J Hum Genet (2000) 45:192–195

SHORT COMMUNICATION

Takanori Azuma · Naohiko Seki · Tsutomu Yoshikawa
Toshiyuki Saito · Yasuhiko Masuho
Masa-aki Muramatsu

cDNA cloning, tissue expression, and chromosome mapping of human homolog of *SOX18*

Received: January 13, 2000 / Accepted: February 17, 2000

Abstract The *SRY* (sex-determining region Y) gene encodes a transcription factor characterized by a DNA-binding motif termed the HMG (high mobility group) domain. The *SOX* (Sry-box) genes comprise a large family related by homology to the HMG-box region. We isolated a cDNA clone with an open reading frame encoding a putative protein of 384 amino acids, which shared 83% identity to the mouse Sox18 protein. Northern blot analysis revealed that a 1.9-kb band of human *SOX18* messenger RNAs was predominantly expressed in heart, although weak signals were seen in brain, liver, testis, and leukocyte. By polymerase chain reaction (PCR)-based analyses with both a human/rodent monochromosomal hybrid cell panel and a radiation hybrid panel, the gene was mapped to the chromosome 20q13.33 region.

Key words *SOX18* · HMG-box · RH mapping · Chromosome 20q13.33

Introduction

The *SRY* (sex-determining region Y) gene is located on the Y chromosome of mammals, and *SRY* encodes a DNA-binding motif known as the HMG (high mobility group) box, defined by a 79-amino acid region. To date, more than 100 HMG box-containing proteins have been reported. The HMG box-containing proteins are classified into two distinct subgroups according to the sequence-specificity of the binding, the number of the DNA binding domains, and phylogenetic considerations (Laudet et al. 1993). One important subgroup includes *SRY* and *SRY box-related (SOX)* genes.

The *SOX* genes comprise a large family which has sequence identity of more than 60% within the HMG-box region of *SRY*. At present, about 30 members have been reported, which can be classified into seven subgroups based on their structural similarity (Cremazy et al. 1998; Wegner 1999). SOX proteins are transcription factors that have a critical role in the regulation of numerous developmental processes. Mouse Sox1, Sox2, and Sox3 are expressed in the developing nervous system and urogenital ridge (Collignon et al. 1996). In addition, mouse Sox11 has been implicated in neural development, because Sox11 is prominently expressed in the developing central nervous system (Uwanogho et al. 1995). Recently, we have isolated a full-length *SOX11* cDNA which is transcriptionally induced during neural differentiation (Azuma et al. 1999).

Here we report the entire sequence, expression profile, and chromosome mapping of the human *SOX18* gene.

T. Azuma · N. Seki · T. Yoshikawa · Y. Masuho · M. Muramatsu (✉)
Helix Research Institute, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel. 81-438-52-3966; Fax 81-438-52-3952
e-mail: mmasaaki@hri.co.jp

N. Seki · T. Saito
Genome Research Group, National Institute of Radiological Sciences, Chiba, Japan

M. Muramatsu
Department of Biological Cybernetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

The nucleotide sequence data reported in this paper have been deposited to DDBJ, EMBL, and GenBank database under the accession number AB033888.
The first two authors contributed equally to this report.

Source and isolation of human *SOX18* gene

In the present study, we searched human cDNA tags for previously unknown *SOX* gene family members in the public database, using the tBLASTN program (www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform = 1) with an HMG-box sequence as a query sequence. We found a human partial sequence for a putative SOX protein (accession number X65664). To obtain a full-length cDNA for the new protein, we performed 5′- and 3′-rapid amplification of cDNA ends (5′- and 3′-RACE), using the partial sequence, and isolated cDNA clones from a

human heart cDNA library. 5′-RACE experiments were performed using nested primers, 5′-AAG GGC CGC TTC TCC GCC GCG-3′ (corresponding to nucleotides 546 to 526) and 5′GCT CCT TCC ACG CTT TGC CCA-3′ (corresponding to nucleotides 521 to 501), with a Marathon cDNA amplification kit (Clontech, Palo Alto, CA, USA). 3′-RACE experiments were performed using nested primers, 5′-CTG GGC AAA GCG TGG AAG GAG-3′ (corresponding to nucleotides 500 to 520) and 5′-CGC GGC GGA GAA GCG GCC CTT-3′ (corresponding to nucleotides 526 to 546). Several independent cDNA clones were sequenced by a primer walking method, using an ABI377 sequencer (Perkin Elmer, Norwalk, CT, USA) according to the supplier's instructions. The resultant consensus sequence of the cDNA was 1,339 bp in length, and it encoded a putative protein of 384 amino acid residues, with the HMG-box DNA-binding domain spanning residues 84–162 (Fig. 1a). Putative nuclear localization signals RRKK and RKARR were found at residues 157–160 and 163–167, respectively (Fig. 1a). A search of the protein data banks indicated that the deduced amino acid sequence conformed with mouse Sox18 protein (accession number, L35032; 83% identity, 85% similarity); thus, the isolated cDNA represents the human SOX18 protein. The nucleotide sequence of the human *SOX18* gene will appear in GenBank/EMBL/DDBJ databases under the accession number, AB033888.

Fig. 1a Nucleotide sequence and deduced amino acid sequence of the human *SOX18* gene. *Asterisk* denotes the stop codon. The HMG box is *single-underlined.* The putative nuclear localization signals are *double-underlined.* The nucleotide sequence of the *SOX18* gene is deposited in GenBank/EMBL/DDBJ databases under the accession number, AB033888

```
gggaggaagcgctgcagggaccaccgccgtccccaccgccatccgccctcccggcctggc      60
ctgccttgcgcccggctccccagtgcccgccgcccgcccgccgcgctcccgcgctccgt     120
tccgcccaggccgcgcccagctggaATGCAGAGATCGCCGCCCGGCTACGGCGCACAGGA     180
                          M  Q  R  S  P  P  G  Y  G  A  Q  D       12
CGACCCGCCCGCCCGCCGCGACTGTGCATGGGCCCCGGGACACGGGGCCGCCGCTGACAC     240
 D  P  P  A  R  R  D  C  A  W  A  P  G  H  G  A  A  A  D  T      32
GCGCGGCCTCGCCGCCGGCCCCGCCGCCCTCGCCGCGCCCGCCGCGCCCGCCTCGCCGCC     300
 R  G  L  A  A  G  P  A  A  L  A  A  P  A  A  P  A  S  P  P      52
CAGCCCGCAGCGCAGTCCCCCGCGCAGCCCCGAGCCGGGGCGCTATGGCCTCAGCCCGGC     360
 S  P  Q  R  S  P  P  R  S  P  E  P  G  R  Y  G  L  S  P  A      72
CGGCCGCGGGGAACGCCAGGCGGCAGACGAGTCGCGCATCCGGCGGCCCATGAACGCCTT     420
 G  R  G  E  R  Q  A  A  D  E  S  R  I  R  R  P  M  N  A  F      92
CATGGTGTGGGCAAAGGACGAGCGCAAGCGGCTGGCTCAGCAGAACCCGGACCTGCACAA     480
 M  V  W  A  K  D  E  R  K  R  L  A  Q  Q  N  P  D  L  H  N     112
CGCGGTGCTCAGCAAGATGCTGGGCAAAGCGTGGAAGGAGCTGAACGCGGCGGAGAAGCG     540
 A  V  L  S  K  M  L  G  K  A  W  K  E  L  N  A  A  E  K  R     132
GCCCTTCGTGGAGGAAGCCGAACGGCTGCGCGTGCAGCACTTGCGCGACCACCCCAACTA     600
 P  F  V  E  E  A  E  R  L  R  V  Q  H  L  R  D  H  P  N  Y     152
CAAGTACCGGCCGCGCCGCAAGAAGCAGGCGCGCAAGGCCCGGCGGCTGGAGCCCGGCCT     660
 K  Y  R  P  R  R  K  K  Q  A  R  K  A  R  R  L  E  P  G  L     172
CCTGCTCCCGGGATTAGCGCCCCCGCAGCCACCGCCCGAGCCTTTCCCCGCGGCGTCTGG     720
 L  L  P  G  L  A  P  P  Q  P  P  P  E  P  F  P  A  A  S  G     192
CTCGGCTCGCGCCTTCCGCGAGCTGCCCCCGCTGGGCGCCGAGTTCGACGGCCTGGGGCT     780
 S  A  R  A  F  R  E  L  P  P  L  G  A  E  F  D  G  L  G  L     212
GCCCACGCCCGAGCGCTCGCCTCTGGACGGCCTGGAGCCCGGCGAGGCTGCCTTCTTCCC     840
 P  T  P  E  R  S  P  L  D  G  L  E  P  G  E  A  A  F  F  P     232
ACCGCCCGCGGCGCCCGAGGACTGCGCGCTGCGGCCCTTCCGCGCGCCCTACGCGCCCAC     900
 P  P  A  A  P  E  D  C  A  L  R  P  F  R  A  P  Y  A  P  T     252
CGAGTTGTCGCGGGACCCCGGCGGTTGCTACGGGGCTCCCCTGGCGGAGGCGCTCAGGAC     960
 E  L  S  R  D  P  G  G  C  Y  G  A  P  L  A  E  A  L  R  T     272
CGCGCCCCCCGCGGCGCCGCTCGCTGGCCTGTACTACGGCACCCTGGGCACGCCCGGCCC    1020
 A  P  P  A  A  P  L  A  G  L  Y  Y  G  T  L  G  T  P  G  P     292
GTACCCCGGCCCGCTGTCGCCGCCGCCCGAGGCCCCGCCGCTGGAGAGCGCCGAGCCGCT    1080
 Y  P  G  P  L  S  P  P  P  E  A  P  P  L  E  S  A  E  P  L     312
GGGGCCCGCCGCCGATCTGTGGGCCGACGTGGACCTCACCGAGTTCGACCAGTACCTCAA    1140
 G  P  A  A  D  L  W  A  D  V  D  L  T  E  F  D  Q  Y  L  N     332
CTGCAGCCGGACTCGGCCCGACGCCCCCGGGCTCCCGTACCACGTGGCACTGGCCAAACT    1200
 C  S  R  T  R  P  D  A  P  G  L  P  Y  H  V  A  L  A  K  L     352
GGGCCCGCGCGCCATGTCCTGCCCAGAGGAGAGCAGCCTGATCTCCGCGCTGTCGGACGC    1260
 G  P  R  A  M  S  C  P  E  E  S  S  L  I  S  A  L  S  D  A     372
CAGCAGCGCGGTCTATTACAGCGCGTGCATCTCCGGCTAGgccgccggcgccgcccgggt    1320
 S  S  A  V  Y  Y  S  A  C  I  S  G  *                         384
ccctgcagcgcttcctccc                                             1339
```

```
human_SOX18  MQRSPPGYGAQDDPPARRDCAWAPGHGAAADTRGLAAGPAALAAPAAPASPPSPQRSPPR
mouse_SOX18  MQRSPPGYGAQDDPPSRRDCAWAPGIGAAAEARGL...PVTNVSPTSPASPSSLPRSPPR

human_SOX18  SPEPGRYGLSPAGRGERQAADESRIRRPMNAFMVWAKDERKRLAQQNPDLHNAVLSKMLG
mouse_SOX18  SPESGRYGF...GRGERQTADELRIRRPMNAFMVWAKDERKRLAQQNPDLHNAVLSKMLG

human_SOX18  KAWKELNAAEKRPFVEEAERLRVQHLRDHPNYKYRPRRKKQARKARRLEPGLLLPGLAPP
mouse_SOX18  KAWKELNTAEKRPFVEEAERLRVQHLRDHPNYKYRPRRKKQERKVRRLEPGLLLPGLVQP

human_SOX18  QPPPEPFPAASGSARAFRELPPLGAEFDGLGLPTPERSPLDGLEPGEAAFFPPPAAPEDC
mouse_SOX18  SAPPEAFAAASGSARSFRELPTLGAEFDGLGLPTPERSPLDGLEPGEASFFPPPLAPEDC

human_SOX18  ALRPFRAPYAPTELSRDPGGCYGAPLAEALRTAPPAAPLAGLYYGTLGTPGPYPGPLSPP
mouse_SOX18  ALRAFRAPYAP.ELARDPSFCYGAPLGEALRTAPPAAPLAGLYYGTLGTPGPXPNPLSPP

human_SOX18  PEAPPLESAEPLGPAADLWADVDLTEFDQYLNCSRTRPDAPGLPYHVALAKLGPRAMSCP
mouse_SOX18  PESPSLEGTEQLEPTADLWADVDLTEFDQYLNCSRTRPDATTLPYHVXLAKLGPRAMSCP

human_SOX18  EESSLISALSDASSAVYYSACISG
mouse_SOX18  EESSLISALSDASSAVYYSACISG
```

Fig. 1b Alignment of human SOX18 (accession number, AB033888) and mouse Sox 18 (accession number, L35032) proteins. Identities are indicated by *black background*, and similar residues are *shadowed*

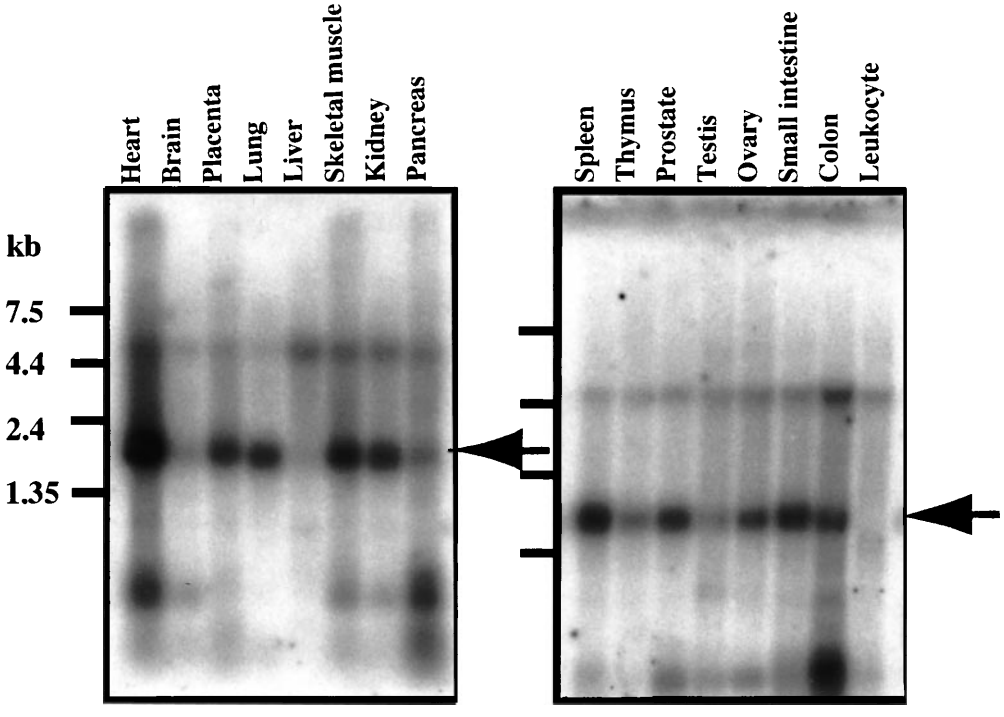

Fig. 2. Northern blot analysis of human *SOX18*. Northern blot filters containing adult human poly (A)+ RNAs (2μg/lane) were purchased from Clontech Laboratories (Palo Alto, CA, USA), and hybridization and washing were performed following the manufacturer's instructions. The 1.9-kb cDNA fragment containing the entire open reading frame was labelled with [α-^{32}P] dCTP and used as a hybridization probe. A major 1.9-kb band seen in various tissues is shown by *arrows*. Size markers (*left*) are in kilobases

The alignment of predicted amino acid sequences of human and mouse SOX18/Sox18 is shown in Fig. 1b.

Expression profile of Human *SOX18* gene

Northern blots of poly (A)+ RNA from various human tissues were hybridized with the radio-labelled human *SOX18* cDNAs, and the gene expression profiles in various tissues were analyzed. The autoradiograms of the blots are shown in Fig. 2. A major band of approximately 1.9kb in length (Fig. 2, arrows) was detected abundantly in heart, and moderately in placenta, lung, skeletal muscle, kidney, spleen, prostate, and small intestine. In contrast, a very weak signal was observed in brain, liver, testis, and leukocyte. Other minor signals (approximately 5.0 and 0.7kb) were also observed. The longer signals were transcribed ubiquitously in various tissues, and the shorter signals appeared in heart, pancreas, and colon. At present, we do not know whether the longer and shorter signals are derived from an alternative form of human *SOX18* mRNA or whether they represent other related gene transcripts.

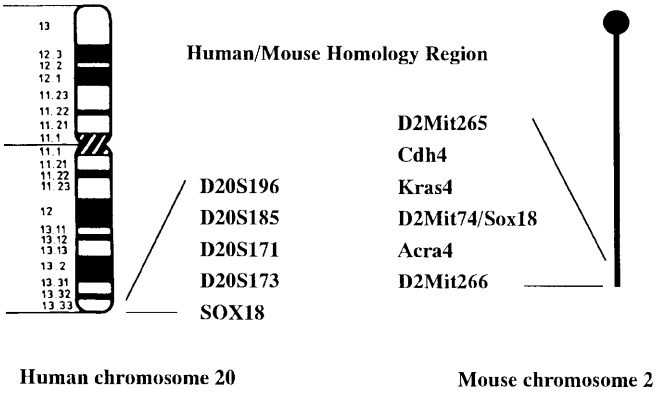

Fig. 3. Chromosomal placement of the human *SOX18* gene at a relative distance to framework markers on the WICGR radiation hybrid map of the human genome (http://carbon.wi.mit.edu:8000/cgi-bin/contig/phys_map). The approximate corresponding cytogenetic location of the genes on the chromosome 20q13.33 region is indicated. Human/mouse homology regions around the *SOX18/Sox18* genes are indicated

Chromosome mapping of human *SOX18* gene

Chromosomal assignment of the human *SOX18* gene was done by polymerase chain reaction (PCR) analysis of a human/rodent somatic cell hybrid panel (National Institute of General Medicine Service, Coriell Cell Repositories, Camden, NJ, USA) and a radiation hybrid panel (Genebridge 4; Research Genetics, Huntsville, AL, USA), as described previously (Seki et al. 1999; Seki et al. 2000). The human *SOX18*-specific PCR primers (5′-GCC AGC AGC GCG GTC TAT TAC-3′, nucleotides 1259 to 1279; 5′-AGG AAG CGC TGC AGG GAC CCG-3′, nucleotides 1316 to 1336), gave rise to an amplified 78-bp product by genomic PCR. First, the specific amplified product for the human gene was detected only from the hybrid containing human chromosome 20 (data not shown). Further mapping analysis was done, using a radiation hybrid panel with the same primer set. Statistical analysis of the radiation hybrid data was performed using the RHMAPPER software package (http://carbon.wi.mit.edu:8000/cgi-bin/contig/rhmapper.pl). The data vector for the human *SOX18* gene was 1000010000 0000000011 1001010000 0001000000 1200100000 0111111011 1100001110 0010010101 1000001000 000, and the consequent report indicated that the gene was placed to 32.30 cR distal to the marker D20S173 (LOD score for linkage 11). The region including the marker was cytogenetically mapped to the 20q13.33 region (Fig. 3). This region shows homologous organization to the mouse chromosome 2 terminal region (www.ncbi.nlm.nih.gov/Homology) (Fig. 3). Mouse *Sox18* was mapped to the D2Mit74 region, using the EUCIB interspecific backcross facility (Greenfield et al. 1996). Thus, it was confirmed that the mouse and human genes encoding *SOX18/Sox18* are localized in a region with conserved linkage homology between these species.

Some members of the *SOX* family have been shown to be associated with human diseases. For example, *SOX9* is associated with the skeletal malformation syndrome, campomelic dysplasia, while *SOX10* is associated with Waardenburg-Hirschsprung disease (Wagner et al. 1994; Foster et al. 1994; Southard-Smith et al. 1998; Herbarth et al. 1998; Kuhlbrodt et al. 1998; Pingault et al. 1998). At present, no inherited disease loci appear to be mapped to chromosome 20qter in which *SOX18* is located. The precise chromosomal position and expression profile of the *SOX18* gene may contribute toward the ongoing positional candidate approaches for disease genes linked to this genomic locus.

References

Azuma T, Ao S, Saito Y, Yano K, Seki N, Wakao H, Masuho Y, Muramatsu M (1999) Human *Sox11*, an upregulated gene during the neural differentiation, has a long 3′ untranslated region. DNA Res 6:357–360

Cremazy F, Soullier, S, Berta P, Jay P (1998) Further complexity of the human *SOX* gene family revealed by the combined use of highly degenerate primers and nested PCR. FEBS Lett 438:311–314

Collignon J, Sockanathan S, Hacker A, Cohen-Tannoudji M, Norris D, Rastan S, Stevanovic M, Goodfellow PN, Lovell-Badge R (1996) A comparison of the properties of *Sox-3* with *Sry* and two related genes, *Sox-1* and *Sox-2*. Development 122:509–520

Foster JW, Dominguez-Stelich MA, Guioli S, Kowk G, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN, Brook JD, Schafer AJ (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an *SRY*-related gene. Nature 372:525–530

Greenfield A, Dunn T, Muscat G, Koopman P (1996) The *Sry*-related gene *Sox18* maps to distal mouse chromosome 2. Genomics 36:558–559

Herbarth B, Pingault V, Bondurand N, Kuhlbrodt K, Hermans-Borgmeyer I, Puliti A, Lemort N, Goossens M, Wegner M (1998) Mutation of the *Sry*-related *Sox10* gene in dominant megacolon, a mouse model for human Hirschsprung disease. Proc Natl Acad Sci USA 95:5161–5165

Kuhlbrodt K, Schmidt C, Sock E, Pingault V, Bondurand N, Goossens M, Wegner M (1998) Functional analysis of *Sox10* mutations found in human Waardenburg-Hirschsprung patients. J Biol Chem 273:23033–23038

Laudet V, Stehelin D, Clevers H (1993) Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res 21:2493–2501

Pingault V, Bondurand N, Kuhlbrodt K, Goerich DE, Prehu M-O, Puliti A, Herbarth B, Hermans-Borgmeyer I, Legius E, Matthijs G (1998) *SOX10* mutations in patients with Waardenburg-Hirschsprung disease. Nature Genet 18:171–173

Seki N, Hattori A, Hayashi A, Kozuma S, Ohira M, Hori T, Saito T (1999) Structure, expression profile and chromosomal location of an isolog of DNA-PKcs interacting protein (*KIP*) gene. Biochim Biophys Acta 1444:143–147

Seki N, Ueki N, Yano K, Saito T, Masuho Y, Muramatsu M (2000) cDNA cloning of a novel human gene *NAKAP95*, neighbor of A-kinase anchoring protein 95 (AKAP95) on chromosome 19p13.11–p13.12 region. J Hum Genet 45:31–37

Southard-Smith EM, Kos L, Pavan WJ (1998) *Sox10* mutation disrupts neural crest development in Dom Hirschsprung mouse model. Nature Genet 18:60–64

Uwanogho D, Rex M, Cartwright EJ, Pearl G, Healy C, Scotting PJ, Sharpe PT (1995) Embryonic expression of the chicken *Sox2*, *Sox3*, and *Sox11* genes suggests an interactive role in neuronal development. Mech Dev 49:23–36

Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the *SRY*-related gene *SOX9*. Cell 79:1111–1120

Wegner M (1999) From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res 27:1409–1420