 Nucleic Acids Research, 1999, Vol. 27, No. 12 **2503–2510**

Identification of a novel *Sry*-related gene and its germ cell-specific expression

Emiko Osaki, Yukio Nishina[2], Johji Inazawa[3], Neal G. Copeland[4], Debra J. Gilbert[4], Nancy A. Jenkins[4], Miho Ohsugi[1], Tohru Tezuka[1], Mitsuaki Yoshida and Kentaro Semba*

Department of Cellular and Molecular Biology and [1]Department of Oncology, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan, [2]Department of Biology, Graduate School of Integrated Science and Faculty of Science, Yokohama City University, Seto 22-2, Kanazawa-ku, Yokohama 236-0027, Japan, [3]Department of Molecular Cytogenetics, Division of Genetics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan and [4]Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI–Frederick Cancer Research and Development Center, Frederick, MD 21702, USA

Received February 18, 1999; Revised and Accepted April 23, 1999 DDBJ/EMBL/GenBank accession nos AB022083, AB022441

ABSTRACT

Sox family proteins are characterized by a unique DNA-binding domain, a HMG box which shows at least 50% sequence similarity with mouse Sry, the sex-determining factor. At present almost 30 *Sox* genes have been identified. Members of this family have been shown to be conserved during evolution and to play key roles during animal development. Some are involved in human diseases, including sex reversal. Here we report the isolation of a novel member of the *Sox* gene family, *Sox30*, which may constitute a distinct subgroup of this family. Using a bacterially expressed DNA-binding domain of Sox30, we show that it is able to specifically recognize the ACAAT motif. Furthermore, Sox30 is capable of activating transcription from a synthetic promoter containing the ACAAT motif. The specific expression of *Sox30* in normal testes, but not in maturing germ cell-deficient testes, suggests the involvement of *Sox30* in differentiation of male germ cells. Mapping analyses revealed that the *Sox30* gene is located on human chromosome 5 (5q33) and on mouse chromosome 11.

INTRODUCTION

The high mobility group (HMG) box was originally identified in the RNA polymerase I transcription factor UBF as a region of homology to the HMG-1 protein (1). This motif consists of ~80 amino acid residues and binds to the minor groove of DNA, resulting in the induction of a dramatic bend within the DNA. HMG box-containing proteins can be classified into two major groups based on the degree of sequence specificity of the DNA binding and the number of HMG boxes within a protein. One group includes UBF, HMG-1 and MT-TF1, which have multiple HMG boxes and recognize DNA with low or no sequence specificity. The other group includes LEF-1, STE-11 and Sox proteins, which possess a single HMG box and show sequence-specific DNA binding (2,3). The Sox (SRY-related HMG box) family proteins are defined by a HMG box which shows >50% sequence similarity with the founding member of this group, mouse Sry. At present almost 30 members have been reported which can be classified into seven subgroups based on their structural similarity (4,5). In addition to the DNA bending activity of the HMG box, several Sox proteins, such as Sox4, Sox9, Sox17, Sox18 and Sox24, have been shown to activate transcription via their own activation domains (6–11).

The *Sox* genes are conserved during evolution and have been found to play key roles in decisions of cell fate during diverse developmental processes (4). The first demonstration of this concept came from the cloning of *SRY*, the gene responsible for human sex reversal (12–15). Shortly thereafter, regulation of male development by expression of *Sry* was clearly demonstrated in karyotypic female mice carrying the mouse *Sry* gene (16). Other members of the *Sox* family, *Sox9* and *Sox10*, have been shown to be associated with the skeletal malformation syndrome campomelic dysplasia and one cause of Waardenburg–Hirschsprung disease, respectively (17–22). *Drosophila Sox* gene *Dichaete* (*fishhook*) is involved in segmentation and central nervous system development (23–26) and *COG-2*, a *Sox* gene in *Caenorhabditis elegans* is required to establish a functional vulval–uterine connection (27). In vertebrates, *Sox5*, *Sox6* and *Sox9* cooperatively regulate chondrogenesis whereas *Sox4* is involved in B cell and endocardial ridge differentiation (28,29). *Xenopus* homologs of *Sox17* are involved in endoderm differentiation in embryos (30). Some *Sox* genes, including *Sox5* and *Sox17*, are differentially expressed during spermatogenesis and thus are believed to regulate differentiation of male germ cells (9,31).

*To whom correspondence should be addressed. Tel: +81 3 5449 5278; Fax: +81 3 5449 5421; Email: ksemba@ims.u-tokyo.ac.jp

Sox proteins are known to activate transcription synergistically with other transcription factors. In lens differentiation, several Sox proteins seem to require the presence of another lens factor in order to regulate transcription of the δ1-crystallin gene (32–34). Sox2 is also involved in the regulation of FGF-4 transcription via protein–protein interaction with Oct-3 (35,36). This Sox–Oct interaction has also been implicated in glial cell differentiation and seems to be conserved during evolution (26,37,38). Sox9 and steroidgenic factor-1 (SF-1) also synergistically activate the anti-Müllerian hormone gene whose product is crucial for proper male sexual differentiation (39).

Here we report the identification of a novel member of the *Sox* gene family, *Sox30*, which is exclusively expressed in testis. Sequence and biochemical analyses indicate that the *Sox30* gene encodes a sequence-specific transcriptional activator which is relatively divergent from other Sox proteins. The germ cell-specific expression of *Sox30* implicates its involvement during spermatogenesis.

MATERIALS AND METHODS

Materials

pLexA-WT1(+/+), a bait for yeast two-hybrid screening, was constructed by inserting the *Sac*II–*Pst*I fragment of pRcCMV-WT1(+/+) containing the WT1 coding sequence (a gift from T. Akiyama) into pBTM116. pRN3myc was generated by inserting a synthesized oligonucleotide containing a myc tag sequence, initiation ATG and *Xho*I site into pRN3 (a gift from P. Lemaire and J.B. Gurdon) (40). For plasmid vectors used in *in vitro* transcription and expression in mammalian cells, the sequence surrounding the initiation ATG in the human *Sox30* cDNA was mutated to a *Sal*I site using Kunkel's method (41). pRN3myc-Sox30 was generated by cloning the *Sal*I–*Not*I fragment containing the coding sequence of human *Sox30* into pRN3myc. Mammalian expression vectors were constructed by cloning the *Sal*I–*Not*I fragment into pME18S-mycII or pME18S-FlagII which carries the SRα promoter (a gift from J. Inoue). For reporter gene assay, one, four or eight copies of Sox30 binding sequence were inserted upstream of dN-luc, which contains an enhancerless promoter of human T cell leukemia virus type 1. The oligonucleotides used were 5'-GGGGAGACAATGGGACAAT-GGCGAGACAATGGGACAAT-3' and 5'-CCCATTGTCCCATTGTCTCGCCATTGTCCCATTGTCTC-3'. A reporter plasmid containing one copy of the Sox30 binding site was isolated by insertion of a partial oligonucleotide, GAGACAATGGGACAGGG. A glutathione *S*-transferase (GST) fused Sox30 expression vector was constructed by inserting the 362 bp *Eco*RI–*Sfi*I fragment of m125-31 cDNA into pGEX-4T3 (Amersham Pharmacia Biotech). The resulting construct encodes a GST fusion of 121 amino acid residues of mouse Sox30 which does not show significant homology with any other proteins. This fusion protein was bacterially expressed, purified using glutathione–Sepharose (Amersham Pharmacia Biotech) and then used for immunization of rabbits and subsequent production of Sox30-specific antiserum.

Library screening and sequencing

Yeast strain L40 was transformed with pLexA-WT1(+/+) and subsequently with a Gal4 activation domain-fused mouse testis cDNA library. A total of 3×10^6 clones were screened and four clones were found to interact specifically with WT1 in yeast. The insert of one clone, m125, was used to screen a human testis 5'-stretch cDNA library (Clontech). The sequences of both strands were determined with an automated DNA sequencer 373A (PE Applied Biosystems) and analyzed by the Genetyx-Mac (Software Development Co., Japan) and Clustal X multiple sequence alignment programs.

Interspecific mouse backcross mapping

Interspecific backcross progeny were generated by mating (C57BL/6J × *Mus spretus*) F_1 females and C57BL/6J males as described (42). A total of 205 N_2 mice were used to map the *Sox30* locus (see text for details). The probes and RFLPs for the loci linked to *Sox30*, including *Adra1a* and *Il3*, have been reported previously (43). Recombination distances were calculated using Map Manager v.2.6.5. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

Northern blot analysis

Poly(A)$^+$ RNA was isolated from organs of adult mice using Isogen (Nippongene, Japan) and Oligotex-dT30 (Takara Biomedicals). Poly(A)$^+$ RNAs (2 μg) or total RNAs (5 or 10 μg) were electrophoresed in 1% agarose gels containing 2.2 M formaldehyde and subjected to blot hybridization.

RT–PCR analysis

RNA was isolated from embryonic tissues and treated with DNase I. cDNA was synthesized using SuperScript II reverse transcriptase (Gibco BRL). PCR reactions were carried out in 20 μl containing cDNA prepared from 10 ng of RNA with primer pairs for mouse *Sox30* or *Hprt*. *Sox30* primers were 5'-CGGTTCTCCTTTCATCACCC-3' and 5'-CCAAGGCTCCAATGTCCAGA-3'. *Hprt* primers were 5'-CCTGCTGGATTACATTAAAGCACTG-3' and 5'-GTCAAGGGCATATCCAACAACAAAC-3'. Cycle conditions for amplification of *Sox30* were: 95°C for 9 min; 35 or 38 cycles of 95°C for 30 s, 57°C for 1 min, 72°C for 1 min; 72°C for 10 min. Cycle conditions for amplification of *Hprt* were: 95°C for 9 min; 35 cycles of 95°C for 30 s, 59°C for 1 min, 72°C for 1 min; 72°C for 10 min. These conditions were optimized for AmpliTaq Gold (PE Applied Biosystems). In the initial experiments, cycles were changed from 25 to 40 to determine the appropriate number of cycles for semi-quantitative PCR.

Determination of the binding consensus sequence

According to Inaba *et al.* (44), binding sequences for GST–Sox30 were selected from a pool of random DNA sequences with a modified binding buffer [10 mM HEPES–KOH pH 7.7, 60 mM KCl, 1 mM EDTA, 1.5 mM $MgCl_2$, 1 mM DTT, 0.1% NP-40, 5% glycerol, 0.16 mg/ml BSA, 0.16 mg/ml poly(dG-dC)·poly(dG-dC)]. In a parallel experiment, GST and GST–SRY were also used as controls. The random sequence library was prepared by PCR amplification of CGCGGATCCTGCAGCTCGAG(A/G/C/T)$_{30}$GTCGACAAGCTTCTAGAGCA with primers of the first and last 20 bases. After six

repetitions of selection/amplification, amplified DNAs were digested with *Bam*HI and *Hin*dIII and cloned into pBluescript SK(–). Individual clones were sequenced and aligned.

Electrophoretic mobility shift assay (EMSA)

Full-length Sox30 protein was prepared using mMESSAGE mMACHINE (Ambion) and the rabbit reticulocyte lysate system (Promega). The oligonucleotides used were: wt sense, 5'-GGGGCTAAACTGAGGGTAACAATGGTCATT-3'; wt antisense, 5'-CTAGAATGACCATTGTTACCCTCAGTTTAG-3'; mu sense, 5'-GGGGCTAAACTGAGGGTAGTGGCGGTCATT-3'; mu antisense, 5'-CTAGAATGACCGCCACTACCCTCAGTTTAG-3'. Fifteen nanograms of annealed wild-type fragment were end-labeled with [γ-^{32}P]ATP. For EMSA, 3 µl of programmed reticulocyte lysate and end-labeled probe (30 000 c.p.m.) were incubated in 15 µl of binding buffer [10 mM HEPES–KOH pH 7.7, 60 mM KCl, 1 mM DTT, 12% glycerol, 0.16 mg/ml BSA and 0.16 mg/ml poly(dG-dC)·poly(dG-dC)] on ice for 30 min. Samples of 3 µl were resolved on a non-denaturing 4% polyacrylamide gel (acrylamide:bisacrylamide 30:1) in 0.5× Tris/borate/EDTA buffer at room temperature. After electrophoresis, gels were dried and exposed to Kodak XAR film for 4 days with an intensifying screen at –70°C.

Luciferase assays

CV-1 cells were transfected by the calcium phosphate co-precipitation method with 2 µg of reporter plasmid and 5 µg of expression vector. Sonicated salmon testis DNA was added to adjust the total amount of DNA to 15 µg. Two days later, cell extracts were prepared using Cell Culture Lysis Reagent (Promega). The luciferase activities were measured with a Berthold Lumat luminometer, LB9501. Values were normalized for protein concentration using Bio-Rad Protein Assay reagent (Bio-Rad).

Immunoblotting

Nuclear and cytoplasmic extracts corresponding to the same number of cells were resolved on a 10% SDS–polyacrylamide gel and blotted onto Immobilon (Millipore) as previously described (45). Tagged Sox30 protein was detected with anti-Flag monoclonal antibody (Kodak) and an ECL detection system (Amersham Pharmacia Biotech).

RESULTS

Structural features of a novel member of the Sox gene family, Sox30

During the screening and characterization of cDNAs encoding WT1 (Wilms' tumor suppressor gene)-associated proteins by yeast two-hybrid screening, we isolated a clone, m125, from a mouse testis cDNA library which seemed to encode a novel HMG box-containing protein. We subsequently screened a human testis cDNA library using this murine cDNA and obtained overlapping clones, h125-13 and h125-17, which appeared to contain the entire coding sequence. Another cDNA clone, h125-15, revealed the possibility of alternative splicing in this gene. Further screening and RT–PCR analysis confirmed that two types (types I and II) of h125 mRNAs exist in humans, encoding polypeptides of 753 and 501 amino acid residues, respectively (Fig. 1A). Sequence analysis of the genomic structure of h125 revealed that exon skipping occurs in type II, resulting in a frameshift (data not shown). The h125 HMG box showed <50% similarity to the HMG boxes of other Sox proteins and no significant homology was found outside the HMG box. We therefore concluded that h125 is a novel member of the *Sox* gene family and termed this gene *Sox30*. To date, the classification of *Sox* genes into seven groups is based upon the structural resemblance of the HMG boxes within a subgroup (over 75% similarity) (2,4,5,46). This suggests that *Sox30* constitutes a novel subgroup within the *Sox* gene family. Alignment of the coding sequences confirmed that *Sox30* is distantly related to other *Sox* genes (Fig. 1B).

Chromosomal localization of the *Sox30* gene

To investigate whether *Sox30* may be involved in any known diseases or mutations, the chromosomal localizations of human and mouse *Sox30* were determined. For mouse, interspecific backcross analysis with progeny derived from matings of (C57BL/6J × *M.spretus*) F_1 × C57BL/6J mice was utilized. The interspecific backcross mapping panel has been typed for over 2400 loci that are distributed among all autosomes as well as the X chromosome (42). C57BL/6J and *M.spretus* DNAs were digested with several enzymes and analyzed by Southern blot hybridization of informative restriction fragment length polymorphisms using a mouse *Sox30* cDNA probe. Fragments of 7.9 and 4.6 kb were detected in *Eco*RV-digested C57BL/6J DNA and fragments of 7.2 and 4.1 kb were detected in *Eco*RV-digested *M.spretus* DNA. Thus, the 7.2 and 4.1 kb *Eco*RV *M.spretus* RFLPs were used to follow segregation of the *Sox30* locus in backcross mice (data not shown). The mapping results indicated that *Sox30* is located in the central region of mouse chromosome 11 linked to *Adra1a* and *Il3*. A total of 114 mice were analyzed for all markers as shown in the segregation analysis; however, up to 167 mice were typed for specific pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice were determined for each pair of loci and the most likely gene order is as follows: centromere–*Adra1a*–2/167–*Sox30*–8/127–*Il3*. The recombination frequencies (expressed as genetic distances in cM ± SE) are –*Adra1a*–1.2 ± 0.8–*Sox30*–6.3 ± 2.2–*Il3* (Fig. 2A).

The chromosomal localization of the human *Sox30* gene was determined by fluorescent *in situ* hybridization using human *Sox30* cDNA as a probe (47). In more than 50 cells analyzed, the only site of specific fluorescent signal (>15%) was observed on 5q33 (Fig. 2B). The 5q region of the human chromosome has been shown to share homology with the central region of mouse chromosome 11. Therefore, the human chromosomal location of *Sox30* is consistent with that of the mouse.

Expression of the *Sox30* gene

The expression of *Sox30* was first analyzed by northern blot hybridization using mouse tissue RNAs. A 3 kb transcript was observed only with testis RNA (Fig. 3A). A testis-restricted pattern of expression was also seen using human multiple tissue cDNA panels (Clontech; data not shown). No expression of *Sox30* was observed in placenta, thymus, colon or peripheral blood leukocytes in humans. *Sox30* transcripts were scarcely

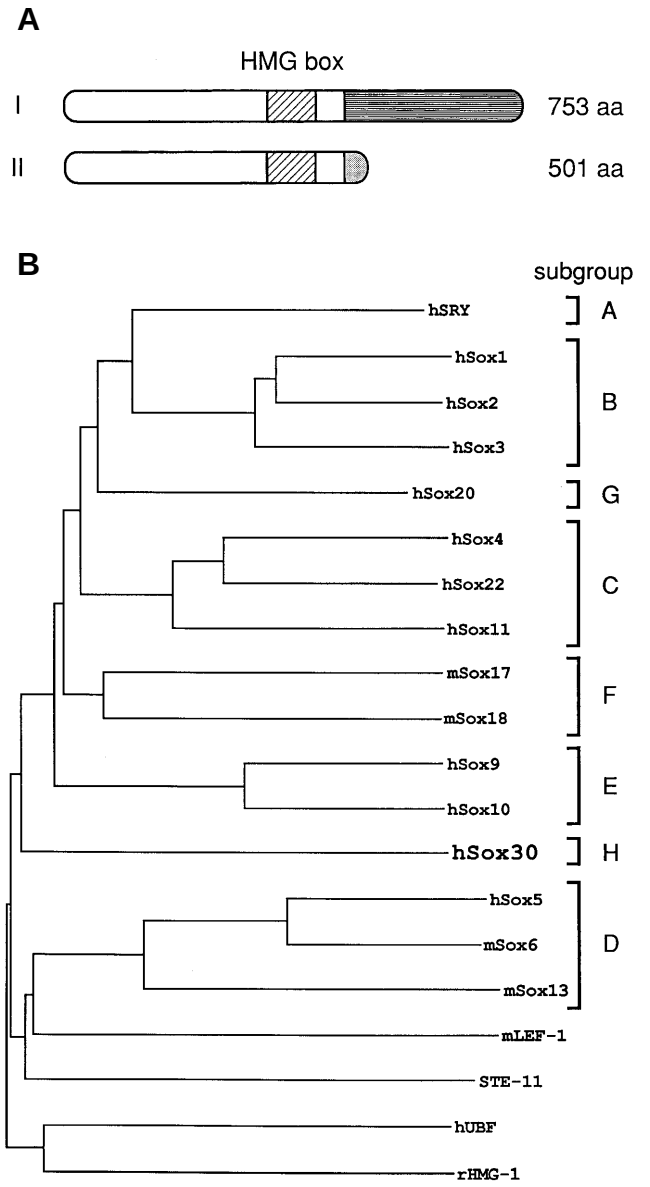


Figure 1. Structure of a novel member of the Sox family, Sox30. (**A**) Schematic illustration of the human Sox30 proteins. Two splice variants, types I and II, were identified by cDNA screening and RT–PCR. The N-terminal regions and the HMG box (indicated by a shaded box) are identical. The C-terminal sequences differ due to a frameshift (see text). The nucleotide sequences of *Sox30* have been submitted to the DDBJ, EMBL and GenBank databases with accession nos. AB022083 (type I) and AB022441 (type II). (**B**) Phylogenetic tree of the Sox family. Multiple alignment was performed by means of Clustal X (v.1.64b from ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/). The phylogenetic tree was drawn using njplot, included in the software package Clustal X. Full-length amino acid sequences of human (h) and mouse (m) Sox proteins were compared. Mouse sequences were used when the human orthologous sequences were not available. Note that Cremazy *et al*. (5) categorized Sox20 into subgroup B2 because of its significant similarity with Sox proteins in group B.

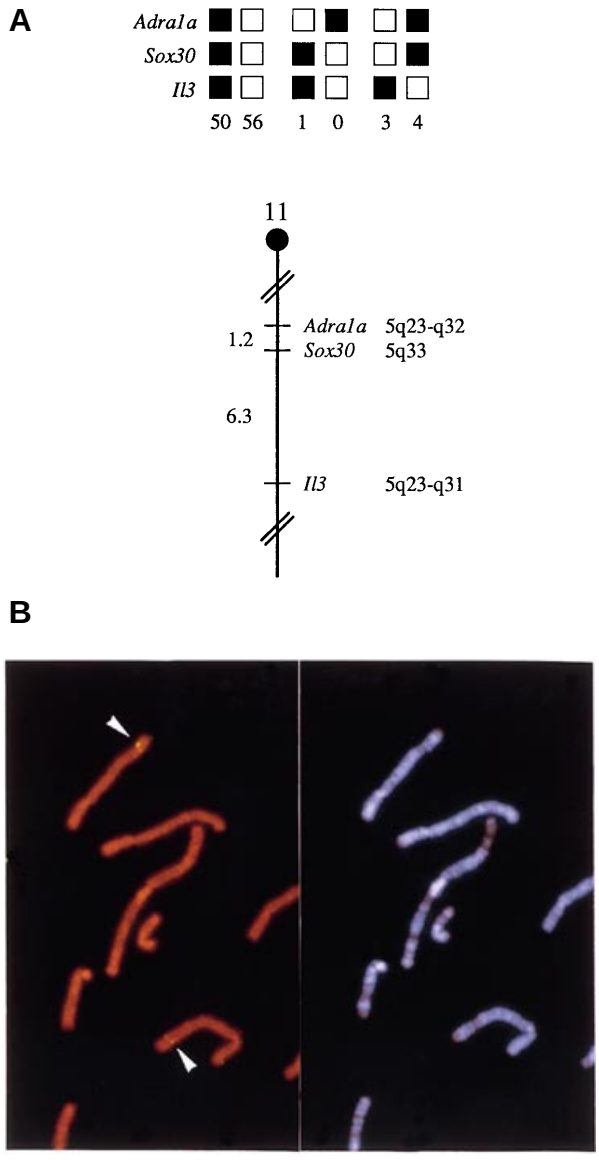


Figure 2. Chromosomal assignment of the *Sox30* gene. (**A**) Chromosomal mapping of mouse *Sox30*. The segregation patterns of *Sox30* and flanking genes in 114 backcross animals typed for all loci are shown at the top of the figure. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M.spretus*) F_1 parent. The shaded box represents the presence of a C57BL/6J allele and white boxes represent the presence of a *M.spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 11 linkage map showing the location of *Sox30* in relation to linked genes is shown at the bottom of the figure. Recombination distances between loci in cM are shown to the left of the chromosome and the positions of loci in human chromosomes are shown to the right. (**B**) Chromosomal mapping of human *Sox30*. Metaphase chromosomes were stained with a propidium iodide-labeled human *Sox30* cDNA probe. Twin spot signals on the long arm of chromosome 5 are indicated by arrows (left). The G-banding patterns of the same chromosomes are also shown (right).

detectable in testes of mice aged 0–4 and 8 days. The level of expression increased significantly by 16 days after birth and continued to increase thereafter. No transcripts were detected in either artificial cryptorchid or *jsd*/*jsd* mouse testes (Fig. 3B). Up until post-natal day 8, mouse germ cells are still at the stage of spermatogonia with no differentiating cells present. After

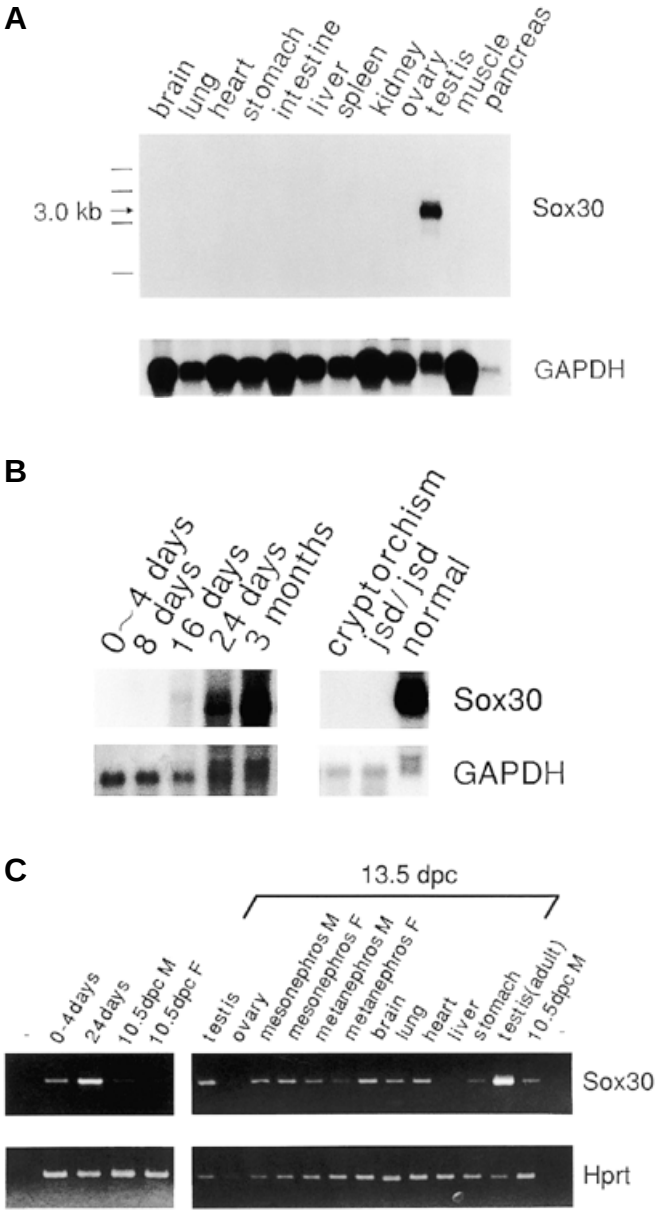


Figure 3. Expression of *Sox30* mRNA. (**A**) Expression in adult mouse tissues. Aliquots of 2 µg of poly(A)$^+$ RNAs were subjected to blot hybridization. Bars indicate mobility of molecular weight markers of 7.46, 4.40, 2.37 and 1.35 kb (RNA Ladder; Gibco BRL). The arrow indicates the mouse *Sox30* transcript. The autoradiogram was exposed for 5 h with an intensifying screen at –70°C. (**B**) Expression in normal and mutant mouse testis. *Sox30* expression was analyzed during male germ cell development (left) and in maturing germ cell-deficient testes (right). Aliquots of 10 (left) or 5 µg (right) of total RNAs were subjected to blot hybridization, respectively. The autoradiogram was exposed for 15 h with an intensifying screen at −70°C. (**C**) Expression in mouse embryonic tissues. Total RNAs from embryonic tissues were subjected to reverse transcription. Separate PCR reactions were performed using primer pairs for *Sox30* (upper) or *Hprt* (lower). Cycles were 35 except for the right panel of Sox30 (38 cycles). M, male; F, female.

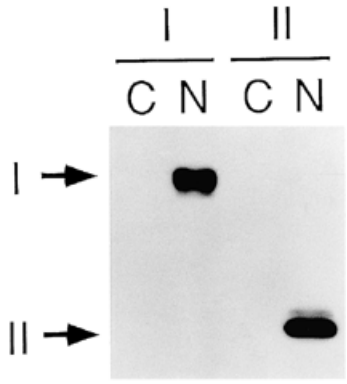


Figure 4. Subcellular localization of type I and type II Sox30 proteins. The amount of Flag-tagged Sox30 protein in nuclear and cytoplasmic extracts was determined by immunoblotting using anti-Flag antibodies. Nuclear (N) and cytoplasmic (C) extracts loaded on the gel are from the same number of cells.

that, primary spermatocytes appear and reach an early pachtene stage by day 14 (48). In artificial cryptorchid testes and in *jsd*/*jsd* mouse testes, only type A spermatogonia and Sertoli cells are found in seminiferous tubules due to the inability of type A spermatogonia to differentiate (49). These results suggested that *Sox30* is specifically expressed in male germ cells and that the increase in its expression may correlate with the later stage of pachtene spermatocytes in which meiosis occurs.

We also analyzed the expression of *Sox30* in 10.5 and 13.5 d.p.c. embryos by RT–PCR (Fig. 3C). *Sox30* transcripts could be detected in embryonic testis, kidney (mesonephros and metanephros), brain, lung, heart and stomach at 13.5 d.p.c. and in the whole embryo at 10.5 d.p.c. However, the level of expression was considerably lower compared with that observed in neonatal testis at 0–4 days (see also Fig. 3B).

Sequence-specific DNA-binding activity of Sox30

We characterized the biochemical properties of Sox30 protein and found that, similarly to other Sox proteins, Sox30 protein is exclusively localized in cell nuclei, as shown by western blotting analysis (Fig. 4). Next we tested whether Sox30 exhibits sequence-specific DNA binding. After incubation of random oligonucleotides with GST–Sox30 fusion protein, oligonucleotides which bound to GST–Sox30 were collected by glutathione–Sepharose beads. After initial amplification of the selected oligonucleotides by PCR, this binding and collection step was repeated. The selection process was repeated six times and subsequently the amplified oligonucleotides were cloned and sequenced. An ACAAT sequence was observed in all of the sequenced clones (Table 1). In parallel experiments, use of GST–SRY resulted in the amplification of an AACAAT motif, consistent with previous reports (data not shown; 50). In contrast to other Sox members, Sox30 appears to prefer a guanine residue surrounding the core ACAAT motif (9,31,50). We tested the sequence-specific DNA-binding ability of the full-length Sox30 protein by EMSA using one of the amplified oligonucleotides as a probe. Incubation with either the GST–Sox30 fusion or the full-length *in vitro* translation product of *Sox30* resulted in a specific DNA–protein complex (Fig. 5 and data not shown). This complex formation was inhibited by addition of an excess amount of wild-type oligonucleotide but

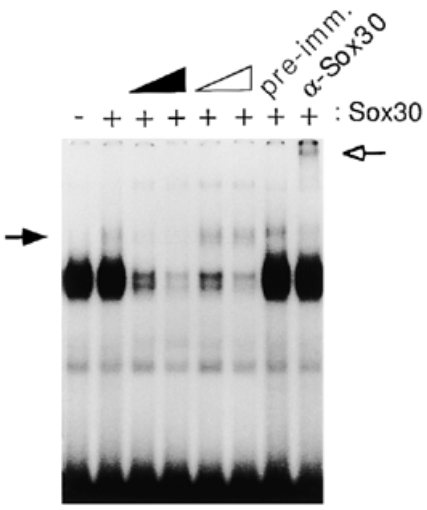


Figure 5. Sequence-specific binding by Sox30 to the ACAAT motif. *In vitro* translated full-length Sox30 protein was subjected to EMSA. Unlabeled competitor oligonucleotides were included (indicated by triangles) in binding reactions at 100- or 300-fold molar excess over probe oligonucleotide. An aliquot of 1 µl of either antiserum or preimmune serum was added before addition of probe. A filled triangle indicates addition of wild-type ACAAT-containing oligonucleotide, while an open triangle indicates addition of mutant oligonucleotide (see Materials and Methods). A filled arrow indicates a Sox30–DNA complex and an open arrow indicates a supershifted Sox30–DNA complex resulting from addition of anti-Sox30 antibody.

not by the mutant. Furthermore, anti-Sox30 antibody was able to induce a supershift of this complex (Fig. 5). These results indicated that Sox30 specifically binds the ACAAT motif. This binding was observed when poly(dG-dC) was used as a non-specific competitor but it was inhibited by poly(dI-dC) (data not shown). Such inhibition has also been reported in the case of Sox9 (7). This may be due to the fact that the HMG box contacts several A-T pairs in the minor groove of the DNA helix, as I-C pairs are known to mimic T-A pairs in the minor groove, whereas G-C pairs do not.

Table 1. Consensus sequence for the Sox30 binding site

	–3	–2	–1	A	C	A	A	T	1	2	3
A	18	17	20	**55**		**55**	**55**		7	6	15
G	20	**27**	**30**						**37**	**36**	22
C	10	7	3		**55**				9	11	10
T	7	4	2					**55**	2	2	8

The core binding sequence ACAAT and its flanking sequences were compared and scored.

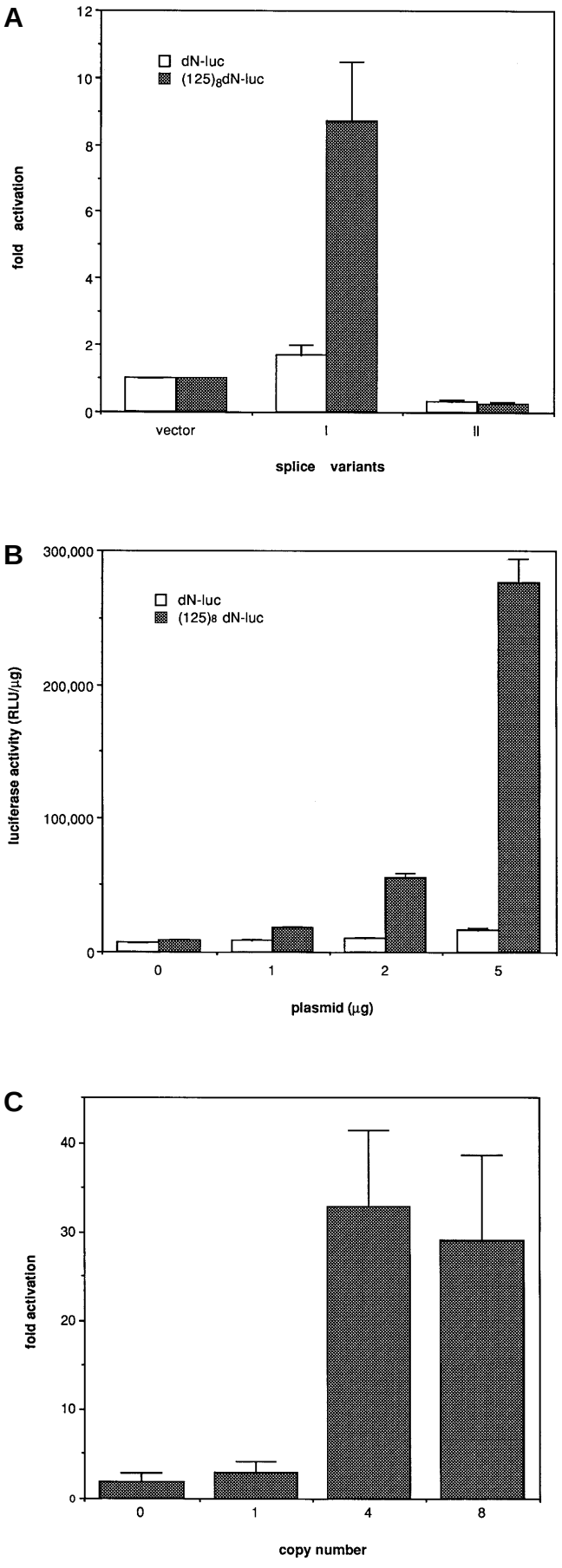
Transcriptional activation by Sox30

We then tested the transcriptional activity of Sox30. For this purpose, we used reporter plasmids containing one, four and eight copies of the Sox30-binding sequence. Co-transfection of a reporter containing eight copies of the Sox30-binding sequence with either type I or type II Sox30 expression vector into CV-1 cells revealed that only type I was able to activate transcription from the reporter (Fig. 6A). This activation appears dependent on Sox30-binding sequence as transcription from the control reporter was not activated significantly. The level of activation was dependent on the amount of the expression vector and on the copy number of the Sox30-binding sequence (Fig. 6B and C). Four copies of the Sox30-binding site were sufficient to observe activation by Sox30 (Fig. 6C). Transcriptional activation by Sox30 type I was observed not only in CV-1 cells but also in other cells such as CHO and JEG-3 cells (data not shown). In contrast, the shorter variant of Sox30 (type II) was not capable of activating transcription, although its expression level and nuclear localization were comparable to type I (Fig. 4B), indicating that the C-terminus of Sox30 is required for its activity. However, the C-terminal sequence alone was not sufficient to promote transcriptional activation when tested in the form of a fusion to the Gal4 DNA-binding domain. Furthermore, neither expression of the Gal4-fused N-terminal sequence alone nor co-expression of N- and C-terminal sequences could activate transcription (data not shown), indicating a requirement for an intact overall structure for activation of transcription by Sox30.

DISCUSSION

In this paper we isolated and characterized a novel *SRY*-related gene, *Sox30*, which encodes a sequence-specific transcriptional activator. To date, *Sox* genes are categorized into seven subgroups on the basis of their homology within the HMG box (4,5). Due to the heterogeneity of the Sox30 HMG box which distinguishes it from other Sox proteins and the lack of any significant similarity outside of the HMG box, we classified *Sox30* into a novel subgroup H: A, *SRY*; B, *Sox1–3* and *14*, zebrafish *Sox19*, *Dichaete* (*fishhook*); C, *Sox4*, *11*, *19* and *22*, mouse *Sox12*, rainbow trout *Sox24*; D, *Sox5*, *6* and *13*, rainbow trout *Sox23* and *SoxLZ*, *Xenopus Sox12*; E, *Sox8–10* and *21*; F, *Sox7*, *17* and *18*; G, *Sox12*, *15*, *16* and *20*; H, *Sox30*. The nomenclature is somewhat misleading and it should be noted that human *Sox12* belongs to subgroup G while mouse *Sox12* is in C and *Xenopus Sox12* in D. Mouse *Sox19* belongs to subgroup C, but zebrafish *Sox19* to B.

We assigned the human and mouse *Sox30* to chromosomes 5q33 and 11, respectively. The *Sox* genes have been shown to be distributed throughout the genome with apparent lack of gene clusters. We compared our interspecific map of chromosome 11 with a composite mouse linkage map that shows the location of many uncloned mouse mutations (provided from the Mouse Genome Database maintained at The Jackson Laboratory, Bar Harbor, ME). *Sox30* maps in a region of the composite map that so far lacks mouse mutations with a defect of spermatogenesis, which might be expected for an alteration of this locus (data not shown). However, in humans, recent deletion mapping revealed two types of chromosome 5 abnormalities in male germ cell tumors, genetic monosomy and regional deletion at 5p15.1–15.2, 5q11 and 5q34–35 (51). Therefore loss of genetic information on chromosome 5 appears closely linked to tumor development. It remains to be determined whether deletions or mutations of *Sox30* are present in male germ cell tumors.



The Sox30 protein was capable of binding DNA in a sequence-specific manner and activating transcription from a promoter containing its target sequence. Our oligonucleotide selection data showed that Sox30 prefers the ACAAT motif. The amplified consensus binding sequence for Sox30 differed slightly from other Sox proteins, in that preference for a guanine residue was observed at the flanking sequences of the core ACAAT motif (Table 1). Nevertheless, Sox30 was also capable of activating transcription from an AACAATA-containing promoter as well as a GACAATG-containing promoter, $(125)_8$dN-luc (data not shown). This might be explained by subtle differences in binding specificity which are not significant for transactivation *in vivo*, but become critical under the repetitive selection conditions.

During mammalian spermatogenesis, male germ cells undergo three major developmental events: the mitotic proliferation of spermatogonia, the meiotic division of spermatocytes and the subsequent formation of haploid spermatids and their morphogenic change into sperm. Northern blot analysis strongly suggested that Sox30 is highly expressed in spermatocytes or more differentiated cells. Other Sox family proteins, Sox5 and Sox 17, seem to be involved in distinct steps during spermatogenesis (9,31). Sox17 is expressed in spermatogonia and its expression declines at the early pachytene spermatocyte stage and onward. Sox5 is highly expressed in early post-meiotic cells (round spermatids). It is interesting that these three Sox family proteins seem to be differentially expressed during spermatogenesis. Sox proteins are known to function cooperatively with other transcription factors (34,35,38,39) and, accordingly, a search for such partner factors and target genes should reveal the function of these genes in more detail.

ACKNOWLEDGEMENTS

We thank H. Hirai, T. Suzuki, J. Fujimoto (IMSUT), N. Takamatsu (Kitasato University) and T. Akiyama (University of Tokyo) for materials and helpful discussions, Y. Nishimune (Osaka University) for *jsd/jsd* mice, P. Bartel, S. Fields and S. Hollenberg for the yeast two hybrid system and A. Reuss for excellent technical assistance. This research was supported in part by the Uehara Memorial Foundation and the National Cancer Institute, DHHS, under a contract with ABL.

Figure 6. Transcriptional activities of Sox30 proteins. (**A**) Transcriptional activities of type I and type II Sox30 proteins. Aliquots of 2 μg of reporter plasmids were transfected into CV-1 cells together with 5 μg of type I or type II Sox30 expression vector. The fold activations were calculated by setting the luciferase activity of each reporter plasmid in the absence of Sox30 to 1.0 and then dividing the activity in the presence of Sox30 by that in its absence. All data are the averages with standard deviations of two independent assays. Each assay was performed in duplicate. (**B**) Dose-dependent transactivation by Sox30 type I. Aliquots of 2 μg of reporter plasmids were transfected into CV-1 cells together with various amounts of Sox30 expression vector. The luciferase activities were normalized for protein concentrations. All data are the averages with standard deviations of triplicate samples. (**C**) ACAAT motif-dependent transactivation by Sox30 type I. Aliquots of 2 μg of reporter plasmids containing various copy numbers of the ACAAT motif were transfected into CV-1 cells together with 5 μg of Sox30 expression vector. All data are the averages with standard deviations of two independent assays. Each assay was performed in duplicate.

REFERENCES

1. Jantzen,H.M., Admon,A., Bell,S.P. and Tjian,R. (1990) *Nature*, **344**, 830–836.
2. Pevny,L.H. and Lovell-Badge,R. (1997) *Curr. Opin. Genet. Dev.*, **7**, 338–344.
3. Grosschedl,R., Giese,K. and Pagel,J. (1994) *Trends Genet.*, **10**, 94–100.
4. Wegner,M. (1999) *Nucleic Acids Res.*, **27**, 1409–1420.
5. Cremazy,F., Soullier,S., Berta,P. and Jay,P. (1998) *FEBS Lett.*, **438**, 311–314.
6. Hosking,B.M., Muscat,G.E., Koopman,P.A., Dowhan,D.H. and Dunn,T.L. (1995) *Nucleic Acids Res.*, **23**, 2626–2628.
7. Lefebvre,V., Huang,W., Harley,V.R., Goodfellow,P.N. and de Crombrugghe,B. (1997) *Mol. Cell. Biol.*, **17**, 2336–2346.
8. Ng,L.J., Wheatley,S., Muscat,G.E., Conway-Campbell,J., Bowles,J., Wright,E., Bell,D.M., Tam,P.P., Cheah,K.S. and Koopman,P. (1997) *Dev. Biol.*, **183**, 108–121.
9. Kanai,Y., Kanai-Azuma,M., Noce,T., Saido,T.C., Shiroishi,T., Hayashi,Y. and Yazaki,K. (1996) *J. Cell Biol.*, **133**, 667–681.
10. van de Wetering,M., Oosterwegel,M., van Norren,K. and Clevers,H. (1993) *EMBO J.*, **12**, 3847–3854.
11. Kanda,H., Kojima,M., Miyamoto,N., Ito,M., Takamatsu,N., Yamashita,S. and Shiba,T. (1998) *Gene*, **211**, 251–257.
12. Sinclair,A.H., Berta,P., Palmer,M.S., Hawkins,J.R., Griffiths,B.L., Smith,M.J., Foster,J.W., Frischauf,A.M., Lovell-Badge,R. and Goodfellow,P.N. (1990) *Nature*, **346**, 240–244.
13. Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Munsterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) *Nature*, **346**, 245–250.
14. Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990) *Nature*, **348**, 448–450.
15. Jager,R.J., Anvret,M., Hall,K. and Scherer,G. (1990) *Nature*, **348**, 452–454.
16. Koopman,P., Gubbay,J., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1991) *Nature*, **351**, 117–121.
17. Foster,J.W., Dominguez-Steglich,M.A., Guioli,S., Kowk,G., Weller,P.A., Stevanovic,M., Weissenbach,J., Mansour,S., Young,I.D., Goodfellow,P.N., Brook,J.D. and Schafer,A.J. (1994) *Nature*, **372**, 525–530.
18. Southard-Smith,E.M., Kos,L. and Pavan,W.J. (1998) *Nature Genet.*, **18**, 60–64.
19. Herbarth,B., Pingault,V., Bondurand,N., Kuhlbrodt,K., Hermans-Borgmeyer,I., Puliti,A., Lemort,N., Goossens,M. and Wegner,M. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 5161–5165.
20. Kuhlbrodt,K., Schmidt,C., Sock,E., Pingault,V., Bondurand,N., Goossens,M. and Wegner,M. (1998) *J. Biol. Chem.*, **273**, 23033–23038.
21. Pingault,V., Bondurand,N., Kuhlbrodt,K., Goerich,D.E., Préhu,M.O., Puliti,A., Herbarth,B., Hermans-Borgmeyer,I., Legius,E., Matthijs,G., Amiel,J., Lyonnet,S., Ceccherini,I., Romeo,G., Smith,J.C., Read,A.P., Wegner,M. and Goossens,M. (1998) *Nature Genet.*, **18**, 171–173.
22. Wagner,T., Wirth,J., Meyer,J., Zabel,B., Held,M., Zimmer,J., Pasantes,J., Bricarelli,F.D., Keutel,J., Hustert,E., Wolf,U., Tommerup,N., Scempp,W. and Scherer,G. (1994) *Cell*, **79**, 1111–1120.
23. Ma,Y., Niemitz,E.L., Nambu,P.A., Shan,X., Sackerson,C., Fujioka,M., Goto,T. and Nambu,J.R. (1998) *Mech. Dev.*, **73**, 169–182.
24. Nambu,P.A. and Nambu,J.R. (1996) *Development*, **122**, 3467–3475.
25. Russell,S.R., Sanchez-Soriano,N., Wright,C.R. and Ashburner,M. (1996) *Development*, **122**, 3669–3676.
26. Soriano,N.S. and Russell,S. (1998) *Development*, **125**, 3989–3996.
27. Hanna-Rose,W. and Han,M. (1998) *Development*, **126**, 169–179.
28. Schilham,M.W., Oosterwegel,M.A., Moerer,P., Ya,J., de Boer,P.A., van de Wetering,M., Verbeek,S., Lamers,W.H., Kruisbeek,A.M., Cumano,A. and Clevers,H. (1996) *Nature*, **380**, 711–714.
29. Lefebvre,V., Li,P. and de Crombrugghe,B. (1998) *EMBO J.*, **17**, 5718–5733.
30. Hudson,C., Clements,D., Friday,R.V., Stott,D. and Woodland,H.R. (1997) *Cell*, **91**, 397–405.
31. Denny,P., Swift,S., Connor,F. and Ashworth,A. (1992) *EMBO J.*, **11**, 3705–3712.
32. Kamachi,Y., Sockanathan,S., Liu,Q., Breitman,M., Lovell-Badge,R. and Kondoh,H. (1995) *EMBO J.*, **14**, 3510–3519.
33. Kamachi,Y., Uchikawa,M., Collignon,J., Lovell-Badge,R. and Kondoh,H. (1998) *Development*, **125**, 2521–2532.
34. Kamachi,Y., Cheah,K.S.E. and Kondoh,H. (1999) *Mol. Cell. Biol.*, **19**, 107–120.
35. Ambrosetti,D.C., Basilico,C. and Dailey,L. (1997) *Mol. Cell. Biol.*, **17**, 6321–6329.
36. Yuan,H., Corbi,N., Basilico,C. and Dailey,L. (1995) *Genes Dev.*, **9**, 2635–2645.
37. Kuhlbrodt,K., Herbarth,B., Sock,E., Hermans-Borgmeyer,I. and Wegner,M. (1998) *J. Neurosci.*, **18**, 237–250.
38. Kuhlbrodt,K., Herbarth,B., Sock,E., Enderich,J., Hermans-Borgmeyer,I. and Wegner,M. (1998) *J. Biol. Chem.*, **273**, 16050–16057.
39. De Santa Barbara,P., Bonneaud,N., Boizet,B., Desclozeaux,M., Moniot,B., Sudbeck,P., Scherer,G., Poulat,F. and Berta,P. (1998) *Mol. Cell. Biol.*, **18**, 6653–6665.
40. Lemaire,P., Garrett,N. and Gurdon,J.B. (1995) *Cell*, **81**, 85–94.
41. Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
42. Copeland,N.G. and Jenkins,N.A. (1991) *Trends Genet.*, **7**, 113–118.
43. McKenzie,A.N.J., Li,X., Largaespada,D.A., Sato,A., Kaneda,A., Zurawski,S.M., Doyle,E.L., Milatovich,A., Francke,U., Copeland,N.G., Jenkins,N.A. and Zurawski,G. (1993) *J. Immunol.*, **150**, 5436–5444.
44. Inaba,T., Shapiro,L.H., Funabiki,T., Sinclair,A.E., Jones,B.G., Ashmun,R.A. and Look,A.T. (1994) *Mol. Cell. Biol.*, **14**, 3403–3413.
45. Sakamoto,Y., Yoshida,M., Semba,K. and Hunter,T. (1997) *Oncogene*, **15**, 2001–2012.
46. Wright,E.M., Snopek,B. and Koopman,P. (1993) *Nucleic Acids Res.*, **21**, 744.
47. Inazawa,J., Saito,H., Ariyama,T., Abe,T. and Nakamura,Y. (1993) *Genomics*, **17**, 153–162.
48. R.,B.A., Cavicchia,J.C., Millette,C.F., O'Brien,D.A., Bhatnagar,Y.M. and Dym,M. (1977) *J. Cell Biol.*, **74**, 68–85.
49. Beamer,W.G., Cunliffe-Beamer,T.L., Shultz,K.L., Langley,S.H. and Roderick,T.H. (1988) *Biol. Reprod.*, **38**, 899–908.
50. Harley,V.R., Lovell-Badge,R. and Goodfellow,P.N. (1994) *Nucleic Acids Res.*, **22**, 1500–1501.
51. Murty,V.V., Reuter,V.E., Bosl,G.J. and Chaganti,R.S. (1996) *Oncogene*, **12**, 2719–2723.