Molecular cloning of chicken *TET* family genes and role of chicken *TET1* in erythropoiesis

Yuya Okuzaki, Hidenori Kaneoka*, Ken-ichi Nishijima, Seitaro Murakami, Yuki Ozawa, and Shinji Iijima

Department of Biomolecular Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan

*Corresponding author. Address: Department of Biomolecular Engineering, Graduate School of Engineering, Nagoya University, Chikusa-ku, Nagoya, 464-8603, Japan

Fax: +81 52 789 3221

E-mail: kaneoka@chembio.nagoya-u.ac.jp

ABSTRACT

Ten-eleven translocation (TET) methylcytosine dioxygenase has potential as an active eraser to regulate the genomic DNA methylation status. We herein cloned chicken *TET* (c*TET)* family genes, and confirmed their functions. Quantitative reverse-transcription PCR showed that c*TET1* was strongly expressed in erythrocytes throughout development. This c*TET1* expression pattern, together with the results of

methylated or hydroxymethylated DNA immunoprecipitation, suggests that cTET1 contributes to demethylation around the promoter region of the definitive-type β-globin gene *βA* in erythroid cells. The knockdown of c*TET1* in T2ECs chicken erythroid progenitor cells suppressed the induction of *βA* expression under differentiation conditions. These results suggest that cTET1 plays an important role in erythroid cell differentiation.

Keywords

Chicken; Ten-eleven translocation (TET); 5-Hydroxymethylcytosine (5hmC); β-globin; Erythropoiesis

1. Introduction

The methylation of cytosine residues in genomic DNA is one of the most important epigenetic modifications together with histone modifications, which regulate embryogenesis and cell differentiation through the modulation of gene expression [1]. DNA methylation has been extensively studied and two classes of DNA methyltransferases (DNMTs) (maintenance and *de novo*) have been identified [1]; [2]. In contrast, the enzymes involved in demethylation remain elusive. The hydroxylation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), and its further oxidation to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) were recently

shown to be catalyzed by ten-eleven translocation (TET) proteins [3]; [4]. TET proteins have been suggested to play passive and active roles in the demethylation of DNA. Maintenance-type DNA methyltransferase (DNMT1) modifies a newly synthesized unmethylated strand of hemimethylated CpG during DNA replication, but does not recognize 5hmC, which leads to passive demethylation [5]. On the other hand, 5fC and 5caC are repaired through a base excision repair pathway to unmodified cytosine, which results in active demethylation [6]; [7].

In the early stage of mouse development, mouse *TET1* (m*Tet1*) and *2* are expressed in ES cells, the inner cell mass, and primordial germ cells, while m*Tet3* is expressed in zygotes [8]; [9]; [10]. TETs maintain pluripotency and control epigenetic reprogramming in these cells. TETs are also involved in the late stage of development and cell differentiation. TET1 regulates neural development [11] and maintains intestinal stem cells [12]. TET2 is broadly expressed in hematopoietic cells and regulates hematopoiesis [13]; [14]; [15]; [16]; [17]; [18], and TET3 is also involved in hematopoiesis and neural development in mammals. However, limited information is currently available on the functions of TETs in the chicken.

Chicken α- and β-globin genes have been well-studied in relation to the epigenetic modifications such as histone modification and DNA methylation [19]; [20]; [21]; [22]; [23]; [24]. The chicken α-globin gene cluster consists of three functional genes: π (*HBZ*), αD (*HBM*), and αA (*HBA1*), while the β-globin gene cluster has four: ρ (*HBG1*),

βH (*HBE1*), *βA* (*HBG2*), and *ε* (*HBE*). The expression patterns of these globin genes are known to change during development [25]. In β-globin genes (*ρ* and *βA*), a clear inverse relationship has been reported between the methylation states of promoters and gene expression [21]; [22].

In the present study, we cloned chicken *TET* family genes, and detected *in vivo* activities. We demonstrated that chicken *TET1* was prominently expressed in erythrocytes and may be involved in β-globin gene regulation.

2. Materials and methods

2.1. Cell lines

293FT, HeLa, and DF-1 cells were maintained in DMEM high glucose (Sigma-Aldrich) containing 10% fetal bovine serum (FBS) (Biological Industries), 100 U/ml penicillin G, and 100 μg/ml streptomycin (P&S). DT40 cells were maintained in RPMI1640 (Nissui Pharmaceutical) containing 7.5% FBS, 2.5% normal chicken serum (Thermo Fisher Scientific), and P&S.

2.2. Chickens and eggs

Chickens (White Leghorn) and fertilized eggs were obtained from Nisseiken or Takeuchi Farm. Organs were minced and washed with phosphate-buffered saline, and

RNA was isolated. Blood samples were obtained from embryos as well as newly hatched and adult chickens using a fine glass needle or syringe. Leukocytes and erythrocytes were separated using Histodenz (Sigma) density gradient centrifugation. Progenitor of erythroid cells (T2ECs (TGF-α / TGF-β-induced erythroid cells)) were established from the bone marrow cells of newborn chicks as described previously [26]. T2ECs were maintained in α-MEM (Invitrogen) containing 10% FBS, 1% normal chicken serum, P&S, 0.1 mM β-mercaptoethanol (Wako), 5 ng/ml human TGF-α (PeproTech), 1 ng/ml human TGF-β1 (PeproTech), and 1 μM dexamethasone (Dex) (Wako). In order to induce differentiation, TGF-α, TGF-β1, and Dex were removed from the culture medium, while 10 ng/ml insulin (Wako) and 100 IU/ml Epojin (Chugai Pharmaceutical) were added at the indicated times. All animal experiments were performed according to the ethical guidelines for animal experimentation of Nagoya University.

2.3. Cloning of TET genes

In the cloning of chicken *TET1*, *2*, *and 3* (*cTET1*, *2*, and *3*) full-length (FL) genes and DNA fragments corresponding to the catalytic domain (CD) (cTET1: 1486–2244 amino acids (AA), cTET2: 1101–1955 AA, and cTET3: 568–1524 AA), DNAs were amplified by PCR using KOD-Fx-Neo (Toyobo) with the chicken blastoderm or brain cDNA as templates. The sequences of primers used were shown in Table. S1. Primers

were designed following the NCBI database of *cTET1*: XM_015278732, c*TET2*: NM_001277794, and c*TET3*: XM_015297468.1. Amplified DNA fragments were cloned into pFLAG-CMV2 (Sigma). m*Tet1* CD (1367–2039 AA) was amplified by PCR with the cDNA of NIH3T3 mouse fibroblast cells, and cloned into pFLAG-CMV2. Primers were designed following the NCBI database of m*Tet1*: NM_001253857.

2.4. In vivo quantification of 5hmC

293FT cells were seeded at 5 x 10^5 cells per 35-mm dish. After 24 hr, cells were transfected with 4 μg of TETs CD expression plasmids using Lipofectamine 2000 (Invitrogen) according to the supplier's recommendations. Cells were cultured for 24 hr. Genomic DNAs were purified by MagExtractor -Genome- (Toyobo). In the dot blot assay, genomic DNAs were spotted on Hybond-XL membranes (GE Healthcare), and then fixed by baking at 80°C for 3 hr. 5hmC was detected by immunoblotting with a rabbit anti-5hmC antibody (Active Motif) and goat anti-rabbit IgG-HRP (Santa Cruz Biotechnology). FLAG-TETs and β-ACTIN were detected by Western blotting with a mouse anti-FLAG antibody (Wako) or mouse anti-β-ACTIN antibody (Medical & Biological Laboratories) and goat anti-mouse IgG-HRP (Santa Cruz Biotechnology).

2.5. Quantitative PCR (qPCR) and quantitative reverse-transcription PCR (qRT-PCR)

Total RNA extraction, cDNA synthesis, and qPCR were performed as previously

described [27]. Primers for PCR were shown in Table. S1.

2.6. Methylated DNA immunoprecipitation (MeDIP) and hydroxymethylated DNA immunoprecipitation (hMeDIP)

Genomic DNAs from chicken erythrocytes, T2ECs, and DT40 cells were purified by the QIAamp DNA Mini Kit (QIAGEN). MeDIP and hMeDIP were then performed as previously described [8].

2.7. Knockdown of cTET1

Regarding knockdown, c*TET1*-specific (Hokkaido System Science) or control (siTrio; B-Bridge International Inc.) siRNA (150 pmol each) was electroporated to 1 x 10^6 cells of T2ECs using the NEPA21 electroporator (NEPAGENE). Cells were then cultured under standard conditions for 24 hr, and differentiation was induced for 48 hr. The sequences of siRNAs were shown in Table. S1.

3. Results

3.1. Cloning of chicken TET1, 2, and 3, and detection of biological activity

Three *Tet* family genes have been reported in the mouse, and all of their protein products have been shown to catalyze the oxidization of methylated DNA [8]. c*TET1*, *2*,

and *3* have also been predicted for the chicken in the NCBI database. We cloned them as N-terminal FLAG fusions and confirmed their nucleotide sequences. The cloned c*TET1*, *2*, and *3* genes encoded proteins with 2244 AA, 1955 AA, and 1524 AA, respectively (Fig. 1A). Some nucleotide replacements were observed in all three c*TET*s, whereas no amino acid mutations were noted in c*TET2*. In c*TET1*, lysine 1832 was changed to proline. c*TET3* had three single-nucleotide insertions (C was inserted between nucleotides 550 and 551, G was inserted between 1292 and 1293, and C was inserted between 1363 and 1364 of the *cTET3* sequence shown in the NCBI database) (Fig. S1), which introduced a frameshift and resulted in the coding of a unique protein. Since the amino acid sequences shown in the present study well matched quail TET3 in the NCBI database (XP_015738510.1), we assumed that this clone was functional, at least in our chicken strain. cTET1 contained the CXXC zinc finger domain near the N-terminal portion and Cys-rich region just beside the catalytic domain containing CD1 and 2. cTET2 and 3 also contained the Cys-rich regions, CD1 and 2, but lacked the CXXC zinc finger domain (Fig. 1A). In other species, TET3 has the CXXC domain, which binds unmethylated and methylated CpGs [28].

Fragments of the active domain have been used to detect the activities of TETs [8]. A part of TET protein containing the Cys-rich regions, CD1 and 2 was fused to the FLAG tag and expressed in 293FT cells. The accumulation of 5hmC was observed in c*TET1*-, c*TET3*-, and m*Tet1*-transfected cells, but not in c*TET2* cells (Fig. 1B). We

confirmed the expression of TETs, except for cTET2, by Western blotting with the anti-FLAG antibody. The cTET2 protein was not detected at the expected molecular mass (Fig. 1C). Similar results were obtained when cTET2 was expressed in HeLa cells or DF-1 chicken fibroblast cells (data not shown). A previous study reported that the TET2 catalytic domain was ubiquitinated and degraded by the ubiquitin-proteasome pathway in human cells [29]. Hence, cells were cultured in the presence of the proteasome inhibitor MG132 and genomic DNA was extracted. We detected the accumulation of 5hmC by dot blotting and the cTET2 protein by Western blotting (Fig. S2). Overall, these results revealed that cTET1, 2, and 3 were catalytically active, and converted 5mC to 5hmC in an *in vivo* assay.

3.2. Analyses of TETs expression in chickens

In chicken embryos, the accumulation of 5hmC was not observed in pre-primitive streak embryos, whereas a strong 5hmC signal was detected after 6-somite stage embryos (28-hr embryos) [30]. Thus, we measured the expression levels of c*TET*s in embryos under different developmental stages and adult chickens. In blastoderms (EK stage X embryos [31]), the expression of all c*TET*s was relatively weak (Fig. 2A). Certain expression levels of c*TET*s were observed in embryos (5-, 10-, and 15-day-old (-d)) and adults; however, the extent of expression differed between organs (Fig. S3 and Fig. 2B-D). In adult organs, all three c*TET*s were more strongly expressed in the lung,

spleen, intestine, oviduct, and leukocytes than in the other organs tested. Unexpectedly, c*TET1* was prominently expressed in adult erythrocytes. Thus, we analyzed the expression levels of c*TET*s in erythrocytes from various developmental stages. The expression level of c*TET1* was increased through development, particularly after 5 days, whereas those of c*TET2* and *3* were low and not changed (Fig. 2E).

3.3. 5mC and 5hmC in the β-globin gene cluster

Our qRT-PCR analyses indicated that c*TET1* was strongly expressed in erythrocytes. We then analyzed the physiological role of cTET1 in erythrocytes. The expression of β-globin genes is known to change during embryogenesis. In the chicken, primitive erythroid cells from 2-d to 5-d embryos express embryonic β-globin genes (ρ and ε). After 5-d embryos, definitive erythroid cells, which express fetal and adult β-globin genes (βH and βA), gradually increased, whereas embryonic-type genes were silenced in these erythroid cells [25]. We compared β-globin expression between 5-d and 8-d embryonic erythrocytes, and confirmed that embryonic-type globin transcripts (ρ and ε) decreased, whereas fetal- and adult-type transcripts (βH and βA) increased as reported previously (Fig. 3A).

A relationship was previously reported between hydroxymethylation and β-globin gene expression in baboon models [16]. In the chicken, the accumulation of 5hmC in αD was also demonstrated during the differentiation of erythroid cells [24]. However,

alterations in and the function of the hydroxymethylation of β-globin genes during chicken erythropoiesis have not yet been examined, although the expression of *ρ* and *βA* is known to be regulated by DNA methylation during development [21]; [22]. In order to analyze the hydroxymethylation of *ρ* and *βA* promoters, we conducted MeDIP and hMeDIP with chicken erythrocytes from various developmental stages. Genomic DNAs were isolated from erythrocytes and the chicken pre-B cell line DT40. DT40 cells were used as a non-erythroid control, which weakly expressed c*TET*s (Fig. S4). 5mC and 5hmC modifications to the *ρ* and *βA* proximal promoter regions within 500 bp of the transcriptional start sites were analyzed by immunoprecipitation with anti-5mC and anti-5hmC antibodies, respectively. As expected, 5mC was not detected in the *ρ* promoter region of 5-d erythrocytes, whereas its content increased in 8-d erythrocytes and was maintained at a certain level in 10-d embryos and adult chickens (Fig. 3B). This change appeared to parallel decreases in transcription. Increased levels of 5mC were also observed with DT40, which did not express the *ρ* gene. Unexpectedly, 5hmC increased between 5-d and 8-d erythrocytes. After that, it gradually decreased as differentiation progressed; however, a certain level of 5hmC remained in adult erythrocytes (Fig. 3C). On the other hand, 5hmC was not detected in the *ρ* promoter of DT40 cells. Regarding the *βA* promoter, 5mC was accumulated in 5-d erythrocytes, markedly decreased between 5-d and 8-d, and was almost undetectable in adult erythrocytes. A certain level of 5mC was detected in DT40 cells. 5hmC levels were

high in 5-d and 8-d erythrocytes, and then gradually decreased during development. We also analyzed HS4, which is known as a strong insulator and is hypomethylated, even in non-erythroid cells [32]; [33]. In this region, 5mC was maintained at low levels under all developmental stages of erythrocytes and DT40 cells, and the levels of 5hmC were also low. Overall, the status of 5mC appeared to be closely related with the expression patterns of the *ρ* and *βA* genes; however, the reason why 5hmC modifications in the *ρ* gene transiently increased in 8-d and 10-d erythrocytes remains elusive.

3.4. cTET1 promotes the expression of βA in differentiated T2ECs

In order to demonstrate that the demethylation of the *βA* promoter by cTET1 is associated with the transcriptional activation of *βA*, we established T2ECs from newborn chick bone marrow cells [26]. We confirmed the differentiation of T2ECs by detecting hemoglobin with tetramethylbenzidine staining. After a 7-day cultivation in differentiation medium, the intensity of staining significantly increased, which showed that T2ECs successfully differentiated (Fig. S5A). We also examined the expression of α- and β-globin genes by qRT-PCR. By inducing differentiation, the expression of adult-type globin genes (*αA* and *βA*) gradually increased, whereas that of embryonic-type globin genes (*π* and *ρ*) did not change (Fig. 4A and Fig. S5B). In addition, the expression of c*TET1* increased, whereas that of c*TET2* and c*TET3* remained unchanged (Fig. 4B). We then examined alterations in 5mC and 5hmC in the

βA promoter region before and after the induction of differentiation. Genomic DNAs were purified from T2ECs, followed by MeDIP or hMeDIP. 5mC in the *βA* promoter region significantly decreased after differentiation (Fig. 4C), while 5hmC significantly increased (Fig. 4D). These results suggested that the *βA* promoter was demethylated, possibly through cTET1 activity, during differentiation.

We then prepared two different siRNAs for the c*TET1* gene and introduced them into T2ECs. Despite extensive trials under various conditions, the efficiency of the knockdown (at approximately 50%) was not sufficiently high because of low transfection efficiency. The *βA* transcript was decreased by this knockdown with either of the two different siRNAs (Fig. 4E), suggesting that cTET1 was involved in *βA* gene activation, possibly through the demethylation of 5mC.

4. Discussion

In the present study, we cloned the chicken *TET* family genes, c*TET1*, *2*, and *3*, and showed the accumulation of 5hmC by the overexpression of these genes *in vivo* (Fig. 1B and Fig. S2A). Previous studies reported that the enzyme activity of TETs is conserved through vertebrates [8]; [15]; [28]; [30]. The present results suggest that TETs activity is also preserved, and cTETs may be involved in the demethylation of 5mC through 5hmC as an intermediate in the chicken.

qRT-PCR experiments revealed that the expression of *cTETs* was relatively low in the blastoderm (Fig. 2A). This result is consistent with previous findings showing that 5hmC was not observed in this stage of chicken embryos [30]. Although extensive *in vivo* analyses are needed in order to obtain conclusive results, the present results suggest that cTETs are not required for the early stage development of chickens, in contrast to mammals.

c*TET1* was expressed at a higher level in erythrocytes, and its expression increased between 5-d and 8-d embryos when primitive erythroid cells change to definitive erythroid cells (Fig. 2E). MeDIP and hMeDIP revealed that 5mC rapidly decreased between 5-d and 8-d embryos, then almost disappeared in adult erythrocytes in the *βA* promoter, as reported previously (Fig. 3B) [20]; [22]. The transient accumulation of 5hmC was also simultaneously observed (Fig. 3C). We measured the expression of *DNMT*s in 3-d and 15-d embryonic erythrocytes, and found that they slightly increased (data not shown), suggesting that reductions in 5mC in the *βA* promoter are not due to passive demethylation by decrements in *DNMT1* expression. Similar changes in 5mC and 5hmC were observed in the *in vitro* differentiation of T2ECs (Fig. 4C and D). These results, together with knockdown experiments (Fig. 4E), suggest that cTET1 regulates the expression of *βA* during erythropoiesis by directly demethylating the *βA* promoter sequence in the chicken. However, we cannot rule out the possibility that cTET1 controls *βA* expression by other mechanisms. For example, zebrafish Tet2 regulates the

expression of the erythroid lineage-specific transcription factors *scl*, *gata-1*, and *cmyb* through the demethylation of promoters of these genes, and the knockout of *tet2* leads to defects in erythropoiesis [15]. In humans, a TET2 deficiency has been shown to disrupt the 5hmC patterns of transcription factor-binding sites and hampers erythroid differentiation [18].

Our experiments indicated that c*TET1* is a major *TET* species in chicken erythroid cells and involved in regulating erythropoiesis. However, in other species such as the zebrafish, baboon, and humans, *TET2* and *3* mainly contribute to erythropoiesis [15]; [16]; [17]; [18]. Thus, it currently remains unclear why c*TET1*, but not c*TET2* or *3* is mainly expressed in chicken erythroid cells and involved in the control of globin expression.

The chicken β-globin locus has been studied extensively as a model of transcriptional regulation, epigenetic modifications, and chromatin organization. This study initially showed that active demethylation by cTET1 contributes to β-globin expression.

Conflicts of Interest

The authors declare no conflicts of interest associated with this manuscript.

Acknowledgements

This work was partly supported by Grant-in-Aid for Scientific Research (KAKENHI) on Innovative Areas, 'Mechanisms regulating gamete formation in animals' [Grant Number 16H01253].

References

[1] Z.D. Smith, A. Meissner, DNA methylation: roles in mammalian development, Nat. Rev. Genet. 14 (2013) 204–20.

[2] M.G. Goll, T.H. Bestor, Eukaryotic cytosine methyltransferases, Annu. Rev. Biochem. 74 (2005) 481–514.

[3] M. Tahiliani, K.P. Koh, Y. Shen, et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, Science 324 (2009) 930–5.

[4] S. Ito, L. Shen, Q. Dai, et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science 333 (2011) 1300–3.

[5] V. Valinluck, L.C. Sowers, Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1, Cancer Res. 67 (2007) 946–50.

[6] A. Maiti, A.C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites, J. Biol. Chem. 286 (2011) 35334–8.

[7] Y.-F. He, B.-Z. Li, Z. Li, et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, Science 333 (2011) 1303–7.

[8] S. Ito, A.C. D'Alessio, O. V. Taranova, et al., Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification, Nature 466 (2010) 1129–33.

[9] S. Yamaguchi, K. Hong, R. Liu, et al., Tet1 controls meiosis by regulating meiotic gene expression, Nature 492 (2012) 443–7.

[10] T.P. Gu, F. Guo, H. Yang, et al., The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes, Nature 477 (2011) 606–10.

[11] M. Santiago, C. Antunes, M. Guedes, et al., TET enzymes and DNA hydroxymethylation in neural development and function - How critical are they?, Genomics 104 (2014) 334–40.

[12] R. Kim, K.L. Sheaffer, I. Choi, et al., Epigenetic regulation of intestinal stem cells by Tet1-mediated DNA hydroxymethylation, Genes Dev. 30 (2016) 2433-42.

[13] S. Montagner, C. Leoni, S. Emming, et al., TET2 regulates mast cell differentiation and proliferation through catalytic and non-catalytic activities, Cell Rep. 15 (2016) 1566–79.

[14] M. Ko, J. An, W.A. Pastor, et al., TET proteins and 5-methylcytosine oxidation in hematological cancers, Immunol. Rev. 263 (2015) 6–21.

[15] L. Ge, R.-P. Zhang, F. Wan, et al., TET2 plays an essential role in erythropoiesis by regulating lineage-specific genes via DNA oxidative demethylation in a zebrafish model, Mol. Cell. Biol. 34 (2014) 989–1002.

[16] M.A. Ruiz, A. Rivers, V. Ibanez, et al., Hydroxymethylcytosine and demethylation of the γ-globin gene promoter during erythroid differentiation, Epigenetics 10 (2015) 397–407.

[17] H. Yan, Y. Wang, X. Qu, J. Li, et al., Distinct roles for TET family proteins in regulating human erythropoiesis, Blood 129 (2017) 2002-12.

[18] J. Madzo, H. Liu, A. Rodriguez, et al., Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis, Cell Rep. 6 (2014) 231–44.

[19] M.D. Litt, M. Simpson, M. Gaszner, et al., Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus, Science 293 (2001) 2453–5.

[20] J.D. Mcghee, G.D. Ginder, Specific DNA methylation sites in the vicinity of the chicken β-globin genes, Nature 280 (1979) 419–20.

[21] R. Singal, R. Ferris, J.A. Little, et al., Methylation of the minimal promoter of an embryonic globin gene silences transcription in primary erythroid cells, Proc. Natl. Acad. Sci. USA 94 (1997) 13724–9.

[22] K. Ramachandran, J. van Wert, G. Gopisetty, R. Singal, Developmentally regulated demethylase activity targeting the βA-globin gene in primary avian erythroid cells, Biochemistry 46 (2007) 3416–22.

[23] R. Singal, J.M. vanWert, L. Ferdinand, Methylation of α-type embryonic globin gene απ represses transcription in primary erythroid cells, Blood 100 (2002) 4217–22.

[24] C.V. Quezada, C.A. Canon, Y.F. Guzmán, et al., CTCF demarcates chicken embryonic α-globin gene autonomous silencing and contributes to adult stage-specific gene expression, Epigenetics 8 (2013) 827–38.

[25] G.A.P. Bruns, V.M. Ingram, Erythropoiesis in the developing chick embryo, Dev. Biol. 459 (1973) 455–9.

[26] O. Gandrillon, U. Schmidt, H. Beug, J. Samarut, TGF-β cooperates with TGF-α to induce the self-renewal of normal erythrocytic progenitors: Evidence for an autocrine mechanism, EMBO J. 18 (1999) 2764–81.

[27] Y. Okuzaki, S. Kidani, H. Kaneoka, et al., Characterization of chicken interferon-inducible transmembrane protein-10, Biosci. Biotechnol. Biochem. 81 (2017) 1–8.

[28] Y. Xu, C. Xu, A. Kato, et al., Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development, Cell 151 (2012) 1200–13.

[29] Y.W. Zhang, Z. Wang, W. Xie, et al., Acetylation enhances TET2 function in protecting against abnormal DNA methylation during oxidative stress, Mol. Cell 65 (2017) 323–35.

[30] R.D. Almeida, M. Loose, V. Sottile, et al., 5-hydroxymethyl-cytosine enrichment of non-committed cells is not a universal feature of vertebrate development, Epigenetics 7 (2012) 383–89.

[31] H.E. Giladi, S. Kochav, From cleavage to primitive streak formation: A complementary normal table and a new look at the first stages of the development of the chick. I. General morphology, Dev. Biol. 49 (1976) 321–37.

[32] J.H. Chung, M. Whiteley, G. Felsenfeld, A 5' element of the chicken β-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila, Cell 74 (1993) 505–14.

[33] M.N. Prioleau, P. Nony, M. Simpson, G. Felsenfeld, An insulator element and condensed chromatin region separate the chicken β-globin locus from an independently regulated erythroid-specific folate receptor gene, EMBO J. 18 (1999) 4035-48.

Figure legends

Fig. 1. Cloning of c*TET*s and measurement of biological activity

A: Domain structures of cTET1, 2, and 3. CXXC: CXXC-type Zn finger domain. Cys-rich & CD1: cysteine-rich region and catalytic domain 1. CD2: catalytic domain 2. B: Genomic DNAs were purified from 293FT cells that expressed FLAG-tagged TET CD, and 200 ng of DNAs were spotted on the membrane, followed by immunoblotting using an anti-5hmC antibody. C: The expression levels of each TETs CD were examined by Western blotting with anti-FLAG antibody. β-ACTIN was used as an internal control.

Fig. 2. Expression of c*TET1, 2, and 3*

A: RNAs from blastoderms were subjected to qRT-PCR. Expression levels of c*TET1*, *2*, and *3* are represented as relative expression levels against *GAPDH*. Data are the mean ± standard error of three different chickens. B-D: RNAs from adult chicken organs were subjected to qRT-PCR. Expression levels of c*TET1* (B), c*TET2* (C), and c*TET3* (D) are represented as relative expression levels against *GAPDH*. Data are the mean ± standard error of six different chickens (three males and three females), except for the testis, oviduct and ovary (N = 3), leukocytes and erythrocytes (N = 4). E: Blood samples from each stage (embryo, newborn, and adult chickens) were isolated, and erythrocytes were obtained by density gradient centrifugation. RNAs from purified erythrocytes were

subjected to qRT-PCR. The expression levels of c*TET1*, *2*, and *3* are represented as relative expression levels against *GAPDH*. Data are the mean ± standard error of four different chickens.

Fig. 3. MeDIP and hMeDIP of the β-globin locus in erythrocytes

A: RNAs were purified from erythrocytes of 5-d and 8-d embryos and subjected to qRT-PCR. The expression levels of ρ, βH, βA, and ε are represented as relative expression levels against *GAPDH*. Data are the mean ± standard error of four different samples. B, C: Genomic DNAs purified from each developmental stage of erythrocytes or DT40 cells were immunoprecipitated by an anti-5mC antibody (B) or anti-5hmC antibody (C). Precipitated DNAs were subjected to qPCR. Data are the mean ± standard error of three (B) or four (C) different samples.

Fig. 4. cTET1 promotes the expression of the βA gene during erythroid cell differentiation

A, B: RNAs from T2ECs, which were differentiated for various periods by Epojin (EPO), were subjected to qRT-PCR. The expression levels of ρ and βA (A) as well as *TET1*, *2*, and *3* (B) are represented as relative expression levels of *GAPDH*. Data are the mean ± standard error of six different samples. C, D: Genomic DNAs purified from T2ECs, differentiated for 48 hr (EPO+) or an undifferentiated control (EPO-), were

immunoprecipitated by an anti-5mC antibody (C) or anti-5hmC antibody (D). Precipitated DNAs were subjected to qPCR. Data are the mean ± standard error of four different samples. * and ** indicate significant differences by the Student's *t*-test ($p<0.05$ and $p<0.01$, respectively). E: siRNA for c*TET1* was electroporated to T2ECs. The expression levels of c*TET1*, *2*, *3*, and *βA* were examined by qRT-PCR. Expression levels in control siRNA-treated cells are set as 1. Data are the mean ± standard error of four different samples. * indicates significant differences by the Student's *t*-test ($p<0.05$); NS, not significant.

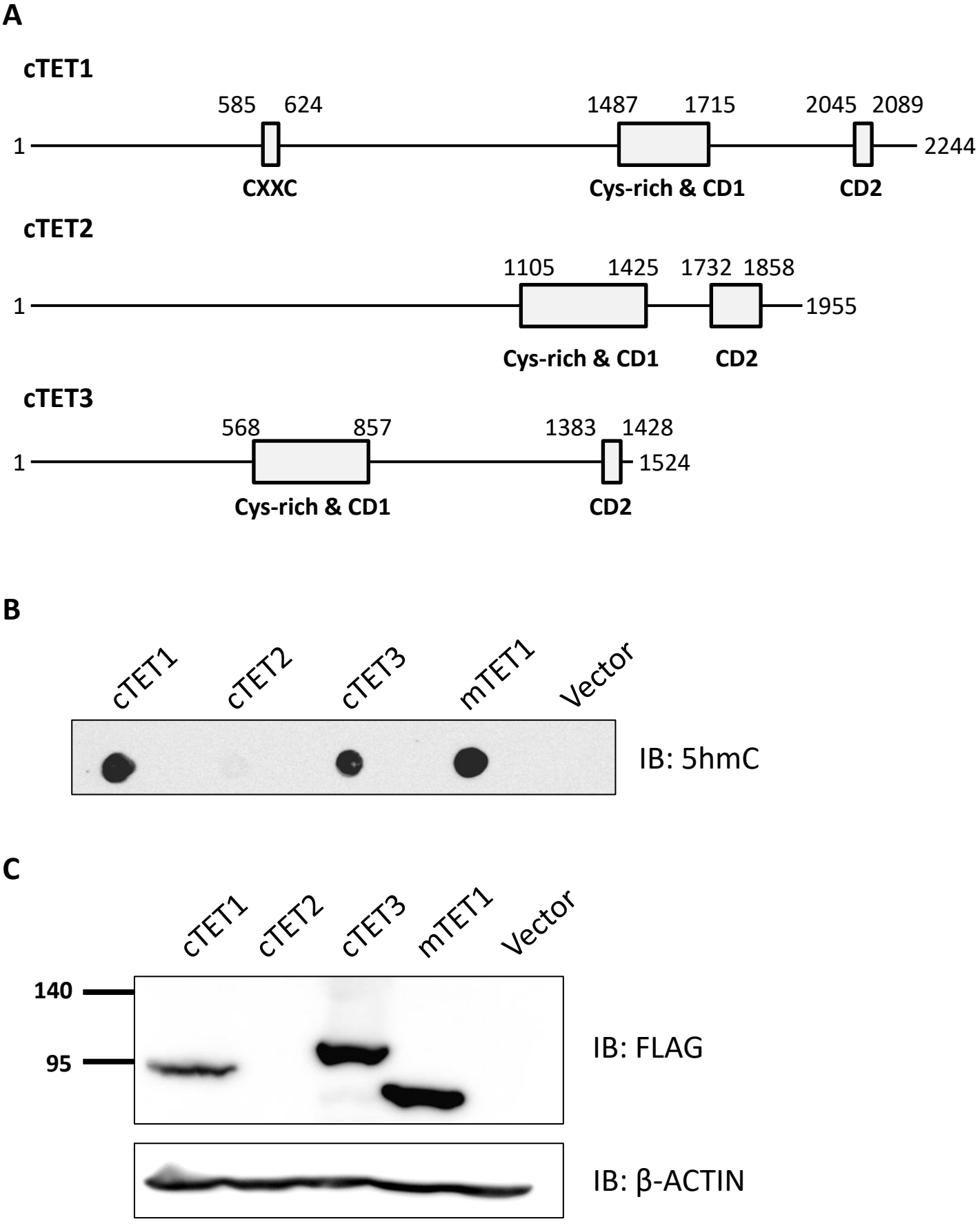

Fig. 1. Okuzaki et al.

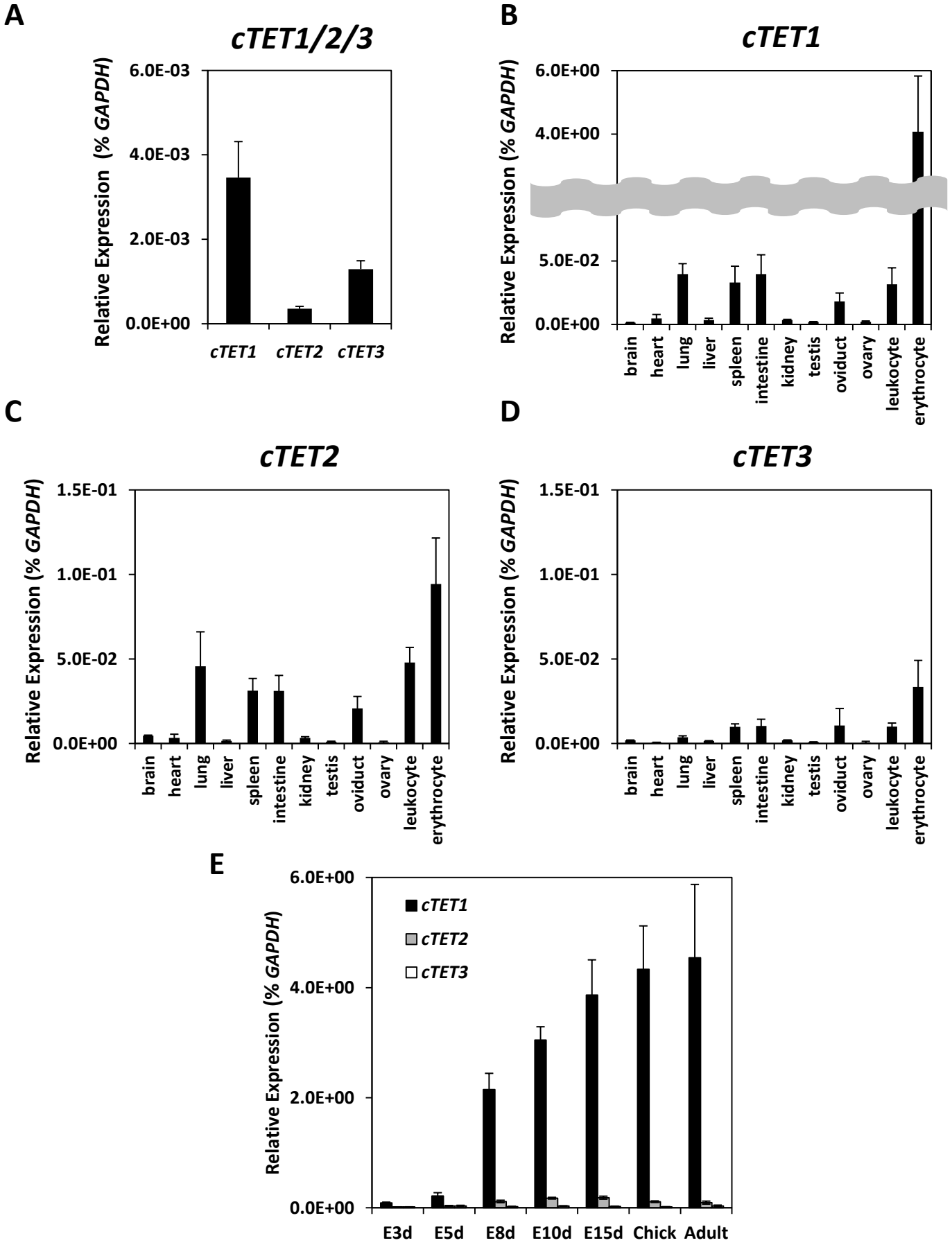

Fig. 2. Okuzaki et al.

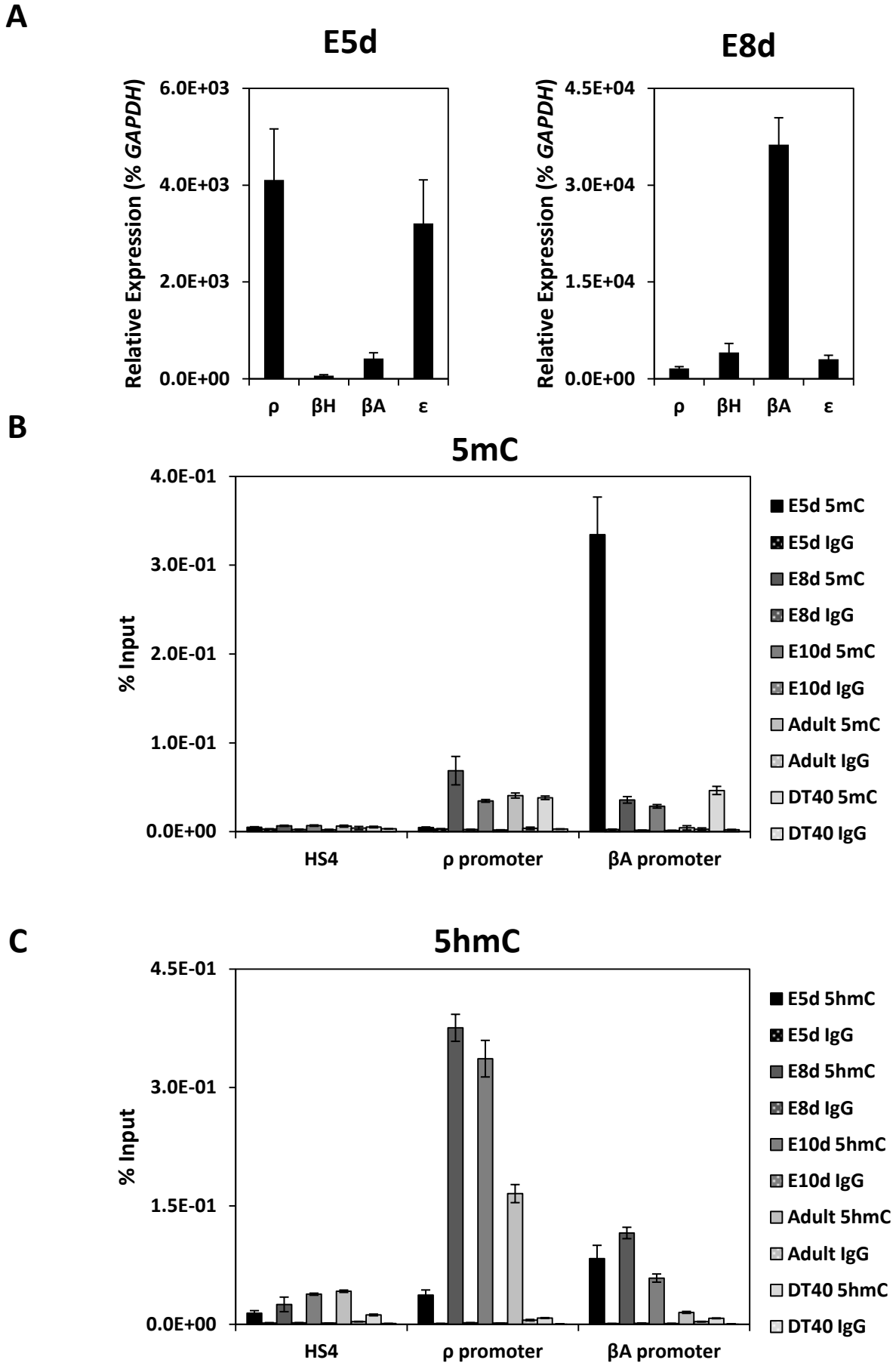

Fig. 3. Okuzaki et al.

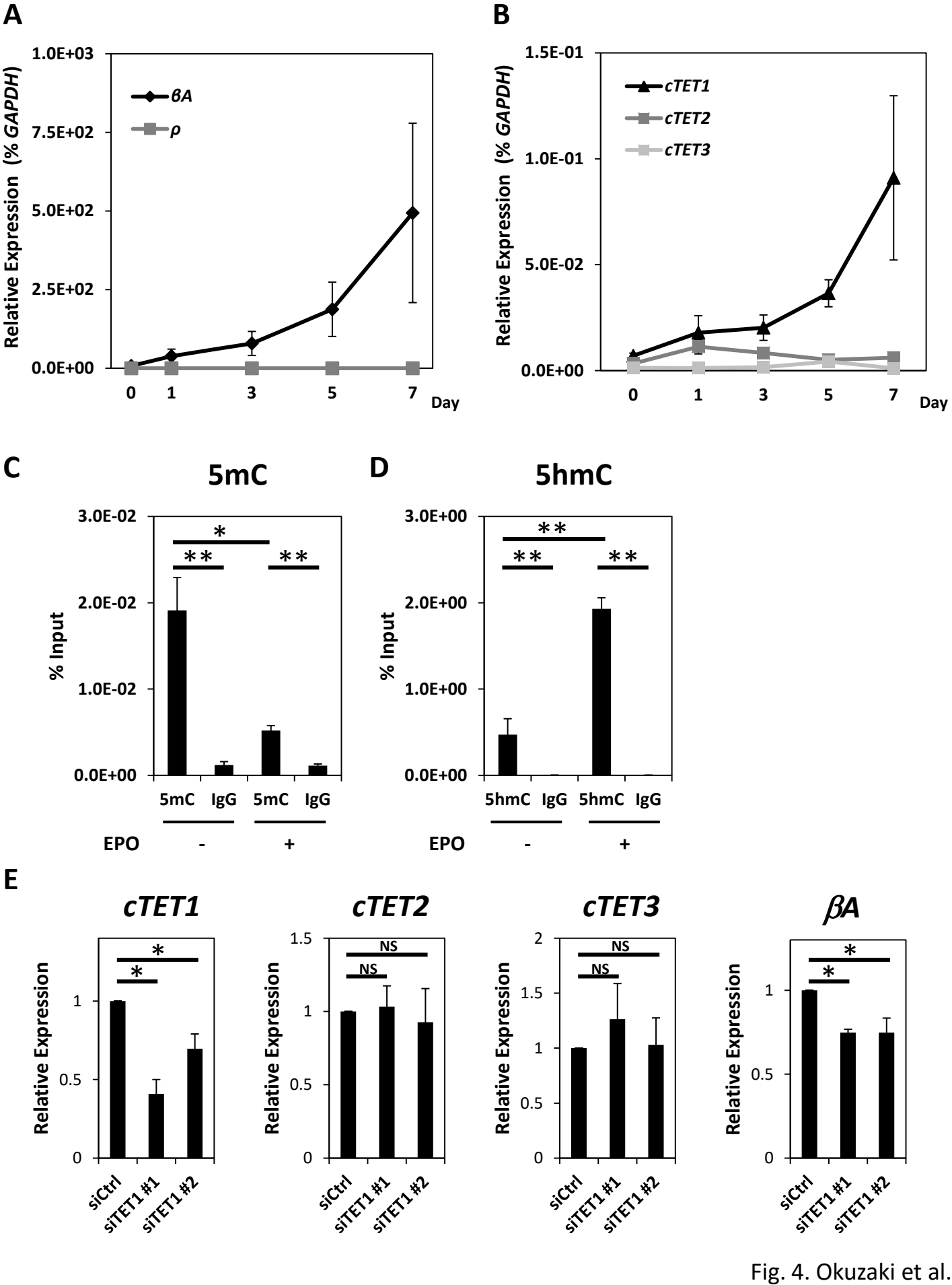

Fig. 4. Okuzaki et al.

A

```
Cloned          1    ATGGCTCACCACGCAAGGCCTTCCAGATTAGCAAAAAAGGAAGAGATAGGCAAAAGAAAAACTAACCAGGGTAAAAAGAAATCTAGCCAAGTGAGAAAGA  100
                     ********************************************************************************************************
NCBI database   1    ATGGCTCACCACGCAAGGCCTTCCAGATTAGCAAAAAAGGAAGAGATAGGCAAAAGAAAAACTAACCAGGGTAAAAAGAAATCTAGCCAAGTGAGAAAGA  100

              101    AAACCACAAAGACTTCCACAAAAGCTGTTAGTTCTGGAAAAGGCAAAAAGCCGGCACAAGAAAAAGATGTTAAGAAAAAGCAGCAAGAAAAGAAACCAAT  200
                     ********************************************************************************************************
              101    AAACCACAAAGACTTCCACAAAAGCTGTTAGTTCTGGAAAAGGCAAAAAGCCGGCACAAGAAAAAGATGTTAAGAAAAAGCAGCAAGAAAAGAAACCAAT  200

              201    CATGAGCTCTTCAGGTAGATTTCTCAGGAGCAGTGTCACCAGAACCTTGTCTGGAACACCTTGGGTGAGTTTGGAGAAAGCTGACAATATCCTCTTTCAT  300
                     ********************************************************************************************************
              201    CATGAGCTCTTCAGGTAGATTTCTCAGGAGCAGTGTCACCAGAACCTTGTCTGGAACACCTTGGGTGAGTTTGGAGAAAGCTGACAATATCCTCTTTCAT  300

              301    AGCCAGGAAGCTTTCAACATCAATGGTTTTACAATGTCTCTGCGTAACCGATCGTTCTCCCGTAGATTTTCTCAAGCTCCAACAATTGCAAAACCCAGAA  400
                     ********************************************************************************************************
              301    AGCCAGGAAGCTTTCAACATCAATGGTTTTACAATGTCTCTGCGTAACCGATCGTTCTCCCGTAGATTTTCTCAAGCTCCAACAATTGCAAAACCCAGAA  400

              401    AAGCTGCAGCACAAAAAAAGCTAGAAACAAAGGAAAAACATGAACAAGCTCTGGCAGTAGTAGAAGAAAAAAATGTTGAAGCATCTGAAAAAGTGTTGGA  500
                     ********************************************************************************************************
              401    AAGCTGCAGCACAAAAAAAGCTAGAAACAAAGGAAAAACATGAACAAGCTCTGGCAGTAGTAGAAGAAAAAAATGTTGAAGCATCTGAAAAAGTGTTGGA  500

              501    ACAGAATTTAGCACAAGATGAATTAGCTTCTCTGCCTGTAGAGGATATTCCATGCAGTGTAGCTGAAGAGCCTGCTGCCACATGTGCTGGTCAAGAAGAG  600
                     ********************************************************************************************************
              501    ACAGAATTTAGCACAAGATGAATTAGCTTCTCTGCCTGTAGAGGATATTCCATGCAGTGTAGCTGAAGAGCCTGCTGCCACATGTGCTGGTCAAGAAGAG  600

              601    GAGGGAGAAATACCTGAAACGAAGAACTCTGTTCCAAGCACCCAAGTAGCTGATGACAATACAGCAGTGTCAGAGGTGGAGTATGAGCACGAAGACCTGC  700
                     ********************************************************************************************************
              601    GAGGGAGAAATACCTGAAACGAAGAACTCTGTTCCAAGCACCCAAGTAGCTGATGACAATACAGCAGTGTCAGAGGTGGAGTATGAGCACGAAGACCTGC  700

              701    CCTGTGAGCCAACACTCGTTGATCTCAATGCAAGTAGTGCAGCTGAAGCATTTGTTGAAGATGCTGCATTGGTGCTTCCGTCTTCTGATTCCATAACCAC  800
                     ********************************************************************************************************
              701    CCTGTGAGCCAACACTCGTTGATCTCAATGCAAGTAGTGCAGCTGAAGCATTTGTTGAAGATGCTGCATTGGTGCTTCCGTCTTCTGATTCCATAACCAC  800

              801    CTCTGAATCAACCTCGAAGGTGTGTATAGAGACTCCCTTCGACTTGGTGCCTAACAGTAAAGAATCTGACTCTGGTTCTATTGATACCTCAGACCCCAGT  900
                     ********************************************************************************************************
              801    CTCTGAATCAACCTCGAAGGTGTGTATAGAGACTCCCTTCGACTTGGTGCCTAACAGTAAAGAATCTGACTCTGGTTCTATTGATACCTCAGACCCCAGT  900

              901    TCAGCCGTGCTGTTAGAAGAGTTGGTCCCACTTGCTGAATCCCACACTGAAGCAGAGTCAGACCTGGTGCTAGCTAGTGAAGAACACAACTCGAATTCCG 1000
                     ********************************************************************************************************
              901    TCAGCCGTGCTGTTAGAAGAGTTGGTCCCACTTGCTGAATCCCACACTGAAGCAGAGTCAGACCTGGTGCTAGCTAGTGAAGAACACAACTCGAATTCCG 1000

             1001    CACCCACAGTTACCTCTGAATTAAATTCACAAGAATGTTTAGAAGATTTGAGTTCATTTGCAGAGTCCTGCTTAAAAGGAGATGGGGATTTGGAGTCTGA 1100
                     ********************************************************************************************************
             1001    CACCCACAGTTACCTCTGAATTAAATTCACAAGAATGTTTAGAAGATTTGAGTTCATTTGCAGAGTCCTGCTTAAAAGGAGATGGGGATTTGGAGTCTGA 1100

             1101    GAATGTACACCTTGGTTCACATTCCAACTCTGAAGCTGCCACTGGATCAAATTCACTTTTACATTTTGAAAGTGTAAGCTCCCTTGTGGAATCCAACATG 1200
                     ********************************************************************************************************
             1101    GAATGTACACCTTGGTTCACATTCCAACTCTGAAGCTGCCACTGGATCAAATTCACTTTTACATTTTGAAAGTGTAAGCTCCCTTGTGGAATCCAACATG 1200

             1201    AAACTCGTGGATTTTGTGCCCGTTCACAAAGAGCTTGACTCATGTTCAGACTTGAGCCGTACAGTAGAGAATACAGAATCAGACTCAAAAAATATATTTA 1300
                     ********************************************************************************************************
             1201    AAACTCGTGGATTTTGTGCCCGTTCACAAAGAGCTTGACTCATGTTCAGACTTGAGCCGTACAGTAGAGAATACAGAATCAGACTCAAAAAATATATTTA 1300

             1301    CAGCAAGCGAATCAGTTTCTGATACCTCTTTAACCAATATTGAAGTGGAGGATATTGAACAGCTTGTTAAATGTATTGATGCAGAGACATTAATTTTGAA 1400
                     ********************************************************************************************************
             1301    CAGCAAGCGAATCAGTTTCTGATACCTCTTTAACCAATATTGAAGTGGAGGATATTGAACAGCTTGTTAAATGTATTGATGCAGAGACATTAATTTTGAA 1400

             1401    TTCAGGTTATGTCCCTGCTTTATCATCAGATATAGAAAAAAGAGCAGTCGATAATATCTCCAGTGAGAGCTCCAGAGAATTCTCTGAAGGTTTTGTTTCA 1500
                     ********************************************************************************************************
             1401    TTCAGGTTATGTCCCTGCTTTATCATCAGATATAGAAAAAAGAGCAGTCGATAATATCTCCAGTGAGAGCTCCAGAGAATTCTCTGAAGGTTTTGTTTCA 1500

             1501    GAGTTGCCTTCAAGCACGGAAATCTCTGTTCTTGCCTTAGAAAAAGCCATAAATGCAGATTTGCCAGCTGACTTGAGACTGCAGCAGAATGATTATTCAT 1600
                     ********************************************************************************************************
             1501    GAGTTGCCTTCAAGCACGGAAATCTCTGTTCTTGCCTTAGAAAAAGCCATAAATGCAGATTTGCCAGCTGACTTGAGACTGCAGCAGAATGATTATTCAT 1600

             1601    CACAGCTGGAAAGTGTTGGAATAAGCTTGAATTTGGTTCAGGACAACGTGAGCAACAGCACTCAAACAGTTGAGGCTTCAGGTCCATCCTCTCTTACAAA 1700
                     ********************************************************************************************************
             1601    CACAGCTGGAAAGTGTTGGAATAAGCTTGAATTTGGTTCAGGACAACGTGAGCAACAGCACTCAAACAGTTGAGGCTTCAGGTCCATCCTCTCTTACAAA 1700

             1701    TCCAGCTGGTGGTTACCCTCCTGCATATTCATCTGTGCTCCCTATGTTAGAGAAAAAGAAACGAAGGCGTTGTGGAGTCTGTGAGCCCTGTCTGCGGAAA 1800
                     ********************************************************************************************************
             1701    TCCAGCTGGTGGTTACCCTCCTGCATATTCATCTGTGCTCCCTATGTTAGAGAAAAAGAAACGAAGGCGTTGTGGAGTCTGTGAGCCCTGTCTGCGGAAA 1800

             1801    ACAAATTGTGAAGAATGCAGCTGTTGCAGGAAGCGCAAAACCAGTCACCGGATATGTAAAAAGAGAAAATGTGAAGAGCTGAAGAAGCCACCACCACCAA 1900
                     ********************************************************************************************************
             1801    ACAAATTGTGAAGAATGCAGCTGTTGCAGGAAGCGCAAAACCAGTCACCGGATATGTAAAAAGAGAAAATGTGAAGAGCTGAAGAAGCCACCACCACCAA 1900

             1901    TCACACTACCTTTTGAGGTTCTAACTGAAAACAAAAGACCTCAGAGGAGGAAGCAAAGAGTTTTTAAGGCAAATTTAGAAAACAAACCAGTAAATGGCCG 2000
                     ********************************************************************************************************
             1901    TCACACTACCTTTTGAGGTTCTAACTGAAAACAAAAGACCTCAGAGGAGGAAGCAAAGAGTTTTTAAGGCAAATTTAGAAAACAAACCAGTAAATGGCCG 2000

             2001    CAGACCAGAACTCATGGAATGCAGTATTTGCGGTCATGGTGAAAAATACAGGGTGAAAACAAACCAAACACAACCCTTTGAAAATGCACAGTCTAAGGAG 2100
                     ********************************************************************************************************
             2001    CAGACCAGAACTCATGGAATGCAGTATTTGCGGTCATGGTGAAAAATACAGGGTGAAAACAAACCAAACACAACCCTTTGAAAATGCACAGTCTAAGGAG 2100

             2101    AAAGAAAGAATGACAGGCTTAGAAGCAGAGAAGTGGGCACATAACGAGAAGCCATCTTTTGGTGTTCATGTCAATGGAGATACCCATGGAACTGTGACAG 2200
                     ********************************************************************************************************
             2101    AAAGAAAGAATGACAGGCTTAGAAGCAGAGAAGTGGGCACATAACGAGAAGCCATCTTTTGGTGTTCATGTCAATGGAGATACCCATGGAACTGTGACAG 2200

             2201    GCCACGAACACTTGAGAAATGCAGAAGAAAGTGAAAGAGCGGTCTCTCCCAGTCAATTAGCTGAGCCAAAAAAATCATTTGCACAGACCATCAAAAATGG 2300
                     ********************************************************************************************************
             2201    GCCACGAACACTTGAGAAATGCAGAAGAAAGTGAAAGAGCGGTCTCTCCCAGTCAATTAGCTGAGCCAAAAAAATCATTTGCACAGACCATCAAAAATGG 2300

             2301    CATAAAAACCCTACATTACTCATCACCAGAAGCAGTTGCTTCGCTGAAAAAAGCATCTGTAGAGGAAAGGACAGATTTGACAAGTAACTCCCATGCCCAG 2400
                     ********************************************************************************************************
             2301    CATAAAAACCCTACATTACTCATCACCAGAAGCAGTTGCTTCGCTGAAAAAAGCATCTGTAGAGGAAAGGACAGATTTGACAAGTAACTCCCATGCCCAG 2400

             2401    TGGATGAAGGCTACCAATCTGAATGACATAGTCAGCACTCTGACAACTAGCGTTAGCTGTGTTCACCCAGACAAACAAGGGAATCTTTTCATGAAAGAGG 2500
                     ********************************************************************************************************
             2401    TGGATGAAGGCTACCAATCTGAATGACATAGTCAGCACTCTGACAACTAGCGTTAGCTGTGTTCACCCAGACAAACAAGGGAATCTTTTCATGAAAGAGG 2500

             2501    AAAATTGTACTTGCAGCATCTTTCAGGATTCTCATAAGTCAGTATTGCAAACCAGTTCTGTTTTGGATCTGCAAGAAAGCACATGTTGTCCTCCATTGGT 2600
                     ********************************************************************************************************
             2501    AAAATTGTACTTGCAGCATCTTTCAGGATTCTCATAAGTCAGTATTGCAAACCAGTTCTGTTTTGGATCTGCAAGAAAGCACATGTTGTCCTCCATTGGT 2600

             2601    TAAAGACGAAGTGCATGGAGAGGATGGCTTCAAAGTACCTAACACAGAGACACAGCATGAGAATCCCTCGCTCCAGCCAACTTTCTTGTCCCTGATTAAA 2700
                     ********************************************************************************************************
             2601    TAAAGACGAAGTGCATGGAGAGGATGGCTTCAAAGTACCTAACACAGAGACACAGCATGAGAATCCCTCGCTCCAGCCAACTTTCTTGTCCCTGATTAAA 2700
```

Fig. S1

```
Cloned          2701  AACAGAAACCTAACTGTGGAGCAGGTTGTAGCTATTGAAGCTTTAACACAGTTATCTGAAGTCCCTTTAGGAACATCTTCCCCAGCTAAAGCTGACAATA  2800
                      ********************************************************************************************************
NCBI database   2701  AACAGAAACCTAACTGTGGAGCAGGTTGTAGCTATTGAAGCTTTAACACAGTTATCTGAAGTCCCTTTAGGAACATCTTCCCCAGCTAAAGCTGACAATA  2800

                2801  CTGAATGTACGGAAGAAAGAACTTCCAACTTGCTCCATAGTTATAAAAGGGATATTTCATCCTCCTTTACGTCTTCTACGTTAAAAGAAATCAAAGACAC  2900
                      ********************************************************************************************************
                2801  CTGAATGTACGGAAGAAAGAACTTCCAACTTGCTCCATAGTTATAAAAGGGATATTTCATCCTCCTTTACGTCTTCTACGTTAAAAGAAATCAAAGACAC  2900

                2901  TTACTTAGAGAAAGAAGAACAGCTCTTGACTCACTGTGCTCACTCACAAAAGCAGCTCCCAAATAAGCCAGTGTTATACAATGGTCAAAGTTCGATTTCT  3000
                      ********************************************************************************************************
                2901  TTACTTAGAGAAAGAAGAACAGCTCTTGACTCACTGTGCTCACTCACAAAAGCAGCTCCCAAATAAGCCAGTGTTATACAATGGTCAAAGTTCGATTTCT  3000

                3001  GAATTGCGTGATAAACCTCCCAATCTAGCTAGCGAGATGTATAAATTGCAGTTATCCTCGAGAGATACCAAAACTTTATCTGAGTTGAGATCTAATGCAA  3100
                      ********************************************************************************************************
                3001  GAATTGCGTGATAAACCTCCCAATCTAGCTAGCGAGATGTATAAATTGCAGTTATCCTCGAGAGATACCAAAACTTTATCTGAGTTGAGATCTAATGCAA  3100

                3101  AATCATTTTCAGGACTTAATACCAAGAAAAACTATGATCCAGTCACCCTTTCACGGTGTTGCAATGTCCAGCATACAAGCAACAACTTGGATACTCTTGG  3200
                      ********************************************************************************************************
                3101  AATCATTTTCAGGACTTAATACCAAGAAAAACTATGATCCAGTCACCCTTTCACGGTGTTGCAATGTCCAGCATACAAGCAACAACTTGGATACTCTTGG  3200

                3201  CCAGTTAGATCAAAAGCCAACCTGTAATGATTCGAGCTTGGAAGCTGGTTCTTTAATTAAAGAAGGAGGAATTCACAGCCAGGATGAAGAGGATGTAGCT  3300
                      ********************************************************************************************************
                3201  CCAGTTAGATCAAAAGCCAACCTGTAATGATTCGAGCTTGGAAGCTGGTTCTTTAATTAAAGAAGGAGGAATTCACAGCCAGGATGAAGAGGATGTAGCT  3300

                3301  ACTCAACTAACTCAACTTGCAGCATTAATTGAATCAAACCGGGCAAATCCAGCGCAGAAGAATGACATCAAGACATCACTGCTTAACCTAATATCCCATG  3400
                      ********************************************************************************************************
                3301  ACTCAACTAACTCAACTTGCAGCATTAATTGAATCAAACCGGGCAAATCCAGCGCAGAAGAATGACATCAAGACATCACTGCTTAACCTAATATCCCATG  3400

                3401  AGAGGCAAGCAAAACATAACCAAGATCTGCTGCAGAGAAAAGAATCTGTATTTTTTAGGCATAACTGCAGCTCCTTACTGTTAAAGCAAAAACAATCAAC  3500
                      ********************************************************************************************************
                3401  AGAGGCAAGCAAAACATAACCAAGATCTGCTGCAGAGAAAAGAATCTGTATTTTTTAGGCATAACTGCAGCTCCTTACTGTTAAAGCAAAAACAATCAAC  3500

                3501  AAATAAAAAAGGAAATGCTGTACCATGGAAACCAAGGCACAAAAAAAAAGCCTCAAGAAGCACGCTATCAACAAATTAATCGAAACCAACTAGAGTTGTTG  3600
                      ********************************************************************************************************
                3501  AAATAAAAAAGGAAATGCTGTACCATGGAAACCAAGGCACAAAAAAAAAGCCTCAAGAAGCACGCTATCAACAAATTAATCGAAACCAACTAGAGTTGTTG  3600

                3601  TCATGCCAGCACAGTGAGTTACAGGACATTTGGATATCATCAAAACTGCACAAGCTTGGTCAAGCCATGCCACAGGACTCCAGAATAGCTCTGGCTGAAA  3700
                      ********************************************************************************************************
                3601  TCATGCCAGCACAGTGAGTTACAGGACATTTGGATATCATCAAAACTGCACAAGCTTGGTCAAGCCATGCCACAGGACTCCAGAATAGCTCTGGCTGAAA  3700

                3701  AAAAATCACCGAGAACCAAAGTACAGAAAGCACTGAATCAAAATGCGCTAACATTACAGGAAAAACCTAGGTTGTTTCTTCCTCAAACACAGATAAAATT  3800
                      ********************************************************************************************************
                3701  AAAAATCACCGAGAACCAAAGTACAGAAAGCACTGAATCAAAATGCGCTAACATTACAGGAAAAACCTAGGTTGTTTCTTCCTCAAACACAGATAAAATT  3800

                3801  CCATAGGTTATCTGAGGTACCACAGGAGAAAAAAAAAAGACAGGTTGTTTGGCTATGAAGTGGTGAATGAGCAAATCAAAACTGCAGGCTCCAGTGACCCA  3900
                      ********************************************************************************************************
                3801  CCATAGGTTATCTGAGGTACCACAGGAGAAAAAAAAAAGACAGGTTGTTTGGCTATGAAGTGGTGAATGAGCAAATCAAAACTGCAGGCTCCAGTGACCCA  3900

                3901  CCAGGCATCGTTCCAGTGAAAAACGACATTGCTCCGTGTAACCAATATAGTGATCTGCTCTCCCATAATCCTCTGTCAGTTTCACAGTTGAAAACAGCAT  4000
                      ********************************************************************************************************
                3901  CCAGGCATCGTTCCAGTGAAAAACGACATTGCTCCGTGTAACCAATATAGTGATCTGCTCTCCCATAATCCTCTGTCAGTTTCACAGTTGAAAACAGCAT  4000

                4001  CCTTGTTAACCAGACCAAGTGAAAACACACAGATGTTTACAGAAAAATGTAATTCTCAGGTACAGCAAGCAGTGAACGTTGGTCAGACACATCCTTTGCC  4100
                      ********************************************************************************************************
                4001  CCTTGTTAACCAGACCAAGTGAAAACACACAGATGTTTACAGAAAAATGTAATTCTCAGGTACAGCAAGCAGTGAACGTTGGTCAGACACATCCTTTGCC  4100

                4101  TCAAACTTTTAGTCAGTCTAACCTGAGAGCTGGTGAAACATGCGGTGAAGATAAAGCCTATGTCCAGCAGGGTGCTGTTGATCAACGGGTAGATCCGAAG  4200
                      ********************************************************************************************************
                4101  TCAAACTTTTAGTCAGTCTAACCTGAGAGCTGGTGAAACATGCGGTGAAGATAAAGCCTATGTCCAGCAGGGTGCTGTTGATCAACGGGTAGATCCGAAG  4200

                4201  TCGCAGATGCTGCCTGTTTCCTGCAGTGCGGGTCAGGTGCCAGGCACTGTGGAAGCTCTCAGGAACATGGAGTGTGCAGGTGAAGTTACCATTCTAACGT  4300
                      ********************************************************************************************************
                4201  TCGCAGATGCTGCCTGTTTCCTGCAGTGCGGGTCAGGTGCCAGGCACTGTGGAAGCTCTCAGGAACATGGAGTGTGCAGGTGAAGTTACCATTCTAACGT  4300

                4301  CAAGAAGTTTGGGTACTGACCATCCACAGAGCACAGGCGACTCAGGGTGTTCTCCAGCAAAAAACACACTCAGCAGTTTTCTCGAATCACCTATGAAGTT  4400
                      ********************************************************************************************************
                4301  CAAGAAGTTTGGGTACTGACCATCCACAGAGCACAGGCGACTCAGGGTGTTCTCCAGCAAAAAACACACTCAGCAGTTTTCTCGAATCACCTATGAAGTT  4400

                4401  TCTCGATACTCCTACAAAAAATTTAATAGATACACCTACAAAAAAAGGACAGTCTGAATTGCCAACTTGTGACTGTGTTGAGCAAATTATTGAGAAAGAT  4500
                      ********************************************************************************************************
                4401  TCTCGATACTCCTACAAAAAATTTAATAGATACACCTACAAAAAAAGGACAGTCTGAATTGCCAACTTGTGACTGTGTTGAGCAAATTATTGAGAAAGAT  4500

                4501  GAAGGCCCGTATTACACACACCTCGGGACAGGACCAAGTGTTGCTGCTGTGAGAGAAATAATGGAGAACAGGTATGGAGCAAAAGGAAGCGCTGTGAGAA  4600
                      ********************************************************************************************************
                4501  GAAGGCCCGTATTACACACACCTCGGGACAGGACCAAGTGTTGCTGCTGTGAGAGAAATAATGGAGAACAGGTATGGAGCAAAAGGAAGCGCTGTGAGAA  4600

                4601  TAGAGGTAGTCGTTTATACAGGAAAGGAAGGAAAAAGCTCTCAGGGGTGTCCGATTGCCAAGTGGGTGATACGAAGAAGTAGTGATGAGGAGAAGTTGCT  4700
                      ********************************************************************************************************
                4601  TAGAGGTAGTCGTTTATACAGGAAAGGAAGGAAAAAGCTCTCAGGGGTGTCCGATTGCCAAGTGGGTGATACGAAGAAGTAGTGATGAGGAGAAGTTGCT  4700

                4701  ATGCTTGGTACGTCAGCGTGCGGGACACCACTGTCAGACTGCTGTCATAGTGATTCTCATACTGGCGTGGGAGGGGATCCCTCATCTCCTGGCTGATACG  4800
                      ********************************************************************************************************
                4701  ATGCTTGGTACGTCAGCGTGCGGGACACCACTGTCAGACTGCTGTCATAGTGATTCTCATACTGGCGTGGGAGGGGATCCCTCATCTCCTGGCTGATACG  4800

                4801  CTGTACAAAGAGCTGACGCAGAGCCTCAGAAAATACGGCTGTCCCACGAGCCGCAGATGTGCACTGAATGAAGACCGTACATGTGCATGTCAGGGACTTG  4900
                      ********************************************************************************************************
                4801  CTGTACAAAGAGCTGACGCAGAGCCTCAGAAAATACGGCTGTCCCACGAGCCGCAGATGTGCACTGAATGAAGACCGTACATGTGCATGTCAGGGACTTG  4900

                4901  ATCCAGAAACATGTGGAGCATCGTTCTCATTTGGCTGTTCATGGAGTATGTATTTCAATGGATGTAAATTCGCCAGAAGCAAAAACCCCAGGAAATTTAG  5000
                      ********************************************************************************************************
                4901  ATCCAGAAACATGTGGAGCATCGTTCTCATTTGGCTGTTCATGGAGTATGTATTTCAATGGATGTAAATTCGCCAGAAGCAAAAACCCCAGGAAATTTAG  5000

                5001  GCTTCTCACTGATGACCCTAAACAAGAGGAACTTCTTGAACATAATTTGCAAACTTTAGCTACTGATGTGGCTCCGGTTTATAAAAAGCTTGCTCCTGAA  5100
                      ********************************************************************************************************
                5001  GCTTCTCACTGATGACCCTAAACAAGAGGAACTTCTTGAACATAATTTGCAAACTTTAGCTACTGATGTGGCTCCGGTTTATAAAAAGCTTGCTCCTGAA  5100

                5101  GCTTTCCAGAACCAGGTGGAAAATGAACACATGGGCCCAGACTGTCGGCTTGGATCCAAGGACGGTAGGCCTTTCTCTGGTGTCACAGCTTGCATAGATT  5200
                      ********************************************************************************************************
                5101  GCTTTCCAGAACCAGGTGGAAAATGAACACATGGGCCCAGACTGTCGGCTTGGATCCAAGGACGGTAGGCCTTTCTCTGGTGTCACAGCTTGCATAGATT  5200

                5201  TTTGTGCCCATGCCCATAAGGACACACATAACATGCACAATGGAAGCACTGTGGTCTGTACATTAACAAAGGAGGACAACCGGAGAGTCGGGGTGATTCC  5300
                      ********************************************************************************************************
                5201  TTTGTGCCCATGCCCATAAGGACACACATAACATGCACAATGGAAGCACTGTGGTCTGTACATTAACAAAGGAGGACAACCGGAGAGTCGGGGTGATTCC  5300

                5301  AAGTGATGAGCAGCTCCATGTGTTACCTCTCTACAAAATTTCACAAACAGATGAATTTGGCACTGAAGAGGGATTGGAAGCTAAGATAAAAGCCGGAGCC  5400
                      ********************************************************************************************************
                5301  AAGTGATGAGCAGCTCCATGTGTTACCTCTCTACAAAATTTCACAAACAGATGAATTTGGCACTGAAGAGGGATTGGAAGCTAAGATAAAAGCCGGAGCC  5400

                5401  ATACAGGTGCTCACAGCGTTTCCCAGAGAAGTACGGATGTTAGCCGAGCCTCTGAGAGCTACAAAGAAAAAGAAACCTGACACAAGGAGGACCCCGAGTG  5500
                      ********************************************************************************************** *****
                5401  ATACAGGTGCTCACAGCGTTTCCCAGAGAAGTACGGATGTTAGCCGAGCCTCTGAGAGCTACAAAGAAAAAGAAACCTGACACAAGGAGGACCCTGAGTG  5500
```

Fig. S1

```
Cloned          5501  AAAAGCAGCAATTAGTGGATAAGAAATATTCAACACCAATAAAACTGAAAACTGAAGCGCCAGAAAATCTCAGTAACACTTTGCATTGCTTAGGGAACAA  5600
                      **************************************************************************************************
NCBI database   5501  AAAAGCAGCAATTAGTGGATAAGAAATATTCAACACCAATAAAACTGAAAACTGAAGCGCCAGAAAATCTCAGTAACACTTTGCATTGCTTAGGGAACAA  5600

                5601  AGCAGATGCTTTAAAACCTGGGATTAAAACGGAGACTTCTGATCACCTCTGTGCCATGAAACATACTTCAAACACTACGAAAAATTGCTCCCTATTAAAA  5700
                      **************************************************************************************************
                5601  AGCAGATGCTTTAAAACCTGGGATTAAAACGGAGACTTCTGATCACCTCTGTGCCATGAAACATACTTCAAACACTACGAAAAATTGCTCCCTATTAAAA  5700

                5701  CAGTATACGGCTTCCCCCCCTTTTAAGGTGGATAGTTTACATCCTTATTCATCGTTAGCACATAAGCCTGGAATAACAGCAGTGACTAATATTCAACAAG  5800
                      **************************************************************************************************
                5701  CAGTATACGGCTTCCCCCCCTTTTAAGGTGGATAGTTTACATCCTTATTCATCGTTAGCACATAAGCCTGGAATAACAGCAGTGACTAATATTCAACAAG  5800

                5801  ATTTTTCAGTTCCCTACGGGTATTTTGAATGCAGTAGTAAGCAACCTCACGTGACACCTTATATTAATTGTAAGAACTTTGATGTGTCTGTCAAAGATTA  5900
                      **************************************************************************************************
                5801  ATTTTTCAGTTCCCTACGGGTATTTTGAATGCAGTAGTAAGCAACCTCACGTGACACCTTATATTAATTGTAAGAACTTTGATGTGTCTGTCAAAGATTA  5900

                5901  TACTGGAATTTTGCTGAATGATAAGATGAACGGTGTGCCGCCAATTCTTCCAGAGGTCACTGCTCCAGGCCCCCCAGCCCACAAAGACCCCCTGCCCAGT  6000
                      **************************************************************************************************
                5901  TACTGGAATTTTGCTGAATGATAAGATGAACGGTGTGCCGCCAATTCTTCCAGAGGTCACTGCTCCAGGCCCCCCAGCCCACAAAGACCCCCTGCCCAGT  6000

                6001  ATACTTGAACATCAGCCAGACAAGCAGAATTGTCAGCCTCAGTCAGACAGTTCTCCTTCTTCACAGATGATCAGCTCTTGCGATTTGTCGGTACCGTTAA  6100
                      **************************************************************************************************
                6001  ATACTTGAACATCAGCCAGACAAGCAGAATTGTCAGCCTCAGTCAGACAGTTCTCCTTCTTCACAGATGATCAGCTCTTGCGATTTGTCGGTACCGTTAA  6100

                6101  GCTCTCCAGCAAAGGACGCAGTTGGAAATGAAGCCGACTGTTCCCACAGATGCCCGGTGGAAAAGGGTAGCTCCCATCGAGAACAGATGTGTGATTTTGA  6200
                      **************************************************************************************************
                6101  GCTCTCCAGCAAAGGACGCAGTTGGAAATGAAGCCGACTGTTCCCACAGATGCCCGGTGGAAAAGGGTAGCTCCCATCGAGAACAGATGTGTGATTTTGA  6200

                6201  CTGCACTGACGAAAAGCAAAACAGTGCACTGGGACAACCGACTGACTCTGAAGAAAAAGCTGAGGAAATGTGGTCAGACAGCGAGCACAACTTTTTGGAT  6300
                      **************************************************************************************************
                6201  CTGCACTGACGAAAAGCAAAACAGTGCACTGGGACAACCGACTGACTCTGAAGAAAAAGCTGAGGAAATGTGGTCAGACAGCGAGCACAACTTTTTGGAT  6300

                6301  GATGACATCGGCGGCGTCGCTGTGGCACCTTCTCATGGTTCTATCCTAATCGAATGTGCGAGACGTGAACTCCATGCTACCACACCTATTAAGAAACCCA  6400
                      **************************************************************************************************
                6301  GATGACATCGGCGGCGTCGCTGTGGCACCTTCTCATGGTTCTATCCTAATCGAATGTGCGAGACGTGAACTCCATGCTACCACACCTATTAAGAAACCCA  6400

                6401  ACCGCAATCATCCCACGCGGATCTCCTTAGTATTCTACCAACACAAAAATTTAAATGAGCCAAAACATGGTTTAGCCATGTGGGAAGCAAAGATGGCTGA  6500
                      **************************************************************************************************
                6401  ACCGCAATCATCCCACGCGGATCTCCTTAGTATTCTACCAACACAAAAATTTAAATGAGCCAAAACATGGTTTAGCCATGTGGGAAGCAAAGATGGCTGA  6500

                6501  GAGGGCGAAAGAAAAAGAAAAAGAAGCAGAAAGATTAGGAACAGAGAACACTGAACTGAACTCCAGCAGCAGGAAAACAAAGCAAACAAGTGAAAACAGA  6600
                      **************************************************************************************************
                6501  GAGGGCGAAAGAAAAAGAAAAAGAAGCAGAAAGATTAGGAACAGAGAACACTGAACTGAACTCCAGCAGCAGGAAAACAAAGCAAACAAGTGAAAACAGA  6600

                6601  GAGATTTTTTATGAGGACAATGAGTTCAACCAAATTCCATCACGCAGAGCATTAACAGTAACTAAAGATAACGTAATAACAGTATCTTCTTATGCCCTTA  6700
                      **************************************************************************************************
                6601  GAGATTTTTTATGAGGACAATGAGTTCAACCAAATTCCATCACGCAGAGCATTAACAGTAACTAAAGATAACGTAATAACAGTATCTTCTTATGCCCTTA  6700

                6701  CGCGAGTTGCAGGGCCTTACAACCATTGGGCATAG
                      ***********************************
                6701  CGCGAGTTGCAGGGCCTTACAACCATTGGGCATAG
```

Fig. S1

B

```
Cloned          1  ATGGAACAGGACAGAACCATCCATGTTGACGGCAATAGATTGAGTCCATTTTTAATATCACAACCTTCTCACATTTGCCAGGCAGACCCTTCTGCAGTGA  100
                   ********************************************************************************************************
NCBI database   1  ATGGAACAGGACAGAACCATCCATGTTGACGGCAATAGATTGAGTCCATTTTTAATATCACAACCTTCTCACATTTGCCAGGCAGACCCTTCTGCAGTGA  100

              101  AGCTACAGAACGGAAGTCCAGCAACAGAGAGGCCTGAGGTCGAAGTCAACGGCAACCACAAGCGGCTGTTCAATAAGAGCAACTACAGAGAGCCCCACGC  200
                   ********************************************************************************************************
              101  AGCTACAGAACGGAAGTCCAGCAACAGAGAGGCCTGAGGTCGAAGTCAACGGCAACCACAAGCGGCTGTTCAATAAGAGCAACTACAGAGAGCCCCACGC  200

              201  AAAGGGAAGCCCAAACCACCGCATTAGCCCTGACCTTTTACAAGAGAAGAAAGCATGCTCCATATATATGCAAAATGGTGGGATAAAACGCACTTTTAGT  300
                   ********************************************************************************************************
              201  AAAGGGAAGCCCAAACCACCGCATTAGCCCTGACCTTTTACAAGAGAAGAAAGCATGCTCCATATATATGCAAAATGGTGGGATAAAACGCACTTTTAGT  300

              301  GAGCCCTCTCTGTTTGGACTTCAACAGAGCAAGAAAGTGAAACAAGACAAAGAGGTAAATGGAGAAAAAGCTGAGCCAGAGGATAACTATGAAACACCAA  400
                   ********************************************************************************************************
              301  GAGCCCTCTCTGTTTGGACTTCAACAGAGCAAGAAAGTGAAACAAGACAAAGAGGTAAATGGAGAAAAAGCTGAGCCAGAGGATAACTATGAAACACCAA  400

              401  GCATCTCCAATTGCTACAGTGAGAAGAAATCTGAGATGGGACAAGAAAATGAAGCTTTGGAGTTGATGCCGTCAACAAGATACAACAGTGGTGGTTCAGA  500
                   ********************************************************************************************************
              401  GCATCTCCAATTGCTACAGTGAGAAGAAATCTGAGATGGGACAAGAAAATGAAGCTTTGGAGTTGATGCCGTCAACAAGATACAACAGTGGTGGTTCAGA  500

              501  AGACCCTCGTGAACTCCTGATTCAGGATGAGCAGGAGGGGGAAAACATTAATTGCCACAACAGGGACATTGTCTTACTACTCAAGAACAAGGCGGTGCCA  600
                   ********************************************************************************************************
              501  AGACCCTCGTGAACTCCTGATTCAGGATGAGCAGGAGGGGGAAAACATTAATTGCCACAACAGGGACATTGTCTTACTACTCAAGAACAAGGCGGTGCCA  600

              601  ATGCCTAATGGTGCTACAGTTTCTGCCTCTTCCATGGACAGCATGCATGGTGAACTCCTGGAGAAAACACTGTCTCAATATTATCCAGAACATGTTTCCA  700
                   ********************************************************************************************************
              601  ATGCCTAATGGTGCTACAGTTTCTGCCTCTTCCATGGACAGCATGCATGGTGAACTCCTGGAGAAAACACTGTCTCAATATTATCCAGAACATGTTTCCA  700

              701  TAGCAATGCAGAAGAACACATCTCATATCAATGCCATTACCAGTCAGGCTACTAATGAGTTGTCCCACGAGACAACGCATTCATCCCATACCTCAGGGCA  800
                   ********************************************************************************************************
              701  TAGCAATGCAGAAGAACACATCTCATATCAATGCCATTACCAGTCAGGCTACTAATGAGTTGTCCCACGAGACAACGCATTCATCCCATACCTCAGGGCA  800

              801  GATCACTTCCCCACAGACCTCAAACTCTGAGCTGCCTCAAGTGCCAGCTGTAGTGGTTACTGAGGTCTACGGTGCTGATGACTCCAGTAAGCCACCTGTA  900
                   ********************************************************************************************************
              801  GATCACTTCCCCACAGACCTCAAACTCTGAGCTGCCTCAAGTGCCAGCTGTAGTGGTTACTGAGGTCTACGGTGCTGATGACTCCAGTAAGCCACCTGTA  900

              901  TTGCCAGGTAGCTGTTCACTTCAGAAACCAGAACTACAGCTACAGCAGCAGATTCCAGGCTATGATACACACCGGTTACCTTTAGGAAACAGTGCTGTTC 1000
                   ********************************************************************************************************
              901  TTGCCAGGTAGCTGTTCACTTCAGAAACCAGAACTACAGCTACAGCAGCAGATTCCAGGCTATGATACACACCGGTTACCTTTAGGAAACAGTGCTGTTC 1000

             1001  ATGGAAGCGTAGGGCAGGTTCCCAATCAAGACCTCTCTCTAAGTTCCAGCAGTAACCTGCAAGCTCAGAATGCTGCATCAGAAAGGTTTTCTGAGCAAGC 1100
                   ********************************************************************************************************
             1001  ATGGAAGCGTAGGGCAGGTTCCCAATCAAGACCTCTCTCTAAGTTCCAGCAGTAACCTGCAAGCTCAGAATGCTGCATCAGAAAGGTTTTCTGAGCAAGC 1100

             1101  AGAGAAAAATGGTGCTTTCTTTACACAGAACTCAATGTTTCACAAAGATTCCTCAACTCCTCCTGCTCCAGAAATGAACAGTGCACTGTCCGTCATGGTG 1200
                   ********************************************************************************************************
             1101  AGAGAAAAATGGTGCTTTCTTTACACAGAACTCAATGTTTCACAAAGATTCCTCAACTCCTCCTGCTCCAGAAATGAACAGTGCACTGTCCGTCATGGTG 1200

             1201  CGAGAAGGATGCCATTCCTATGACAACAGATGCAATGAAACTCTTCCTGGAGAGATCAAGAACGAAGGGCAACATCAGGGACCAATGCCAGAAAGTCCCG 1300
                   ********************************************************************************************************
             1201  CGAGAAGGATGCCATTCCTATGACAACAGATGCAATGAAACTCTTCCTGGAGAGATCAAGAACGAAGGGCAACATCAGGGACCAATGCCAGAAAGTCCCG 1300

             1301  GCCTCAGCCAACAGCAACTTCACCCCCAGCAAAGGCTTCCGCAGCAGGTGCAAACGTCGCAGCACAAAGTCAGCCACAGCGATCCCCGAGCTGCTGCAGC 1400
                   ********************************************************************************************************
             1301  GCCTCAGCCAACAGCAACTTCACCCCCAGCAAAGGCTTCCGCAGCAGGTGCAAACGTCGCAGCACAAAGTCAGCCACAGCGATCCCCGAGCTGCTGCAGC 1400

             1401  CGCCTCCATTCAGCAGCACCCAGAAGAAATGCCGCCGCCGCCATCAGAGCCTCCCCTCCAAAACCTGCATGCGTGCGGAAGCGATGGTGAGTTGCCTCAA 1500
                   ********************************************************************************************************
             1401  CGCCTCCATTCAGCAGCACCCAGAAGAAATGCCGCCGCCGCCATCAGAGCCTCCCCTCCAAAACCTGCATGCGTGCGGAAGCGATGGTGAGTTGCCTCAA 1500

             1501  CTGTGTCAGCGTTTCCCAGGACAGAGAGAACCTGAGATTCCTCCTGACAAAGAAAAGGACCAAGTGAAAGAGTCTGTGCAACAGCCTCAGCGTTACTCAA 1600
                   ********************************************************************************************************
             1501  CTGTGTCAGCGTTTCCCAGGACAGAGAGAACCTGAGATTCCTCCTGACAAAGAAAAGGACCAAGTGAAAGAGTCTGTGCAACAGCCTCAGCGTTACTCAA 1600

             1601  AGCCAGCCTGGATAGAATTGGTTTCCACGCCGTTCCGGCAGGGAGAGCTTCCCCACAAGCCAAACGAAGCATTACTGCGATCAATTCTTCAGTACCAGGC 1700
                   ********************************************************************************************************
             1601  AGCCAGCCTGGATAGAATTGGTTTCCACGCCGTTCCGGCAGGGAGAGCTTCCCCACAAGCCAAACGAAGCATTACTGCGATCAATTCTTCAGTACCAGGC 1700

             1701  AAATACATCCAAAGCAGCTTATATGAAACAGTATGCTGGAAGTCCTGATGCATTAAAGGGGCCGTCGGGACAGCCCCAGAGCCAGAAGATAATGCAACAA 1800
                   ********************************************************************************************************
             1701  AAATACATCCAAAGCAGCTTATATGAAACAGTATGCTGGAAGTCCTGATGCATTAAAGGGGCCGTCGGGACAGCCCCAGAGCCAGAAGATAATGCAACAA 1800

             1801  GAACAAATTCCCCTGCAGTACAAAAGCGAGAGCTCCCAGATGCAGCAGCATCCCACAGCTGACCTGCAGCTGCTGTTCCAAAAGCACTCACCGCAGCCAC 1900
                   ********************************************************************************************************
             1801  GAACAAATTCCCCTGCAGTACAAAAGCGAGAGCTCCCAGATGCAGCAGCATCCCACAGCTGACCTGCAGCTGCTGTTCCAAAAGCACTCACCGCAGCCAC 1900

             1901  AGCTCACAAAGATGGATTCCCTGCTCAAGTCCCGAGTGCAGCAACACCCTCCACAGCAGCTCCATTTCCAGCAACAACTTGAACAACAAACTGAACAGCC 2000
                   ********************************************************************************************************
             1901  AGCTCACAAAGATGGATTCCCTGCTCAAGTCCCGAGTGCAGCAACACCCTCCACAGCAGCTCCATTTCCAGCAACAACTTGAACAACAAACTGAACAGCC 2000

             2001  TTTAGGGGCCCCGCTGAAACAGCAGCACTTGAATCCCCAGCCAGGGGAAAGTGAACAGTTCTTGCATTCACACATTTTGCAACAGATGCTCCAAAAGCAG 2100
                   ********************************************************************************************************
             2001  TTTAGGGGCCCCGCTGAAACAGCAGCACTTGAATCCCCAGCCAGGGGAAAGTGAACAGTTCTTGCATTCACACATTTTGCAACAGATGCTCCAAAAGCAG 2100

             2101  ACACAGCAGACACAGATGCTGTGCAGTCCGCAGCTAACTCCAAACCAGCAACAGGCTCTGCAAATGAAAAGTAAAGAACCGCCCCAAACTATTCCCCACT 2200
                   ********************************************************************************************************
             2101  ACACAGCAGACACAGATGCTGTGCAGTCCGCAGCTAACTCCAAACCAGCAACAGGCTCTGCAAATGAAAAGTAAAGAACCGCCCCAAACTATTCCCCACT 2200

             2201  CCCAAAGCAACGCGGAGCAGCAGCCAGACAGGACATCCTTCAGTCAGCCCAAAGCAGATGAGTGCTTTCAAACCGGGAATAAGTACATGAAACCAACCGC 2300
                   ********************************************************************************************************
             2201  CCCAAAGCAACGCGGAGCAGCAGCCAGACAGGACATCCTTCAGTCAGCCCAAAGCAGATGAGTGCTTTCAAACCGGGAATAAGTACATGAAACCAACCGC 2300

             2301  ATTCCCACTGCATAGCCCTCAGCAAGGGCTAGAGCAGGTACAGAGCATGAACAACAAAACTCCCCTTTACAGCCAGAAAACAAGCACTGGTCTGCAGCAT 2400
                   ********************************************************************************************************
             2301  ATTCCCACTGCATAGCCCTCAGCAAGGGCTAGAGCAGGTACAGAGCATGAACAACAAAACTCCCCTTTACAGCCAGAAAACAAGCACTGGTCTGCAGCAT 2400

             2401  CCCTGCCCAAACAACGTGCACTTGATGTCAGAGAAGACGGAGAATGCCGCAAATCTTGAACACTTCGGAGCCAACAAAGCGCGTGACTTGCAACACGTGC 2500
                   ********************************************************************************************************
             2401  CCCTGCCCAAACAACGTGCACTTGATGTCAGAGAAGACGGAGAATGCCGCAAATCTTGAACACTTCGGAGCCAACAAAGCGCGTGACTTGCAACACGTGC 2500

             2501  AGTATTTCTCAAATAACTTGCCCCCCAAGCAAGATGTGAATCACTGTTTTCAAGAGCAAGAGCAACAGACGCAACAAGCTTCAGTTATACAGCTGCCACA 2600
                   ********************************************************************************************************
             2501  AGTATTTCTCAAATAACTTGCCCCCCAAGCAAGATGTGAATCACTGTTTTCAAGAGCAAGAGCAACAGACGCAACAAGCTTCAGTTATACAGCTGCCACA 2600

             2601  AGGCTATGGTGGTAGCCTCAGTCAAGATCCCCCGTGCCAACAGGCCGCACCGATGCCCCAGCGGTACTTACCGCACAGCCAGCAAACTCCTGCACACTCA 2700
                   ********************************************************************************************************
             2601  AGGCTATGGTGGTAGCCTCAGTCAAGATCCCCCGTGCCAACAGGCCGCACCGATGCCCCAGCGGTACTTACCGCACAGCCAGCAAACTCCTGCACACTCA 2700
```

Fig. S1

```
Cloned          2701  CAAGATCAGAGAGGCTGTCATTTGCAGAGCCAAGCCCCAAAGGATTTTCACAAGCACGCTGCTCTAAGGTGGCATCTCTTACAGAAACAGGAGCAACAAG  2800
                      ********************************************************************************************************
NCBI database   2701  CAAGATCAGAGAGGCTGTCATTTGCAGAGCCAAGCCCCAAAGGATTTTCACAAGCACGCTGCTCTAAGGTGGCATCTCTTACAGAAACAGGAGCAACAAG  2800

                2801  CATACCAGCAACCCAAAACCGAGACTGGCGCCGGTGCAGCACGCAAGCCTATAAAAATTGAGGCTGGAGCAAAGTCTAACTTTTGCATGCGTCTGTCAGC  2900
                      ********************************************************************************************************
                2801  CATACCAGCAACCCAAAACCGAGACTGGCGCCGGTGCAGCACGCAAGCCTATAAAAATTGAGGCTGGAGCAAAGTCTAACTTTTGCATGCGTCTGTCAGC  2900

                2901  TGGGCAGCTGGAAAACAAAATGTGGAAAAAAACAATTAAACAAGAGAATCAGCACTTTGGCTGTGAGAACACACAACAAAAGAGCATCATCGAGACAATG  3000
                      ********************************************************************************************************
                2901  TGGGCAGCTGGAAAACAAAATGTGGAAAAAAACAATTAAACAAGAGAATCAGCACTTTGGCTGTGAGAACACACAACAAAAGAGCATCATCGAGACAATG  3000

                3001  GAACAGCAGCTAAAACAGATACAGGTCAAATCACTGTTTGATCACAAGACTTTTACTGTCAAATCACCTAAACACGTGAAGGTTGAAACAGCAGGCCCTA  3100
                      ********************************************************************************************************
                3001  GAACAGCAGCTAAAACAGATACAGGTCAAATCACTGTTTGATCACAAGACTTTTACTGTCAAATCACCTAAACACGTGAAGGTTGAAACAGCAGGCCCTA  3100

                3101  TTACCATCCTATCCAGAAACACCAGTGCTGCAGAATTTGATACTCACACCCCGATCTTAGAACAGCAAGCAAATGTGTCTGCTGAGAAAACCCCGACCAA  3200
                      ********************************************************************************************************
                3101  TTACCATCCTATCCAGAAACACCAGTGCTGCAGAATTTGATACTCACACCCCGATCTTAGAACAGCAAGCAAATGTGTCTGCTGAGAAAACCCCGACCAA  3200

                3201  AAGAACAGCTGGAACTGTTCTCAATAATTTTTTAGACTCACCTTCCAAGTTATTGGATACTCCTGTAAAAAATTTATTGGACACACCTGCCAAAACCCAG  3300
                      ********************************************************************************************************
                3201  AAGAACAGCTGGAACTGTTCTCAATAATTTTTTAGACTCACCTTCCAAGTTATTGGATACTCCTGTAAAAAATTTATTGGACACACCTGCCAAAACCCAG  3300

                3301  TATGATTTCCCATCTTGCAGCTGTGTTGAGCAAATTATTGAAAAAGATGAAGGTCCTTTCTATACCCACCTAGGAGCCGGTCCTAATGTGGCAGCTATTA  3400
                      ********************************************************************************************************
                3301  TATGATTTCCCATCTTGCAGCTGTGTTGAGCAAATTATTGAAAAAGATGAAGGTCCTTTCTATACCCACCTAGGAGCCGGTCCTAATGTGGCAGCTATTA  3400

                3401  GAGAAATCATGGAAGAAAGATTTGGACAGAAGGGTAAAGCTATAAGGATTGAGAGGGTTGTCTACACTGGGAAAGAAGGCAAAAGTTCTCAAGGATGTCC  3500
                      ********************************************************************************************************
                3401  GAGAAATCATGGAAGAAAGATTTGGACAGAAGGGTAAAGCTATAAGGATTGAGAGGGTTGTCTACACTGGGAAAGAAGGCAAAAGTTCTCAAGGATGTCC  3500

                3501  AATTGCTAAATGGGTAGTCCGCAGAAGCAGTCAGGAGGAAAAGCTGCTCTGCCTGGTGCGTGAGCGAGCGGGCCACACGTGCGAGACGGCGGTGATTGTG  3600
                      ********************************************************************************************************
                3501  AATTGCTAAATGGGTAGTCCGCAGAAGCAGTCAGGAGGAAAAGCTGCTCTGCCTGGTGCGTGAGCGAGCGGGCCACACGTGCGAGACGGCGGTGATTGTG  3600

                3601  ATCCTCATCCTGGTTTGGGAGGGAATCCCAACCAGCCTGGCTGACAAGCTCTATTCCGAACTCACCGACACTCTCAGAAAGTACGGCACGCTCACGAACC  3700
                      ********************************************************************************************************
                3601  ATCCTCATCCTGGTTTGGGAGGGAATCCCAACCAGCCTGGCTGACAAGCTCTATTCCGAACTCACCGACACTCTCAGAAAGTACGGCACGCTCACGAACC  3700

                3701  GGCGCTGCGCCCTCAACGAAGAACGGACTTGTGCTTGTCAAGGGCTGGACCCTGAAACTTGTGGTGCTTCATTTTCCTTTGGTTGCTCCTGGAGCATGTA  3800
                      ********************************************************************************************************
                3701  GGCGCTGCGCCCTCAACGAAGAACGGACTTGTGCTTGTCAAGGGCTGGACCCTGAAACTTGTGGTGCTTCATTTTCCTTTGGTTGCTCCTGGAGCATGTA  3800

                3801  CTACAATGGTTGTAAGTTTGCCAGAAGCAAGATTCCAAGAAAGTTTAAGCTGATGGGGGATGATCCTAAAGAGGAAGAAAAACTAGAATCCCATTTGCAG  3900
                      ********************************************************************************************************
                3801  CTACAATGGTTGTAAGTTTGCCAGAAGCAAGATTCCAAGAAAGTTTAAGCTGATGGGGGATGATCCTAAAGAGGAAGAAAAACTAGAATCCCATTTGCAG  3900

                3901  AATCTGTCAACCCTGATGGCACCCACCTACAAGAAGCTTGCACCTGATGCATATAACAACCAGATCGAGTACGAACACAGAGCGCCCGAGTGTCGCCTGG  4000
                      ********************************************************************************************************
                3901  AATCTGTCAACCCTGATGGCACCCACCTACAAGAAGCTTGCACCTGATGCATATAACAACCAGATCGAGTACGAACACAGAGCGCCCGAGTGTCGCCTGG  4000

                4001  GTTTAAAAGAAGGTCGCCCATTCTCAGGGGTCACTGCCTGCCTGGACTTCTGCGCCCATGCTCACAGAGACTTGCACAATATGCAGAACGGGAGTACACT  4100
                      ********************************************************************************************************
                4001  GTTTAAAAGAAGGTCGCCCATTCTCAGGGGTCACTGCCTGCCTGGACTTCTGCGCCCATGCTCACAGAGACTTGCACAATATGCAGAACGGGAGTACACT  4100

                4101  GGTTTGCACACTAACTAGAGAAGACAATCGTGAAATTGGCCAAACACCTGAAGATGAGCAGCTCCACGTGCTCCCCTTGTACAAAGTCTCTGATGTGGAT  4200
                      ********************************************************************************************************
                4101  GGTTTGCACACTAACTAGAGAAGACAATCGTGAAATTGGCCAAACACCTGAAGATGAGCAGCTCCACGTGCTCCCCTTGTACAAAGTCTCTGATGTGGAT  4200

                4201  GAGTTCGGAAGCACTGAAGGCCAGGAGGAGAAGAAGAGGAATGGCAGCATCCAGGTCCTTACCTCCTTTCGTCGGAAAGTAAGGATGTTAGCAGAGCCGG  4300
                      *************************************************************************************************
                4201  GAGTTTGGAAGCACTGAAGGCCAGGAGGAGAAGAAGAGGAATGGCAGCATCCAGGTCCTTACCTCCTTTCGTCGGAAAGTAAGGATGTTAGCAGAGCCGG  4300

                4301  TTAAGACGTGCCGGCAAAGGAAGCTAGAAGCAAAGAAAGCAGCTGCAGAAAAGCTTTCCTCCTTGGAGAATGGGTCTAGCAAAGCTGAAAGAGACAAGTC  4400
                      ********************************************************************************************************
                4301  TTAAGACGTGCCGGCAAAGGAAGCTAGAAGCAAAGAAAGCAGCTGCAGAAAAGCTTTCCTCCTTGGAGAATGGGTCTAGCAAAGCTGAAAGAGACAAGTC  4400

                4401  TGCTGCAGCACGCAACAAACAAGGCAACTCTGAGGCAGCAGGTCATGCAAAGCAGCTAGCAGATCTTTTACGTCTTTCAGGACCAGCGACACAACAGCAG  4500
                      ********************************************************************************************************
                4401  TGCTGCAGCACGCAACAAACAAGGCAACTCTGAGGCAGCAGGTCATGCAAAGCAGCTAGCAGATCTTTTACGTCTTTCAGGACCAGCGACACAACAGCAG  4500

                4501  CAGCAGCATCCACAGCGTACTCTCCCTAACAACCCTCAGTCAAATGCTATTAACACTTACTCGGGTTCAGGTTCTGCAAATCTGTATGTAAGGTTGCCTA  4600
                      ********************************************************************************************************
                4501  CAGCAGCATCCACAGCGTACTCTCCCTAACAACCCTCAGTCAAATGCTATTAACACTTACTCGGGTTCAGGTTCTGCAAATCTGTATGTAAGGTTGCCTA  4600

                4601  ATCCAGCCAGTGCTTATCCAAGCTCTTCATACACTTCAGATCCCTATGGAGGGTCTGGTGCCATGAACCTCTATACCACCTCATCACAGCCTGCGGGGTC  4700
                      ********************************************************************************************************
                4601  ATCCAGCCAGTGCTTATCCAAGCTCTTCATACACTTCAGATCCCTATGGAGGGTCTGGTGCCATGAACCTCTATACCACCTCATCACAGCCTGCGGGGTC  4700

                4701  TTATTTAAATTCTTCCAGTCCCATGAACCCTTATTCTGGATCATTAAGTCAAAATAACCAGTATCCACCCTATCAATGCAATGGAAACATACAGATGGAC  4800
                      ********************************************************************************************************
                4701  TTATTTAAATTCTTCCAGTCCCATGAACCCTTATTCTGGATCATTAAGTCAAAATAACCAGTATCCACCCTATCAATGCAATGGAAACATACAGATGGAC  4800

                4801  AACTGCCCCTCTTACTTGGGCTCTTACCCTTCCCAGCATCAGCACATGGACTTGTATAATTGTCAGAGCCAAGACCCTATGTCCAAACTAAGCCTACCAC  4900
                      ***************************************************************** ************************************
                4801  AACTGCCCCTCTTACTTGGGCTCTTACCCTTCCCAGCATCAGCACATGGACTTGTATAATTGTCAGAGTCAAGACCCTATGTCCAAACTAAGCCTACCAC  4900

                4901  CCATTCAAACATTATACCAGCATAGGTTTGGGAATAACCAGAGTTTTGGTCCCAAGTACTTGAATTACGGAAACCAAAATATGCAGGTAGACTCTTTCAG  5000
                      ********************************************************************************************************
                4901  CCATTCAAACATTATACCAGCATAGGTTTGGGAATAACCAGAGTTTTGGTCCCAAGTACTTGAATTACGGAAACCAAAATATGCAGGTAGACTCTTTCAG  5000

                5001  TAATTGCACCATTAGACCAAATGTACACCACGTAGGGTCTTTTTCTTCTTACTCCACCCACGAGGCTGACGGTCATTTTATGGAGGTTGCCTCAAGGCTA  5100
                      ********************************************************************************************************
                5001  TAATTGCACCATTAGACCAAATGTACACCACGTAGGGTCTTTTTCTTCTTACTCCACCCACGAGGCTGACGGTCATTTTATGGAGGTTGCCTCAAGGCTA  5100

                5101  AAATCTAATCTGAGTAATCCAAGCATGGACTATGCCTCCATGAGTAAAACCACTGAACACCATCACGTGCAACCCCCTCCACATTTAGCACGCGACTACC  5200
                      ********************************************************************************************************
                5101  AAATCTAATCTGAGTAATCCAAGCATGGACTATGCCTCCATGAGTAAAACCACTGAACACCATCACGTGCAACCCCCTCCACATTTAGCACGCGACTACC  5200

                5201  ATTCTGCTTCGAGTATGTTTAGCGGTCCTCCTAATTCGTTGCATCTCCAAAATAAGGATAGTGAAATGATTTCACATGCAGTAAACGGCTTGTCTAACAC  5300
                      *************************************** ************************************************************
                5201  ATTCTGCTTCGAGTATGTTTAGCGGTCCTCCTAATTCATTGCATCTCCAAAATAAGGATAGTGAAATGATTTCACATGCAGTAAACGGCTTGTCTAACAC  5300

                5301  GATTCCAGGTCAGAACCACGATAGGACTACACCCCAAGGTGGTTTAGATAAAACAGATGTGCTGAATCCTGAAAAAGCTGAGGATCCCGATGAAGTCTGG  5400
                      ********************************************************************************************************
                5301  GATTCCAGGTCAGAACCACGATAGGACTACACCCCAAGGTGGTTTAGATAAAACAGATGTGCTGAATCCTGAAAAAGCTGAGGATCCCGATGAAGTCTGG  5400

                5401  TCAGATAGTGAACAGAGCTTCCTGGATCCGGAAATTGGAGGAGTGGCAGTTGCTCCATCTCACGGGTCAATTCTCATAGAGTGTGCGAAACGTGAGCTCC  5500
                      ********************************************************************************************************
                5401  TCAGATAGTGAACAGAGCTTCCTGGATCCGGAAATTGGAGGAGTGGCAGTTGCTCCATCTCACGGGTCAATTCTCATAGAGTGTGCGAAACGTGAGCTCC  5500
```

Fig. S1

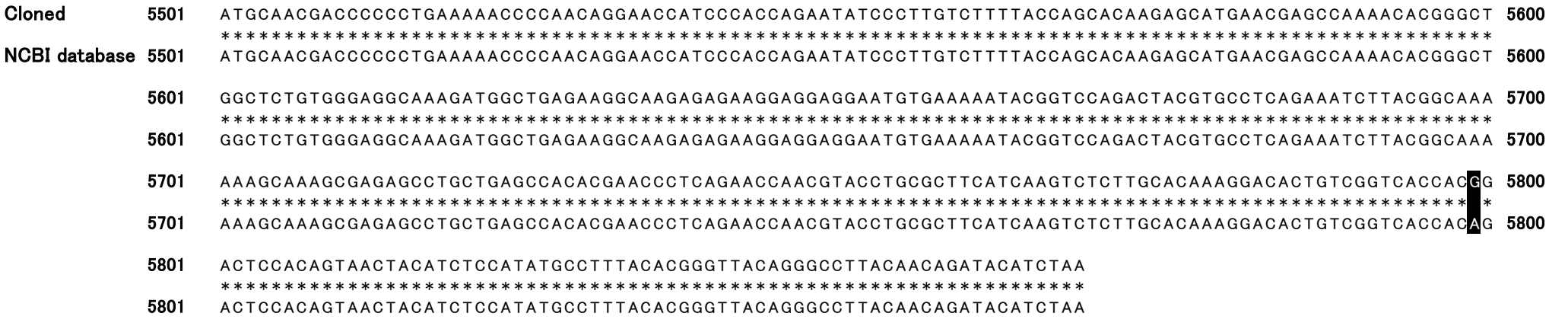
Cloned 5501 ATGCAACGACCCCCCTGAAAAACCCCAACAGGAACCATCCCACCAGAATATCCCTTGTCTTTTACCAGCACAAGAGCATGAACGAGCCAAAACACGGGCT 5600
NCBI database 5501 ATGCAACGACCCCCCTGAAAAACCCCAACAGGAACCATCCCACCAGAATATCCCTTGTCTTTTACCAGCACAAGAGCATGAACGAGCCAAAACACGGGCT 5600
5601 GGCTCTGTGGGAGGCAAAGATGGCTGAGAAGGCAAGAGAGAAGGAGGAGGAATGTGAAAAATACGGTCCAGACTACGTGCCTCAGAAATCTTACGGCAAA 5700
5601 GGCTCTGTGGGAGGCAAAGATGGCTGAGAAGGCAAGAGAGAAGGAGGAGGAATGTGAAAAATACGGTCCAGACTACGTGCCTCAGAAATCTTACGGCAAA 5700
5701 AAAGCAAAGCGAGAGCCTGCTGAGCCACACGAACCCTCAGAACCAACGTACCTGCGCTTCATCAAGTCTCTTGCACAAAGGACACTGTCGGTCACCACGG 5800
5701 AAAGCAAAGCGAGAGCCTGCTGAGCCACACGAACCCTCAGAACCAACGTACCTGCGCTTCATCAAGTCTCTTGCACAAAGGACACTGTCGGTCACCACAG 5800
5801 ACTCCACAGTAACTACATCTCCATATGCCTTTACACGGGTTACAGGGCCTTACAACAGATACATCTAA
5801 ACTCCACAGTAACTACATCTCCATATGCCTTTACACGGGTTACAGGGCCTTACAACAGATACATCTAA

Fig. S1

C

```
Cloned          1  ATGGCTGCCCGGCCCCCCGCCGGCCCCGCGCGCCTGGAGGATGCCCGCAACCTGGTGGCCTTCTCGGCGGTGGCCGAAGCGGTGTCCTCCTACCGGCTGC  100
                   ****************************************.***********************************************************
NCBI database   1  ATGGCTGCCCGGCCCCCCGCCGGCCCCGCGCGCCTGGAGGACGCCCGCAACCTGGTGGCCTTCTCGGCGGTGGCCGAAGCGGTGTCCTCCTACCGGCTGC  100

              101  CCGCCCCGGCTTCGCCCTCGCTGCTCTACAAGAAGTTTGACACCGAGATGAGCCGGGGGGGGGCTGGCGCCGCGGGACGGCGTGCCGAGGGGGGAGGACCT  200
                   ********************************************************************************************************
              101  CCGCCCCGGCTTCGCCCTCGCTGCTCTACAAGAAGTTTGACACCGAGATGAGCCGGGGGGGGGCTGGCGCCGCGGGACGGCGTGCCGAGGGGGGAGGACCT  200

              201  GCACGCCCTCAAGGCCGCCCTCGCCCTGGCCAAGCACGGCATGAAGCCCCCCAACTGCAACTGCGACGGCCCCGAGTGCCCCGACTACCTGGAGTGGCTG  300
                   ********************************************************************************************************
              201  GCACGCCCTCAAGGCCGCCCTCGCCCTGGCCAAGCACGGCATGAAGCCCCCCAACTGCAACTGCGACGGCCCCGAGTGCCCCGACTACCTGGAGTGGCTG  300

              301  GAGCAGAAGATCCAGGCGGCCCTGGGAGAAGGCCCGCTGCCCCCCGCCCCGCCGCCCCCCGACAAAGGTGCTATCGCTCCCCCCGCGGTGCTGGAGGCCG  400
                   ********************************************************************************************************
              301  GAGCAGAAGATCCAGGCGGCCCTGGGAGAAGGCCCGCTGCCCCCCGCCCCGCCGCCCCCCGACAAAGGTGCTATCGCTCCCCCCGCGGTGCTGGAGGCCG  400

              401  CCGAGCCGTGCCCGTCGGACGGCCTCCCCTTTTCCCAAAGTGCACTGAACATCGCCAAGGAGAAGAACATCAGCCTGCAGACGGCCATCGCCATCGAGGC  500
                   ********************************************************************************************************
              401  CCGAGCCGTGCCCGTCGGACGGCCTCCCCTTTTCCCAAAGTGCACTGAACATCGCCAAGGAGAAGAACATCAGCCTGCAGACGGCCATCGCCATCGAGGC  500

              501  GCTCACGCAGCTCTCGGCCGCGCTCCCTCAACCCGGTGCTGATGGGCCGCCCCCCCCGCCACCGCCGCCGCCGCCGCCCCCCTCGGTGCCCCCCACTCCC  600
                   ********************************************.**********.************************************************
              501  GCTCACGCAGCTCTCGGCCGCGCTCCCTCAACCCGGTGCTGATGGGCCG-CCCCCCCGCCGCCGCCGCCGCCGCCCCCCTCGGTGCCCCCCACTCCC  599

              601  CAGGAGCCGTCCATCAGCGCGGTGCCTCCGTCCGGGACCGCCGACCCTATGGCGGAGCTGGAGCAGCTCCTGGGCAGCACCGAATACATCGCCACGGCCT  700
                   ********************************************************************************************************
              600  CAGGAGCCGTCCATCAGCGCGGTGCCTCCGTCCGGGACCGCCGACCCTATGGCGGAGCTGGAGCAGCTCCTGGGCAGCACCGAATACATCGCCACGGCCT  699

              701  TCAAGCGGCCCGAGGCCGGCAGAGCCCCCCCCGCGGCAGCCCCAAAGCCCCCGGATCGGACGCCGGGCAAGGAGGCGGCGAGCAGCCCCCGCGTGCTGCA  800
                   ********************************************************************************************************
              700  TCAAGCGGCCCGAGGCCGGCAGAGCCCCCCCCGCGGCAGCCCCAAAGCCCCCGGATCGGACGCCGGGCAAGGAGGCGGCGAGCAGCCCCCGCGTGCTGCA  799

              801  GGAGCCCGACCTGCACAGGAAGACGCAGCTGGTCCTGCAGCAGCACCTGCACCACAAGCGCAGCCTTTTCCTCGAGCAGAACCTCTCCGCGCCCCCCTCG  900
                   ********************************************************************************************************
              800  GGAGCCCGACCTGCACAGGAAGACGCAGCTGGTCCTGCAGCAGCACCTGCACCACAAGCGCAGCCTTTTCCTCGAGCAGAACCTCTCCGCGCCCCCCTCG  899

              901  GAGCGCCCGCCCGGCTGGTGGACCCCCAGCACGCCCAAACCCTTTGAGAAGCAGGCCAAGGAGAAGAAGAAAAGGATGCAGCCGGACAAGCCGGCCCGCA 1000
                   ********************************************************************************************************
              900  GAGCGCCCGCCCGGCTGGTGGACCCCCAGCACGCCCAAACCCTTTGAGAAGCAGGCCAAGGAGAAGAAGAAAAGGATGCAGCCGGACAAGCCGGCCCGCA  999

             1001  AGCAGGTGCAGATCAAGAAGCCCAAGCAGAAGGATTCGCAGCCGCTCTTCCTGCCCTTCTGGCAAATTAGCCTGGAGGGGCTGCGGGCCCCCGCCGAGCC 1100
                   ********************************************************************************************************
             1000  AGCAGGTGCAGATCAAGAAGCCCAAGCAGAAGGATTCGCAGCCGCTCTTCCTGCCCTTCTGGCAAATTAGCCTGGAGGGGCTGCGGGCCCCCGCCGAGCC 1099

             1101  CCCCGCCTCAGCGCCGCAGTCCGAACCCCCACCGCCCACCCCGCTCCCACCCACCGCTCCCGACTCTCAGGAAAGGGGTACCCCCGGGGGGGGACACCCAC 1200
                   ***.****************************.***************************************************************
             1100  CCCTGCCTCAGCGCCGCAGTCCGAACCCCCGCCGCCCACCCCGCTCCCACCCACCGCTCCCGACTCTCAGGAAAGGGGTACCCCCGGGGGGGGACACCCAC 1199

             1201  AACCACCCGCCGGCACCGGGCCCGGAGGGAGCGGCGCCCGCCGTGGTGGACGACAAGCTGGAGGAGCTCATCCGGCAGTTCGAGGCCGAGTTCGGGGACA 1300
                   ********************************************************************************************.******
             1200  AACCACCCGCCGGCACCGGGCCCGGAGGGAGCGGCGCCCGCCGTGGTGGACGACAAGCTGGAGGAGCTCATCCGGCAGTTCGAGGCCGAGTTC-GGGACA 1298

             1301  CCTTCAGCCTGCCGCCCAGCGGGGCCACCGCGCTGCCGGAGGGGCCGCCGCCAGCCCCACAGGTGCCCCCCACCGCCGCCAGCACCGCCGCCGTGGCCCC 1400
                   ********************************************************.***********************************
             1299  CCTTCAGCCTGCCGCCCAGCGGGGCCACCGCGCTGCCGGAGGGGCCGCCGCCAGCCCCACAGGTG-CCCCCACCGCCGCCAGCACCGCCGCCGTGGCCCC 1397

             1401  CAGCAGCGCCGCCCCGGGGGGCTCCGCGCTGTCGCCGGGGAAGGGACCTCCTCCGGAGCATCTCTTCTCCGTGCGCTCCCCCAAGCAGATCAAAATCGAG 1500
                   ********************************************************************************************************
             1398  CAGCAGCGCCGCCCCGGGGGGCTCCGCGCTGTCGCCGGGGAAGGGACCTCCTCCGGAGCATCTCTTCTCCGTGCGCTCCCCCAAGCAGATCAAAATCGAG 1497

             1501  TCCTCTGGTGCTATCACCGTGGTGTCCACCACGTGCTTTTATTCCGAGGAGAGCCAGAACGCCGACGAGGCGGAGGGGACGCCCACCAAGGATGAGGTGC 1600
                   ********************************************************************************************************
             1498  TCCTCTGGTGCTATCACCGTGGTGTCCACCACGTGCTTTTATTCCGAGGAGAGCCAGAACGCCGACGAGGCGGAGGGGACGCCCACCAAGGATGAGGTGC 1597

             1601  CGCTGACCCCCACCCTCAGCGGGTTCCTGGAGTCGCCGCTCAAATACCTGGACACGCCGACCAAGAGCCTGCTGGACACCCCCGCCAAGCGAGCGCAGGC 1700
                   ********************************************************.***********************************
             1598  CGCTGACCCCCACCCTCAGCGGGTTCCTGGAGTCGCCGCTCAAATACCTGGACACGCCAACCAAGAGCCTGCTGGACACCCCCGCCAAGCGAGCGCAGGC 1697

             1701  GGAGTTCCCCACCTGCGACTGCGTGGAGCAAATCGTGGAGAAGGACGAGGGGCCGTACTACACCCACCTGGGCTCGGGGCCCACCGTGGCGTCCATCCGG 1800
                    ********************************************************************************************************
             1698  CGAGTTCCCCACCTGCGACTGCGTGGAGCAAATCGTGGAGAAGGACGAGGGGCCGTACTACACCCACCTGGGCTCGGGGCCCACCGTGGCGTCCATCCGG 1797

             1801  GAGCTGATGGAGGAGCGGTACGGCGAGAAGGGCAAAGCCATCCGCATCGAGAAGGTCATCTACACTGGGAAGGAGGGGAAGAGCTCCCGGGGCTGCCCCA 1900
                   ********************************************************************************************************
             1798  GAGCTGATGGAGGAGCGGTACGGCGAGAAGGGCAAAGCCATCCGCATCGAGAAGGTCATCTACACTGGGAAGGAGGGGAAGAGCTCCCGGGGCTGCCCCA 1897

             1901  TCGCCAAGTGGGTGATCCGGAGACACAACCAAGAGGAGAAGCTGCTGTGCCTGGTGCGTCACCGAGCTGGCCACCACTGCCAGAATGCTGTCATCATCAT 2000
                   ********************************************************************************************************
             1898  TCGCCAAGTGGGTGATCCGGAGACACAACCAAGAGGAGAAGCTGCTGTGCCTGGTGCGTCACCGAGCTGGCCACCACTGCCAGAATGCTGTCATCATCAT 1997

             2001  CCTGATCCTGGCCTGGGAGGGCATCCCCCGCACTCTGGGCGACACACTGTACCAGGAGCTCACTGACACCCTCACCAAGTATGGCAACCCCACCAGCCGC 2100
                   ********************************************************************************************************
             1998  CCTGATCCTGGCCTGGGAGGGCATCCCCCGCACTCTGGGCGACACACTGTACCAGGAGCTCACTGACACCCTCACCAAGTATGGCAACCCCACCAGCCGC 2097

             2101  CGCTGCGGCTTGAATGACGACCGGACTTGTGCGTGCCAAGGCAAGGACCCCAACACCTGCGGTGCTTCCTTCTCCTTCGGCTGCTCTTGGAGCATGTATT 2200
                   ********************************************************************************************************
             2098  CGCTGCGGCTTGAATGACGACCGGACTTGTGCGTGCCAAGGCAAGGACCCCAACACCTGCGGTGCTTCCTTCTCCTTCGGCTGCTCTTGGAGCATGTATT 2197

             2201  TTAATGGCTGCAAATATGCCCGAAGCAAAACTCCACGGAAGTTCAGGCTGGTGGGGGACAATCCCAAGGAGGAAGAGCTGCTCCGGAAAAGCTTTCAGGA 2300
                   ********************************************************************************************************
             2198  TTAATGGCTGCAAATATGCCCGAAGCAAAACTCCACGGAAGTTCAGGCTGGTGGGGGACAATCCCAAGGAGGAAGAGCTGCTCCGGAAAAGCTTTCAGGA 2297

             2301  CTTGGCCACTGAGGTTGCTCCACTTTACAAGAGGTTGGCACCGCAGGCCTACCAGAATCAGGTTACCAATGAGGACATCGCAATAGACTGCCGTCTGGGC 2400
                   ********************************************************************************************************
             2298  CTTGGCCACTGAGGTTGCTCCACTTTACAAGAGGTTGGCACCGCAGGCCTACCAGAATCAGGTTACCAATGAGGACATCGCAATAGACTGCCGTCTGGGC 2397

             2401  TTGAAGGAGGGGAGGCCGTTTTCAGGGGTGACAGCGTGCATGGACTTCTGTGCTCATGCTCACAAAGATCAACATAACCTCTACAATGGCTGCACAGTGG 2500
                   ********************************************************.***********.************************
             2398  TTGAAGGAGGGGAGGCCGTTTTCAGGGGTGACAGCGTGCATGGACTTCTGTGCTCACGCTCACAAAGATCAGCATAACCTCTACAATGGCTGCACAGTGG 2497

             2501  TCTGCACGCTGACAAAGGAAGACAATCGAGTGGTGGGGAAGATCCCCGAAGATGAGCAGCTGCACGTCCTCCCCCTCTACAAGATGTCCAGCACAGATGA 2600
                   ********************************************************************************************************
             2498  TCTGCACGCTGACAAAGGAAGACAATCGAGTGGTGGGGAAGATCCCCGAAGATGAGCAGCTGCACGTCCTCCCCCTCTACAAGATGTCCAGCACAGATGA 2597

             2601  GTTTGGCAGCGAGGAGAACCAAAACGCAAAGGTGGGCAGCGGGGCCATCCAGGTGCTCACATCCTTCCCCCGTGAGGTGCGTAAGCTGCCTGAGCCTGCC 2700
                   ********************************************************************************************************
             2598  GTTTGGCAGCGAGGAGAACCAAAACGCAAAGGTGGGCAGCGGGGCCATCCAGGTGCTCACATCCTTCCCCCGTGAGGTGCGTAAGCTGCCTGAGCCTGCC 2697
```

Fig. S1

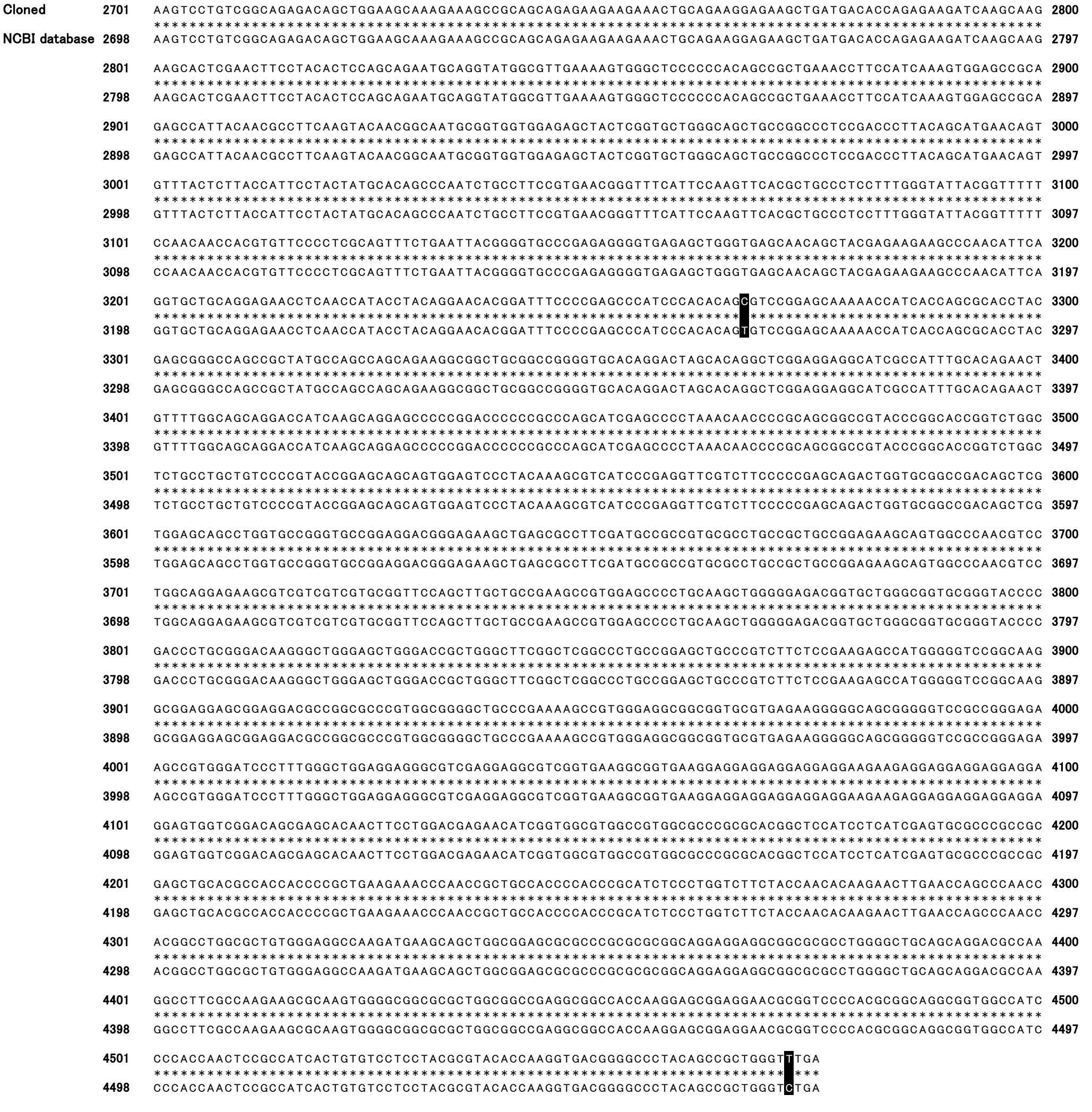
Cloned
NCBI database
2701 AAGTCCTGTCGGCAGAGACAGCTGGAAGCAAAGAAAGCCGCAGCAGAGAAGAAGAAACTGCAGAAGGAGAAGCTGATGACACCAGAGAAGATCAAGCAAG 2800
2698 AAGTCCTGTCGGCAGAGACAGCTGGAAGCAAAGAAAGCCGCAGCAGAGAAGAAGAAACTGCAGAAGGAGAAGCTGATGACACCAGAGAAGATCAAGCAAG 2797
2801 AAGCACTCGAACTTCCTACACTCCAGCAGAATGCAGGTATGGCGTTGAAAAGTGGGCTCCCCCCACAGCCGCTGAAACCTTCCATCAAAGTGGAGCCGCA 2900
2798 AAGCACTCGAACTTCCTACACTCCAGCAGAATGCAGGTATGGCGTTGAAAAGTGGGCTCCCCCCACAGCCGCTGAAACCTTCCATCAAAGTGGAGCCGCA 2897
2901 GAGCCATTACAACGCCTTCAAGTACAACGGCAATGCGGTGGTGGAGAGCTACTCGGTGCTGGGCAGCTGCCGGCCCTCCGACCCTTACAGCATGAACAGT 3000
2898 GAGCCATTACAACGCCTTCAAGTACAACGGCAATGCGGTGGTGGAGAGCTACTCGGTGCTGGGCAGCTGCCGGCCCTCCGACCCTTACAGCATGAACAGT 2997
3001 GTTTACTCTTACCATTCCTACTATGCACAGCCCAATCTGCCTTCCGTGAACGGGTTTCATTCCAAGTTCACGCTGCCCTCCTTTGGGTATTACGGTTTTT 3100
2998 GTTTACTCTTACCATTCCTACTATGCACAGCCCAATCTGCCTTCCGTGAACGGGTTTCATTCCAAGTTCACGCTGCCCTCCTTTGGGTATTACGGTTTTT 3097
3101 CCAACAACCACGTGTTCCCCTCGCAGTTTCTGAATTACGGGGTGCCCGAGAGGGGTGAGAGCTGGGTGAGCAACAGCTACGAGAAGAAGCCCAACATTCA 3200
3098 CCAACAACCACGTGTTCCCCTCGCAGTTTCTGAATTACGGGGTGCCCGAGAGGGGTGAGAGCTGGGTGAGCAACAGCTACGAGAAGAAGCCCAACATTCA 3197
3201 GGTGCTGCAGGAGAACCTCAACCATACCTACAGGAACACGGATTTCCCCGAGCCCATCCCACACAGCGTCCGGAGCAAAAACCATCACCAGCGCACCTAC 3300
3198 GGTGCTGCAGGAGAACCTCAACCATACCTACAGGAACACGGATTTCCCCGAGCCCATCCCACACAGTGTCCGGAGCAAAAACCATCACCAGCGCACCTAC 3297
3301 GAGCGGGCCAGCCGCTATGCCAGCCAGCAGAAGGCGGCTGCGGCCGGGGTGCACAGGACTAGCACAGGCTCGGAGGAGGCATCGCCATTTGCACAGAACT 3400
3298 GAGCGGGCCAGCCGCTATGCCAGCCAGCAGAAGGCGGCTGCGGCCGGGGTGCACAGGACTAGCACAGGCTCGGAGGAGGCATCGCCATTTGCACAGAACT 3397
3401 GTTTTGGCAGCAGGACCATCAAGCAGGAGCCCCCGGACCCCCCGCCCAGCATCGAGCCCCTAAACAACCCCGCAGCGGCCGTACCCGGCACCGGTCTGGC 3500
3398 GTTTTGGCAGCAGGACCATCAAGCAGGAGCCCCCGGACCCCCCGCCCAGCATCGAGCCCCTAAACAACCCCGCAGCGGCCGTACCCGGCACCGGTCTGGC 3497
3501 TCTGCCTGCTGTCCCCGTACCGGAGCAGCAGTGGAGTCCCTACAAAGCGTCATCCCGAGGTTCGTCTTCCCCCGAGCAGACTGGTGCGGCCGACAGCTCG 3600
3498 TCTGCCTGCTGTCCCCGTACCGGAGCAGCAGTGGAGTCCCTACAAAGCGTCATCCCGAGGTTCGTCTTCCCCCGAGCAGACTGGTGCGGCCGACAGCTCG 3597
3601 TGGAGCAGCCTGGTGCCGGGTGCCGGAGGACGGGAGAAGCTGAGCGCCTTCGATGCCGCCGTGCGCCTGCCGCTGCCGGAGAAGCAGTGGCCCAACGTCC 3700
3598 TGGAGCAGCCTGGTGCCGGGTGCCGGAGGACGGGAGAAGCTGAGCGCCTTCGATGCCGCCGTGCGCCTGCCGCTGCCGGAGAAGCAGTGGCCCAACGTCC 3697
3701 TGGCAGGAGAAGCGTCGTCGTCGTGCGGTTCCAGCTTGCTGCCGAAGCCGTGGAGCCCCTGCAAGCTGGGGGAGACGGTGCTGGGCGGTGCGGGTACCCC 3800
3698 TGGCAGGAGAAGCGTCGTCGTCGTGCGGTTCCAGCTTGCTGCCGAAGCCGTGGAGCCCCTGCAAGCTGGGGGAGACGGTGCTGGGCGGTGCGGGTACCCC 3797
3801 GACCCTGCGGGACAAGGGCTGGGAGCTGGGACCGCTGGGCTTCGGCTCGGCCCTGCCGGAGCTGCCCGTCTTCTCCGAAGAGCCATGGGGGTCCGGCAAG 3900
3798 GACCCTGCGGGACAAGGGCTGGGAGCTGGGACCGCTGGGCTTCGGCTCGGCCCTGCCGGAGCTGCCCGTCTTCTCCGAAGAGCCATGGGGGTCCGGCAAG 3897
3901 GCGGAGGAGCGGAGGACGCCGGCGCCCCGTGGCGGGGCTGCCCGAAAAGCCGTGGGAGGCGGCGGTGCGTGAGAAGGGGGCAGCGGGGGTCCGCCGGGAGA 4000
3898 GCGGAGGAGCGGAGGACGCCGGCGCCCCGTGGCGGGGCTGCCCGAAAAGCCGTGGGAGGCGGCGGTGCGTGAGAAGGGGGCAGCGGGGGTCCGCCGGGAGA 3997
4001 AGCCGTGGGATCCCTTTGGGCTGGAGGAGGGCGTCGAGGAGGCGTCGGTGAAGGCGGTGAAGGAGGAGGAGGAGGAGGAAGAAGAGGAGGAGGAGGAGGA 4100
3998 AGCCGTGGGATCCCTTTGGGCTGGAGGAGGGCGTCGAGGAGGCGTCGGTGAAGGCGGTGAAGGAGGAGGAGGAGGAGGAAGAAGAGGAGGAGGAGGAGGA 4097
4101 GGAGTGGTCGGACAGCGAGCACAACTTCCTGGACGAGAACATCGGTGGCGTGGCCGTGGCGCCCGCGCACGGCTCCATCCTCATCGAGTGCGCCCGCCGC 4200
4098 GGAGTGGTCGGACAGCGAGCACAACTTCCTGGACGAGAACATCGGTGGCGTGGCCGTGGCGCCCGCGCACGGCTCCATCCTCATCGAGTGCGCCCGCCGC 4197
4201 GAGCTGCACGCCACCACCCCGCTGAAGAAACCCAACCGCTGCCACCCCACCCGCATCTCCCTGGTCTTCTACCAACACAAGAACTTGAACCAGCCCAACC 4300
4198 GAGCTGCACGCCACCACCCCGCTGAAGAAACCCAACCGCTGCCACCCCACCCGCATCTCCCTGGTCTTCTACCAACACAAGAACTTGAACCAGCCCAACC 4297
4301 ACGGCCTGGCGCTGTGGGAGGCCAAGATGAAGCAGCTGGCGGAGCGCGCCCCGCGCGCGGCAGGAGGAGGCGGCGCGCCTGGGGCTGCAGCAGGACGCCAA 4400
4298 ACGGCCTGGCGCTGTGGGAGGCCAAGATGAAGCAGCTGGCGGAGCGCGCCCCGCGCGCGGCAGGAGGAGGCGGCGCGCCTGGGGCTGCAGCAGGACGCCAA 4397
4401 GGCCTTCGCCAAGAAGCGCAAGTGGGGCGGCGCGCTGGCGGCCGAGGCGGCCACCAAGGAGCGGAGGAACGCGGTCCCCACGCGGCAGGCGGTGGCCATC 4500
4398 GGCCTTCGCCAAGAAGCGCAAGTGGGGCGGCGCGCTGGCGGCCGAGGCGGCCACCAAGGAGCGGAGGAACGCGGTCCCCACGCGGCAGGCGGTGGCCATC 4497
4501 CCCACCAACTCCGCCATCACTGTGTCCTCCTACGCGTACACCAAGGTGACGGGGCCCTACAGCCGCTGGGTTTGA
4498 CCCACCAACTCCGCCATCACTGTGTCCTCCTACGCGTACACCAAGGTGACGGGGCCCTACAGCCGCTGGGTCTGA

Fig. S1

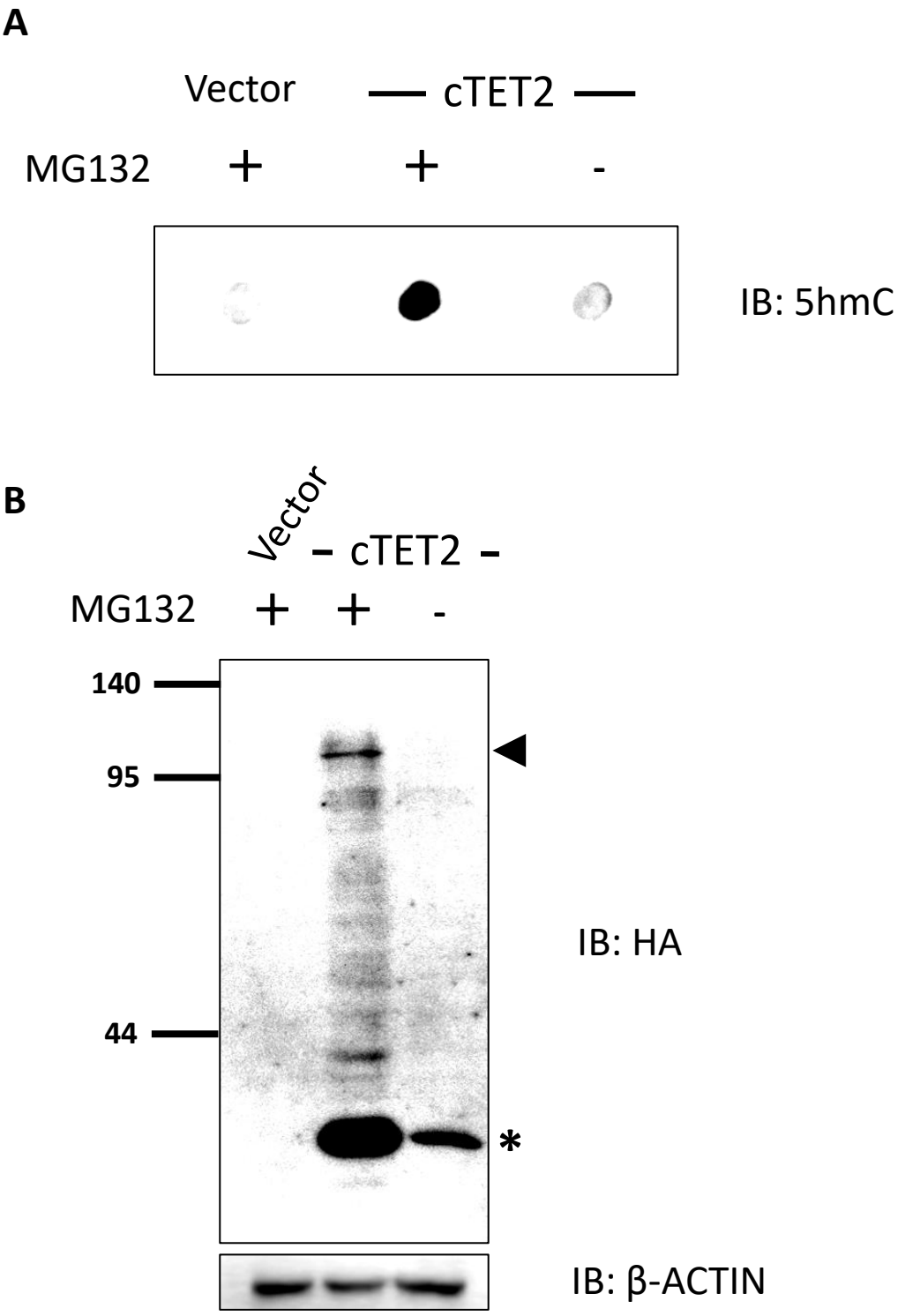
A
Vector
cTET2
MG132
+
+
-
IB: 5hmC
B
Vector
cTET2
MG132
+
+
-
140
95
44
IB: HA
IB: β-ACTIN

Fig. S2

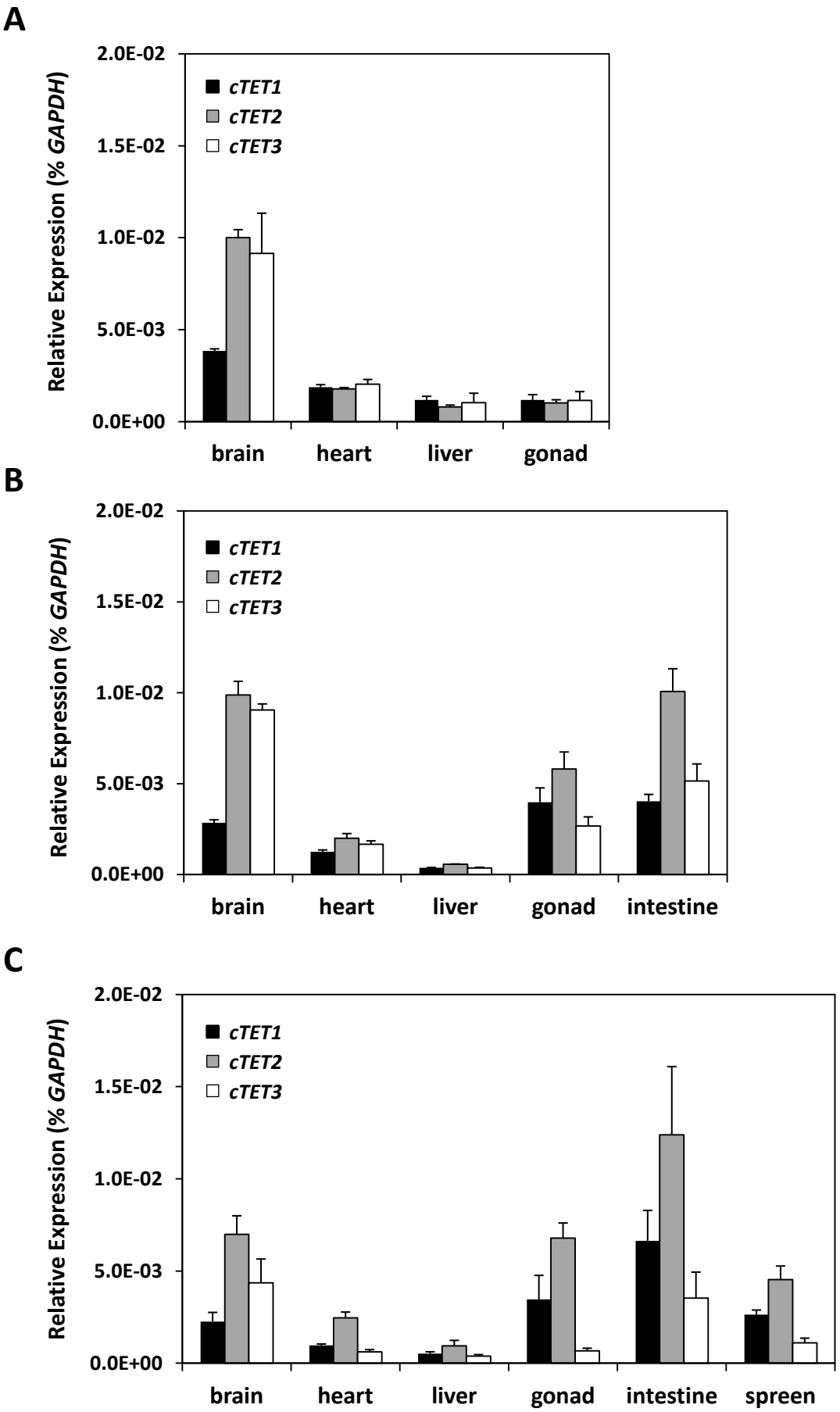
A
2.0E-02
1.5E-02
1.0E-02
5.0E-03
0.0E+00
Relative Expression (% GAPDH)
cTET1
cTET2
cTET3
brain
heart
liver
gonad
B
2.0E-02
1.5E-02
1.0E-02
5.0E-03
0.0E+00
Relative Expression (% GAPDH)
cTET1
cTET2
cTET3
brain
heart
liver
gonad
intestine
C
2.0E-02
1.5E-02
1.0E-02
5.0E-03
0.0E+00
Relative Expression (% GAPDH)
cTET1
cTET2
cTET3
brain
heart
liver
gonad
intestine
spreen

Fig. S3

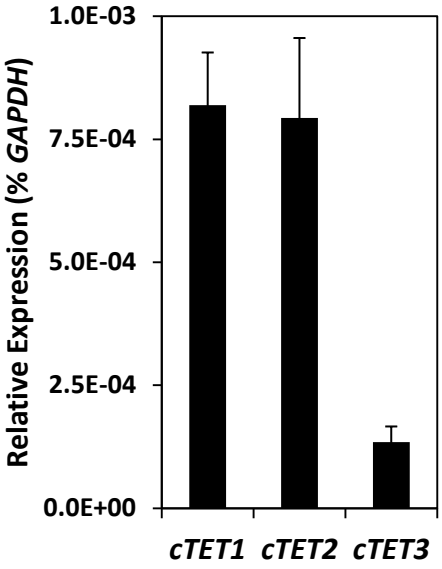
1.0E-03
7.5E-04
5.0E-04
2.5E-04
0.0E+00
Relative Expression (% GAPDH)
cTET1
cTET2
cTET3

Fig. S4

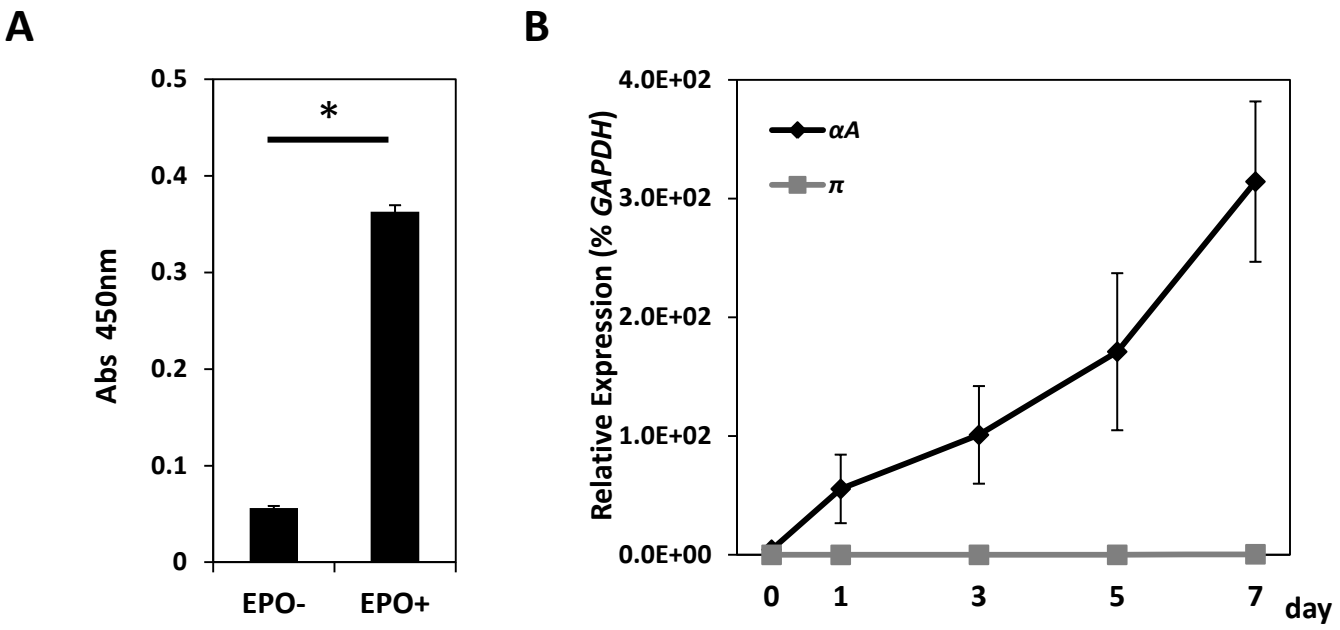
A
B
0.5
0.4
0.3
0.2
0.1
0
Abs 450nm
*
EPO-
EPO+
4.0E+02
3.0E+02
2.0E+02
1.0E+02
0.0E+00
Relative Expression (% GAPDH)
αA
π
0
1
3
5
7
day

Fig. S5

Primer List

Primers for Cloning		
c*TET1* FL	Dir	CATATCGATAATGGCTCACCACGCAAGGC
c*TET1* CD	Dir	CATATCGATAGAATTGCCAACTTGTGACTGTGTTG
c*TET1* Rev	Rev	CATTCTAGACTATGCCCAATGGTTGTAAGGCCC
c*TET2* FL	Dir	CATGAATTCAAATGGAACAGGACAGAACCATCCATG
c*TET2* CD	Dir	CATGAATTCAAGATTTCCCATCTTGCAGCTG
c*TET2* Rev	Rev	CATGCGGCCGCTTAGATGTATCTGTTGTAAAGGCC
c*TET3* FL	Dir	CATGAATTCAATGGCTGCCCGGCCCCCCGCCGGCC
c*TET3* CD	Dir	CATATCGATAGAGTTCCCCACCTGCGA
c*TET3* Rev	Rev	CATTCTAGATCAAACCCAGCGGCTG
m*Tet1* CD	Dir	CATATCGATAGAAGCTGCACCCTGTGACTGTGATG
	Rev	CATGGATCCTTAGACCCAACGATTGTAGGGTCCC

(Restriction enzymes recognition sites were underlined)

Primers for qPCR (expression analysis)		
GAPDH	Dir	GGGCACGCCATCACTATC
	Rev	GTGAAGACACCAGTGGACTCC
cTET1	Dir	CAGGAAGCGCAAAACCAGTC
	Rev	CCTCAAAAGGTAGTGTGA
cTET2	Dir	TCGAGTACGAACACAGAGCG
	Rev	TGCAAACCAGTGTACTCCCG
cTET3	Dir	CAGAATGCAGGGTATGGCGT
	Rev	TCATGCTGTAAGGGTCGGAG
π	Dir	TCACTGGAGAGGCTTTTTGCC
	Rev	GTGGGAAAGCAGCTTGAAGTT
αA	Dir	CCCTGGAAAGGATGTTCACC
	Rev	GGCCCAGGAGTTTGAAGTTG
ρ	Dir	CTTCAGGCTCCTGGGGAACA
	Rev	TCACACTGTGTCCTGCTCTG
βH	Dir	GAGAACTTCAGGCTCCTGGG
	Rev	GAGCATCTCCAAGTGGCTGT
βA	Dir	ACTTCAGGCTCCTGGGTGA
	Rev	GTGATCTTTGGTGCTGGTGC
ε	Dir	GAACTTCAGGCTCCTTGGGG
	Rev	CAACGTTGACCAGCTTCTGC

Primers for qPCR (MeDIP, hMeDIP)		
HS4	Dir	CGGGGAAGGTGGCACG
	Rev	AGCTTTTTCCCCGTATCCCC
ρ promoter	Dir	TGCAGTGAGGACAGCAAGAT
	Rev	TGTGCACAAGGTGTGGTCTT
βA promoter	Dir	CCTCTGGAGATGCAGCCAAT
	Rev	TCTTGCTCCCGTGGGGATA

siRNA

siTET1 #1	guide	UUAUCAUUCAGCAAAAUUCCA
	passenger	GAAUUUUGCUGAAUGAUAAGA
siTET1 #2	guide	CAUCCUUUGCCUCAAACUUUU
	passenger	AAGUUUGAGGCAAAGGAUGUG