Mammalian Genome 4, 577–584 (1993).

Mammalian Genome

Characterization and mapping of the human SOX4 gene

Christine J. Farr,[1] David J. Easty,[2] Jiannis Ragoussis,[3] Jerome Collignon,[4] Robin Lovell-Badge,[4] Peter N. Goodfellow[1]

[1]Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK
[2]Department of Anatomy, St. George's Medical School, Cranmer Terrace, London SW17 ORE, UK
[3]Division of Medical and Molecular Genetics, UMDS, Guy's Hospital, London Bridge, London SE1 9RT, UK
[4]Laboratory of Eukaryotic Molecular Genetics, MRC National Institute of Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Received: 22 April 1993 / Accepted: 2 July 1993

Abstract. The SOX genes comprise a large family related by homology to the HMG-box region of the testis-determining gene SRY. We have cloned and sequenced the human SOX4 gene. The open reading frame encodes a 474 amino acid protein, which includes an HMG-box. The non-box sequence is particularly rich in serine residues and has several polyglycine and polyalanine stretches. With somatic cell hybrids, human SOX4 has been mapped to Chromosome (Chr) 6p distal to the MHC region. There is no evidence for clustering of other members of the SOX1, -2, and -3 or SOX4 gene families around the SOX4 locus.

Introduction

The testis-determining gene SRY encodes a protein with a DNA-binding motif known as an HMG-box (Gubbay et al. 1990; Sinclair et al. 1990). The same motif is shared both with proteins that bind DNA without sequence specificity (Jantzen et al. 1990) and with several sequence-specific DNA-binding proteins, such as the T cell-specific factors TCF-1 and TCF-1α/LEF-1 (Travis et al. 1991; van de Wetering et al. 1991; Waterman et al. 1991). Four *Sry*-related genes called *Sox* genes (for *S*RY-b*ox*) were originally isolated from the mouse on the basis of their homology to SRY in the HMG-box region. These genes are autosomal or X-linked and can be divided into two subfamilies. The sequences of mouse *Sox-1*, *-2*, and *-3* are almost identical in the conserved motif (95–99% nucleotide homology), whereas *Sox-4* seems to have diverged independently (78% homology to *Sox-1*, *-2*, or *-3* at the DNA level). The conservation of these two subfamilies in birds has recently been reported (Griffiths 1991), and several other members of the *Sox* gene family have now been identified (Denny et al. 1992a,b).

The SRY gene has been shown to encode a protein with sequence-specific DNA-binding capacity (Harley et al. 1992), and it is probable that SRY functions as a transcription factor in the embryo, regulating the genes that determine testicular development. The related *Sox-1*, *-2*, *-3*, and *Sox-4* genes are expressed during embryogenesis in the mouse [these genes were referred to as a1, a2, a3, and a4 respectively in (Gubbay et al. 1990)], and by analogy to *Sry* these genes might be expected to play some role in regulating developmental processes. Recent reports on other members of the *Sry*-related gene family indicate that these genes also encode DNA-binding proteins that recognize similar sequences. The mouse *Sox-5* gene, which is highly expressed during spermatogenesis, has been shown to have a DNA-binding specificity overlapping that of *Sry*, while a rat *Sox* gene has been isolated on the basis of its binding to an insulin-response element in the 5′ flanking region of the human glyceraldehyde-3-phosphate-dehydrogenase gene (*GAPDH;* Nasrin et al. 1991). It has recently been demonstrated that HMG-box proteins like SRY can induce significant bending of DNA upon binding (Ferrari et al. 1992; Giese et al. 1992), suggesting that transcription activation by these factors might be mediated by the promotion of protein-protein interactions.

Here we describe the isolation and characterization of the human SOX4 gene. SOX4 maps to 6p distal to the MHC region.

The nucleotide sequence data reported in this paper have been submitted to the EMBL Data Library and have been assigned the accession number X70683.

Correspondence to: C.J. Farr

Materials and methods

Library screening

The LT5.1 cDNA library was custom-made by Stratagene. The normal melanocyte library was kindly provided by B. Kwon (Kwon et al. 1987). The YAC clones analyzed in this study were isolated from the ICRF reference YAC library (Larin et al. 1991), as described previously (Ragoussis et al. 1992). Molecular biology techniques were as in Sambrook and colleagues (1989) unless otherwise stated.

Culture of melanocytes and melanomas

Normal melanocytes were isolated from newborn foreskins and cultured according to Eisinger and Marko (1982), with modifications reported by Bennett and coworkers (1985). The melanoma cell line DX3-LT5.1 is a subclone of SK-MEL-93-DX3 and was kindly provided by Dr Ian Hart (London). DX3-LT5.1 was derived from the poorly metastatic parental cell line after brief exposure to 5-azacytidine and repeated passage through the lungs of nude mice to derive a highly metastatic line (Omerod et al. 1986). DX3-LT5.1 was grown in Dulbecco's modification of Eagle's medium supplemented with 5% FCS.

Southern and Northern blot analysis

DNA was extracted and digested with restriction enzymes according to standard procedures. Fragments were separated by agarose gel electrophoresis and transferred to nylon membrane (Hybond N+, Amersham). Filters were probed with DNA fragments labeled by the random-primer method (Feinberg and Vogelstein 1984). Southern hybridizations were done in a Hybaid oven and washed at high stringency (0.1 × SSPE/0.1%SDS at 65°C) unless otherwise stated. Aliquots (10 μg) of poly(A)$^+$ RNA were processed for agarose gel electrophoresis (Aviv and Leder 1972; Chirgwin et al. 1979) and transferred to nitrocellulose. Prehybridization (4 h) and hybridization (18 h) were at 42°C in 5 × SSC, 5 × Denhardts, 50% formamide, 1% SDS, and 20 μg/ml salmon sperm DNA. The final wash stringency was at 0.1 × SSC, 0.1% SDS at 65°C. After washing, filters were autoradiographed for several days at −70°C. The size of the hybridizing bands was determined by comparison with molecular weight markers (Pharmacia). To ensure equal RNA loading, blots were rehybridized with the housekeeping gene *GAPDH* (a gift from Professor Mike Clemens) and exposed for 3 h. For tissue distribution, RNA was extracted from the mouse strain MF1.

Polymerase chain reaction

A SOX4-box probe was generated by PCR with the following primers: HMG BoxF CCGAATTCTGAACGCCTTCATGGTGTGGTC and HMG BoxR2 CCGAATTCCGTTGCCCGACTTCACCTTCTT. PCR was carried out with Promega *Taq* polymerase and reaction buffer as follows: 5 min/94°C (1 min/94°C; 1 min/45°C; 1 min/72°C) 30 cycles plus 5 min/72°C. Two human SOX4-specific primers from the 3′UNT region were used to screen genomic DNAs: SOX4F5 CCGCGCTCCCATCCCCCACC and SOX4R2 GTGGTGATGCAGGAAGGGGGAGAC. PCR conditions were: 5 min/94°C (1 min/94°C; 1 min/65°C; 1 min/72°C) 35 cycles in 1 × Promega reaction buffer plus 10% dimethylsulphoxide.

Sequencing

cDNAs were subcloned into pUC18/19 and sequenced on both strands of double-stranded template using T7 DNA polymerase and conditions as described in the Sequenase kit (USB). Electrophoresis was in 6% acrylamide, 7 M urea gels.

Results

Identification and analysis of human SOX4

To identify the human homologs of the mouse *Sox* genes, cDNA probes, which excluded the highly conserved HMG-box, were used to screen cDNA libraries. With a mouse *Sox-4* probe (*Sma*I-*Eco*RI 350bp fragment, Fig. 1), clones were obtained from a λgt10 cDNA library of the human melanoma cell line LT5.1 (approx. 15 positives were recovered per 1×10^6 pfu at a stringency of 0.1 × SSPE/65°C). These incomplete clones were used to rescreen the library, and additional clones were isolated from a λgt11 normal melanocyte cDNA library. Figure 1 shows the overlapping cDNA clones from which sequence was obtained.

2794 bps of sequence were obtained by double-strand sequencing (Fig. 2). A G–A transition at position 945 was identified in cDNA LT4, which creates a *Pst*I site. This base change appears to be a cloning artifact since this *Pst*I site is not present in the other cDNAs, and no polymorphism was detected on screening a panel of random human DNAs (data not shown).

The SRY-box in human SOX4 is 94% identical at the nucleotide level to the mouse *Sox-4* sequence (14 differences out of 237 nucleotides), resulting in only one change at the amino acid level (arginine–human, glutamine–mouse at position 54 of the box; Fig. 3). The open-reading frame (ORF) which includes the box encodes a protein of 474 amino acids (predicted molecular mass 47.3 kDa) and is presented in Fig. 2. The first methionine in this ORF (and in the cDNA) has an excellent match to the Kozak consensus. The predicted SOX4 protein is particularly rich in serine residues (18% overall) with several polyserine stretches on the C-terminal side of the SRY-box. There are also several stretches of glycine and of alanine residues in this region of the protein. Outside the SRY-box, SOX4 is not closely related to sequences in either the EMBL (release 33) or the protein sequence Database Owl18_0.

The ORF is preceded by a long (347 bp) 5′-untranslated leader sequence, and in addition 1022 bps of 3′untranslated DNA have been sequenced. No polyA tail or polyadenylation site have been identified.

Genomic organization of the SOX4 gene family

Southern blot analysis of human, mouse, and hamster genomic DNA, cut with *Eco*RI or *Hin*dIII and probed with the box-containing fragment from human SOX4, revealed three strongly hybridizing bands and several fainter signals at high stringency (0.1 × SSPE/65°C; Fig. 4). The 4-kb *Hin*dIII and 6.5-kb *Eco*RI fragments in human hybridize specifically to a 250-bp *Xba*I-*Eco*RI fragment from the SOX4 3′-untranslated DNA (Fig. 1). These specific bands are also detected with a mouse *Sox-4* cDNA probe which excludes the box (the *Sma*I-*Eco*RI 350-bp fragment), indicating that the human homolog of *Sox-4* has been isolated rather than

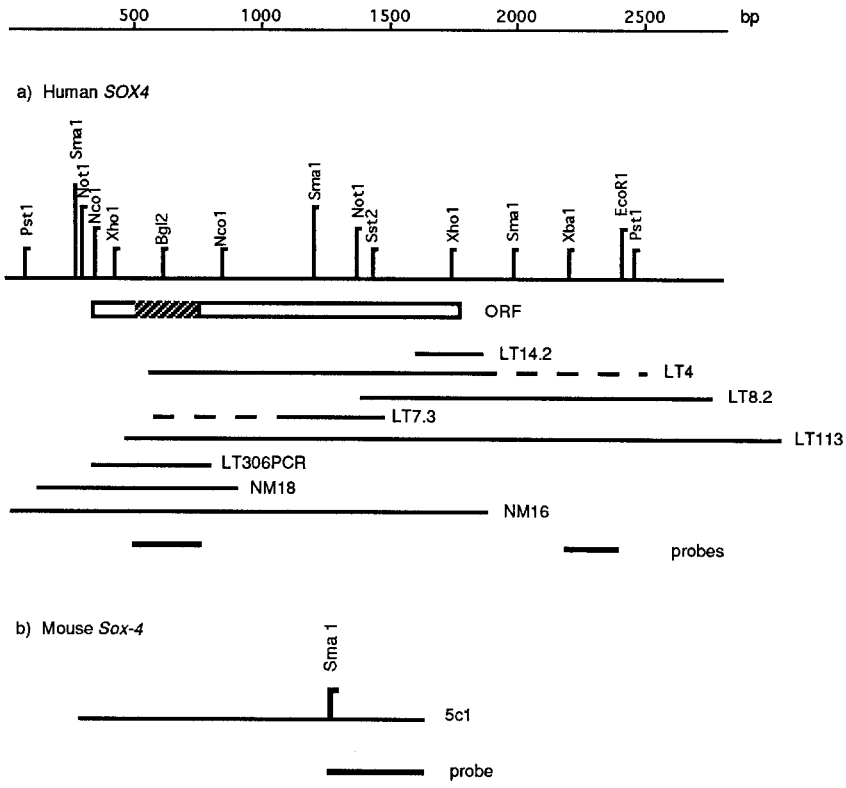

Fig. 1. **(a)** Overlapping cDNA clones were isolated from a λgt10 human melanoma cDNA library (LT) and from a λgt11 normal melanocyte cDNA library (NM). The **dashed lines** represent unrelated coligations. Clone LT306PCR was isolated by PCR from phage 306 with a SOX4-specific oligonucleotide and the λgt10 reverse primer. The **open box** indicates the longest ORF, which encompasses the HMG-box motif **(shaded region).** DNA fragments used as probes are indicated; **(b)** The mouse *Sox-4* cDNA 5c.1 (an *Eco*RI clone). The *Sma*I-*Eco*RI 350-bp fragment was used for the initial library screens and also for Southern blot analysis.

another closely related member of this multigene family.

The possibility that SOX genes may be clustered in the genome was tested by isolating two YAC clones (350 and 550 kb; ICRFy900EO999 and ICRFy900603140) with the human 3′-untranslated *Xba*I-*Eco*RI DNA fragment. Restriction mapping and probing with box probes from human SOX4 and mouse *Sox-1* revealed only the cognate SOX4 hybridizing band after washing at low stringency (1 × SSPE/65°C) and overnight autoradiography (data not shown).

SOX4 expression

The expression pattern of SOX4 was analyzed by Northern blotting. mRNA from the melanoma cell line DX3-LT5.1 and from normal melanocytes was probed with a SOX4-specific DNA fragment (*Xba*I-*Eco*RI, 250 bp). A single transcript of about 5.2 kb was detected after 2 days' exposure (Fig. 5A). Expression has also been detected in poly(A)$^+$ RNA from adult human testis, where the probe also hybridized with a minor band of 3.9 kb (data not shown). The SOX4 transcript is larger than the cDNA sequence, suggesting that approximately 2 kb of 3′-untranslated DNA remain to be analyzed.

Since only a restricted analysis of SOX4 tissue distribution was possible (because of the limited availability of RNA from human tissues and the lack of an RNA-specific PCR assay), a wider survey was carried out with the mouse *Sox-4* cDNA 5c.1 and a panel of poly(A)$^+$ RNAs from adult mouse tissues. A 5.2-kb *Sox-4* transcript was detected in brain, testis, and heart (Fig. 5B). A very low level of expression was also detectable in skeletal muscle after prolonged exposure. The signal for kidney is difficult to interpret owing to some degradation of the RNA. No transcripts were detectable in liver. As with the human gene, a minor 3.9-kb transcript was again apparent in testis and also in brain after extended autoradiography (data not shown). The smaller transcript detected in some tissues may represent use of an alternative polyadenylation site or could be derived from a closely related gene. As both the 5.2-kb and the 3.9-kb transcripts were detected on blots with a SOX4-specific probe from the 3′-untranslated DNA, the latter is unlikely.

Mapping of the human SOX4 gene

The human-specific *Xba*I-*Eco*RI probe was used to map the SOX4 gene against a somatic cell hybrid panel (Table 1). This indicated that the gene is present on human Chr 6. In order to localize it further, hybrids retaining fragments of Chr 6 only were screened by PCR (Fig. 6). SOX4 is present in cell line 5647cl22, which retains 6p21.1 to pter, but absent in line MCP6, which contains 6qter to 6p21.3 only. This maps SOX4 to 6p21.3-pter and distal to the HLA region.

Discussion

SRY is a Y-located gene required for normal male sex determination. The major structural feature of the SRY protein is an HMG-box. This box confers se-

Table 1. Chromosome assignment of human SOX4 using a panel of somatic cell hybrids.

	Chr no.																						Probe:
Hybrid	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	SOX4
CTP34B4	1	2	3		5	6	7	8				12		14		16	17	18					X	+
CTP41A2		2	3			6	7							14			17						X	+
SIF4A31			3	4	5	6								14		16							X	+
TWIN19D12	1		3	4		6		8			11	12		14		16	17	18		20				+
MOG34A4	1		3	4	5	6	7	8		10	11	12	13	14		16		18	19		21		X	+
DT1.2.4			3								11				15		17	18		20	21		X	−
3W4C15							7			10	11	12		14	15		17				21		X	−
DUR4.3			3		5					10	11	12	13	14	15		17	18		20	21	22	X	−
SIR74ii	1	2	3	4								12	13	14			17	tr			21	tr	X	−
F4SC13Cl12	1								9					14									X	−

Symbols: tr, trace.

quence-specific DNA-binding activity on SRY (Harley et al. 1992), and the region of the gene encoding this motif is the only part showing conservation between species [(Gubbay et al. 1990) and unpublished observations]. Recently sequences closely related to SRY in the HMG-box region have been cloned (Denny et al. 1992a,b; Griffiths 1991; Gubbay et al. 1990). In many cases these DNAs have been isolated by PCR-based approaches to amplify the highly conserved 237-bp HMG-box. Since many of these PCR products encode very similar proteins, it is difficult at present to estimate the number of SOX genes and pseudogenes in the genome.

We have cloned and sequenced human SOX4. Multiple overlapping clones for SOX4 were isolated from normal melanocyte and melanoma cDNA libraries. When the SOX4-box region was used as a probe for Southern blot analysis of human genomic DNA, multiple hybridizing bands were detected, whereas only one of these bands was detected with a 3′-untranslated portion of SOX4. At the stringency used—0.1 × SSPE/65°C—members of the SOX1, -2, and -3 gene family (which in the mouse show 80% nucleotide homology with *Sox-4* within the HMG-box) do not cross-hybridize to the SOX4 box probe. These data are consistent with there being three (or more) genes that share a SOX4-like HMG-box. DNA sequence from the box region of several genes belonging to the SOX4 family has recently been reported from a range of species (Fig. 3). Our human SOX4 encodes a protein very similar to the SOX4-box reported by Denny and coworkers (1992a). In both proteins there is a change of glutamine to arginine at position 54 of the HMG-box compared with the mouse sequence, but the two human SOX4 proteins differ themselves by one amino acid (proline/glutamine) at box position 15. Until more information is available, it is not possible to resolve whether these sequences represent the same gene or different members of the SOX4 family. In Fig. 3, amino acid sequences of the HMG-boxes from several members of the SOX4 family are aligned. The high degree of conservation within this domain, from human to reptiles and birds, is striking and suggests that SOX4 function is strongly conserved.

In some gene families the genes are clustered in the genome, and the clustering may be related to function (Farr and Goodfellow 1992). To test for clustering of SOX4 with other closely related genes (arising perhaps through tandem duplication), we isolated two YACs containing the human SOX4 locus. No additional SOX genes were detected around the SOX4 locus when the YACs were probed with the SOX4 or *Sox-1* box DNAs and washed at low stringency, suggesting that SOX4 is not closely linked to other members of the SOX1, -2, -3, or SOX4 families.

The deduced amino acid sequence of SOX4 shows several structural similarities to other de facto transcription factors outside the DNA-binding HMG-box. First, several poly-alanine and poly-glycine stretches are present. Glycine-alanine-rich stretches have been described in members of the homeodomain gene family (He et al. 1991; Suzuki et al. 1990) and in the human helix-loop-helix protein TFEB (Carr and Sharp 1990). Although the potential functional significance of glycine-alanine-rich sequences has not been established, alanine-rich subsegments are found in proteins known to be transcriptional repressors, such as engrailed, even-skipped, kruppel (Han et al. 1989; Licht et al. 1990; Zuo et al. 1991) and in *AEF-1* (Falb and Maniatis 1992). A detailed analysis of the Drosophila kruppel protein has revealed that an 85-amino acid, alanine-rich region is sufficient to repress transcription at a distance in mammalian cotransfection assays (Licht et al. 1990). Second, regions rich in serine residues are found and the predicted SOX4 protein contains a number of putative casein kinase and histone kinase phosphorylation sites, which may be involved in the regulation of this protein's function, while a putative AP2 site has been identified in the 5′-untranslated DNA. This serine-rich sequence with multiple potential phosphorylation sites may confer transactivation activity as shown for the human placental CREB-327 protein (Lee et al. 1990). Furthermore, the SOX4 protein has a serine-rich (21%) acidic (23%) 77-amino acid carboxyl-tail. This feature shows some similarity to the very acidic, serine-rich C terminus of the yeast transcription factor CDC68 (Rowley et al. 1991) and to the RNA polymerase 1 transcription factors UBF1 and UBF2, in which it is postulated to function as an activating domain (O'Mahony et al. 1992).

A comparison of the human and mouse SOX4/Sox-4 proteins outside the HMG-box (Hans Clevers personal communication) reveals identity over the N-terminal 56 amino acids and over the serine-rich

Fig. 2. SOX4 cDNA and deduced amino acid sequence. The beginning of the sequence is the 5′ end of cDNA clone NM16. The 79-amino acid HMG-box domain is **open-boxed.** An in-frame stop codon is indicated with an **asterisk.**

carboxyl-tail (a stretch of 89 amino acids). However, from amino acid 136 to 385, human SOX4 has several insertions not present in the mouse protein, as well as some single amino acid substitutions. In general, the insertions consist of strings of additional glycine and/or alanine residues, which account for the greater size of the human SOX4 protein compared with mouse Sox-4 (474/440 amino acids). This suggests that the glycine-alanine-rich region of the protein may serve primarily as a spacer and as such is able to tolerate variation in overall length.

Finally, somatic cell hybrids have been used to assign the SOX4 gene to human Chr 6 distal to the MHC. Metaphase fluorescence in situ hybridization places

SOX4-like HMG-boxes

```
SOX4(this paper)            GHIKRPMNAF MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIREAERLR LKHMADYPDY KYRPRKKVK
SOX4(Denny et al.,1992a)                 SEIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIREAERLR LKHMADY
Sox-4(Gubbay et al.,1990)   GHIKRPMNAF MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIQEAERLR LKHMADYPDY KYRPRKKVK
AMAES4                        VKRPMNAF MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIREAERLR LKHMADYPDY KYRP
AMAES1                        VKRPMNAF MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKGSDKI PFIREAERLR LKHMADYPDY KYRP
IRE ABP(Nasrin et al.,1991) GHIKRPMNAF MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIQEAEGLR LKHMADYPDY KYRPRKKVK
Lf4(Griffiths,1991)                      MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIREAERLR LKHMAD
Lf6(Griffiths,1991)                      MVWSKIERRK IMEQSPDMHN AEISKRLGKR WKMLKDSEKI PFIREAERLR LKHMAD
```

Fig. 3. Amino acid sequence alignment of the HMG-boxes from different SOX4-like proteins. Abbreviations are as follows: SOX4, human SOX4 protein; Sox-4, murine Sox-4 protein; AMAES, alligator SRY-related sequences, EMBL accession no. M86313 and M86315; IRE-ABP, insulin response element-Abinding protein from rat; Lf, lesser black-backed gull (*Larus fuscus*) SRY-related sequences. **Boxes** indicate amino acid differences.

SOX4 at 6p23 (unpublished data). This localization is of interest because of the recent cloning of a rat *SOX* family member (*IRE-ABP*) on the basis of its binding to an insulin-response element (Alexander et al. 1992a,b; Nasrin et al. 1991). IRE-ABP is 98% identical to mouse *Sox-4* in the HMG-box, which suggests a possible role for the SOX4 protein in regulating transcription in response to insulin. Since linkage of insulin-dependent diabetes mellitus (IDDM) to or near the HLA region has long been established (Todd 1990), the localization of human SOX4 to 6p21.3-pter raises the possibility that SOX4 may be an IDDM susceptibility gene. However, until sequence outside the HMG-box is available, it remains unclear whether *IRE-ABP* is the rat homolog of human SOX4 or another member of this closely related gene family.

The deduced amino acid sequence of the human SOX4 gene is consistent with its being a transcription factor. By analogy with SRY it may have some putative role in differentiation and development. For unraveling the function of SOX4, disruption of the gene in the mouse germ line or in vitro will ultimately be required.

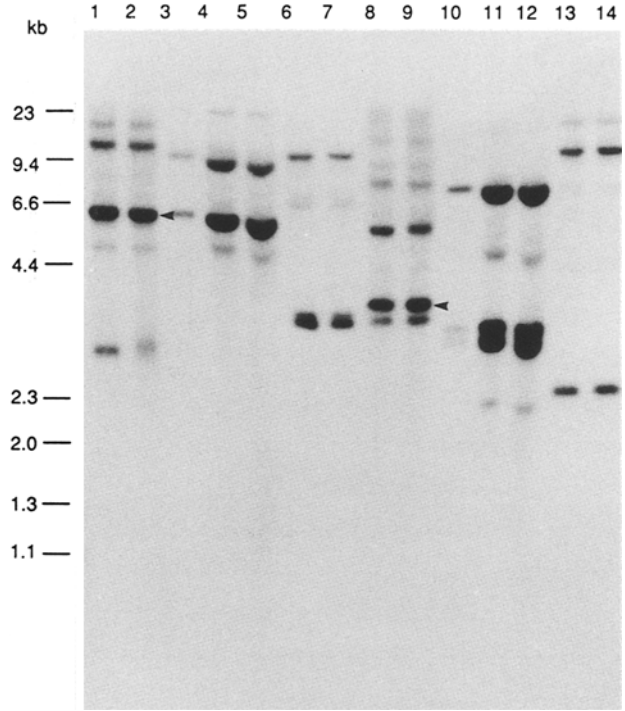

Fig. 4. Southern blot analysis of genomic DNAs with a PCR-generated SOX4-box probe, washed at high stringency (65°C/0.1 × SSPE). In lanes 1–7 genomic DNA has been cut with *Eco*RI, while lanes 8–14 are *Hin*dIII digests. The DNAs are as follows: 1 and 8, PGF (a male lymphoblastoid cell line); 2 and 9, WT49 (a female human lymphoblastoid cell line); 3 and 10, 853 (a hamster–human hybrid containing only the human Y chromosome), these lanes are relatively underloaded; 4 and 11, clone 2D (a hamster–human hybrid, containing the human X Chr only); 5 and 12, Wg3H (hamster); 6 and 13, BALB/c male DNA; and 7 and 14, BALB/c female DNA. The **arrows** indicate the SOX4-specific bands detected when the *Xba*I-*Eco*RI 3′-untranslated fragment is used as a probe.

Acknowledgments The authors are particularly indebted to Dot Bennett, Ross Hawkins, Milena Stevanovic, Jamie Foster, Nigel Spurr and Lesley Rooke, Ketan Patel and Fiona Priest, Mark Ross,

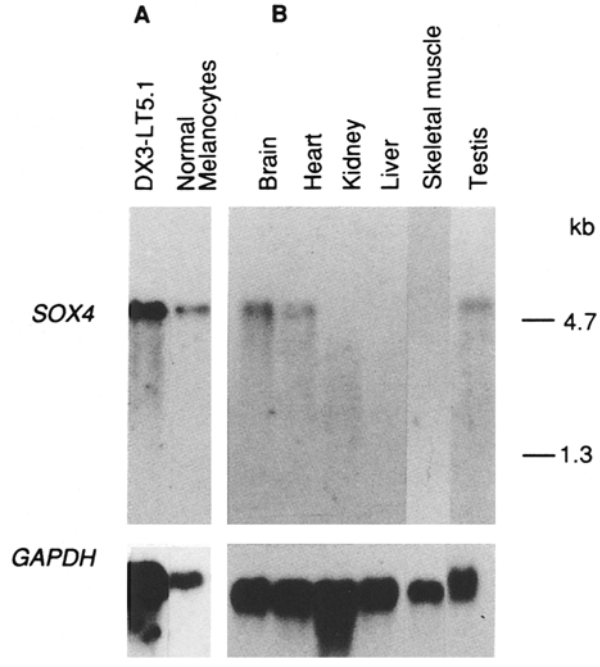

Fig. 5. Northern blot analysis of **(A)** 10 μg of poly(A)$^+$ RNA from normal melanocytes and from a melanoma cell line (DX3-LT5.1) hybridized with a SOX4-specific probe (the 3′-untranslated *Xba*I-*Eco*RI fragment). **(B)** 10 μg of poly(A)$^+$ RNA from various adult mouse tissues, probed with the mouse *Sox-4* cDNA 5c.1. In the lower panels the same filters were subsequently rehybridized with a GAPDH probe as a control for RNA quality and loading.

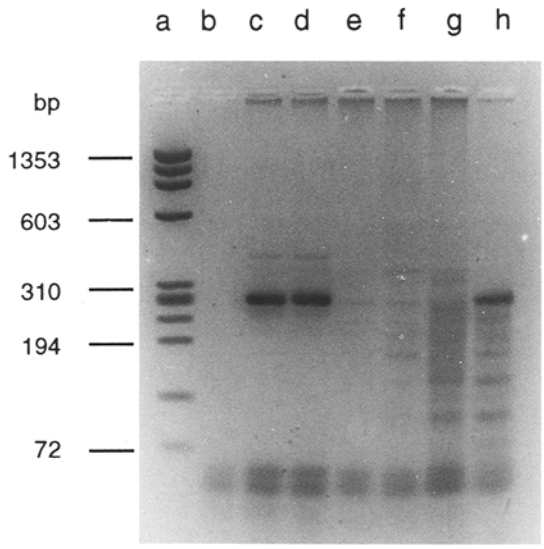

Fig. 6. Analysis of DNA products amplified by PCR with primers specific for SOX4. DNA samples are: Lane (a) Molecular weight markers (ϕX174 *Hae*III); (b) water control; (c) WT49 (human 46XX); (d) PGF (human 46XY); (e) 129/J female; (f) 129/J male; (g) mouse–human hybrid MCP6 (retains human 6p21.3-qter (Nicholas et al. 1973)); (h) mouse–human hybrid 5647 clone 22 (retains human 6p21.1-pter (Nagarajan et al. 1986)).

Androulla Economou, and to other members of the various labs for helpful discussions. We thank the Imperial Cancer Research Fund, the Medical Research Council of Great Britain, and The Wellcome Trust for their support. C.J. Farr is funded by a Medical Research Council HGMP Senior Research Fellowship.

References

Alexander, B.M., Dugast, I., Ercolani, L., Kong, X.F., Giere, L., Nasrin, N. (1992a). Multiple insulin-responsive elements regulate transcription of the GAPDH gene. Adv. Enzyme Regul. 32, 149–59.

Alexander, B.M., Ercolani, L., Kong, X.F., Nasrin, N. (1992b). Identification of a core motif that is recognized by three members of the HMG class of transcriptional regulators: IRE-ABP, SRY, and TCF-1 alpha. J. Cell Biochem. 48, 129–135.

Aviv, H., Leder, R. (1972). Purification of biologically active globin mRNA by chromatography on oligo thymidylic acid cellulose. Proc. Natl. Acad. Sci. USA 69, 1408–1412.

Bennett, D.C., Bridges, K., McKay, I.A. (1985). Clonal separation of mature melanocytes from premelanocytes in a diploid strain: spontaneous and induced pigmentation of premelanocytes. J. Cell Sci. 77, 167–183.

Carr, C.S., Sharp, P.A. (1990). A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. Mol. Cell. Biol. 10, 4384–4388.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294–5299.

Denny, P., Swift, S., Brand, N., Dabhade, N., Barton, P., Ashworth, A. (1992a). A conserved family of genes related to the testis determining gene, SRY. Nucleic Acids Res. 20, 2887.

Denny, P., Swift, S., Connor, F., Ashworth, A. (1992b). An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. EMBO J. 11, 3705–3712.

Eisinger, M., Marko, O. (1982). Selective proliferation of normal human melanocytes from premelanocytes in a human diploid strain. Proc. Natl. Acad. Sci. USA 79, 2018–2022.

Falb, D., Maniatis, T. (1992). Drosophila transcriptional repressor protein that binds specifically to negative control elements in fat body enhancers. Mol. Cell. Biol. 12, 4093–4103.

Farr, C.J., Goodfellow, P.N. (1992). Hidden messages in genetic maps. Science 258, 49.

Feinberg, A.P., Vogelstein, V. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 137, 266–267.

Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R., Bianchi, M.E. (1992). SRY, like HMG1, recognises sharp angles in DNA. EMBO J. 11, 4497–4506.

Giese, K., Cos, J., Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell 69, 185–195.

Griffiths, R. (1991). The isolation of conserved DNA sequences related to the human sex-determining region Y gene from the lesser black-backed gull (*Larus fuscus*). Proc. R. Soc. Lond. Biol. 244, 123–128.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., Lovell, B.R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature 346, 245–250.

Han, K., Levine, M.S., Manley, J.L. (1989). Synergistic activation and repression of transcription by Drosophila homeobox proteins. Cell 56, 573–583.

Harley, V.R., Jackson, D.I., Hextall, P.J., Hawkins, J.R., Berkovitz, G.D., Sockanathan, S., Lovell, B.R., Goodfellow, P.N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. Science 255, 453–456.

He, X., Gerrero, R., Simmons, D.M., Park, R.E., Lin, C.J., Swanson, L.W., Rosenfeld, M.G. (1991). Tst-1, a member of the POU domain gene family, binds the promoter of the gene encoding the cell surface adhesion molecule Po. Mol. Cell. Biol. 11, 1739–1744.

Jantzen, H.-M., Admon, A., Bell, S.P., Tjian, R. (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature 344, 830–836.

Kwon, B.S., Haq, A.K., Pomerantz, S.H., Halaban, R. (1987). Isolation and sequence of a cDNA clone for human tyrosinase that maps at the mouse c-albino locus. Proc. Natl. Acad. Sci. USA 84, 7473–7477.

Larin, Z., Monaco, A.P., Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA 88, 4123–4127.

Lee, C.Q., Yun, Y.D., Hoeffler, J.P., Habener, J.F. (1990). Cyclic-AMP-responsive transcriptional activation of CREB-327 involves interdependent phosphorylated subdomains. EMBO J. 9, 4455–4465.

Licht, J.D., Grossel, M.J., Figge, J., Hansen, U.M. (1990). Drosophila Kruppel protein is a transcriptional repressor. Nature 346, 76–79.

Nagarajan, L., Louie, E., Tsujimoto, Y., Ar-Rushdi, A., Huebner, K., Croce, C.M. (1986). Localization of the human pim oncogene (PIM) to a region of chromosome 6 involved in translocations in acute leukemias. Proc. Natl. Acad. Sci. USA 83, 2556–2560.

Nasrin, N., Buggs, C., Kong, X.F., Carnazza, J., Goebl, M., Alexander, B.M. (1991). DNA-binding properties of the product of the testis-determining gene and a related protein. Nature 354, 317–320.

Nicholas, J.F., Dubois, P., Jakob, H., Gaillard, J., Jacob, F. (1973). Teratocarcinome de la souris: differenciation en culture d'une lignee de cellules primitives a potentialites multiples. Ann. Microbiol. 126, 3–22.

O'Mahony, D.J., Smith, S.D., Xie, W., Rothblum, L.I. (1992). Analysis of the phosphorylation, DNA-binding and dimerization properties of the RNA polymerase I transcription factors UBF1 and UBF2. Nucleic Acids Res. 20, 1301–1308.

Omerod, E.J., Everett, C.A., Hart, I.A. (1986). Enhanced experimental metastatic capacity of a human tumour line following treatment with 5-azacytidine. Cancer Res. 46, 884–890.

Ragoussis, R., Senger, G., Mockridge, I., Sanseau, P., Ruddy, S., Dudley, K., Sheer, D., Trowsdale, J. (1992). A testis-expressed Zn finger gene (ZN76) in human 6p21.3 centromeric to the MHC is closely linked to the human homolog of the t-complex gene tcp-11. Genomics 14, 673–679.

Rowley, A., Singer, R.A., Johnston, G.C. (1991). CDC68, a yeast

gene that affects regulation of cell proliferation and transcription, encodes a protein with a highly acidic carboxyl terminus. Mol. Cell. Biol. 11, 5718–5726.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). *Molecular Cloning: A Labotatory Manual,* 2nd ed. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell, B.R., Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240–244.

Suzuki, N., Rohdewohld, H., Neuman, T., Gruss, P., Scholer, H.R. (1990). Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain. EMBO J. 9, 3723–3732.

Todd, J.A. (1990). The role of MHC class II genes in Type 1 diabetes. Curr. Top. Microbiol. Immunol. 164, 17–40.

Travis, A., Amsterdam, A., Belanger, C., Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. Genes Dev. 5, 880–894.

van de Wetering, M., Oosterwegel, M., Dooijes, D., Clevers, H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J. 10, 123–132.

Waterman, M.L., Fischer, W.H., Jones, K.A. (1991). A thymus-specific member of the HMG protein family regulates the human T cell receptor Cα enhancer. Genes Dev. 5, 656–669.

Zuo, P., Stanojevic, D., Colgan, J., Han, K., Levine, M., Manley, J.L. (1991). Activation and repression of transcription by the gap proteins hunchback and Kruppel in cultured Drosophila cells. Genes Dev. 5, 254–264.