCAM1P*, is considered to be a pseudogene[30], albeit mouse *Tcam1* encodes a cell adhesion protein[29]. *TCAM1P* was pseudogenized since human-mouse divergence[30], and therefore, it is of significance to examine whether expression patterns and regulatory mechanisms at the human *TCAM1P* locus are conserved or not. The mechanisms are also notable in terms of the gene regulation by multifunctional genomic elements, because the *TCAM1P* gene is linked to the *hGH* gene cluster, a famous model for tissue-specific gene activation by multifunctional elements[15,17,19,21,125].

In this chapter, I investigated the expression and regulation of human *TCAM1P* and *SMARCD2* genes. My results indicated that the expression pattern of *TCAM1P* was slightly different from that of mouse *Tcam1* and that hCNS1 could function as DPE, although the molecular

mechanism was different from mCNS1. To my knowledge, this is the first report of the presence of DPE in human genome. This study suggests that the human *TCAM1P* locus would serve as an important model for exploring changes in gene regulatory mechanisms that accompany pseudogenization.

Results

Tissue distribution of human *SMARCD2* and *TCAM1P* mRNAs

I first investigated the tissue-specificity of human *SMARCD2* and *TCAM1P* genes by searching for the next generation sequencing data on Integrative Genomics Viewer (IGV) (http://www.broadinstitute.org/igv). Transcript signals at the *TCAM1P* gene were most abundantly present in the testis among eleven human tissues, while significant levels were observed at the *SMARCD2* gene in all the tissues (Fig. 2-1A). To know what types of cells expressed *TCAM1P* mRNA in the human testis, I referred to microarray data of fractionated testicular germ cells and of the whole testis from patients whose spermatogenesis was arrested at various stages. Both data showed that *TCAM1P* mRNA was exclusively expressed in germ cells, mainly in spermatocytes, but weak signals were also detected in spermatogonia and spermatids[126,127]. These indicated that the cell type-specificity of human *TCAM1P* was different from that of mouse *Tcam1* which was expressed only in spermatocytes (Fig. 1-1)[29].
Because the *TCAM1P* gene has been considered to be a pseudogene, I next investigated what kinds of transcripts were actually transcribed. According to the Ensembl database

(http://www.ensembl.org/index.html), seven transcripts were transcribed from the *TCAM1P* gene as shown in Fig. 2-1B. To distinguish these transcripts from one another, I designed four primer sets (Fig. 2-1B, P1-P4) and performed RT-PCR with human testis and liver RNAs. As a result, I could not amplify any specific products with P1, P2, and P4 in either tissue, but detected three specific bands in the testis with P3 (data not shown, Fig. 2-1C). I named the detected transcripts, *TCAM1P-I*, *TCAM1P-II*, and *TCAM1P-III* (Fig. 2-1B, C).

Determination of full-length sequences of *TCAM1P* transcripts

To determine full-length sequences of the three *TCAM1P* transcripts, I conducted 5'RACE and 3'RACE analyses using the human testis RNA. 5'RACE was performed with two primer sets: one for detecting TSS of *TCAM1P-I* and *TCAM1P-III* (Fig. 2-2B), and the other for TSS of *TCAM1P-I* and *TCAM1P-II* (Fig. 2-2C). I identified one nucleotide at the position +401 (which was 401 bp downstream of TSS in the Ensembl database) as TSS of *TCAM1P-I* and *TCAM1P-III* (Fig. 2-2B). In contrast, I found five different TSSs of *TCAM1P-I* and *TCAM1P-II*, at the positions +494, +401, +377, +329, and +5100 (Fig. 2-2C). The nucleotide at +401 might be derived from *TCAM1P-I*. These results indicate that there was one TSS of *TCAM1P-I* and *TCAM1P-III* at +401, and that *TCAM1P-II* had at least four TSSs.

3'RACE was performed using a primer set that could detect the transcriptional termination site (TTS) of all the transcripts. Two TTSs were identified at the position of +10555 and +10837 (Fig. 2-2D). Because 8 out of 10 subclones contained the nucleotide at +10837 as TTS, it was considered to be a major TTS. To see whether all the three transcripts were actually terminated at the major TTS, I conducted RT-PCR by using a primer at the final exon for reverse transcription and a primer pair specific to each transcript for PCR. As a result, a specific signal

was observed for each of the three transcripts (data not shown), which indicated that all the *TCAM1P* transcripts contained +10837 as their TTSs. I also examined whether the three transcripts were terminated at the minor TTS, but I could not detect any specific signals (data not shown).

Taken together, *TCAM1P-II* had at least four TSSs and one TTS, and *TCAM1P-I* and *TCAM1P-III* had one TSS and one TTS. Therefore, I concluded that at least six *TCAM1P* transcripts were present in the human testis as shown in Fig. 2-2E, and that the *TCAM1P* gene was actually composed of five exons (Fig. 2-2A). However, considering that a continuous region of the entire *TCAM1P* gene was transcribed in the human testis by the IGV data (Fig. 1A), many types of transcripts are probably transcribed from this gene. Types of *TCAM1P* transcripts may be different between individuals or may be dependent on ages or diseases.

Although the *Tcam1* gene in rodents and rhesus monkey encodes a protein related to a cell adhesion molecule, the human *TCAM1P* gene has been considered to be a pseudogene[30]. Since I determined full-length sequences of six *TCAM1P* transcripts, I searched for their open reading frames (ORFs). The maximum ORFs predicted from the nucleotide sequences were 99, 58, and 49 amino acids for *TCAM1P-I*, *TCAM1P-II*, and *TCAM1P-III*, respectively. However, amino acid sequences of those ORFs showed no homology to the mouse TCAM1 protein. Therefore, the human *TCAM1P* gene could be noncoding RNAs or could encode polypeptides that were not related to the TCAM1 protein of other species.

Enhancer and promoter activity of hCNS1

I then investigated a regulatory mechanism at the human *SMARCD2-TCAM1P* locus by examining whether hCNS1 could function as DPE similarly to mCNS1.

I first investigated enhancer activity of hCNS1. Because TSS of *TCAM1P-I* and *TCAM1P-III* was +401 and three TSSs of *TCAM1P-II* were close to +401 (Fig. 2-2E), I cloned about 2-kb sequence upstream of +401 as a major promoter of the *TCAM1P* gene and connected it to the luciferase gene (TCAM1P-Pro-luc). hCNS1 was amplified by genome PCR and cloned to upstream of the *TCAM1P* promoter (hCNS1-Pro-luc) or to downstream of the luciferase gene (Pro-luc-hCNS1 and Pro-luc-rev hCNS1). Enhancer activity of hCNS1 was assessed by reporter gene assay with three cell lines, GC2-spd(ts), Hepa1-6, and HEK293T, which were derived from mouse spermatocytes, mouse hepatocytes, and human embryonic kidney, respectively.

The activity of hCNS1-Pro-luc was significantly increased in all three cell lines compared to the construct without hCNS1 (Fig. 2-3A). When I connected hCNS1 to 3'of the luciferase gene, the activity was also significantly increased in all cells irrespective of its orientation (Fig. 2-3A). Fold-increases were 1.8-3.5 in most experiments, but a 4.5-fold increase was observed in GC-2spd(ts) cells when hCNS1 was connected to 3' of the luciferase gene in forward orientation (Fig. 2-3A, Pro-luc-hCNS1). I do now know the reason for the variable activity, but there were several reports of enhancers showing different activity depending on its position or direction by reporter gene assay[128–131]. I also examined the luciferase activity in linearized constructs. By linearizing Pro-luc-hCNS1 and Pro-luc-rev hCNS1 constructs, hCNS1 still significantly increased *TCAM1P* promoter activity in HEK293T cells (Fig. 2-3B). All of these data showed that hCNS1 could function as an enhancer for the *TCAM1P* gene.

I next investigated whether hCNS1 had promoter activity or not. Since hCNS1 encompassed exon 1 and an upstream region of *SMARCD2*, it was regarded as a promoter of *SMRACD2*. To assess the promoter activity of hCNS1, I connected it directly to the luciferase gene in both orientations (Fig. 2-4). By reporter gene assay, hCNS1 showed promoter activity in both directions, but the activity was stronger in a reverse direction in all the cell lines (Fig. 2-4). This indicated that hCNS1 had bidirectional promoter activity, but the activity was stronger for the

SMARCD2 gene.

Intergenic transcription at the human *SMARCD2-TCAM1P* locus

At the mouse *Smarcd2-Tcam1* locus, mCNS1 can drive not only *Smarcd2* but also *lncRNA-Tcam1* (Chapter 1). Because hCNS1 showed bidirectional promoter activity by reporter gene assay (Fig. 2-4), I tested the possibility that hCNS1 could be a promoter for an lncRNA as well as *SMARCD2*. I performed RT-PCR with the human testis RNA and successfully amplified the hCNS2 sequence (data not shown) which was positioned close to hCNS1 (Fig. 2-5A). To determine the full-length sequence of the transcript, I performed 5'RACE and 3'RACE analyses and detected a single TSS and a single TTS (data not shown). Consequently, I identified a 1.1-kb transcript including one intron (Fig. 2-5A) and named it *lncRNA-TCAM1P*, because it was classified as an lncRNA by the coding potential calculator[68]. However, the *lncRNA-TCAM1P* promoter region did not overlap with hCNS1 (Fig. 2-5A), which indicated that hCNS1 was not a promoter for *lncRNA-TCAM1P*.

I then investigated the possible presence of other transcripts from the intergenic region between *SMARCD2* and *TCAM1P* genes. By RT-PCR with six primer pairs (Fig. 2-5A, 1-6) and human testis and liver RNAs, I detected specific signals from all the examined regions in the testis, but not in the liver (Fig. 2-5B). This indicated that, consistent with Fig. 2-1A, most of the intergenic region between *SMARCD2* and *TCAM1P* genes were specifically transcribed in the testis. Therefore, it is possible that there was a noncoding transcript driven by hCNS1 in the opposite direction to *SMARCD2*.

Discussion

hCNS1 is a potential DPE

One aim of this chapter was to determine whether human genome contained DPE. By reporter gene assay, it was demonstrated that hCNS1 had enhancer activity (Fig. 2-3). This suggests that hCNS1 is an enhancer for *TCAM1P* similarly to mCNS1. However, a mechanism underlying the enhancer activity seemed to be different, because hCNS1 enhancer activity was observed in all the cell lines I examined (Fig. 2-3), whereas mCNS1 enhancer activity was specific to spermatocyte-derived GC-2spd(ts) cells (Fig. 1-5A). This indicates that different transcription factors function in different nucleotides between mCNS1 and hCNS1. In the human testis, a transcription factor common to various cell types may bind to hCNS1 and up-regulate the *TCAM1P* expression. In this case, the testis-specific *TCAM1P* gene activation may be controlled by the epigenetic state or by other *cis*-/*trans*-elements.

hCNS1 also showed bidirectional promoter activity, but the activity was stronger for *SMARCD2* than for the opposite direction (Fig. 2-4). This is in contrast to mCNS1, whose promoter activity in each direction was similar (Fig. 1-7A). Because I failed to detect any transcripts driven by hCNS1 extending to the opposite direction of *SMARCD2* (Fig. 2-5), hCNS1 may be a unidirectional promoter. However, hCNS1 promoter activity opposite to *SMARCD2* was still much higher than that of *TCAM1P* promoter (Fig. 2-4), and most sequences of the intergenic region between these two genes were transcribed (Figs. 2-1A, 2-5). Further analysis will be necessary to reveal whether hCNS1 is a bidirectional promoter or not, but my data clearly indicate that hCNS1 is at least a promoter of the *SMARCD2* gene.
Collectively, hCNS1 could enhance *TCAM1P* promoter activity and functioned as a promoter of

the *SMARCD2* gene. Therefore, similarly to mCNS1, hCNS1 can be DPE in the human testis.

Evolutionary changes occurring in parallel with pseudogenization

Another aim of this work was to learn more about evolutionary changes in gene regulatory mechanisms that would have accompanied pseudogenization. Pseudogenes are generally classified as unitary, duplicated, and retrotransposed pseudogenes, which were generated by spontaneous mutations in a single copy gene, duplication of a genomic region, and retrotransposition of processed mRNAs into other genomic regions, respectively[132]. *TCAM1P* is a unitary pseudogene[30], and it is possible to directly compare the human *TCAM1P* locus with the mouse one.

In both human and mouse, I confirmed ubiquitous *Smarcd2/SMARCD2* expression and DPE activity of mCNS1/hCNS1, but otherwise, many differences were found by my analyses. Firstly, human *TCAM1P* was testis-specific, but it was expressed not only in spermatocytes but also in spermatogonia and spermatids, in contrast to mouse *Tcam1* mRNA that is specific to spermatocytes (Fig. 1-1)[126,127]. Secondly, functions of *TCAM1P* and *Tcam1* are obviously different. *TCAM1P* does not encode a cell adhesion molecule, and instead, encodes smaller peptides or functions as lncRNAs. Thirdly, hCNS1 showed enhancer activity in more cell types than mCNS1 does, indicating different molecular mechanisms between these two species as discussed above. Fourthly, hCNS1 enhancer activity was lower than that of mCNS1. By reporter gene assay, mCNS1 mostly showed about 5-fold higher activity than *Tcam1* promoter (Figs. 1-5, 1-6), while the activity of hCNS1 was 1.8-4.5-fold higher than *TCAM1P* promoter (Fig. 2-3). Fifthly, as discussed above it is unlikely that hCNS1 is a bidirectional promoter. Lastly, more lncRNAs were transcribed at the human locus.

Taken together, the expression of *TCAM1P* and lncRNAs was changed during evolution, and the gene regulation at this locus, especially the molecular mechanism underlying DPE activity, was also changed. These data provide an important example of evolutionary changes, especially in gene regulatory mechanisms, in parallel with pseudogenization.

Regulation of the *SMARCD2-TCAM1P-hGH* locus by multifunctional genomic elements

Finally, it is interesting to note gene regulation by multifunctional genomic elements because the human *TCAM1P* gene is linked to the *hGH* gene cluster. The *hGH* cluster contains one pituitary-specific and four placenta-specific paralogues[14,53], and their tissue-specific expression is related to some multifunctional elements. Firstly, the *hGH* cluster is controlled by the LCR, which overlaps with the *SCN4A* gene encoding a sodium channel protein in skeletal muscle[15,16,133]. Secondly, the *hGH* LCR activates the cluster genes by different mechanisms in pituitaty and placenta[18,19,22,23,134]. Thirdly, the *CD79b* gene, which is located between the LCR and the cluster, functions as a noncoding RNA to activate the *hGH-N* gene in the pituitary, but it is translated to a protein in the B cell[19,20,135]. In addition to these, my current data suggest that DPE controls *SMARCD2* and *TCAM1P* gene expression at this locus. Therefore, the *SMARCD2-TCAM1P-hGH* locus is a very important model to study multifunctional genomic elements and tissue-specific gene activation.

Materials and methods

RNA Analyses

Total RNAs from the human liver and testis were purchased from Clontech Laboratories, Inc. and RT-PCR was done as previously described[117]. Primer sequences are shown in Table 2-1. The primer pair for *β-ACTIN* was used by another group[136].

5' RACE and 3'RACE

5'RACE and 3'RACE were performed as described in chapter 1. All the amplified products were subcloned into a pBluescript vector (Stratagene) by the TA-cloning method, and at least 10 subclones were sequenced for each product. All the primer sequences used for these analyses are listed in Table 2-1.

Reporter constructs

A promoter region of the *TCAM1P* gene was amplified by PCR using KOD FX (Toyobo) with human sperm genomic DNA (Clontech Laboratories, Inc.) and a primer pair listed in Table 2-1. The 2012-bp product was ligated to upstream of the luciferase gene in a pGL3-Basic vector (Promega Corporation) at the *Sma*I site, and I named the resulting construct TCAM1P-Pro-luc.

To obtain the hCNS1 sequence, I had to perform two rounds of PCR reactions, presumably due to the high GC content of the region encompassing hCNS1. The first round PCR was performed by KOD FX (Toyobo) with human genomic DNA using the 'hCNS1 first' primer pair (Table 2-1), and the 1340-bp product was purified and used as a template for the second PCR with the 'hCNS1 nested' primer pair (Table 2-1). By inserting the 450-bp product into the blunted *Nhe*I site of TCAM1P-Pro-luc, I obtained the hCNS1-Pro-luc construct. To generate the Pro-luc-hCNS1 and Pro-luc-reversed (rev) hCNS1 constructs, I cloned the hCNS1 fragment into the blunted *Bam*HI site of TCAM1P-Pro-luc. To linearize the constructs, I digested them by *Sal*I. The hCNS1-luc and rev hCNS1-luc constructs were generated by inserting the hCNS1 sequence into the *Sma*I site of a pGL3-Basic vector. All the constructs were subject to the sequencing analysis prior to transfection studies.

Cell culture and the reporter gene assay

GC-2spd(ts), Hepa1-6, and HEK293T cells were cultured in DMEM containing 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin, and 292 μg/ml L-glutamine (Invitrogen). The reporter gene assay was performed as previously described[110].

Statistical analysis

The results were expressed as means ± S.D. One-way ANOVA followed by Tukey-Kramer's test was performed using Microsoft Excel statistical analysis functions (Microsoft Corp.). $P < 0.05$ was considered statistically significant.

Accession number

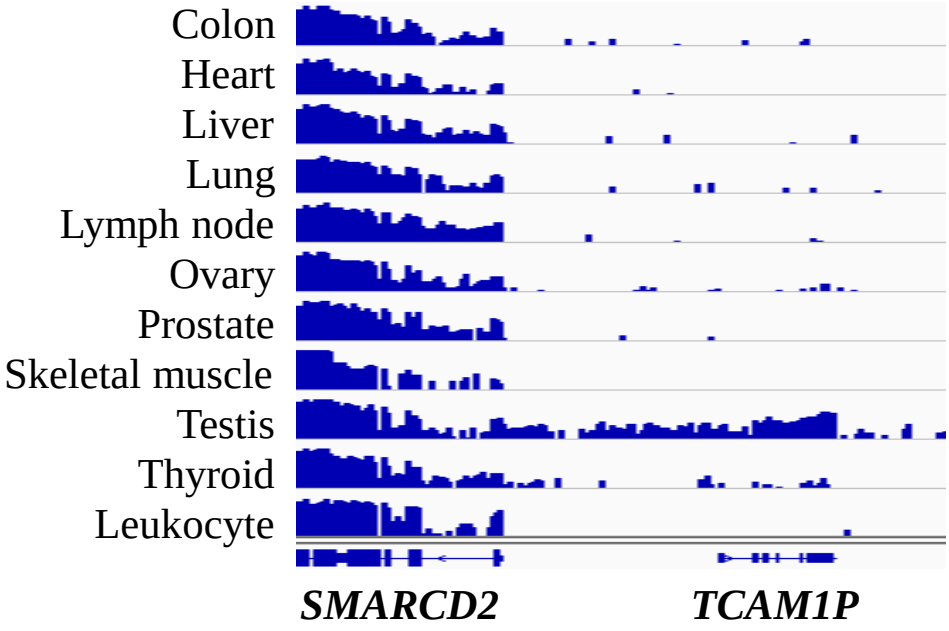
Nucleotide sequence data for the six *TCAM1P* transcripts and *lncRNA-TCAM1P* are available in DDBJ/EMBL/GenBank databases under the accession numbers LC005096 (*TCAM1P-I*), LC005097 (*TCAM1P-IIa*), LC005098 (*TCAM1P-IIb*), LC005099 (*TCAM1P-IIc*), LC005100 (*TCAM1P-IId*), LC005101 (*TCAM1P-III*), LC005102 (*lncRNA-TCAM1P*)

Table 2-1. Oligonucleotide primers used in Chapter 2

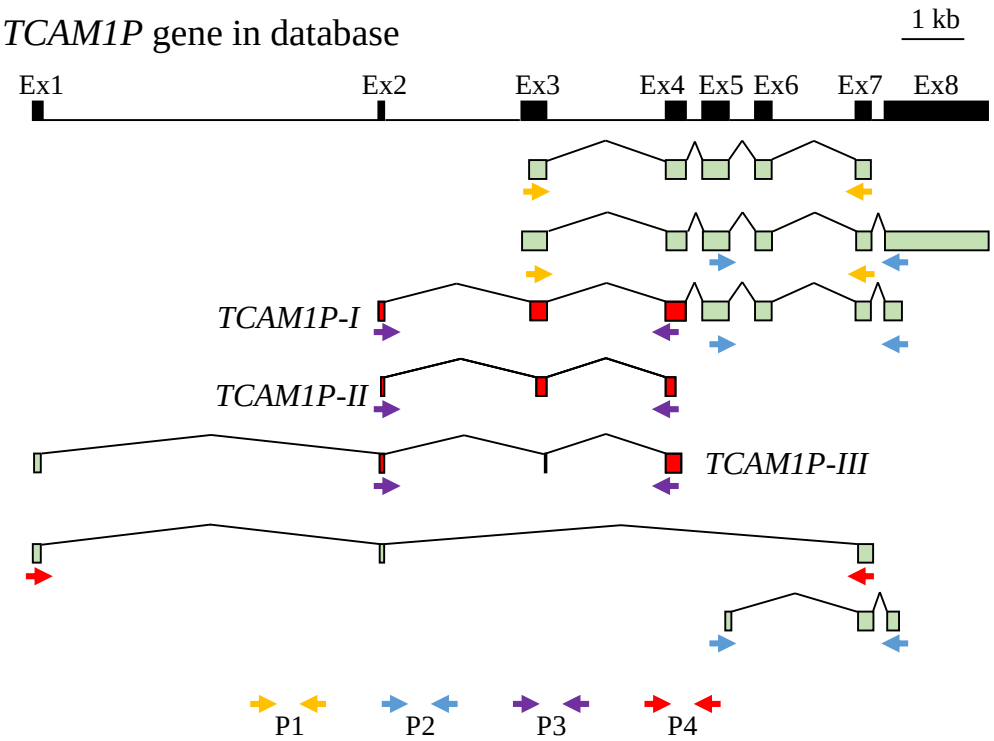
Designation	Forward	Reverse
RT-PCR		
TCAM1P P3	5'-TTCTGTTTGGTGTCTGGACC-3'	5'-GGGTACATGACACTTCACTG-3'
lncRNA-TCAM1P	5'-CTCCGGGATTTCCTAATCCA-3'	5'-GGGAGGTTTCTCATAAGCCA-3'
1	5'-CCCTCTGATGGTCTGGATCT-3'	5'-TAGCTTGCCCCTTTTCTAGG-3'
2	5'-CCCTCCCATCCCTGTAAAGT-3'	5'-AGAAAAAGATCCTGCCTTCG-3'
3	5'-GCCTCTCTGTTCTCTCACAC-3'	5'-CCTCCCACTACACTCAAACA-3'
4	5'-TCTCCCTTGATCCCTGTTGT-3'	5'-GGGAATATTGCCTCCACTTC-3'
5	5'-CCACCGATTTTTGGTAGCTG-3'	5'-AGCTGGGACAGGCTCTGATA-3'
6	5'-GAAGGACTGGATGCCAAGA-3'	5'-CGGCTGTATGAGGAAAACAA-3'
ACTB	5'-AAGAGAGGCATCCTCACCCT-3'	5'-TACATGGCTGGGGTGTTGAA-3'
5'RACE		
AAP	5'-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3'	
AUAP	5'-GGCCACGCGTCGACTAGTAC-3'	
TCAM1P for RT-1		5'-GGGTACATGACACTTCACTG-3'
TCAM1P for GSP1-1		5'-TCCAGACACCAAACAGAAGC-3'
TCAM1P for GSP2-1		5'-CTGATAAGAAGGGCCAAAGG-3'
TCAM1P for RT-2		5'-AAGCACTGCAGGATGGAA-3'
TCAM1P for GSP1-2		5'-CTTTGCCCACCTGAGTTTTC-3'
TCAM1P for GSP2-2		5'-TCCAGACACCAAACAGAAGC-3'
lncRNA-TCAM1P for RT		5'-AACCGGGTCTGTGAGGAACT-3'
lncRNA-TCAM1P GSP1		5'-CTGTGAGGAACTCGCGAACT-3'
lncRNA-TCAM1P GSP2		5'-TGGATTAGGAAATCCCGGAG-3'
3’RACE		
AP		5'-CTGATCTAGAGGTACCGGATCC-3'

TCAM1P for GSP3	5'-TTCTGTTTGGTGTCTGGACC-3'	
TCAM1P for GSP4	5'-GATTCAAAAGGACACAAGCC-3'	
lncRNA-TCAM1P GSP3	5'-CCACACTCCCCGAAAATAGT-3'	
lncRNA-TCAM1P GSP4	5'-AAAGACTCAGCTGCTCTGTG-3'	
Reporter constructs		
TCAM1P Promoter	5'-CAGCACAGACTCTGGGAATT-3'	5'-TGAATGCACGAACCAACGCA-3'
hCNS1 first	5'-AGTTCCTCATGCATTCCCCA-3'	5'-GTTGAGTTGGCTGCTGTTCT-3'
hCNS1 nested	5'-CTCCGGGATTTCCTAATCCA-3'	5'-AGAGAGTAGTAGCAACGGCT-3'

(A)


Colon
Heart
Liver
Lung
Lymph node
Ovary
Prostate
Skeletal muscle
Testis
Thyroid
Leukocyte
SMARCD2
TCAM1P

(B)


TCAM1P gene in database
1 kb
Ex1
Ex2
Ex3
Ex4
Ex5
Ex6
Ex7
Ex8
TCAM1P-I
TCAM1P-II
TCAM1P-III
P1
P2
P3
P4

(C)

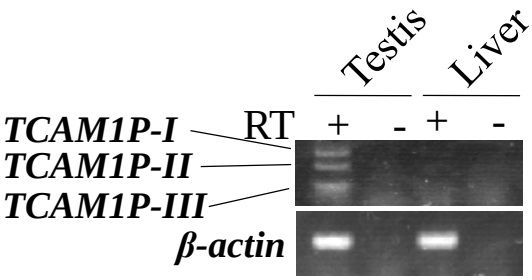

Testis
Liver
RT
+
-
+
-
TCAM1P-I
TCAM1P-II
TCAM1P-III
β-actin

Fig. 2-1. Ubiquitous distribution of *SMARCD2* mRNA and testis-specific expression of *TCAM1P* mRNA in human tissues. (A) Expression of *SMARCD2* and *TCAM1P* mRNAs in eleven human tissues were searched on IGV. Transcript signals were indicated as blue bars and exon-intron structures of these genes are depicted at the bottom. The signals were observed in all eleven tissues at the *SMARCD2* gene, while those at the *TCAM1P* locus were mostly present in the testis. (B) Putative seven transcripts of the human *TCAM1P* gene are shown according to the Ensembl database. We searched for the database when we started this study in 2009. The predicted *TCAM1P* gene consisted of 8 exons. To distinguish the seven transcripts, RT-PCR was performed using four primer pairs whose positions are indicated by arrows (P1, P2, P3, and P4). cDNAs were prepared by reverse transcription with the oligo(dT) primer using testis and liver total RNAs. By PCR reactions, specific signals were amplified only with P3. The amplified exons were painted with red, and the transcripts were named *TCAM1P-I*, *TCAM1P-II*, and *TCAM1P-III*. (C) The result of RT-PCR using a primer pair, P3, and human testis and liver RNAs is shown. Three transcripts were detected at different positions only in the testis. *β-ACTIN* was amplified as an internal control.

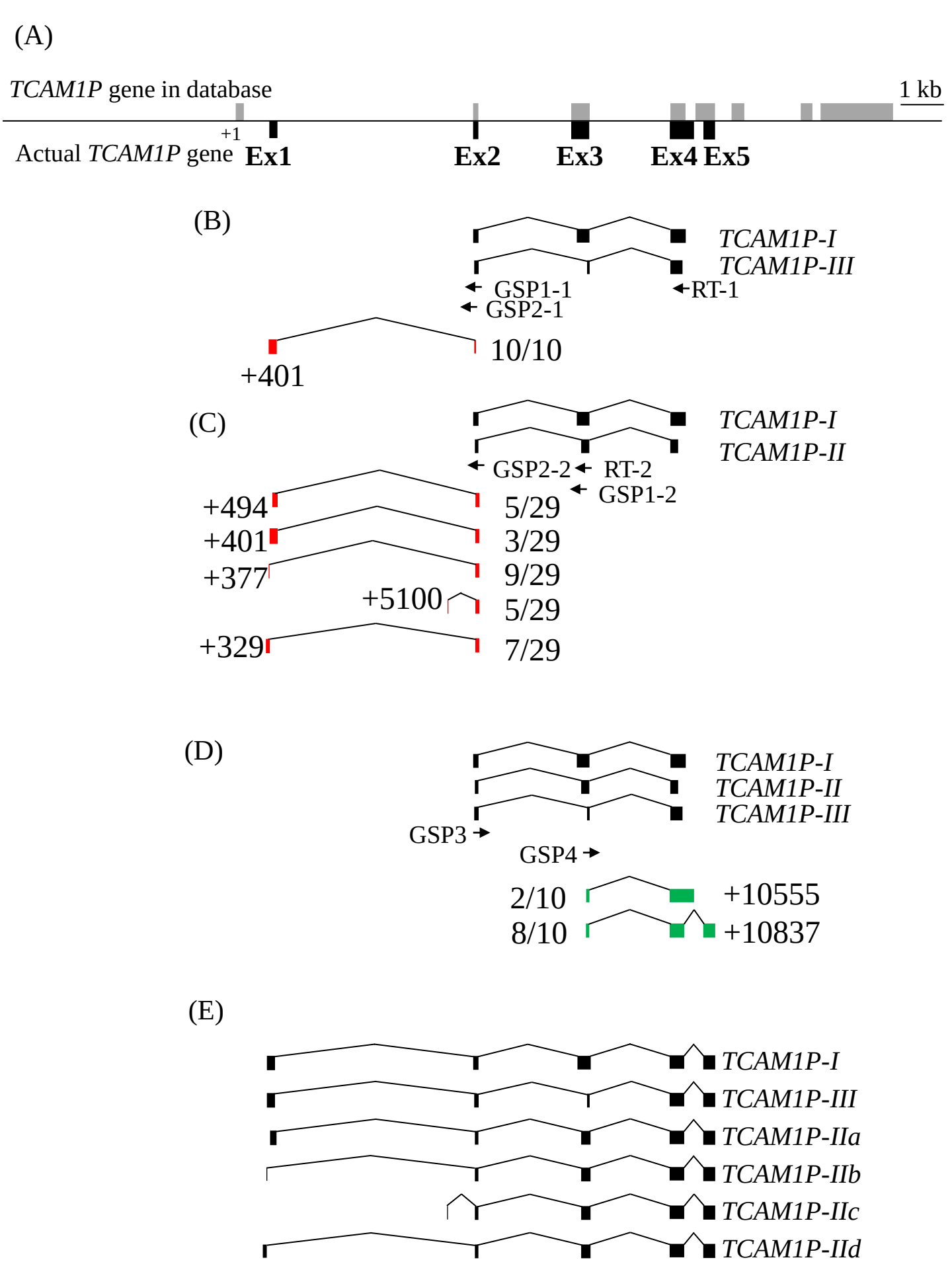

(A)
TCAM1P gene in database
1 kb
+1
Actual TCAM1P gene
Ex1
Ex2
Ex3
Ex4
Ex5
(B)
TCAM1P-I
TCAM1P-III
GSP1-1
RT-1
GSP2-1
10/10
+401
(C)
TCAM1P-I
TCAM1P-II
GSP2-2
RT-2
GSP1-2
+494
5/29
+401
3/29
+377
9/29
+5100
5/29
+329
7/29
(D)
TCAM1P-I
TCAM1P-II
TCAM1P-III
GSP3
GSP4
2/10
+10555
8/10
+10837
(E)
TCAM1P-I
TCAM1P-III
TCAM1P-IIa
TCAM1P-IIb
TCAM1P-IIc
TCAM1P-IId

Fig. 2-2. Determination of 5’ and 3’ ends of *TCAM1P* transcripts. (A) The human *TCAM1P* gene structure based on the Ensembl database is shown at the top with gray boxes. The actual exon-intron structure determined in this study is shown below with black boxes. Exon numbers of the actual *TCAM1P* gene is indicated as Ex1-Ex5. The first nucleotide of exon 1 in the database is designated as +1 in this figure. (B) 5’RACE was performed to determine TSS of *TCAM1P-I* and *TCAM1P-III* transcripts. cDNA was prepared with the human testis total RNA by reverse transcription using a primer, RT-1. The first and second PCR reactions were performed using GSP1-1 and GSP2-1, respectively. After subcloning of the product, 10 subclones were sequenced, and a single TSS was determined at the position of +401. (C) 5’RACE was performed for TSS of *TCAM1P-I* and *TCAM1P-II* transcripts as in (B) by using RT-2 for reverse transcription, and GSP1-2 and GSP2-2 primers for two rounds of PCRs. After subcloning, 29 subclones were sequenced, and five different TSSs were detected as indicated. The number of subclone, which contained each position as TSS, is shown at the right side of each structure. (D) 3’RACE was performed to determine TTS of *TCAM1P-I*, *TCAM1P-II*, and *TCAM1P-III*. cDNA was prepared with the human testis total RNA by reverse transcription with the oligo(dT) primer connected to an AP. The first and second PCR reactions were performed with GSP3 and GSP4 primers, respectively. After subcloning, 10 subclones were sequenced, and 8 of them contained the position of +10837 as TTS. The other 2 subclones were terminated at +10555. (E) Structures of human *TCAM1P* transcripts determined in this study. RACE analyses revealed that *TCAM1P-II* had at least four TSSs and a single TTS, and *TCAM1P-I* and *TCAM1P-III* had a single TSS and a single TTS. Consequently, six *TCAM1P* transcripts were expressed in the human testis

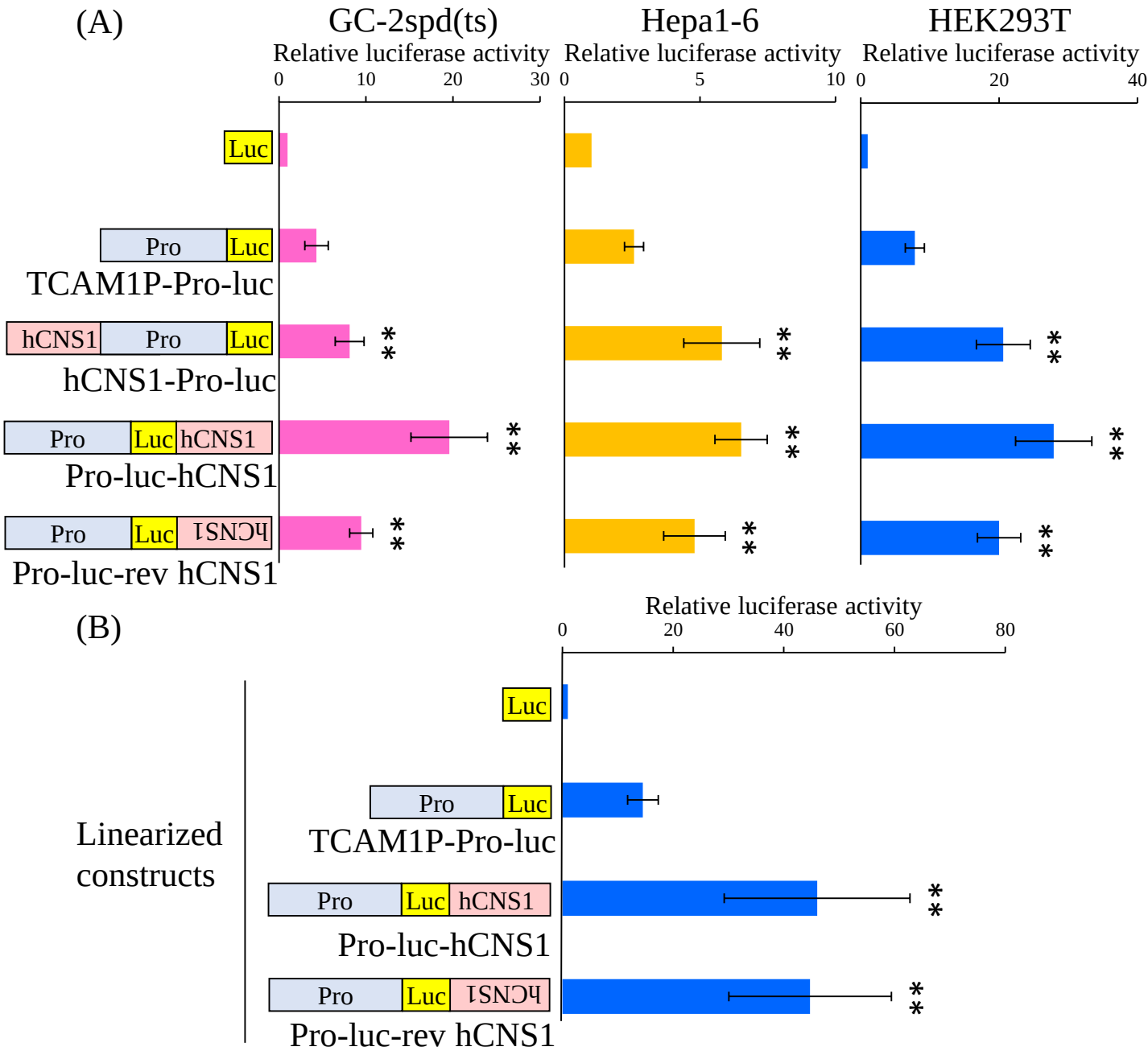


Fig. 2-3. Enhancer activity of hCNS1 by *in vitro* reporter gene analysis. (A) Reporter gene constructs were generated as indicated at the left side of the graph. The constructs were transfected into GC-2spd(ts) (pink bar), Hepa1–6 (yellow bar), or HEK293T cells (blue bar) by GeneJuice transfection reagent, 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. *TCAM1P* promoter activity was enhanced by hCNS1 in all three cell lines irrespective of its position and orientation. The data are presented as mean ± S.D. from at least six independent experiments. n = 6-18. $^{**}P < 0.01$ relative to TCAM1P-Pro-luc construct, one-way ANOVA followed by Tukey-Kramer's test. (B) Enhancer activity of hCNS1 was assessed in linearized constructs. The indicated constructs were linearized by digestion with *Sal*I and transfected into HEK293T cells. hCNS1 significantly enhanced *TCAM1P* promoter activity. The data are presented as mean ± S.D. from six independent experiments. n = 6. $^{**}P < 0.01$ relative to linearized TCAM1P-Pro-luc construct, one-way ANOVA followed by Tukey-Kramer's test.

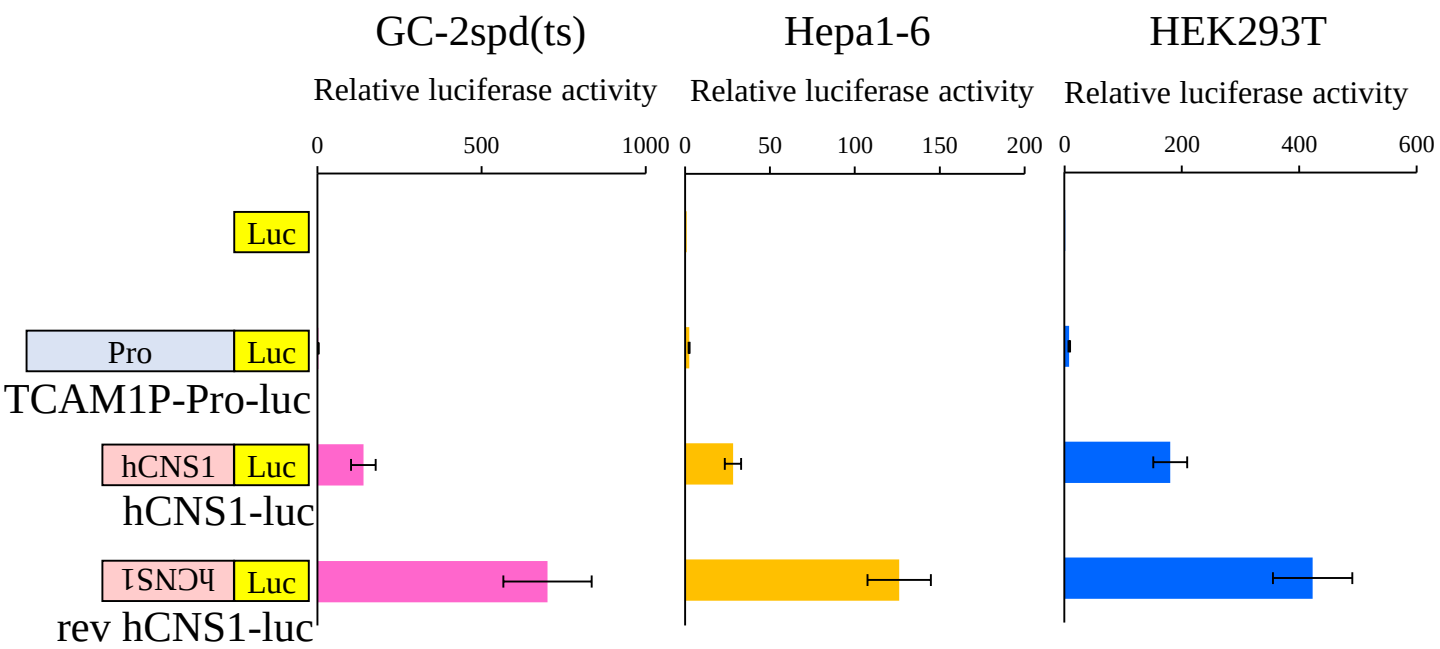


Fig. 2-4. Bidirectional promoter activity of hCNS1 by reporter gene assay. hCNS1 was connected directly to the luciferase gene in both directions, and the reporter gene assay was performed as in Fig. 2-3 using GC-2spd(ts), Hepa1–6, and HEK293T cells. The data are presented as mean ± S.D. from six independent experiments. n = 6.

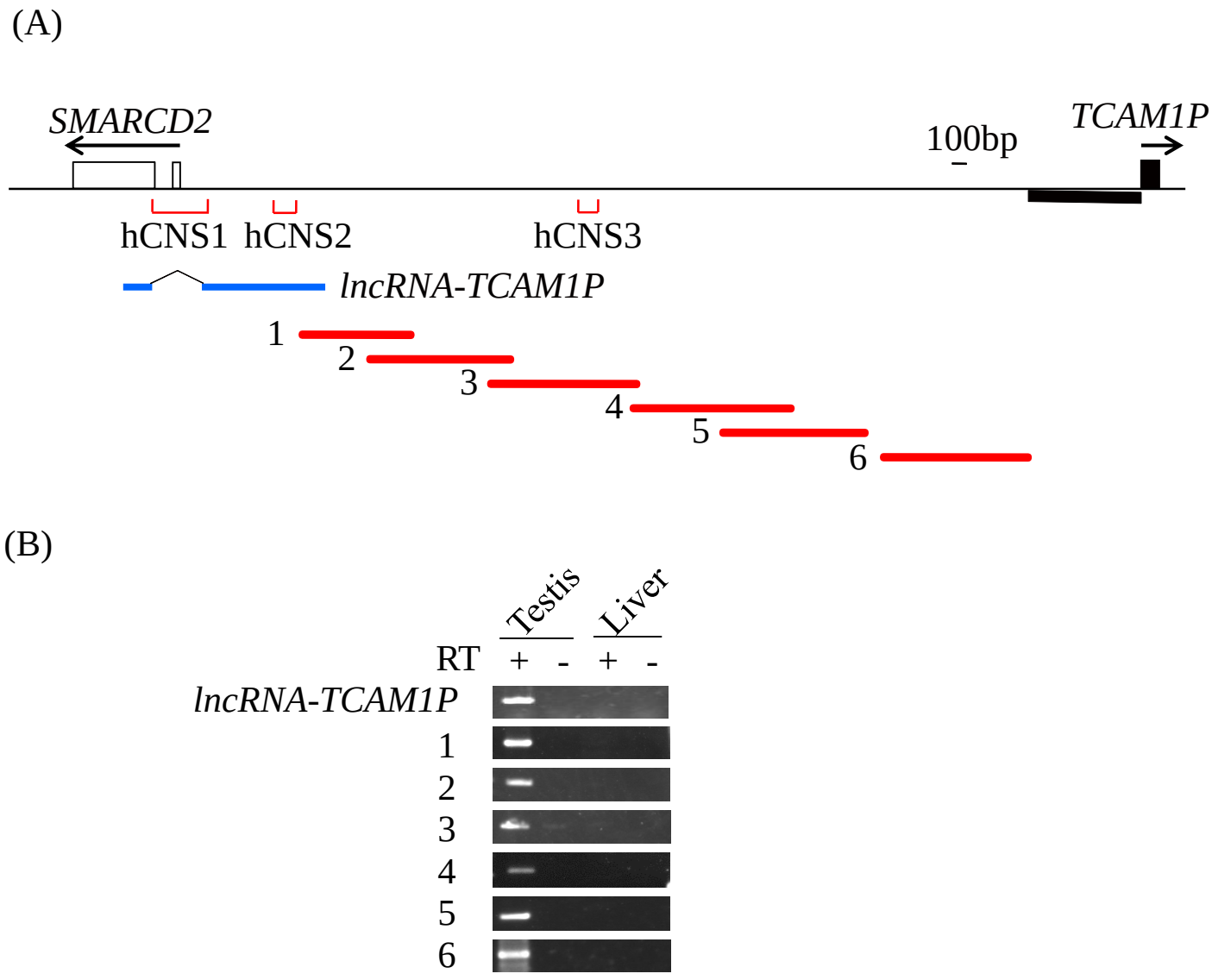


Fig. 2-5. Intergenic transcription between *SAMRCD2* and *TCAM1P* genes. (A) A schematic drawing of the 5'upstream region of the *TCAM1P* gene. Exons 1 and 2 of the *SMARCD2* gene and exon 1 of the *TCAM1P* gene, which I determined, are depicted with white and black rectangles, and arrows indicate their transcriptional directions. Positions of three CNSs (hCNS1, hCNS2, and hCNS3) are shown. A full-length1.1-kb transcript, *lncRNA-TCAM1P*, and six amplicons are indicated by blue and red lines, respectively. (B) Expression of intergenic regions at the *SMARCD2-TCAM1P* locus by RT-PCR. PCR reactions were conducted for six amplicons indicated in (A) and for *lncRNA-TCAM1P*. A specific signal was detected for all the amplicons in the testis, but not in the liver. Because we used the same cDNAs as in Fig. 2-1C, an internal control should be referred to Fig. 2-1C.

GENERAL DISCUSSION

The most important issues in this study are the multifunctional genomic element and the transcriptional regulation through pseudogenization. Both of them are critical to understand the genomic function, but have not been well studied so far. Here, I focus on and discuss these two issues.

Multifunctional genomic elements

While most genomic sequences are annotated as elements that have only one function such as an exon, a promoter, and an enhancer, genome-wide transcription analyses imply that some of them are multifunctional[137–139]. For example, a significant number of transcriptional units overlap with gene regulatory elements, such as enhancers located in introns of coding genes and some noncoding RNAs overlapping with enhancers[40,43,140,141]. Other noncoding RNAs are transcribed from the same or opposite strand of coding genes to activate or repress their expression[139]. Many bidirectional promoters and enhancers controlling the expression of different genes have also been reported to cooperatively regulate multiple genes[34,35,91]. All of these suggest that the genomic function is much more complicated than has been thought.

In this study, I identified and characterized one multifunctional regulatory element, DPE, in mouse and human genome. This is the first indication of DPE in mammals. However, genome-wide investigation revealed that RNA polymerase II is located not only at promoters but also at TTS and at intergenic regions coinciding with enhancers[33]. Furthermore, a large number of noncoding RNAs have been reported to be transcribed from enhancers[40,43]. Although it is not clear whether they are transcribed from enhancers during meiosis, many lncRNAs are expressed in testicular germ cells[142]. Therefore, I assume that a significant number of DPEs exist in mammalian genome and function in various tissues.

If there are many multifunctional genomic elements in our genome, why are they generated in the process of evolution? The fact that all genomic sequences do not necessarily contribute to transcriptional regulation may be one explanation. Eukaryotic genome is packed into nucleus, and the higher-order architecture of genome is non-random. A recent study has revealed that 40% of mammalian genomic regions are associated with nuclear envelope proteins to maintain the architecture[143]. Such genomic regions have the repressive chromatin feature, and cannot directly participate in gene regulation. In addition, eukaryotic genome is organized at multiple length scales into sophisticated higher-order architecture[144], by which many genomic sequences are not accessible by transcription factors. Therefore, the population of genomic sequences that directly engage in transcriptional regulation should be much less than anticipated. This could cause the generation of multifunctional sequences in mammalian genome, which leads to the complicated spatiotemporal gene regulation.

Even if multifunctional genomic elements need to be generated in the process of evolution, they have a bad aspect. When a mutation occurs in a multifunctional sequence, multiple functions are disturbed at a time, resulting in severer effects on organisms. Then, what is the biological significance of multifunctional genomic elements? There should be some good aspects of them, too. In case of a protein-coding gene which is alternatively transcribed, diverse types of proteins can be produced from relatively a small number of exons. In case of multifunctional regulatory elements that regulate multiple gene expression simultaneously, one transcription factor does not necessarily bind to multiple regions. If a multifunctional region is responsive to a hormone, the response to that hormone should be quicker and more effective. Therefore, multifunctional genomic elements are essential to the genome function.

In the case of my study, Sp1 was found to be involved in promoter activity of *Smarcd2* but not in a promoter for *lncRNA-Tcam1* and an enhancer for *Tcam1* (Fig. 1-11). This suggests that mCNS1 promoter activity for *Smarcd2* is separated from the other functions. At present, it is

unclear whether mCNS1 can function as a promoter for *lncRNA-Tcam1* and an enhancer for *Tcam1* simultaneously, but their expression patterns were very similar (Figs. 1-1, 1-8), showing the possibility that they are controlled by the same transcription factors. Indeed, it was proposed that a DPE at the chicken lysozyme gene locus functioned as a promoter for LINoCR and an enhancer for the gene simultaneously, through the binding of the same transcription factors[100]. This means that one stimulation or one transcription factor can affect the function of DPE, which may result in the activation of multiple genes. A similar mechanism may exist for the regulation of *lncRNA-Tcam1* and *Tcam1*.

Finally, I discuss the significance of multifunctional elements in meiosis. While gene expression in somatic cells usually occurs during interphase[145], a large number of genes are expressed in testicular germ cells during meiosis[146]. However, the chromatin in meiotic cells is condensed, and transcription factors cannot bind to many genomic regions. Therefore, to utilize the limited number of regulatory sequences, multifunctional genomic elements should play important roles in the gene regulation during meiosis.

Changes in the *TCAM1P/Tcam1* regulation

Marques *et al.* analyzed unitary pseudogenes that were specifically pseudogenized in rodent lineage, and found that correlation of 17 mouse unitary pseudogenes with their human orthologues was higher than that of randomly sampled pairs of murine protein-coding genes[147]. However, there is no study of unitary pseudogenes that were specifically pseudogenized in human lineage, and I provided the data about it for the first time.

Because 76 functional genes have been reported to lose their primary functions and to be

fixed as unitary pseudogenes throughout primate evolution[30], I examined whether their expression patterns were conserved after pseudogenization by using IGV and BioGPS (http://biogps.org/#goto=welcome) databases. As a result, I found 10 unitary pseudogenes other than *TCAM1P* in IGV. The reason why I could not find the other 65 pseudogenes might be due to the difference in annotated names between IGV and the report. Among the 10 genes, the databases showed that *Cetn4*, *Tex21*, and *Hyal6* genes were expressed specifically in the mouse testis (Fig. D1A). Interestingly, after pseudogenization, *Cetn4* and *Tex21* were expressed not only in the testis but also in other human tissues such as brain, kidney, ovary and thyroid, and *Hyal6* was not detected in any human tissues (Fig. D1B). I also found *Sec1* expressed only in the mouse testis, but the expression pattern of the human *SEC1P* gene was unclear because it overlapped with two other genes in human genome (data not shown). Among the other six genes, the expression patterns of five genes were also changed and one gene was silenced after pseudogenization. These suggest that the expression patterns of many unitary pseudogenes are not conserved between human and mouse, which in turn indicates that the change in gene structure probably occurred in parallel with alteration of the regulatory mechanism.

In this study, I focused on the *Tcam1/TCAM1P* gene, and revealed that the testis-specific expression and the DPE activity of CNS1 were conserved after pseudozenition. However, cell type-specificity in testis of the *Tcam1/TCAM1P* expression was changed, and consistent with that, hCNS1 showed enhancer activity in more cell types than mCNS1. This suggests that different nucleotides between mCNS1 and hCNS1 are core sequences of the enhancer activity and that the binding of divergent transcriptional factors to CNS1 results in the alteration of cell type-specificity of *Tcam1* and *TCAM1P*. Therefore, comparison of CNS1 sequences in various species will help us understand the evolutionary change of the *Tcam1/TCAM1P* regulation.

TCAM1P was reported to be pseudogenized at gorilla-chimpanzee divergence by comparing genome sequences of human, chimpanzee, gorilla, orangutan, rhesus monkey,

marmoset, and tarsier[30]. Therefore, I validated how the regulatory system of *Tcam1/TCAM1P* was altered by searching for the database to obtain and compare the CNS1 sequences from various species. As a result, the concordance rate of hCNS1 with CNS1 in gibbon, which was diverged before gorilla during anthropoid evolution, and rhesus monkey was 96% and 94%, respectively, but that with mouse and rat were around 77% (Table D1). Conversely, mCNS1 was highly homologous to rat CNS1, but the concordance rate of mCNS1 with CNS1 in gibbon and rhesus monkey was 71%, and 74%, respectively (Table D1). Because I could not find any nucleotides in CNS1 that were changed only in human and conserved in the other four species (Fig. D2), the change, which allowed divergent transcriptional factors to bind to CNS1, were likely to occur before pseudogenization of *TCAM1P*. This suggests that cell-type specificity of the *TCAM1* gene was also altered before pseudogenization, and that the *TCAM1* expression in spermatocytes as well as in spermatogonia and spermatids is adaptive in many primates.

Finally, I discuss the new function of *TCAM1P* in chimpanzee and human. Recently, a lot of studies have shown that some pseudogenes regulate the normal growth of an organism and the development of some diseases[123,148–151]. They can be processed into antisense transcripts and miRNA decoys, produce siRNAs, or encode short peptides or proteins, and *TCAM1P* may have a similar function. One report showed that *TCAM1P* and two other testis-specific genes, *STAG3* and *SYCP2*, were specifically up-regulated in human papillomavirus-positive cancers, and *SYCP2* and *TCAM1P* were induced by viral oncogenes[152]. *STAG3* and *SYCP2* are specifically expressed in meiosis, and become components of the synaptonemal complex that facilitate recombination[153,154]. Because *TCAM1P* is mainly expressed in spermatocyte, it is possible that it also contributes to recombination in human meiotic cells.

Table D1. Concordance rates of CNS1 sequences between five mammalian species.

	Homo sapiens	Nomascus leucogenys	Macaca mulatta	Rattus norvegicus	Mus musculus
Homo sapiens		96%	94%	77%	77%
Nomascus leucogenys			93%	76%	71%
Macaca mulatta				80%	74%
Rattus norvegicus					93%
Mus musculus					

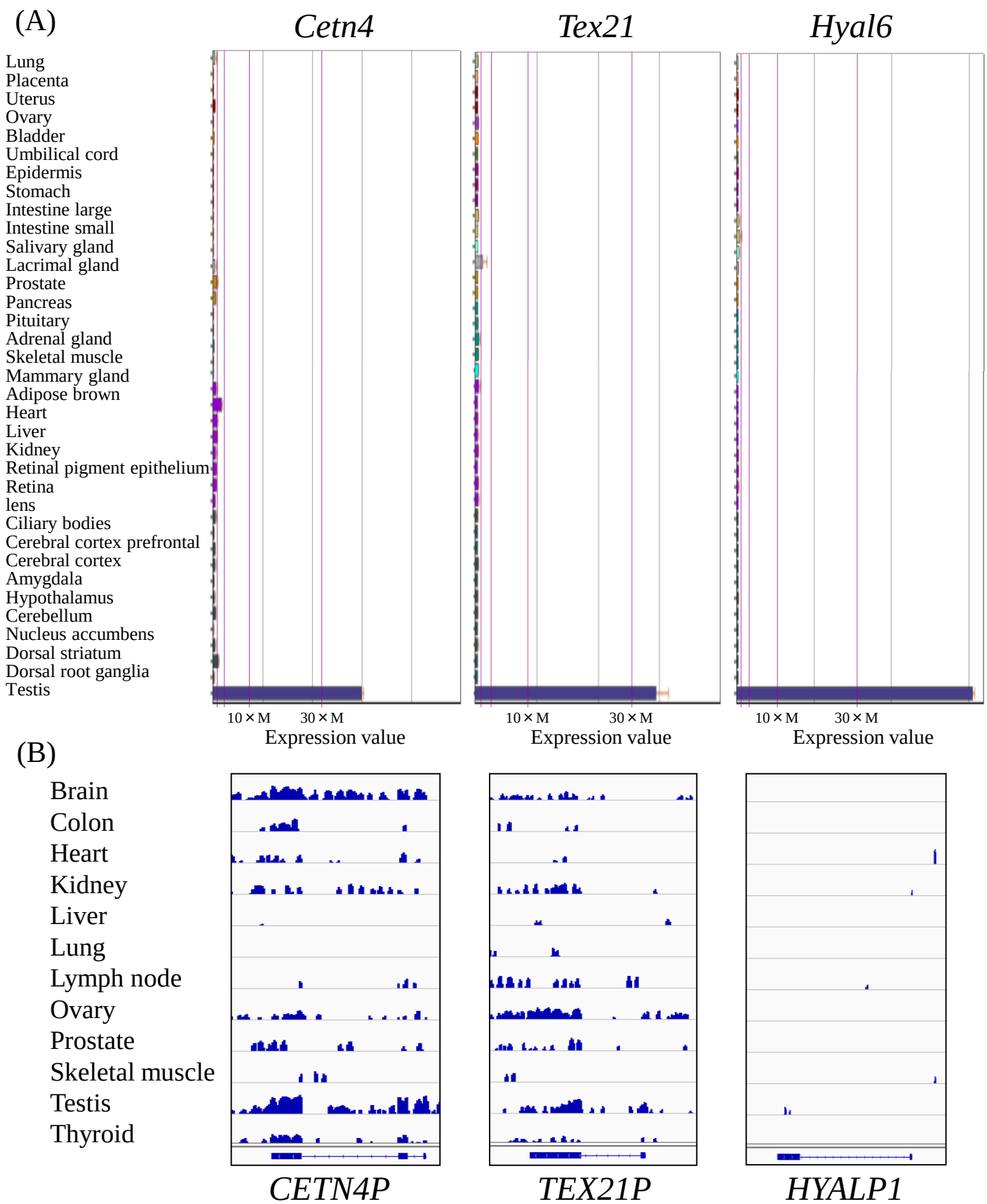


Fig. D1. The expression of unitary pseudogenes in mouse and human tissues. (A) The expression patterns of *Cetn4*, *Tex21*, and *Hyal6* in thirty-five mouse tissues are shown based on the BioGPS database. (B) The expression of *CETN4*, *TEX21*, and *HYAL1P* in eleven human tissues are shown based on the IGV database. Transcript signals are indicated as blue bars, and exon-intron structures of these genes are depicted at the bottom.

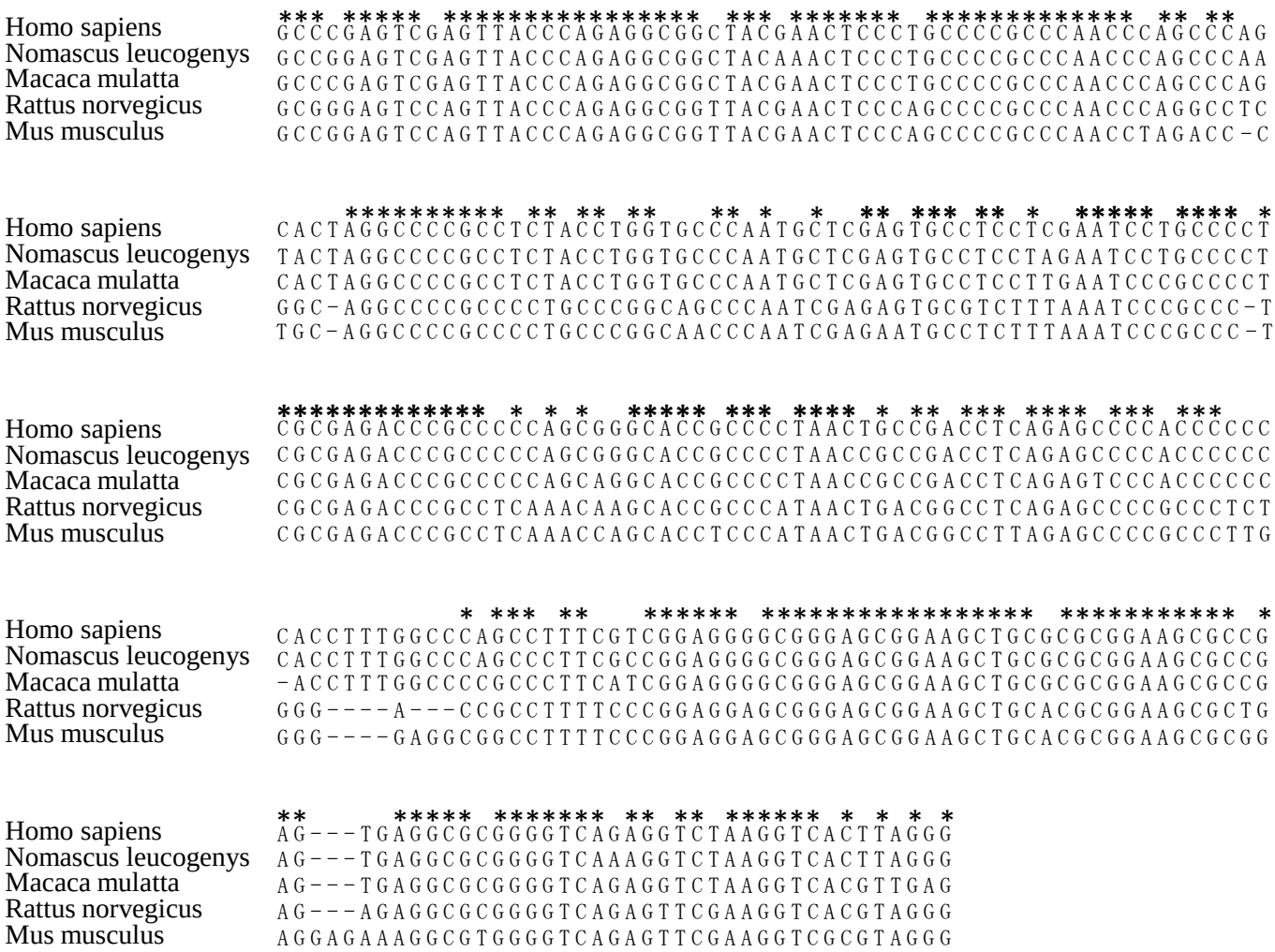


Fig. D2. A multiple alignment of CNS1 from five mammalian species. CNS1 sequences were obtained from *Homo sapiens*, *Nomascus leucogenys*, *Macaca mulatta*, *Rattus norvegicus*, and *Mus musculus* by searching for the Ensembl database, and compared using the DNASIS-Pro software (HITACHI Software Engineering). Conserved nucleotides among all five species are indicated by asterisks.

ACKNOWLEDGEMENTS

I wish to express my grateful acknowledgement to Associate Professor Atsushi P. Kimura for his hard instruction, positive encouragement, and valuable discussion throughout this study. I am deeply grateful to Professor Takashi Shimizu, Professor Masakane Yamashita, Professor Atsushi Kato, and Professor Takayuki Takahashi for valuable advices, suggestions, and encouragement in this study. I am also great thankful to Associate Professor Yoshinao Kastu and Assistant Professor Katsueki Ogiwara for their helpful technical advices, suggestions, and encouragement. I express deep gratitude to Professor Honoo Satake and Dr Akira Shiraishi for their helping in bioinformatic analysis in Chapter 1, Dr. Hiroshi Kimura at Tokyo Institute of Technology for kindly gifting the antibodies against modified histones in Chapter 1, and Dr. Saadi Khochbin at Université Joseph Fourier for his valuable advices in Chapter 2. I would particularly like to thank Dr. Shin Matsubara, Dr. Chika Fujimori, Dr. Ryoma Yoneda, and Ms. Akane Hagiwara for their incredible advices and teaching me the basic experimental techniques in this field. Finally, I would like to offer my special thanks to all the members of reproductive and developmental biology course for their help, discussion, encouragement, and friendship in the daily laboratory life.

REFERENCES

1. Werner, T. (1999). Models for prediction and recognition of eukaryotic promoters. *Mamm. Genome* **10**, 168–75

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

3. 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

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

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

6. Bulger, M. & Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–39

7. Spitz, F. & Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–26

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

9. Kim, A. & Dean, A. (2012). Chromatin loop formation in the β-globin locus and its role in globin gene transcription. *Mol. Cells* **34**, 1–5

10. Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., Shen, Y., Pervouchine, D.D., Djebali, S., Thurman, R.E., Kaul, R., Rynes, E., Kirilusha, A., Marinov, G.K., Williams, B.A., Trout, D., Amrhein, H., Fisher-Aylor, K., Antoshechkin, I., DeSalvo, G., See, L.-H., Fastuca, M., Drenkow, J., Zaleski, C., Dobin, A., Prieto, P., Lagarde, J., Bussotti, G., Tanzer, A., Denas, O., Li, K., Bender, M.A., Zhang, M., Byron, R., Groudine, M.T., McCleary, D., Pham, L., Ye, Z., Kuan, S., Edsall, L., Wu, Y.-C., Rasmussen, M.D., Bansal, M.S., Kellis, M., Keller, C.A., Morrissey, C.S., Mishra, T., Jain, D., Dogan, N., Harris, R.S., Cayting, P., Kawli, T., Boyle, A.P., Euskirchen, G., Kundaje, A., Lin, S., Lin, Y., Jansen, C., Malladi, V.S., Cline, M.S., Erickson, D.T., Kirkup, V.M., Learned, K., Sloan, C.A., Rosenbloom, K.R., Lacerda de Sousa, B., Beal, K., Pignatelli, M., Flicek, P., Lian, J., Kahveci, T., Lee, D., James Kent, W., Ramalho Santos, M., Herrero, J., Notredame, C., Johnson, A., Vong, S., Lee, K., Bates, D., Neri, F., Diegel, M., Canfield, T., Sabo, P.J., Wilken, M.S., Reh, T.A., Giste, E., Shafer, A., Kutyavin, T., Haugen, E., Dunn, D., Reynolds, A.P., Neph, S., Humbert, R., Scott Hansen, R., De Bruijn, M., Selleri, L., Rudensky, A., Josefowicz, S., Samstein, R., Eichler, E.E., Orkin, S.H., Levasseur, D., Papayannopoulou, T., Chang, K.-H., Skoultchi, A., Gosh, S., Disteche, C., Treuting, P., Wang, Y., Weiss, M.J., Blobel, G.A., Cao, X., Zhong, S., Wang, T., Good, P.J., Lowdon, R.F., Adams, L.B., Zhou, X.-Q., Pazin, M.J., Feingold, E.A., Wold, B., Taylor, J., Mortazavi, A., Weissman, S.M., Stamatoyannopoulos, J.A., Snyder, M.P., Guigo, R., Gingeras, T.R., Gilbert, D.M., Hardison, R.C., Beer, M.A. & Ren, B. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–64

11. Lin, S., Lin, Y., Nery, J.R., Urich, M.A., Breschi, A., Davis, C.A., Dobin, A., Zaleski, C., Beer, M.A., Chapman, W.C., Gingeras, T.R., Ecker, J.R. & Snyder, M.P. (2014).

Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci.* **111**, 201413624

12. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigó, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., Karlsson, E.K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W.J., Kirby, A., Kolbe, D.L., Korf, I., Kucherlapati, R.S., Kulbokas, E.J., Kulp, D., Landers, T., Leger, J.P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D.R., Mardis, E.R., Matthews, L., Mauceli, E., Mayer, J.H., McCarthy, M., McCombie, W.R., McLaren, S., McLay, K., McPherson, J.D., Meldrim, J., Meredith, B., Mesirov, J.P., Miller, W., Miner, T.L., Mongin, E., Montgomery, K.T., Morgan, M., Mott, R., Mullikin, J.C., Muzny, D.M., Nash, W.E., Nelson, J.O., Nhan, M.N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M.J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G.,

Pepin, K.H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C.S., Poliakov, A., Ponce, T.C., Ponting, C.P., Potter, S., Quail, M., Reymond, A., Roe, B.A., Roskin, K.M., Rubin, E.M., Rust, A.G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M.S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J.B., Slater, G., Smit, A., Smith, D.R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J.P., Von Niederhausern, A.C., Wade, C.M., Wall, M., Weber, R.J., Weiss, R.B., Wendl, M.C., West, A.P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R.K., Winter, E., Worley, K.C., Wyman, D., Yang, S., Yang, S.-P., Zdobnov, E.M., Zody, M.C. & Lander, E.S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–62

13. Barrera-Saldaña, H.A. (1998). Growth hormone and placental lactogen: biology, medicine and biotechnology. *Gene* **211**, 11–8

14. 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**, 479–97

15. 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

16. Su Y, Liebhaber, S.A. & Cooke, N.E. (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**, 7902–9

17. Elefant, F., Cooke, N.E. & Liebhaber, S.A. (2000). Targeted recruitment of histone acetyltransferase activity to a locus control region. *J. Biol. Chem.* **275**, 13827–34

18. 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

19. 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**, 365–75

20. Yoo, E.J., Cooke, N.E. & Liebhaber, S.A. (2012). An RNA-independent linkage of noncoding transcription to long-range enhancer function. *Mol. Cell. Biol.* **32**, 2020–9

21. 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

22. 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

23. Tsai, Y.-C., Cooke, N.E. & Liebhaber, S.A. (2014). Tissue specific CTCF occupancy and boundary function at the human growth hormone locus. *Nucleic Acids Res.* **42**, 4906–21

24. Lunyak, V. V, Prefontaine, G.G., Núñez, E., Cramer, T., Ju, B.-G., Ohgi, K.A., Hutt, K., Roy, R., García-Díaz, A., Zhu, X., Yung, Y., Montoliu, L., Glass, C.K. & Rosenfeld,

M.G. (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**, 248–51

25. Yoo, E.J., Cajiao, I., Kim, J.-S., Kimura, A.P., Zhang, A., Cooke, N.E. & Liebhaber, S.A. (2006). Tissue-specific chromatin modifications at a multigene locus generate asymmetric transcriptional interactions. *Mol. Cell. Biol.* **26**, 5569–79

26. Komatsu, A., Otsuka, A. & Ono, M. (2002). Novel regulatory regions found downstream of the rat B29/Ig-β gene. *Eur. J. Biochem.* **269**, 1227–36

27. Chayahara, K., Itaya, K. & Ono, M. (2011). Transcriptional and epigenetic effects of deleting large regions, alone or in combination, from their natural context in the chicken Ig-β gene. *Gene* **486**, 1–7

28. Itaya, K., Chayahara, K., Hirai, T., Minbuta, T., Uchikawa, T., Tanaka, T., Masaki, S., Kuroda, K. & Ono, M. (2011). DT40 knock-out and knock-in studies determine the regions necessary and sufficient for transcription and epigenetic conversion of the chicken Ig-β gene. *Genes Cells* **16**, 291–303

29. 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**, 246–53

30. Zhang, Z.D., Frankish, A., Hunt, T., Harrow, J. & Gerstein, M. (2010). Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol.* **11**, R26

31. Banerji, J., Olson, L. & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**, 729–40

32. Muller, A.J., Young, J.C., Pendergast, A.M., Pondel, M., Landau, N.R., Littman, D.R. & Witte, O.N. (1991). BCR first exon sequences specifically activate the BCR/ABL tyrosine kinase oncogene of Philadelphia chromosome-positive human leukemias. *Mol. Cell. Biol.* **11**, 1785–92

33. 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

34. Grosveld, F., van Assendelft, G.B., Greaves, D.R. & Kollias, G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–85

35. Mohrs, M., Blankespoor, C.M., Wang, Z.E., Loots, G.G., Afzal, V., Hadeiba, H., Shinkai, K., Rubin, E.M. & Locksley, R.M. (2001). Deletion of a coordinate regulator of type 2 cytokine expression in mice. *Nat. Immunol.* **2**, 842–7

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

37. 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

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

39. 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**, 622–34

40. 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

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

42. Koch, F., Jourquin, F., Ferrier, P. & Andrau, J.-C. (2008). Genome-wide RNA polymerase II: not genes only! *Trends Biochem. Sci.* **33**, 265–73

43. 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

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

45. Ørom, U.A. & Shiekhattar, R. (2013). Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell* **154**, 1190–3

46. 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**, 117–39

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

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

49. Palstra, R., de Laat, W. & Grosveld, F. (2008). Chapter 4 β−globin regulation and long-range interactions. *Adv. Genet.* **61**, 107–42

50. 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

51. 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**, 823–31

52. 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**, 1224–34

53. 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**, 40–5

54. 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

55. 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**, 957–62

56. Osano, K. & Ono, M. (2003). State of histone modification in the rat Ig-beta/growth hormone locus. *Eur. J. Biochem.* **270**, 2532–9

57. 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

58. 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

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

60. Berger, S.L. (2002). Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**, 142–8

61. 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

62. 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

63. 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

64. 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

65. 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

66. 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

67. 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

68. 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

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

70. 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

71. 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

72. 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

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

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

75. 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

76. 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

77. 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

78. 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

79. 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

80. vanWert, J.M., Wolfe, S.A. & Grimes, S.R. (1995). Testis-specific expression of the rat histone H1t gene in transgenic mice. *Biochemistry* **34**, 8733–43

81. 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

82. 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

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

84. 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

85. 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

86. 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

87. 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

88. 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**, 958–70

89. 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**, 576–89

90. 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

91. 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

92. 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

93. 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

94. 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**, 171–6

95. 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**, 524–35

96. 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

97. 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

98. 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

99. 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

100. 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

101. 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

102. 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

103. Batista, P.J. & Chang, H.Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–307

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

105. 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**, 717–26

106. 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

107. 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

108. 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**, 409–19

109. 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**, 648–55

110. 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

111. 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

112. 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

113. 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

114. 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

115. 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

116. 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

117. 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

118. 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**, 114–23

119. Liao, B.-Y. & Zhang, J. (2006). Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol. Biol. Evol.* **23**, 530–40

120. Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., Patapoutian, A., Hampton, G.M., Schultz, P.G. & Hogenesch, J.B. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4465–70

121. Balakirev, E.S. & Ayala, F.J. (2003). Pseudogenes: are they "junk" or functional DNA? *Annu. Rev. Genet.* **37**, 123–51

122. Han, Y.J., Ma, S.F., Yourek, G., Park, Y.-D. & Garcia, J.G.N. (2011). A transcribed pseudogene of MYLK promotes cell proliferation. *FASEB J.* **25**, 2305–12

123. Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W.-O., Corcoran, M., Grandér, D. & Morris, K. V (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.* **20**, 440–6

124. Groen, J.N., Capraro, D. & Morris, K. V (2014). The emerging role of pseudogene expressed non-coding RNAs in cellular functions. *Int. J. Biochem. Cell Biol.* **54**, 350–5

125. Elefant, F., Su, Y., Liebhaber, S.A. & Cooke, N.E. (2000). Patterns of histone acetylation suggest dual pathways for gene activation by a bifunctional locus control region. *EMBO J.* **19**, 6814–22

126. Chalmel, F., Lardenois, A., Evrard, B., Mathieu, R., Feig, C., Demougin, P., Gattiker, A., Schulze, W., Jégou, B., Kirchhoff, C. & Primig, M. (2012). Global human tissue

profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum. Reprod.* **27**, 3233–48

127. Chalmel, F., Rolland, A.D., Niederhauser-Wiederkehr, C., Chung, S.S.W., Demougin, P., Gattiker, A., Moore, J., Patard, J.-J., Wolgemuth, D.J., Jégou, B. & Primig, M. (2007). The conserved transcriptome in human and rodent male gametogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8346–51

128. Iynedjian, P.B., Marie, S., Wang, H., Gjinovci, A. & Nazaryan, K. (1996). Liver-specific enhancer of the glucokinase gene. *J. Biol. Chem.* **271**, 29113–20

129. Surinya, K.H., Cox, T.C. & May, B.K. (1998). Identification and characterization of a conserved erythroid-specific enhancer located in intron 8 of the human 5-aminolevulinate synthase 2 gene. *J. Biol. Chem.* **273**, 16798–809

130. Fujiwara, J., Kimura, T., Ayusawa, D. & Oishi, M. (1994). A novel regulatory sequence affecting the constitutive expression of tissue plasminogen activator (tPA) gene in human melanoma (Bowes) cells. *J. Biol. Chem.* **269**, 18558–62

131. Madhusudhan, K.T., Naik, S.S. & Patel, M.S. (1995). Characterization of the promoter regulatory region of the human pyruvate dehydrogenase .beta. gene. *Biochemistry* **34**, 1288–94

132. Li, W., Yang, W. & Wang, X.-J. (2013). Pseudogenes: pseudo or real functional elements? *J. Genet. Genomics* **40**, 171–7

133. Bennani-Baiti, I.M., Jones, B.K., Liebhaber, S.A. & Cooke, N.E. (1995). Physical linkage of the human growth hormone gene cluster and the skeletal muscle sodium channel alpha-subunit gene (SCN4A) on chromosome 17. *Genomics* **29**, 647–52

134. Hogan, K.A., Jefferson, H.S., Karschner, V.A. & Shewchuk, B.M. (2009). Expression of Pit-1 in nonsomatotrope cell lines induces human growth hormone locus control region histone modification and hGH-N transcription. *J. Mol. Biol.* **390**, 26–44

135. Cajiao, I., Zhang, A., Yoo, E.J., Cooke, N.E. & Liebhaber, S.A. (2004). Bystander gene activation by a locus control region. *EMBO J.* **23**, 3854–63

136. Oshiumi, H., Sasai, M., Shida, K., Fujita, T., Matsumoto, M. & Seya, T. (2003). TIR-containing adapter molecule (TICAM)-2, a bridging adapter recruiting to toll-like receptor 4 TICAM-1 that induces interferon-beta. *J. Biol. Chem.* **278**, 49751–62

137. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engström, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. & Wahlestedt, C. (2005). Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–6

138. Ge, X., Wu, Q., Jung, Y.-C., Chen, J. & Wang, S.M. (2006). A large quantity of novel human antisense transcripts detected by LongSAGE. *Bioinformatics* **22**, 2475–9

139. Kapranov, P., Willingham, A.T. & Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* **8**, 413–23

140. Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P. & Kohtz, J.D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470–84

141. Onoguchi, M., Hirabayashi, Y., Koseki, H. & Gotoh, Y. (2012). A noncoding RNA regulates the neurogenin1 gene locus during mouse neocortical development. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16939–44

142. Bao, J., Wu, J., Schuster, A.S., Hennig, G.W. & Yan, W. (2013). Expression profiling reveals developmentally regulated lncRNA repertoire in the mouse male germline. *Biol. Reprod.* **89**, 107

143. Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., Reinders, M., Wessels, L. & van Steensel, B. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–13

144. Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. *Cell* **128**, 787–800

145. Bertoli, C., Skotheim, J.M. & de Bruin, R.A.M. (2013). Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* **14**, 518–28

146. White-Cooper, H. & Davidson, I. (2011). Unique aspects of transcription regulation in male germ cells. *Cold Spring Harb. Perspect. Biol.* **3**, a002626

147. Marques, A.C., Tan, J., Lee, S., Kong, L., Heger, A. & Ponting, C.P. (2012). Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. *Genome Biol.* **13**, R102

148. Betran, E., Wang, W., Jin, L. & Long, M. (2002). Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. *Mol. Biol. Evol.* **19**, 654–63

149. Kandouz, M., Bier, A., Carystinos, G.D., Alaoui-Jamali, M.A. & Batist, G. (2004). Connexin43 pseudogene is expressed in tumor cells and inhibits growth. *Oncogene* **23**, 4763–70

150. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M. & Hannon, G.J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534–8

151. Korneev, S.A., Kemenes, I., Bettini, N.L., Kemenes, G., Staras, K., Benjamin, P.R. & O'Shea, M. (2013). Axonal trafficking of an antisense RNA transcribed from a pseudogene is regulated by classical conditioning. *Sci. Rep.* **3**, 1027

152. Pyeon, D., Newton, M.A., Lambert, P.F., den Boon, J.A., Sengupta, S., Marsit, C.J., Woodworth, C.D., Connor, J.P., Haugen, T.H., Smith, E.M., Kelsey, K.T., Turek, L.P. & Ahlquist, P. (2007). Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res.* **67**, 4605–19

153. Offenberg, H.H., Schalk, J.A., Meuwissen, R.L., van Aalderen, M., Kester, H.A., Dietrich, A.J. & Heyting, C. (1998). SCP2: a major protein component of the axial elements of synaptonemal complexes of the rat. *Nucleic Acids Res.* **26**, 2572–9

154. Prieto, I., Suja, J.A., Pezzi, N., Kremer, L., Martínez-A, C., Rufas, J.S. & Barbero, J.L. (2001). Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. *Nat. Cell Biol.* **3**, 761–6