HOKKAIDO UNIVERSITY

Title	Characterization of the human TCAM1P pseudogene and its activation by a potential dual promoter-enhancer: Comparison with a protein-coding mouse orthologue
Author(s)	Kurihara, Misuzu; Kimura, Atsushi P.
Citation	FEBS Letters, 589(4), 540-547 https://doi.org/10.1016/j.febslet.2015.01.023
Issue Date	2015-02-13
Doc URL	https://hdl.handle.net/2115/58456
Rights	(C) 2015 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
Type	journal article
File Information	FEBSLet_589_p540-.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

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

Misuzu Kurihara[a] and Atsushi P. Kimura[a,b,*]

[a]Graduate School of Life Science, Hokkaido University, Sapporo 060-0810, Japan

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

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

Abstract

***TCAM1P* is a unitary pseudogene, which was disabled since the human-mouse divergence. Here we found that *TCAM1P* was specifically expressed in the human testis, with different cell type-specificity from mouse *Tcam1*, and characterized its transcripts. At the mouse locus, a multifunctional dual promoter–enhancer (DPE) controls the expression of *Tcam1* and *Smarcd2* genes. This time, the corresponding human sequence was found to potentially function as a DPE, although the molecular mechanism was different from mouse. Interestingly, the change in DPE activity occurred before pseudogenization of *TCAM1P*. These data suggest the presence of a DPE in human genome for the first time, and provide an important model of evolutionary changes in a regulatory mechanism of a pseudogene.**

Keywords: dual promoter–enhancer; human; long noncoding RNA; unitary pseudogene; *TCAM1P*; testis

Abbreviations: ANOVA, analysis of variance; CNS, conserved noncoding sequence; DPE, dual promoter–enhancer; hCNS1, human CNS1; hGH, human growth hormone; IGV, Integrative Genomics Viewer; LCR, locus control region; lncRNA, long noncoding RNA; mCNS1, mouse CNS1; ORF, open reading frame; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-polymerase chain reaction; Smarcd2/SAMRCD2, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2; Tcam1/TCAM1P, testicular cell adhesion molecule 1; TSS, transcriptional start site ; TTS, transcriptional termination site

1. Introduction

Functions and expression patterns of orthologous genes are generally conserved between closely related species [1,2], 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 [3]. Recently, some of the transcribed pseudogenes were revealed to be functional and have novel roles [4–6]. 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.

We previously investigated a regulatory mechanism at the mouse *testicular cell adhesion molecule 1 (Tcam1)* gene locus, and identified mouse conserved noncoding sequence 1 (mCNS1) as a multifunctional dual promoter–enhancer (DPE) [7]. mCNS1 could function as a spermatocyte-specific enhancer for *Tcam1* and a bidirectional promoter of the *SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily d, member 2* (*Smarcd2*) gene and a testis-specific long noncoding RNA (lncRNA), *lncRNA-Tcam1* [7]. This was the first indication of a DPE in mammals, although two examples had been reported in chicken [8,9].

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 [10], albeit mouse *Tcam1* encodes a cell adhesion protein [11]. *TCAM1P* was pseudogenized since the human-mouse divergence [10], and therefore, it is of significance to note 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 *human growth hormone* (*hGH*) gene cluster, a famous model for tissue-specific gene activation by multifunctional elements [12–16].

In this study, we investigated the expression and regulation of human *TCAM1P* and *SMARCD2* genes. Our results indicated that the expression pattern of *TCAM1P* was slightly different from mouse

Tcam1, and that human CNS1 (hCNS1) could function as a DPE, although the molecular mechanism was different from mCNS1. The data provide the first evidence for the presence of DPE in human genome, and represent an important model for gene regulatory mechanisms before and after pseudogenization.

2. Materials and methods

2.1. RNA Analyses

Total RNAs from the human liver and testis were purchased from Clontech Laboratories, Inc. (Mountain View, USA) and reverse transcription-polymerase chain reaction (RT-PCR) was done as previously described [17]. Primer sequences are shown in Table 1. The primer pair for *β-ACTIN* was used by another group [18].

2.2. Rapid amplification of cDNA ends (RACE)

5'RACE and 3'RACE were performed as previously described [7,17]. All the amplified products were subcloned into a pBluescript vector (Stratagene, La Jolla, CA) 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 1.

2.3. Reporter constructs

A promoter region of the *TCAM1P* gene was amplified by PCR using KOD FX (Toyobo, Osaka, Japan) with human sperm genomic DNA (Clontech Laboratories, Inc.) and a primer pair listed in Table

1. The 2012-bp product was ligated to upstream of the luciferase gene in a pGL3-Basic vector (Promega Corporation, Madison, WI) at the *Sma*I site, and we named the resulting construct TCAM1P-Pro-luc. To obtain the hCNS1 sequence, we 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 1), and the 1340-bp product was purified and used as a template for the second PCR with the 'hCNS1 nested' primer pair (Table 1). By inserting the 450-bp product into the blunted *Nhe*I site of TCAM1P-Pro-luc, we obtained the hCNS1-Pro-luc construct. To generate the Pro-luc-hCNS1 and Pro-luc-reversed (rev) hCNS1 constructs, we cloned the hCNS1 fragment into the blunted *Bam*HI site of TCAM1P-Pro-luc. To linearize the constructs, we 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.

2.4. Cell culture and the reporter gene assay

GC-2spd(ts), Hepa1-6, and HEK293T cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum, 100 U/ml penicillin, 100 μg/ml streptomycin, and 292 μg/ml L-glutamine (Invitrogen, Carlsbad, CA). The reporter gene assay was performed as previously described [19].

2.5. Statistical analysis

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

3. Results

3.1. Tissue distribution of human SMARCD2 and TCAM1P mRNAs

We 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. 1A). To know what types of cells expressed *TCAM1P* mRNA in the human testis, we 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 [20,21]. These indicated that the cell type-specificity of human *TCAM1P* was different from mouse *Tcam1* which was expressed only in spermatocytes [7,11].

Because the *TCAM1P* gene was considered to be a pseudogene, we 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. 1B. To distinguish these transcripts, we designed four primer sets (Fig. 1B, P1–P4) and performed RT-PCR with human testis and liver RNAs. As a result, we 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. 1C). We named the detected transcripts, *TCAM1P-I*, *TCAM1P-II*, and *TCAM1P-III* (Fig. 1B, C).

3.2. Determination of full-length sequences of TCAM1P transcripts

To determine full-length sequences of the three *TCAM1P* transcripts, we conducted 5'RACE and 3'RACE analyses using the human testis RNA. 5'RACE was performed with two primer sets: one for detecting the transcriptional start site (TSS) of *TCAM1P-I* and *TCAM1P-III* (Fig. 2B), and the other for TSS of *TCAM1P-I* and *TCAM1P-II* (Fig. 2C). We identified one nucleotide at the position +401, which was 401 bp downstream of TSS in the Ensembl database (Fig. 2A, B), as TSS of *TCAM1P-I* and *TCAM1P-III* (Fig. 2B). In contrast, we found five different TSSs of *TCAM1P-I* and *TCAM1P-II*, each of which was positioned at +494, +401, +377, +329, and +5100 (Fig. 2C). The nucleotide at +401 might be derived from *TCAM1P-I*. These results indicated that there was one TSS of *TCAM1P-I* and *TCAM1P-III* at +401, and *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. 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, we 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. We also examined whether the three transcripts were terminated at the minor TTS, but we 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, we concluded that at least six *TCAM1P* transcripts were present in the human testis as shown in Fig. 2E, and that the *TCAM1P* gene was actually composed of five exons (Fig. 2A).

Although the *Tcam1* gene in rodents and rhesus monkey encodes a protein related to a cell adhesion molecule, the human *TCAM1P* gene is considered to be a pseudogene [10]. Since we

determined full-length sequences of six *TCAM1P* transcripts, we 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.

3.3. Enhancer and promoter activity of hCNS1

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

We 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. 2E), we 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. 3A). When we connected hCNS1 to 3'of the luciferase gene, the activity was also significantly increased in all cells irrespective of its orientation (Fig. 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. 3A, Pro-luc-hCNS1). We 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

[22–25].

We also examined the luciferase activity in linearized constructs. If hCNS1 was actually a DPE, it might drive the transcription of the entire sequence in a circular plasmid, which might affect the luciferase gene expression. To prevent hCNS1 from driving the luciferase gene transcription in the sense direction, we linearized Pro-luc-hCNS1 and Pro-luc-rev hCNS1 constructs by cutting the immediate downstream sequence of hCNS1 and rev hCNS1. As a result, hCNS1 still significantly increased *TCAM1P* promoter activity in HEK293T cells, and the fold-increases were around 3 in both constructs compared to the linearized TCAM1P-Pro-luc construct (Fig. 3B). These data showed that hCNS1 could function as an enhancer for the *TCAM1P* gene.

We 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, we connected it directly to the luciferase gene in both orientations (Fig. 4). By reporter gene assay, hCNS1 showed promoter activity in both directions, but the activity was stronger in the reverse direction in all the cell lines (Fig. 4). This indicated that hCNS1 had bidirectional promoter activity, but the activity was stronger for the *SMARCD2* gene.

3.4. Intergenic transcription at the human SMARCD2-TCAM1P locus

At the mouse *Smarcd2-Tcam1* locus, mCNS1 can drive not only *Smarcd2* but also *lncRNA-Tcam1* [7]. Because hCNS1 showed bidirectional promoter activity by reporter gene assay (Fig. 4), we assessed the possibility that hCNS1 could be a promoter for an lncRNA as well as *SMARCD2*. We performed RT-PCR with the human testis RNA and successfully amplified the hCNS2 sequence (data not shown) which was positioned close to hCNS1 (Fig. 5A). To determine the full-length sequence of the transcript, we performed 5’RACE and 3’RACE analyses. By sequencing 10 subclones each for 5’RACE and 3’RACE, we detected a single TSS and a single TTS (data not shown). Consequently, we identified a 1.1-kb transcript including one intron (Fig. 5A) and named it *lncRNA-TCAM1P*, because it

was classified as an lncRNA by the coding potential calculator [26]. However, the *lncRNA-TCAM1P* promoter region did not overlap with hCNS1 (Fig. 5A), which indicated that hCNS1 was not a promoter for *lncRNA-TCAM1P*.

We 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. 5A, 1-6) and human testis and liver RNAs, we detected specific signals from all the examined regions in the testis, but not in the liver (Fig. 5B). This indicated that, consistent with Fig. 1A, most of the intergenic region between *SMARCD2* and *TCAM1P* genes were specifically transcribed in the testis.

To further characterize the intergenic transcription, we attempted to determine full-length sequences of the transcripts from amplicons 1 and 2. However, because approximately 70% sequence from 3' end of amplicon 2 overlapped with repeat sequences, we performed 5'RACE only for this transcript, while both ends were determined for amplicon 1. As a result, we detected three TSSs and three TTSs for amplicon 1, and two TSSs for amplicon 2 (Fig. 5C, D). Judging from the number of subclones, a major TSS was presumed to be the nucleotide at -6124 bp and -5339 bp relative to TSS of *TCAM1P* for ampicons 1 and 2, respectively, and a major TTS for amplicon 1 was positioned at -4563 bp (Fig. 5C, D). Collectively, we identified two novel lncRNAs: a full-length sequence of *lncTCAM1P-6.1kb* (Fig. 5C) and a partial sequence of *lncTCAM1P-5.3kb* (Fig. 5D). These transcripts were confirmed to be transcribed as single transcripts by RT-PCR performed as described in section 3.2 (data not shown).

As pointed out by many researchers [27,28], it was difficult to directly align the human lncRNAs we detected with mouse *lncRNA-Tcam1*, but they might be functionally homologous. What is important is that promoters of *lncTCAM1P-6.1kb* and *lncTCAM1P-5.3kb* did not overlap with hCNS1, which indicated that hCNS1 could not be promoters of these lncRNAs. The data suggest that hCNS1 may be a unidirectional promoter, although we cannot completely rule out the possibility that it functions bidirectionally.

3.5. Changes of CNS1 activity before pseudogenization of TCAM1P

The above data revealed the functional difference between mCNS1 and hCNS1, and we finally investigated how CNS1 activity was changed during evolution. Zhang *et al.* demonstrated that *TCAM1P* was pseudogenized at the gorilla-chimpanzee divergence by comparing genomic sequences from several primate species [10], and we examined the evolutionary change in CNS1. We collected CNS1 sequences from mouse, rat, an old world monkey (rhesus monkey), a great ape (gibbon), and human by searching for the Ensembl and NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) databases, and compared each sequence with another. As a result, hCNS1 shared more than 90% nucleotide identity with CNS1 of rhesus monkey and gibbon, in contrast to less than 80% with that of rodents (Fig. 6A). Conversely, mCNS1 showed 93% nucleotide identity with rat CNS1, but had less than 80% with other species (Fig. 6A). In addition, we could not find any nucleotides in CNS1 that were changed only in human and conserved in the other four species to allow hCNS1 to possess a specific function to human (Fig. 6B). Therefore, CNS1 is presumed to be functionally changed before pseudogenization of *TCAM1P*, which suggests that cell-type specificity of the *TCAM1* gene was altered before the pseudogenization.

4. Discussion

4.1. hCNS1 is a potential DPE

One aim of this study was to determine whether human genome contained a DPE, and we assessed enhancer activity of hCNS1. Although all the constructs containing hCNS1 increased *TCAM1P* promoter activity (Fig. 3), the strong promoter activity of hCNS1 might affect the luciferase gene expression. Even in the assay with linearized constructs, hCNS1 and rev hCNS1 might drive the

transcription of the luciferase gene in the antisense direction, which might affect the luciferase expression. However, because the promoter activity of hCNS1 was orientation-dependent (Fig. 4), the antisense transcription would affect the luciferase activity to different extents in the Pro-luc-hCNS1 and Pro-luc-rev hCNS1 constructs. Nonetheless, the activity in these constructs was similar (Fig. 3B), which indicated that the antisense transcription was unlikely to affect the luciferase gene expression. Therefore, we conclude that hCNS1 is a potential enhancer for *TCAM1P* similarly to mCNS1.

Interestingly, a mechanism underlying the enhancer activity seemed to be different, because hCNS1 enhancer activity was observed in all the cell lines we examined (Fig. 3), whereas mCNS1 enhancer activity was specific to spermatocyte-derived GC-2spd(ts) cells [7]. 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. 4). This is in contrast to mCNS1, whose promoter activity in each direction was similar [7]. Because we failed to detect any transcripts driven by hCNS1 extending to the opposite direction of *SMARCD2* (Fig. 5), hCNS1 may be a unidirectional promoter. However, hCNS1 promoter activity opposite to *SMARCD2* was still much higher than that of *TCAM1P* promoter (Fig. 4), and there were transcripts, other than the three lncRNAs we identified, transcribed from the intergenic region between *SMARCD2* and *TCAM1P* genes (Fig. 1A, Fig. 5). Further analysis will be necessary to reveal whether hCNS1 is a bidirectional promoter or not, but our data clearly indicate that hCNS1 is at least a promoter of the *SMARCD2* gene.

Taken together, hCNS1 could enhance *TCAM1P* promoter activity and functioned as a promoter of the *SMARCD2* gene. Therefore, similarly to mCNS1, hCNS1 can be a DPE in the human testis.

4.2. Evolutionary changes in parallel with pseudogenization

Another aim of this work was to determine evolutionary changes, especially in gene regulatory mechanisms, in light of the pseudogenization of the human *TCAM1P* gene. 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 [29]. *TCAM1P* is a unitary pseudogene [10], and it is possible to directly compare the human *TCAM1P* locus with the mouse one.

In both human and mouse, we confirmed ubiquitous *Smarcd2/SMARCD2* expression and DPE activity of mCNS1/hCNS1, but otherwise, many differences were found by our analyses. Firstly, human *TCAM1P* was testis-specific, but the cell type expressing it was not only spermatocytes but also spermatogonia and spermatids, whereas mouse *Tcam1* mRNA was specific to spermatocytes [7,20,21]. 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, indicating different molecular mechanisms between these two species as discussed above. Fourthly, hCNS1 enhancer activity was lower than mCNS1. By reporter gene assay, mCNS1 mostly showed about 5-fold higher activity than *Tcam1* promoter [7], while the activity of hCNS1 was 1.8-4.5-fold higher than *TCAM1P* promoter (Fig. 3). Fifthly, hCNS1 may not be a bidirectional promoter as discussed above. Finally, more lncRNAs were transcribed at the human locus. These indicates that the expression of *TCAM1P* and lncRNAs was changed during evolution, and that the gene regulation at this locus, especially the molecular mechanism underlying DPE activity, was also changed.

Furthermore, we found that CNS1 activity was changed before pseudogenization of the *TCAM1P* gene. This suggests that the *TCAM1* gene is expressed in spermatogonia, spermatocytes, and spermatids not only in human but also in rhesus monkey and gibbon that contain the protein-coding *TCAM1* gene. If this is the case, the cell adhesion protein may be necessary in the three types of germ cells during meiosis in rhesus monkey and gibbon, but unnecessary in chimpanzee and human.

Considering that *Tcam1*-knockout mice were normal [11] and the function of TCAM1 as a cell adhesion molecule is probably redundant, the *TCAM1P* pseudogene might gain a new role in chimpanzee and human. In this context, it is of note that human *TCAM1P*, together with other testis-specific genes, was induced in human papillomavirus-positive cancer cells [30]. *TCAM1P* may have some functions related both to cell proliferation and to male meiosis.

In any case, the present data provide an important example of evolutionary changes, especially in gene regulatory mechanisms, in parallel with pseudogenization.

4.3. 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 [31,32], and their tissue-specific expression is related to some multifunctional elements. Firstly, the *hGH* cluster is controlled by the locus control region (LCR), which overlapps with the *SCN4A* gene encoding a sodium channel protein in the skeletal muscle [12,33,34]. Secondly, the *hGH* LCR activates the cluster genes by different mechanisms between pituitaty and placenta [16,35–38]. 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 [16,39,40]. In addition to these, our current data suggest that a 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.

Acknowledgements

We thank Dr. Saadi Khochbin for his valuable advice on *TCAM1P* expression in testicular cells. This work was supported by a Grant-in-aid for Scientific Research 21770068 to A.P.K. from the Ministry of Education, Culture, Sports, Science and Technology, Japan. M.K. was supported by a research fellowship of the Japan Society of the Promotion of Science (24-6230).

References

[1] Liao B.-Y. and Zhang J. (2006) Evolutionary conservation of expression profiles between human and mouse orthologous genes. Mol. Biol. Evol. 23, 530–540.

[2] 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. and Hogenesch J.B. (2002) Large-scale analysis of the human and mouse transcriptomes. Proc. Natl. Acad. Sci. U. S. A. 99, 4465–4470.

[3] Balakirev E.S. and Ayala F.J. (2003) Pseudogenes: are they “junk” or functional DNA? Annu. Rev. Genet. 37, 123–151.

[4] Han Y.J., Ma S.F., Yourek G., Park Y.-D. and Garcia J.G.N. (2011) A transcribed pseudogene of MYLK promotes cell proliferation. FASEB J. 25, 2305–2312.

[5] Johnsson P., Ackley A., Vidarsdottir L., Lui W.-O., Corcoran M., Grandér D. and Morris K.V. (2013) A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. Nat. Struct. Mol. Biol. 20, 440–446.

[6] Groen J.N., Capraro D. and Morris K.V. (2014) The emerging role of pseudogene expressed non-coding RNAs in cellular functions. Int. J. Biochem. Cell Biol. 54, 350–355.

[7] Kurihara M., Shiraishi A., Satake H. and Kimura A.P. (2014) 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. J. Mol. Biol. 426, 3069–3093.

[8] Lefevre P., Witham J., Lacroix C.E., Cockerill P.N. and Bonifer C. (2008) The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Mol. Cell 32, 129–139.

[9] Wilczek C., Chayka O., Plachetka A. and 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, 35314–35324.

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

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

[12] Jones B.K., Monks B.R., Liebhaber S.A. and Cooke N.E. (1995) The human growth hormone gene is regulated by a multicomponent locus control region. Mol. Cell. Biol. 15, 7010–7021.

[13] Elefant F., Cooke N.E. and Liebhaber S.A. (2000) Targeted recruitment of histone acetyltransferase activity to a locus control region. J. Biol. Chem. 275, 13827–13834.

[14] Elefant F., Su Y., Liebhaber S.A. and Cooke N.E. (2000) Patterns of histone acetylation suggest dual pathways for gene activation by a bifunctional locus control region. EMBO J. 19, 6814–622.

[15] Kimura A.P., Liebhaber S.A. and 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, 1018–1032.

[16] Ho Y., Elefant F., Liebhaber S.A. and Cooke N.E. (2006) Locus control region transcription plays an active role in long-range gene activation. Mol. Cell 23, 365–375.

[17] Matsubara S., Kurihara M. and 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, 243–256.

[18] Oshiumi H., Sasai M., Shida K., Fujita T., Matsumoto M. and 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, 49751–49762.

[19] Matsubara S., Takahashi T. and 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.

[20] Chalmel F., Lardenois A., Evrard B., Mathieu R., Feig C., Demougin P., Gattiker A., Schulze W., Jégou B., Kirchhoff C. and 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, 3233–3248.

[21] 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. and Primig M. (2007) The conserved transcriptome in human and rodent male gametogenesis. Proc. Natl. Acad. Sci. U. S. A. 104, 8346–8351.

[22] Iynedjian P.B., Marie S., Wang H., Gjinovci A. and Nazaryan K. (1996) Liver-specific enhancer of the glucokinase gene. J. Biol. Chem. 271, 29113–29120.

[23] Surinya K.H., Cox T.C. and 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, 16798–16809.

[24] Fujiwara J., Kimura T., Ayusawa D. and 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, 18558–18562.

[25] Madhusudhan K.T., Naik S.S. and Patel M.S. (1995) Characterization of the Promoter Regulatory Region of the Human Pyruvate Dehydrogenase .beta. Gene. Biochemistry 34, 1288–1294.

[26] Kong L., Zhang Y., Ye Z.-Q., Liu X.-Q., Zhao S.-Q., Wei L. and Gao G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35, W345–349.

[27] Johnsson P., Lipovich L., Grandér D. and Morris K.V. (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim. Biophys. Acta 1840, 1063–1071.

[28] Mercer T.R., Dinger M.E. and Mattick J.S. (2009) Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10, 155–159.

[29] Li W., Yang W. and Wang X.-J. (2013) Pseudogenes: pseudo or real functional elements? J. Genet. Genomics 40, 171–177.

[30] 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. and 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, 4605–4619.

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

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

[33] Bennani-Baiti I.M., Jones B.K., Liebhaber S.A. and 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, 647–652.

[34] Su Y., Liebhaber S.A. and 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, 7902–7909.

[35] Ho Y., Elefant F., Cooke N. and Liebhaber S. (2002) A defined locus control region determinant links chromatin domain acetylation with long-range gene activation. Mol. Cell 9, 291–302.

[36] Kimura A.P., Sizova D., Handwerger S., Cooke N.E. and Liebhaber S.A. (2007) Epigenetic activation of the human growth hormone gene cluster during placental cytotrophoblast differentiation. Mol. Cell. Biol. 27, 6555–6568.

[37] Hogan K.A., Jefferson H.S., Karschner V.A. and 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.

[38] Tsai Y.-C., Cooke N.E. and Liebhaber S.A. (2014) Tissue specific CTCF occupancy and boundary function at the human growth hormone locus. Nucleic Acids Res. 42, 4906–4921.

[39] Cajiao I., Zhang A., Yoo E.J., Cooke N.E. and Liebhaber S.A. (2004) Bystander gene activation by a locus control region. EMBO J. 23, 3854–3863.

[40] Yoo E.J., Cooke N.E. and Liebhaber S.A. (2012) An RNA-independent linkage of noncoding transcription to long-range enhancer function. Mol. Cell. Biol. 32, 2020–2029.

Figure legends

Fig. 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.

Fig. 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 adaptor sequence. The first and second PCR reactions were performed with GSP3-1 and GSP4-1 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 s 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. Nucleotide sequence data for the six *TCAM1P* transcripts 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*), and LC005101 (*TCAM1P-III*).

Fig. 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. 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. 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. 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 (determined in this study) 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-length 1.1-kb transcript, *lncRNA-TCAM1P*, and six amplicons are indicated by blue and black lines, respectively. TSS and TTS of *lncRNA-TCAM1P* were determined to be positioned at -7023 bp and -5549 bp relative to TSS of *TCAM1P* by 5'RACE and 3'RACE. In each analysis, all 10 subclones contained the same nucleotide as TSS or TTS. (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. 1C, an internal control should be referred to Fig. 1C. (C) Cloning of the full-length sequence of *lncTCAM1P-6.1kb*. The structure of the locus is drawn as in (A). 5' and 3' ends of the transcript from amplicon 1 were determined by 5'RACE and 3'RACE. 5'RACE was performed with RT-3 for reverse transcription and GSP1-3 and GSP2-3 for two rounds of PCRs. After subcloning of the product, 13 subclones were sequenced, and three different TSSs were detected as indicated. For 3'RACE, GSP3-2 and GSP4-2 were used for two rounds of PCRs, and we detected three TTSs by sequencing 12 subclones as indicated. A full-length of the major transcript for *lncTCAM1P-6.1kb* is depicted at the top. (D) Determination of TSSs of *lncTCAM1P-5.3kb*. 5'RACE was performed using RT-4 for reverse transcription and GSP1-4 and GSP2-4 for PCRs. By sequencing 11 subclones, two TSSs were determined as indicated. Data for full-length sequences of *lncRNA-TCAM1P* and *lncTCAM1P-6.1kb* and a partial sequence of *lncTCAM1P-5.3kb* are available in DDBJ/EMBL/GenBank databases under the accession numbers LC005102, LC019129, and LC019128, respectively.

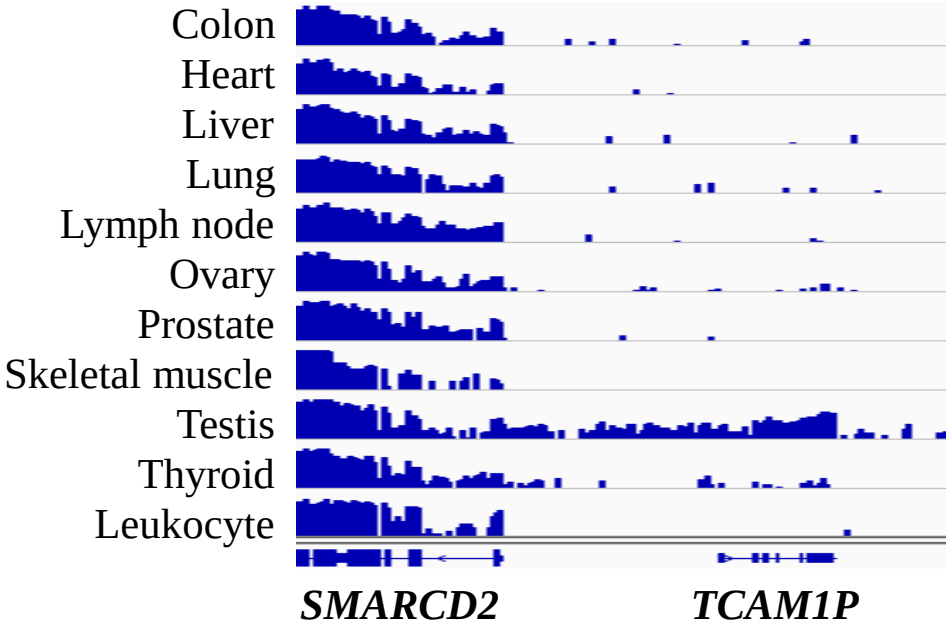
Fig. 6. Comparison of CNS1 sequences from five mammalian species. (A) Nucleotide identities (%) between CNS1 sequences from the indicated species. The CNS1 sequences were obtained from *Homo sapiens* (human), *Nomascus leucogenys* (gibbon), *Macaca mulatta* (rhesus monkey), *Rattus norvegicus* (rat), and *Mus musculus* (mouse) by searching for the Ensembl and NCBI databases, and each sequence was compared with another. (B) A multiple alignment of the CNS1 sequences from five mammalian species. The CNS1 sequences from the indicated species were compared using the DNASIS-Pro software (HITACHI Software Engineering, Yokohama, Japan). Conserved nucleotides among four or five species are shaded.

Table 1. Oligonucleotide primers used in this study

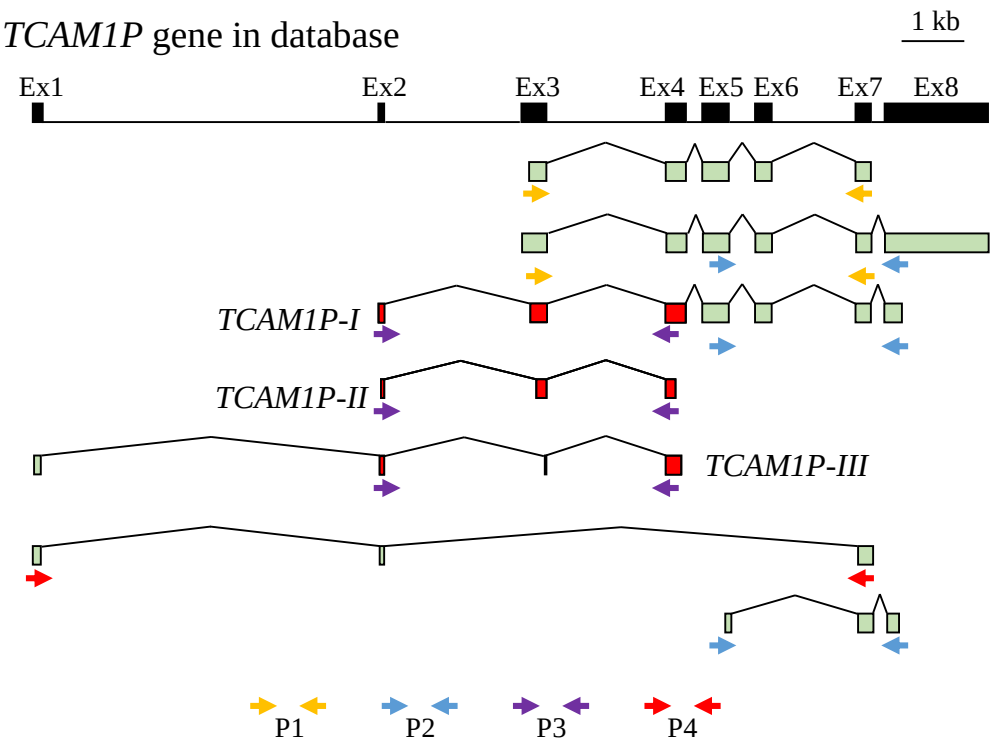
Designation	Forward	Reverse
RT-PCR		
TCAM1P P3	5'-TTCTGTTTGGTGTCTGGACC-3'	5'-GGGTACATGACACTTCACTG-3'
lncRNA-TCAM1P	5'-CTCCGGGATTTCCTAATCCA-3'	5'-GGGAGGTTTCTCATAAGCCA-3'
1	5'-CCCTCTGATGGTCTGGATCT-3'	5'-TAGCTTGCCCCTTTTCTAGG-3'
2	5'-CCCTCCCATCCCTGTAAAGT-3'	5'-AGAAAAAGATCCTGCCTTCG-3'
3	5'-GCCTCTCTGTTCTCTCACAC-3'	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 RT-1		5'-GGGTACATGACACTTCACTG-3'
TCAM1P GSP1-1		5'-TCCAGACACCAAACAGAAGC-3'
TCAM1P GSP2-1		5'-CTGATAAGAAGGGCCAAAGG-3'
TCAM1P RT-2		5'-AAGCACTGCAGGATGGAA-3'
TCAM1P GSP1-2		5'-CTTTGCCCACCTGAGTTTTC-3'
TCAM1P GSP2-2		5'-TCCAGACACCAAACAGAAGC-3'
lncRNA-TCAM1P RT		5'-AACCGGGTCTGTGAGGAACT-3'
lncRNA-TCAM1P GSP1		5'-CTGTGAGGAACTCGCGAACT-3'
lncRNA-TCAM1P GSP2		5'-TGGATTAGGAAATCCCGGAG-3'
lncTCAM1P-6.1kb RT3		5'-CAGGAGGTAAGAAAACAGCA-3'
lncTCAM1P-6.1kb GSP1-3		5'-GTGCTATGATGCAAGGAAGG-3'
lncTCAM1P-6.1kb GSP2-3		5'-TTACTTGGTTGGGGCTGGAA-3'

lncTCAM1P-5.3kb RT4		5'-TAGCTTGCCCCTTTTCTAGG-3'
lncTCAM1P-5.3kb GSP1-4		5'-CGGAGGTATTTTGGAGTAGC-3'
lncTCAM1P-5.3kb GSP2-4		5'-ACTTTACAGGGATGGGAGGG-3'
3’RACE		
AP		5'-CTGATCTAGAGGTACCGGATCC-3'
TCAM1P GSP3-1	5'-TTCTGTTTGGTGTCTGGACC-3'	
TCAM1P GSP4-1	5'-GATTCAAAAGGACACAAGCC-3'	
lncRNA-TCAM1P GSP3	5'-CCACACTCCCCGAAAATAGT-3'	
lncRNA-TCAM1P GSP4	5'-AAAGACTCAGCTGCTCTGTG-3'	
lncTCAM1P-6.1kb GSP3-2	5'-AGTGGAGAATCAGAGTCCTC-3'	
lncTCAM1P-6.1kb GSP4-2	5'-GGTCGCATCAGAAAACACAA-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)

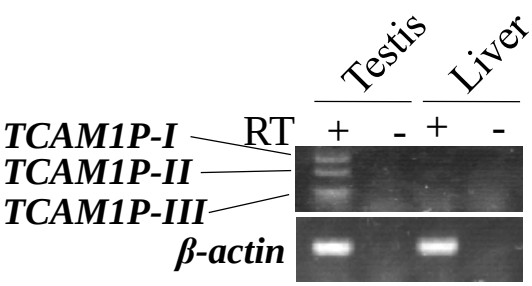

Testis
Liver
RT
+
-
+
-
TCAM1P-I
TCAM1P-II
TCAM1P-III
β-actin

Fig. 1

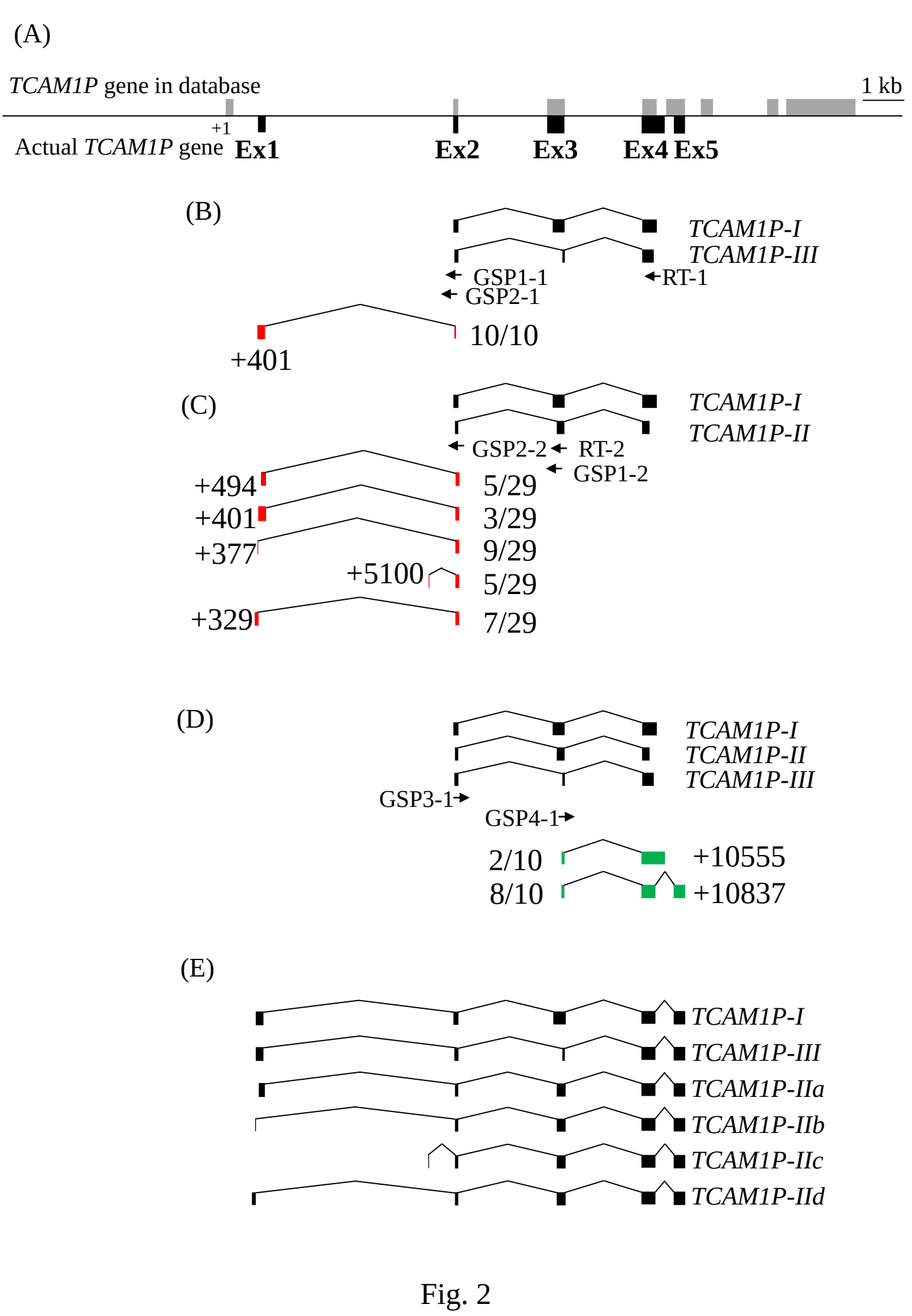
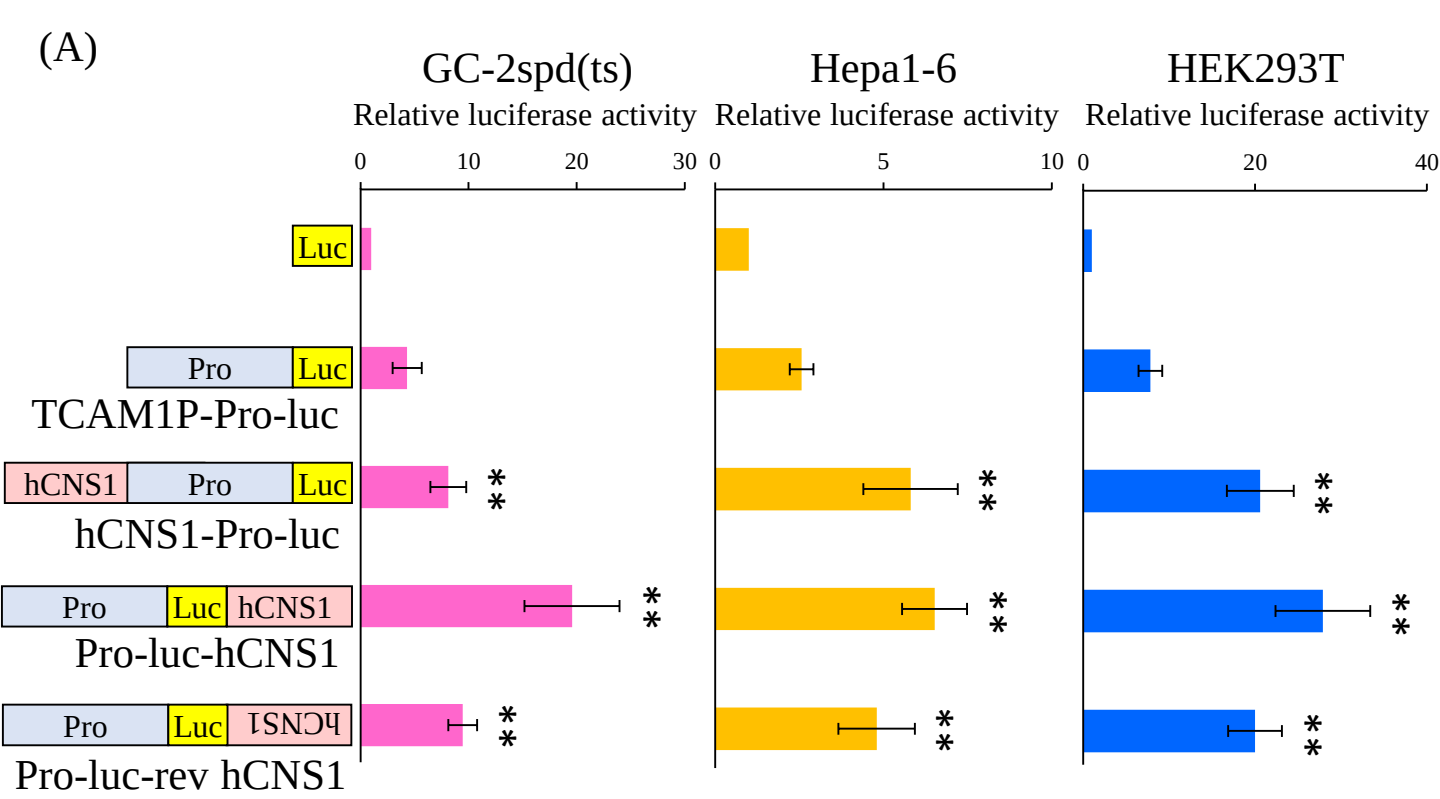

(A)
TCAM1P gene in database
1 kb
+1
Actual TCAM1P gene
Ex1
Ex2
Ex3
Ex4
Ex5
(B)
TCAM1P-I
TCAM1P-III
GSP1-1
RT-1
GSP2-1
10/10
+401
(C)
TCAM1P-I
TCAM1P-II
GSP2-2
RT-2
GSP1-2
+494
5/29
+401
3/29
+377
9/29
+5100
5/29
+329
7/29
(D)
TCAM1P-I
TCAM1P-II
TCAM1P-III
GSP3-1
GSP4-1
2/10
+10555
8/10
+10837
(E)
TCAM1P-I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

Fig. 2


(A)
GC-2spd(ts)
Hepa1-6
HEK293T
Relative luciferase activity
Relative luciferase activity
Relative luciferase activity
0
10
20
30
0
5
10
0
20
40
Luc
Pro
Luc
TCAM1P-Pro-luc
hCNS1
Pro
Luc
hCNS1-Pro-luc
Pro
Luc
hCNS1
Pro-luc-hCNS1
Pro
Luc
hCNS1
Pro-luc-rev hCNS1
**

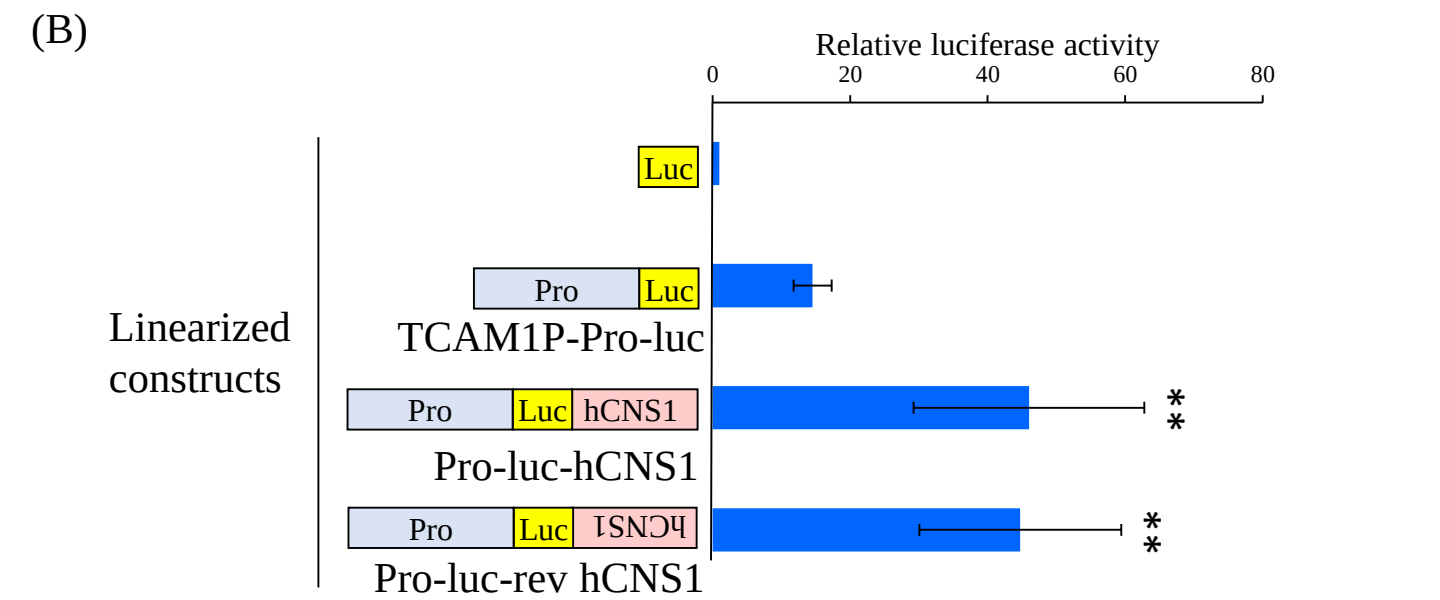

(B)
Relative luciferase activity
0
20
40
60
80
Linearized constructs
Luc
Pro
Luc
TCAM1P-Pro-luc
Pro
Luc
hCNS1
Pro-luc-hCNS1
Pro
Luc
hCNS1
Pro-luc-rev hCNS1
**

Fig. 3

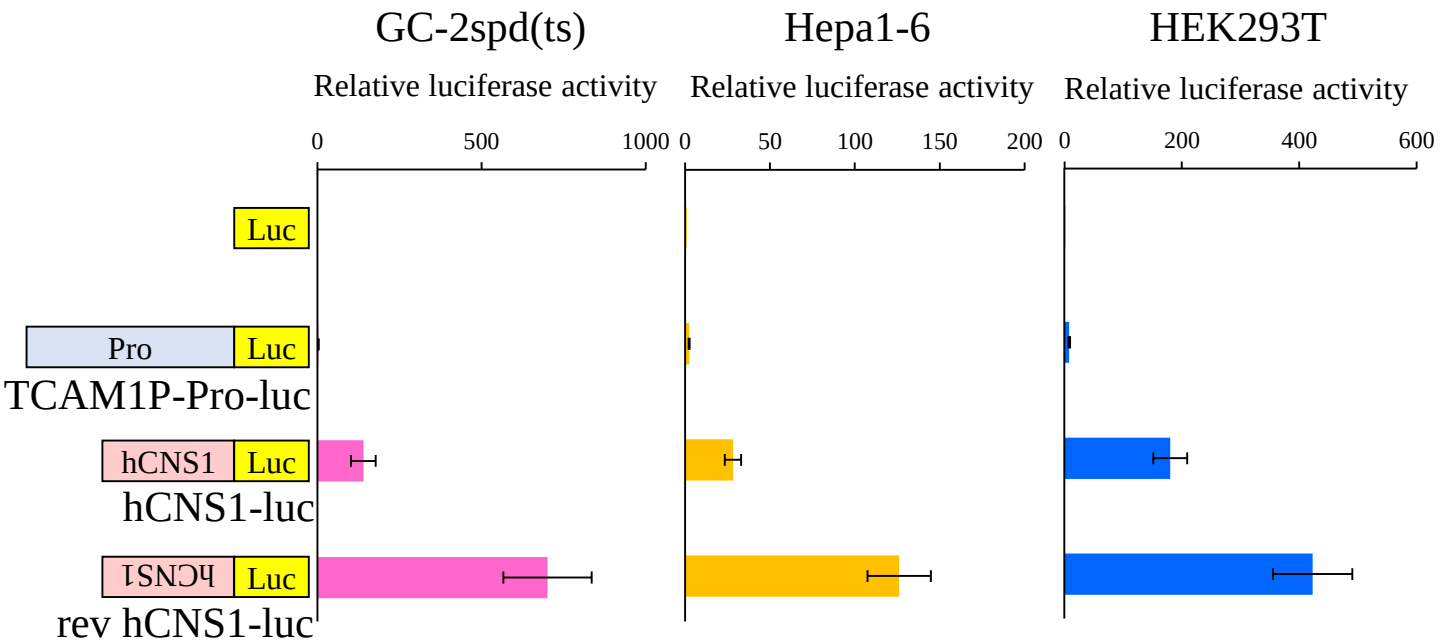

GC-2spd(ts)
Hepa1-6
HEK293T
Relative luciferase activity
Relative luciferase activity
Relative luciferase activity
0
500
1000
0
50
100
150
200
0
200
400
600
Luc
Pro
Luc
TCAM1P-Pro-luc
hCNS1
Luc
hCNS1-luc
ɹSNƆɥ
Luc
rev hCNS1-luc

Fig. 4

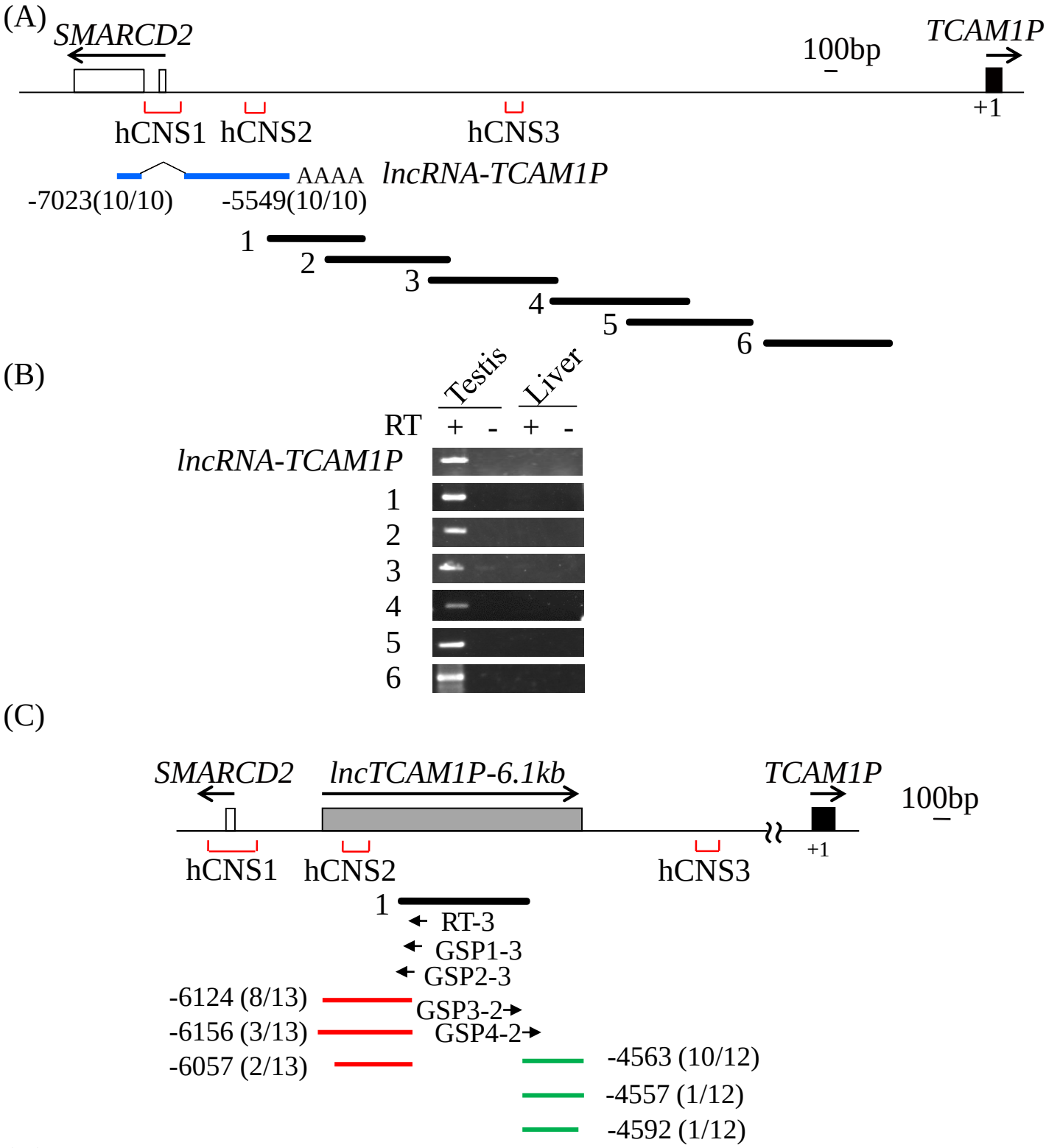

(A)
SMARCD2
TCAM1P
100bp
+1
hCNS1
hCNS2
hCNS3
AAAA
lncRNA-TCAM1P
-7023(10/10)
-5549(10/10)
1
2
3
4
5
6
(B)
Testis
Liver
RT + - + -
lncRNA-TCAM1P
1
2
3
4
5
6
(C)
SMARCD2
lncTCAM1P-6.1kb
TCAM1P
100bp
+1
hCNS1
hCNS2
hCNS3
1
RT-3
GSP1-3
GSP2-3
-6124 (8/13)
GSP3-2
-6156 (3/13)
GSP4-2
-6057 (2/13)
-4563 (10/12)
-4557 (1/12)
-4592 (1/12)
(D)
SMARCD2
lncTCAM1P-5.3kb
TCAM1P
100bp
+1
hCNS1
hCNS2
hCNS3
2
RT-4
GSP1-4
GSP2-4
-5339 (9/11)
-5338 (2/11)

Fig. 5

(A)

	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					

(B)

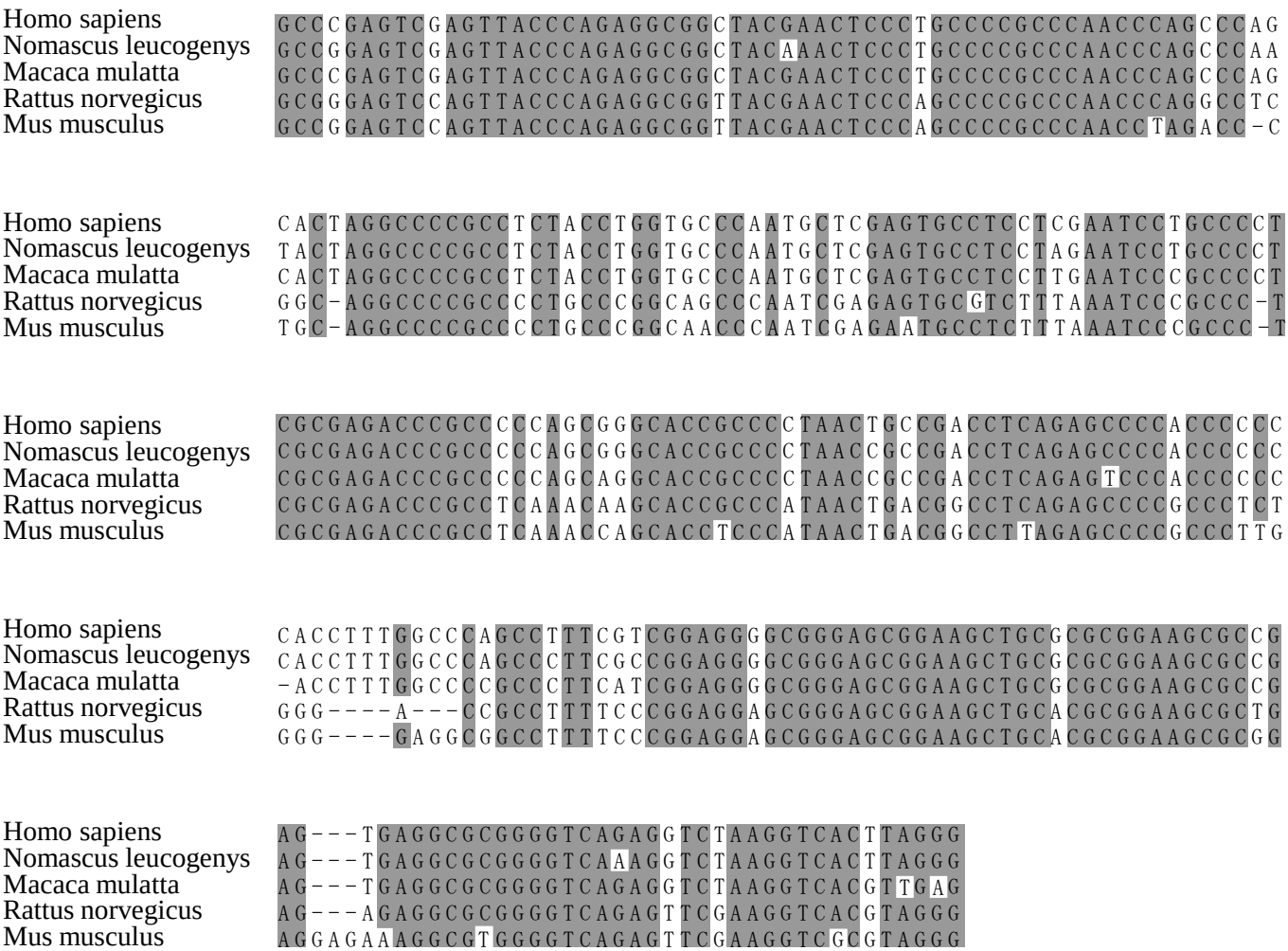


Fig. 6

Highlights

Human *TCAM1P* gene is a unitary pseudogene.

Cell type-specificity of *TCAM1P* in the testis is different from mouse *Tcam1*.

hCNS1 may enhance *TCAM1P* expression by a different mechanism from mouse.

hCNS1 potentially functions as a dual promoter–enhancer in human genome.

This provides an example of changes in gene regulation followed by pseudogenization.