LETTER

The Human Homolog *T* of the Mouse *T(Brachyury)* Gene; Gene Structure, cDNA Sequence, and Assignment to Chromosome 6q27

Y.H. Edwards,[1] W. Putt, K.M. Lekoape,[2] D. Stott,[2] M. Fox, D.A. Hopkinson, and J. Sowden

Medical Research Council (MRC) Human Biochemical Genetics Unit, University College London, London NW1 2HE, England; [2]Animal Molecular Genetics Group, Department of Biochemistry, University of Warwick, Coventry CV4 7AL, England

We have cloned the human gene encoding the transcription factor T. T protein is vital for the formation of posterior mesoderm and axial development in all vertebrates. *Brachyury* mutant mice, which lack T protein, die in utero with abnormal notochord, posterior somites, and allantois. We have identified human T genomic clones and derived the mRNA sequence and gene structure. There is 91% amino acid identity between human and mouse T proteins overall and complete identity across 77 amino acids of the T-box motif within the DNA-binding domain. Human *T* expression is very similar to that found for *T* in other vertebrate species and is confined to cells derived from the notochord. The human *T* gene maps to chromosome 6q27 and is only the second human member of the T-box gene family to be described.

T protein is vital for the formation and differentiation of posterior mesoderm and for axial development in vertebrates. Evidence for this comes from the analysis of *T* mutant mice and zebrafish (Beddington et al. 1992; Schulte-Merker et al. 1994). *Brachyury* mutant mice lack T protein and die in utero. The notochord fails to differentiate and does not extend into the trunk, posterior somites are missing, and the allantois is severely reduced (for review, see Wilson et al. 1993, 1995). In zebrafish the *no tail* mutation (*ntl*) has been identified as the homolog of *Brachyury* (Schulte-Merker et al. 1994). *ntl* embryos die after hatching, lack notochords and tails, and possess abnormal trunk somites (Halpern et al. 1993).

The *T* gene encodes a transcription factor that binds to a specific DNA element via its amino-terminal region (Schulte-Merker et al. 1992; Kispert and Herrmann 1993; Kispert et al. 1995a). As yet, target genes for T have not been identified, but it seems likely that they will be involved in posterior mesoderm formation and notochord differentiation. A protein motif within the DNA-binding domain, the so-called T-box, is highly conserved among *T* homologs from different species and also defines a broader family of T-box genes. These include a human gene *TBX2* (Campbell et al. 1995), the mouse genes *Tbx1*, *Tbx2*, and *Tbx3* (Bollag et al. 1994), the *Drosophila* genes *Trg* (Kispert et al. 1994) and optomotor blind (*omb*) (Pflugfelder et al. 1992), the sea urchin gene *HpTa* (Harada et al. 1995) and four *Caenorhabditis elegans* genes (Agulnik et al. 1995).

The *T* gene has been cloned from a spectrum of vertebrate species: the mouse (Herrmann et al. 1990), the zebrafish (Zf-T; Schulte-Merker et al. 1992), *Xenopus* (Xbra; Smith et al. 1991), chicken (Ch-T; Kispert et al. 1995b), and a lower chordate, *Halocynthia roretzi*. (As-T; Yasuo and Satoh 1993). The observation that *T* is expressed early in development in prospective posterolateral mesoderm and notochord in all vertebrates and is expressed in the primordial notochord cells of the ascidian, suggests that the role of T in axial development is highly conserved. Thus far, the human homologue of *T* has not been described, and this omission reflects the relative inaccessibility of the early stage human embryo material that

[1]Corresponding author.
E-MAIL yedwards@hgmp.mrc.ac.uk; FAX 0171 387 3496.

6:226–233 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

would be an appropriate source of *T* mRNA. We have used the alternative approach of isolating human *T* genomic clones and deriving mRNA and amino acid sequences for human *T*. This study represents the first detailed structural analysis of the *T* gene from any species and complements the more limited structural data reported for the mouse and zebrafish genes (Stott et al. 1993; Schulte-Merker et al. 1994). The human T protein sequence displays high homology to T from the mouse and other vertebrate species, particularly in the DNA-binding domain. We have mapped the human *T* gene to chromosome 6q27, a localization consistent with the previous assignment of other human homologs of mouse genes that lie proximally in the mouse t-complex.

RESULTS

The human *T* gene; isolation and exon/intron structure

Genomic clones spanning the entire human *T* gene were isolated and DNA fragments containing *T* exons were identified by Southern blot analysis of single and double *Bam*HI, *Sac*I, and *Pst*I digests using the mouse cDNA (Herrmann et al. 1990) as probe. The exon-containing fragments were subcloned, and the precise location of the exon/intron boundaries determined by further restriction enzyme mapping and sequence analysis. These studies showed that an 11-kb *Not*I fragment (pNot10) from the cosmid cKL1 contained the entire *T* gene apart from the extreme 3′ end. To recover the missing 3′ sequences, a 1-kb *Eco*RI fragment from intron 7 was used to screen a human genomic phage library. A single λ clone, λHuT. 1, containing a 17-kb insert that overlapped pNot10 by 7 kb at its 5′ end and extended a farther 10 kb into 3′-flanking sequence was isolated and mapped.

A detailed restriction map of the complete human *T* gene was obtained by isolating *Sac*I, *Pst*I, and *Bam*HI fragments from each clone and further digestion with a variety of restriction enzymes; for clarity only the positions of *Pst*I, *Sac*I, *Hin*dIII, and *Bam*HI sites are shown in Figure 1. The *T* gene is 10 kb long and comprises eight exons. The sequences at the 5′ and 3′ junction of each intron and exon are shown in Table 1. The exons vary in size between 62 bp (exons 4 and 5) and 934 bp (exon 8). The introns vary in size from 0.128 kb [intervening sequence 4 (IVS 4)] to 2.1 kb (IVS 7). All of the intron/exon junctions conform to the consensus eukaryotic structure beginning with the dinucleotide gt and ending with ag. The overall structure of human *T* is apparently similar to that reported for the mouse *T* gene (Stott et al. 1993), although direct comparison is not possible as details of the exon/intron boundaries and intron size have not been published for the mouse.

The human and mouse 5′ sequences show 66% identity across the 109 bp of mouse 5′-untranslated region (UTR) for which data are available (Herrmann et al. 1990). From comparison with the mouse *T* mRNA we expect that the 5′ UTR of the human *T* mRNA will extend ~159 bp upstream of the first methionine codon (Stott et al. 1993). In the human sequence the first Met is preceded by 216 nucleotides of open reading frame, in contrast to mouse *T* where a stop codon occurs 42 nucleotides upstream of the first Met. The 5′ sequences of the human *T* gene are G + C rich; across a 1-kb region that encompasses exon 1, intron 1, and the most proximal 5′-flanking sequence, the G + C content is 68% and the CpG

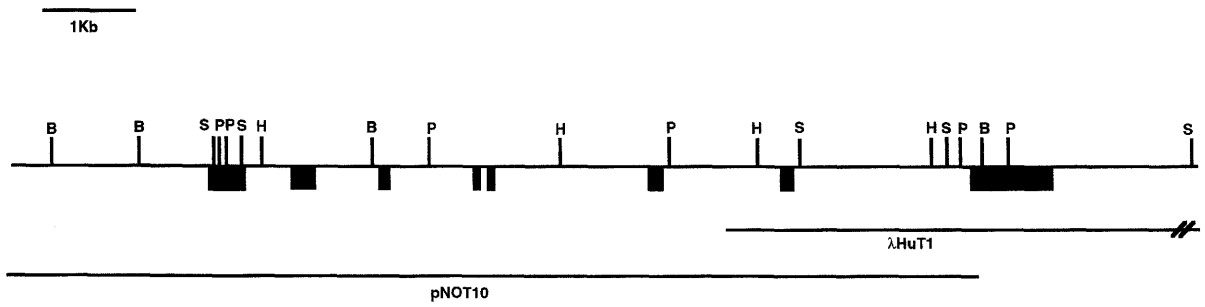

Figure 1 Map of the human *T* gene. Sites for the restriction enzymes *Bam*HI (B); *Sst*I (S); *Pst*I (P); and *Hin*dIII (H) are shown. Exons are indicated as solid boxes, and the extent of the two genomic clones pNot10 and λHuT.1 is also shown.

Table 1. Intron/Exon Structure of the Human *T* Gene

Exon (bp)	3′ exon sequence[a]	Intron (bp)	5′ exon sequence[a]
exon 1 (206 coding plus c. 200 UTR)	67 Thr Lys Asn Gly Ar G ACC AAG AAC GGC AG	gtgggtgcgcgtccggagcc IVS 1 (528) gcccgccccgtgttttccag	71 g Arg Met Phe Pro G AGG ATG TTT CCG GT
exon 2 (265)	155 Asn Gly Gly Gly Gln AAC GGA GGG GGC CAG	gtaggtgtgaggagcccgcc IVS 2 (c775) cctgtatctgtctccctcag	160 Ile Met Leu Arg Ser ATC ATG CTG AAC TCC
exon 3 (135)	200 Tyr Gln Asn Glu Glu TAT CAG AAC GAG GAG	gtgagaaggacagaggcgga IVS 3 (c900) tgcaatttcttgtttttcag	205 Ile Thr Ala Leu Lys ATC ACA GCT CTT AAA
exon 4 (62)	221 Asp Ala Lys Glu Ar T GAT GCA AAG GAA AG	gtgaagaagaattctcttaa IVS 4 (128) ttttttttcattttgatag	225 g Ser Asp His Lys A AGT GAT CAC AAA GA
exon 5 (62)	242 Gly Tyr Ser Gln T CT GGG TAC TCC CAA T	gtacggtttgttgcactctg IVS 5 (c1850) cactgtctttctcacagcag	246 rp Gly Trp Leu Leu GG GGG TGG CTT CTT C
exon 6 (174)	300 Asn Asn Ser Pro T GG AAC AAT TCT CCA A	gtgagtcctcagcctcattc IVS 6 (c1300) ttgttaccttcccccttag	304 hr Tyr Ser Asp Asn CC TAT TCT GAC AAC T
exon 7 (130)	342 Pro Thr Ser Ser Se A CCT ACC AGC TCC AG	gtaatgtatgtccaggacct IVS 7 (c2100) ctggtgtctttctgttcgag	346 r Gln Tyr Pro Ser T CAG TAC CCC AGC CT
exon 8 (274) c.660bp UTR	435 Pro Pro Ser Met * CCA CCT TCC ATG TGA		

[a]Each exon/intron junction is shown with exon sequence in uppercase and intron sequence in lowercase. The corresponding amino acids are numbered as in the complete human *T* amino acid sequence. The sizes of each exon and intron (IVS) are given in parentheses. Where exact numbers are given the intron was sequenced in full; in other cases, partial sequence has been obtained and an estimate of size made from the restriction map.

to GpC ratio is 0.85 (compared to a G + C of 40% and a CpG/GpC ratio of 0.1 for bulk DNA). The elements normally associated with genes transcribed by RNA polymerase II, the TATA box, and the CCAAT box do not occur in the immediate 5′-flanking sequence. The mouse promoter also lacks these elements (D. Stott, unpubl.).

Nucleotide sequence extending 713 nucleotides downstream of the stop codon in exon 8 was determined and compared with the mouse *T* 3′ UTR. The two sequences showed a moderate level of sequence identity (58%). In mouse *T* the 3′ UTR is 626 nucleotides long with a polyadenylation signal AATAAA, 25 nucleotides upstream of the site of polyadenylation (Herrmann et al. 1990). This site is conserved in human *T* and lies at nucleotide 636 within a 69-nucleotide stretch showing 90% identity to the mouse sequence. The human *T* sequence also contains a second potential polyadenylation signal 457 nucleotides downstream of the stop codon, which is not present in the mouse.

The coding sequence of human *T* is 1305 nucleotides long, encoding 435 amino acids, one less than in the mouse *T* protein (Fig. 2). Overall the human *T* coding sequence shows 85% nucleotide sequence identity and 91% amino acid sequence identity to the mouse sequences. Conservation of the DNA-binding domain of the T protein (Kispert and Herrmann 1993) is highest between amino acids 1–223, with only two amino acid substitutions, Asn (human) to Ser (mouse) at amino acid 26 in exon 1 and Ala (human) to Thr (mouse) at amino acid 96 in exon 2. The remaining 212 amino acids of the protein sequences encoded by exons 5–8, are less well conserved, as there are 38 differences (18% difference) between man and mouse. Most of the changes are amino acid substitutions, but in exon 8 the human sequence has two amino acids

```
          exon1
Hu-T      MSSPGTESAGKSLQYRVDHLLSAVENELQAGSEKGDPTERELRVGLEESE  50
M-T       ......................S.....................
                                exon2
Hu-T      LWLRFKELTNEMIVTKNGR^RMFPVLKVNVSGLDPNAMYSFLLDFVAADN  99
M-T       ..................  ........................T...

Hu-T      HRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLT 149
M-T       ..................................................
                    exon3
Hu-T      NKLNGGGQ^IMLNSLHKYEPRIHIVRVGGPQRMITSHCFPETQFIAVTAY 198
M-T       ........ ..........................................
               exon4                    exon5
Hu-T      QNEE^ITALKIKYNPFAKAFLDAKE^RSDHKEMMEEPGDSQQPGYSQ^WG 245
M-T       .... ..................... .N...DV......C....... ..
          exon6
Hu-T      WLLPGTSTLCPPANPHPQFGGALSLPSTHSCDRYPTLRSHRSSPYPSPYA 295
M-T       ..V..AG......SS......S.......G.E...A..N...........
                  exon7
Hu-T      HRNNSP^TYSDNSPACLSMLQSHDNWSSLGMPAHPSSLPVSHNASPPTSS 344
M-T       ...S.. ..A...S.................V.G.T.............G.
           exon8
Hu-T      ^SQYPSLWSVSNPAVTPGSQAAAVSNGLGAQFFRGSPAHYTPLTHPVSAP 393
M-T        ...........GTI.....T.G.....................T...A

Hu-T      --SSSGSPLYEGAAAATDIVDSQYDAAAQGRLIASWTPVSPPSM 435
M-T       TS......M.....TV...S.....-T..SLL............
```

Figure 2 Deduced amino acid sequence of the human T protein (Hu-T) and comparison with the mouse Brachyury T sequence (M-T). Dots indicate amino acids identical to those in Hu-T, and dashes indicate sequence breaks included to give optimal alignments. Exon/intron boundaries are indicated as ^; where a residue is interupted it is shown to the right of the boundary. The DNA-binding domain is shown in bold; two transcriptional activator domains are underlined.

less than the mouse; the extra amino acids Thr and Ser in the mouse sequence lie between residues 393 and 394 of the human sequence. In addition, the human T protein contains an extra amino acid, Ala, at position 417. This carboxy-terminal region of the T protein constitutes two *trans*-activation and two repressor domains (Kispert et al. 1995a) but the even spread of amino acid differences between human and mouse T suggests that none of these domains is any more conserved than another.

The human T protein sequence shows a high level of sequence identity with T homologs from other vertebrate species. For example there is 80% amino acid identity with chicken Ch-T, 75% with *Xenopus laevis* Xbra and 64% with zebrafish Zf-T protein. The tunicate homolog As-T and *Drosophila* Trg (Kispert et al. 1994) both show a lower level of identity (46%). In every case the highest level of conservation is found in the DNA-binding domain (Bollag et al. 1994); 80%–98% when human T is compared to vertebrate sequence and 64% is compared to *Drosphila Trg*.

In contrast with the high level of sequence identity between human T and T from other vertebrates, there appears to be relatively less similarity between T and other T-box protein sequences. Full-length comparisons are not possible, as complete sequences are not yet available for all T-box proteins, but across the human T-box region (amino acids 124–200) there is only 52% identity between human T and human TBX2 and only 58%, 52%, and 55% identity with mouse Tbx1, Tbx2, and Tbx3 sequences, respectively (Fig. 3).

Human *T* Expression in Notochord

In the mouse, *T* expression is high in notochord cells throughout early development and is readily detected in the mature notochord cells of the nucleous pulposus that forms in the central portion of the intervertebral disc at 17.5 days postcoitum (dpc) (Wilkinson et al. 1990). This latter stage of mouse development is approximately equivalent to 14–15 weeks gestation in human development, and we have used RT–PCR to look for the presence of *T* mRNA in 13 and 14 week gestation human fetal intervertebral discs. Four primers were designed from exons 2, 3, 5, and 7 of human *T* and were combined in pairs to amplify between exons 2 and 3 (2F/3R), exons 5 and 7 (5F/7R), and exons 2 and 7 (2F/7R). The predicted products of 201, 289, and 568 nucleotides, respectively, were generated from the cDNA prepared from human fetal intervertebral discs (Fig. 4A). *T* RNA was not detected in several other human tissues, including adult and fetal intestine and muscle, erythroid and intestinal cell lines, and 14-week gestation spinal cord (Fig. 4B). Southern blotting and hybridization with radiolabeled human *T*-specific probes confirmed that the PCR products contained *T* sequences and that *T* RNA was not present in other tissues even in very low amounts (data not shown). The quantity and quality of the RNA samples used in the RT–PCR was assessed in all cases by amplification of the same cDNAs with primers specific for the ubiquitously expressed phosphoglucomutase gene (*PGM1*) (Fig 4B).

Chromosome Assignment of Human *T*

Using fluorescent in situ hybridization (FISH) and pNot10 DNA as probe we have mapped the human *T* gene to the end of the long arm of

```
      124                                                        172
Hu-T  YIHPDSPNFG AHWMKAPVSF SKVKLTNKLN GG-GQIMLNS LHKYEPRIHI
 M-T  .......... .......... .......... ..-....... ..........
Ch-T  .......... .......... .......... ..-....... ..........
Zf-T  .......... .......... .....S.... ..-....... ..........
TBX2  .......AT. EQ..AK..A. H.L....NIS DKH.FTI... M...Q..F..
Tbx2  .......AT. EQ..AK..A. H.L....NIS DKH.FTI... M...Q..F..
Tbx3  .......AT. EQ..SKV.T. H.L....NIS DKH.FTI... M...Q..F..
Tbx1  .......AK. .Q...QI... D.L....N.L DDN.H.I... M.R.Q..F.V

      173                                  200
Hu-T  VRVGGPQRMI T-------SH CFPETQFIAV TAYQN
 M-T  .......... .-------.. .......... .....
Ch-T  .......... .-------.. S......T.. .....
Zf-T  .K...I.K.. S--------.Q S......... .....
TBX2  .---RANDIL KLPYSTFRTY V....D.... .....
Tbx2  .---RANDIL KLPYSTFRTY V....D.... .....
Tbx3  .---RANDIL KLPYSTFRTY L....E.... .....
Tbx1  .Y.APRKDSE KYEEENFKTF V.E..R.T.. .....
```

Figure 3 Comparison of the T-box amino acid sequence among vertebrate T homologs, including human (Hu-T), mouse (M-T), chicken (Ch-T), zebrafish (Zf-T), and various other T-box proteins, including human (TBX2) and mouse (Tbx 1-3). Dots indicate sequence identity; gaps introduced to maximize alignment are indicated by dashes. Residues are numbered according to the human sequence.

chromosome 6, 6q27 (Fig. 5). In the mouse, *T* maps close to the MHC locus and forms part of the so-called t-complex on chromosome 17. It is of interest that other human homologs from the mouse t-complex, TCP10, PLG, IGF2R, TCP1, THBS2, and TCTE3, also map in or close to the tip of 6q (Ziegler et al. 1991; Blanché et al. 1992; Rappold et al. 1992). These genes are grouped together in the proximal region of the mouse t-complex in the order *Tcp10a, Plg, Igf2r, Tcp1, Tcp10b, Tcp10c, Thbs2,* and *Tcte3*. Recently, we have carried out genetic analysis using a polymorphism of a human to examine the position of *T* relative to some of these markers in human 6q27 (Morrison et al. 1996). These studies show that human*T* lies between *TCP1* and *TCP10*.

DISCUSSION

This study reports the isolation and mapping of human genomic clones encoding the transcription factor T. The human *T* gene is homologous to the murine *T(Brachyury)* gene. The human gene consists of eight exons and spans 10 kb, and the open reading frame encodes a protein of 435 amino acids that shares 91% identity overall with mouse T. The human and mouse amino acid sequences are identical within the T-box region of the DNA-binding domain. RT–PCR shows that in the developing human embryo *T* expression is limited to the nucleus pulposus of the intervertebral discs, strongly suggesting that the notochord is an important site of *T* expression in man, as it is in the mouse and other vertebrate species.

T is part of a large new family of transcription factors that share a novel protein motif, the T-box, associated with DNA-binding activity. In addition to *T*, one other human T-box gene has been cloned, *TBX2*, which is expressed in lung,

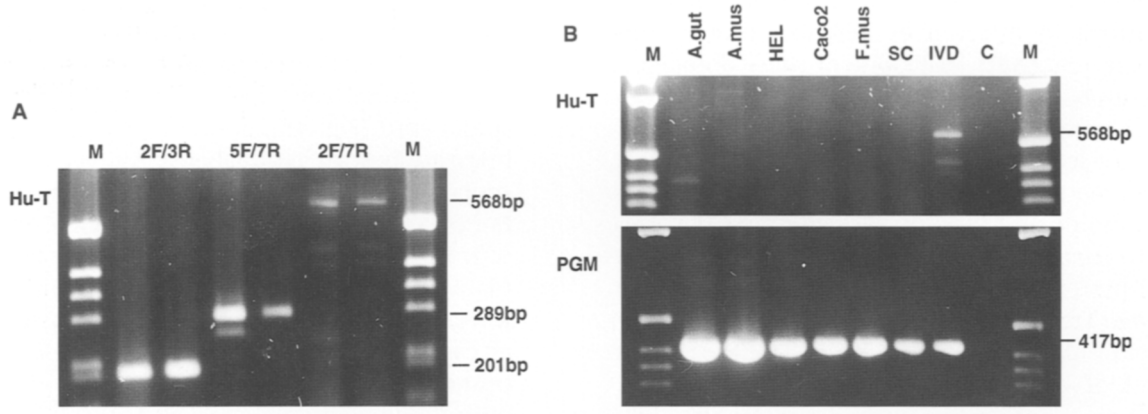

Figure 4 RT–PCR analysis of human *T*. Four primers were designed from exons 2, 3, 5, and 7 of human *T* and combined in pairs to amplify between exons 2 and 3 (2F/3R), exons 5 and 7 (5F/7R), and exons 2 and 7 (2F/7R). The sizes of the products are 201, 289, and 568 nucleotides, repectively. (*A*) Amplification of cDNA prepared from fetal intervertebral discs using the primer pairs 2F/3R, 5F/7R, and 2F/7R; (*B*) Amplification of cDNA from adult intestine (A.gut), skeletal muscle (A.Mus), fetal skeletal muscle (F.mus), spinal cord (SC), intervertebral disc (IVD), erythroid (HEL), and intestinal (Caco2) cell lines using 2F/7R. Amplification of the same cDNAs using primers specific for PGM1 was carried out as a control for RNA quality and quantity.

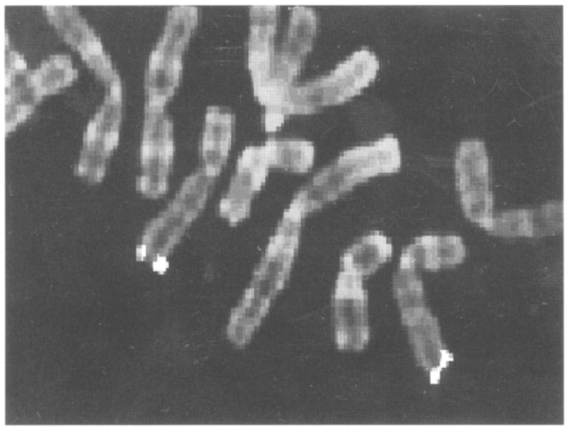

Figure 5 Fluorescent in situ hybridization to human metaphase chromosomes showing the localization of human *T* to 6q27. The genomic clone pNot10 was used as probe.

kidney, and placenta. In addition three T-box genes (*Tbx1–Tbx3*), have been described in the mouse (Bollag et al. 1994; Campbell et al. 1995), four T-box genes have been identified in *C. elegans* (Agulnik et al. 1995), and two from *Drosophila melanogaster* (Pflugfelder et al. 1992; Kispert et al. 1994). Phylogenetic trees based on sequence comparisons suggest that the common early metazoan ancestor possessed three T-box genes (Agulnik et al. 1995): one related to the *T* homologs, another to mouse *Tbx2* and *Tbx3, Drosophila omb* and *C. elegans Ce-tbx-2* and *Ce-tbx-7*, and a third related to mouse *Tbx1*.

We have shown that the human *T* gene is located on chromosome 6q27 within a cluster of genes whose mouse homologs map to the proximal region of the t-complex on mouse chromosome 17. At present it seems unlikely that other human T-box genes will be found in the 6q27 region, as mouse *Tbx1, Tbx2,* and *Tbx3* do not map to chromosome 17 but are dispersed on chromosomes 16, 11, and 5, respectively (Bollag et al. 1994), and human *TBX2* has been mapped to 17q23 (Agulnik et al. 1995). However the finding that four T-box genes of *C. elegans* are closely linked in chromosome III (Agulnik et al. 1995) suggests that clusters of T-box genes may exist in vertebrate species but have not yet been identified.

Although deficiency of *T* has been associated with defective mesoderm formation and abnormal posterior axial development in mouse and zebrafish (Beddington et al. 1992; Schulte-Merker et al. 1994), no clinical condition has been identified with abnormal *T* expression in man. In the 1970s the idea that *T* might be associated with neural tube defects was investigated genetically using HLA (a class 1 MHC locus) as marker (Amos et al. 1975; Bobrow et al. 1975; Fellous et al. 1979). The results from these studies were equivocal, and recently we have begun to reinvestigate this idea using *T* as a genetic marker (Morrison et al. 1996). At present, the only disease loci that have been mapped in the human 6q27 region are tumor suppressor genes associated with ovarian tumours (Orphanos et al. 1995) non-Hodgkin's lymphoma (Chomczynski and Sacchi 1987; Gaidano et al. 1992), and renal cell carcinoma (Morita et al. 1992). The specificity of *T* expression and the lack of tumors in *T* mutant mice argues against the involvement of *T* in any of these neoplasias. However, it is notable that the other human T-box gene, *TBX2*, maps to chromosome 17q23, a region frequently altered in ovarian carcinoma (Campbell et al. 1995).

METHODS

Isolation and Characterization of Genomic Clones for Human *T*

A genomic DNA clone (cosmid cKL1) was isolated from a library constructed in Cos2EMBL (a gift from Dr. A-M Frischauf, Imperial Cancer Research Fund, London, UK) using the mouse *T* cDNA as probe (Herrmann et al. 1990). An 11-kb *Not*I fragment from the 3′ end of cKL1 containing all of the hybridizing signal was subcloned (pNot10) into a Bluescript vector. In addition, a single λ clone, λHuT.1, which contains exons 7 and 8 and 3′-flanking sequences of the *T* gene, was isolated using a 3′ 1-kb *Eco*RI fragment from pNot10 to screen a λ2001 library. pNot10 and λHuT.1 were mapped for *Hin*dIII, *Sma*I, *Eco*RI, *Acc*I, *Bam*HI, *Hin*cII, *Kpn*I, *Sal*I, *Pst*I, *Sst*I, *Xba*I, and *Sph*I sites using single and double digests. Fragments hybridizing to cDNA were subcloned into M13mp18. DNA sequences across exon/intron junctions and in flanking regions were determined by the dideoxy chain termination method (Sanger et al. 1977). The distances between exons were estimated by restriction mapping, and those between exons 1, 2, and 3 confirmed by PCR amplification.

Fluorescent In Situ Hybridization

FISH was carried out on spreads of human lymphocyte metaphase chromosomes with biotinylated pNot10 DNA as probe exactly as described previously (Shortle et al. 1993).

Expression of *T* in Human Tissues

RNA was prepared (Chomczynski and Saachi 1987) from various human tissues, including intervertebral discs taken

from human fetal samples at 13 and 14 weeks gestation and dissected to remove the outer rings of collagen fibers. RNA PCR was carried out exactly as described in Sowden et al. (1995), using random oligonucleotide primers to generate first-strand cDNA and gene-specific primers in the subsequent PCR amplification of cDNA aliquots. Four primers derived from *T* exon sequence were synthesized: 5′-AGATGATGGAGGAACCCGGAGACAG-3′, Hu–Tex5F; 5′-CCAAGGCTGGACCAATTGTCATGGG-3′, Hu–Tex7R; 5′-ACTGGATGAAGGCTCCCGTCTCCTT-3′, Hu–Tex2F; and 5′-CCTCGTTCTGATAAGCAGTCACGGC-3′, Hu–Tex3R. Primers for the ubiquitously expressed *PGM1* gene were the same as those described previously (Edwards et al. 1995). After electrophoresis of PCR products on agarose gels, Southern blotting onto nylon membranes and hybridization using ^{32}P-labeled oligonucleotide was carried out using standard procedures.

ACKNOWLEDGMENTS

We would like to thank Dr. Lesley Wong at the MRC Tissue Bank, Royal Postgraduate Medical School, Hammersmith Hospital, for providing tissue samples. K.M. Leokape was in receipt of a Southern Africa Campus Scholarship from the World University Service. Part of this work was carried out with funding from the European Community Commission under the Biomedicine and Health program.

REFERENCES

Agulnik, S.I., R.J. Bollag, and L.M. Silver. 1995. Conservation of the T-Box gene family from *Mus musculus* to *Caenorhabditis elegans*. *Genomics* **25:** 214–219.

Amos, D.B., N. Ruderman, N. Mendell, and A.H. Johnson. 1975. Linkage between HLA and spinal development. *Transplant. Proc.* **7:** 93–95.

Beddington, R.S.P., P. Rashbass, and V. Wilson. 1992. *Brachyury*—A gene affecting mouse gastrulation and early organogenesis. *Development* (Suppl.): 157–165.

Blanché, H., L.G. Wright, G. Vergnaud, B. de Gouyon, V. Lauthier, L.M. Silver, J. Dausset, H.M. Cann, and R.S. Spielman. 1992. Genetic mapping of three human homologs of murine t-complex genes localizes TCP10 to 6q27, 15 cM distal to TCP1 and PLG. *Genomics* **12:** 826–828.

Bobrow, M., J.G. Bodmer, W.F. Bodmer, H.O. McDevitt, J. Lorber, and P.N. Swift. 1975. The search for a human equivalent of the mouse T-locus. Negative results from a study of HLA types in spina bifida. *Tissue Antigens* **5:** 234–237.

Bollag, R.J., Z. Siegfried, J.A. Cebra-Thomas, N. Garvey, E.M. Davison, and L.M. Silver. 1994. An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the *T* locus. *Nature Genet.* **7:** 383–389.

Campbell, C., K. Goodrich, G. Casey, and B. Beatty. 1995. Cloning and mapping of a human gene (TBX2) sharing a highly conserved protein motif with the Drosophila *omb* gene. *Genomics* **28:** 255–260.

Chomczynski, P. and N. Sacchi. 1987. Single-step of RNA isolation by acid guanidinium thiocyanate phenol chloroform extract. *Anal. Biochem.* **162:** 156–159.

Edwards, Y.H., W. Putt, M. Fox, and J.H. Ives. 1995. A novel human phosphoglucomutase (*PGM5*), maps to the centromeric region of chromosome 9. *Genomics* **30:** 350–353.

Fellous, M., J. Hors, J. Bone, J. Dausset, and F. Jacob. 1979. Are there human analogs of the mouse T-locus in central nervous system malformations? *Birth Defects Orig. Art. Ser.* **XV(3):** 93–104.

Gaidano, G., R.S. Hauptschein, N.Z. Parsa, K. Offit, P.H. Rao, G. Lenoir, D.M. Knowles, R.S. Chaganti, and R. Dalla-Favera. 1992. Deletions involving two distinct regions of 6q in B-cell non-Hodgkin lymphoma. *Blood* **80:** 1781–1787.

Halpern, M.E., R.K. Ho, C. Walker, and C.B. Kimmel. 1993. Induction of muscle pioneers and floor plate is distinguished by the zebrafish *no tail* mutation. *Cell* **75:** 99–111.

Harada, Y., H. Yasuo, and N. Satoh. 1995. A sea urchin homologue of the chordate Brachyury (T) gene is expressed in the secondary mesenchyme founder cells. *Development* **121:** 2747–2754.

Herrmann, B.G., S. Labeit, A. Poustka, T.R. King, and H. Lehrach. 1990. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343:** 617–622.

Kispert, A. and B.G. Herrmann. 1993. The *Brachyury* gene encodes a novel DNA binding protein. *EMBO J.* **12:** 3211–3220.

Kispert, A., B.G. Herrmann, M. Leptin, and R. Reuter. 1994. Homologs of the mouse *Brachyury* gene are involved in the specification of posterior terminal structures in *Drosophila, Tribolium*, and *Locusta*. *Genes & Dev.* **8:** 2137–2150.

Kispert, A., B. Koschorz, and B.G. Herrmann. 1995a. The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J.* **14:** 4763–4772.

Kispert, A., H. Ortner, J. Cooke, and B.G. Herrmann. 1995b. The chick brachyury gene: Developmental expression pattern and response to axial induction by localized activin. *Dev. Biol.* **168:** 406–415.

Morita, R., K. Hikiji, K. Yamakawa, and Y. Nakamura. 1992. Tumour suppressor genes associated with

development of human renal cell carcinoma. *Nippon-Rinsho* **50:** 3100–3105.

Morrison, K., C. Papapetrou, J. Attwood, F. Hol, A. Sampath, J. Burn, J. Sowden, D. Stott, E. Mariman, and Y.H. Edwards. 1996. Genetic mapping of the human homolog of mouse *T(Brachyury)* and a search for allele association between *T* and spina bifida. *Hum. Mol. Genet.* (in press).

Orphanos, V., G. McGown, Y. Hey, M. Thorncroft, M. Santibanez-Koref, S.E. Russell, I. Hickey, R.J. Atkinson, and J.M. Boyle. 1995. Allelic imbalance of chromosome 6q in ovarian tumours. *Br. J. Cancer* **71:** 666–669.

Pflugfelder, G.O., H. Roth, B. Poeck, S. Kerscher, H. Schwarz, B. Jonschker, and M. Heisenberg. 1992. The lethal (1) optomotor-blind gene of *Drosophila melanogaster* is a major organizer of optic lobe development: Isolation and characterization of the gene. *Proc. Natl. Acad. Sci.* **89:** 1199–1203.

Rappold, G.A., J. Trowsdale, and P. Lichter. 1992. Assignment of the human homolog of the mouse t-complex gene TCTE3 to human chromosome 6q27. *Genomics* **13:** 1337–1339.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Schulte-Merker, S., R.K. Ho, B.G. Herrmann, and C. Nüsslein-Volhard. 1992. The protein product of the zebrafish homolog of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo. *Development* **116:** 1021–1032.

Schulte-Merker, S., F.J.M. van Eeden, M.E. Halpern, C.B. Kimmel, and C. Nüsslein-Volhard. 1994. *no tail* (*ntl*) is the zebrafish homolog of the mouse T(*Brachyury*) gene. *Development* **120:** 1009–1015.

Shortle, V.P., P. Malaspina, M. Fox, J.S. Dooley, A. Volz, A. Ziegler, J. Trowsdale, K.M. Morrison, and Y.H. Edwards. 1993. A cosmid library specific for human chromosome regions 6p21.3 and 6q27. *Mamm. Genome* **4:** 493–498.

Smith, J.C., B.M.J. Price, J.B.A. Green, D. Weigel, and B.G. Herrmann. 1991. Expression of a xenopus homolog of *Brachyury* (*T*) is an immediate-early response to mesoderm induction. *Cell* **67:** 79–87.

Sowden, J., W. Putt, K. Morrison, R. Beddington, and Y. Edwards. 1995. The embryonic RNA helicase gene (*ERH*): A new member of the DEAD box family of RNA helicases. *Biochem. J.* **308:** 839–846.

Stott, D., A. Kispert, and B.G. Herrmann. 1993. Rescue of the tail defect of *Brachyury* mice. *Genes & Dev.* **7:** 197–203.

Wilkinson, D.G., S. Bhatt, and B.G. Herrmann. 1990. Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* **343:** 657–659.

Wilson, V., P. Rashbass, and R.S.P. Beddington. 1993. Chimeric analysis of T (Brachyury) gene function. *Development* **117:** 1321–1331.

Wilson, V., L. Manson, W.C. Skarnes, and R.S.P. Beddington. 1995. The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121:** 877–886.

Yasuo, H. and N. Satoh. 1993. Function of vertebrate *T* gene. *Nature* **364:** 582–583.

Ziegler, A., L.L. Field, and A.Y. Sakaguchi. 1991. Report of the committee on the genetic constitution of chromosome 6. *Cytogenet. Cell. Genet.* **58:** 295–336.

Received December 12, 1995, Accepted February 15, 1996.

The human homolog T of the mouse T(Brachyury) gene; gene structure, cDNA sequence, and assignment to chromosome 6q27.

Y H Edwards, W Putt, K M Lekoape, et al.

Genome Res. 1996 6: 226-233
Access the most recent version at doi:10.1101/gr.6.3.226