Mammalian Genome 6, 793–797 (1995).

Mammalian Genome

Identification, characterization, and localization to Chromosome 17q21-22 of the human TBX2 homolog, member of a conserved developmental gene family

D.J. Law,[1] T. Gebuhr,[1] N. Garvey,[2] S.I. Agulnik,[2] L.M. Silver[2]

[1]Human Genome Center and Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109-0674, USA
[2]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544-1014, USA

Abstract. The T-box motif is present in a family of genes whose structural features and expression patterns support their involvement in developmental gene regulation. Previously, sequence comparisons among the T-box domains of ten vertebrate and invertebrate T-box (*Tbx*) genes established a phylogenetic tree with three major branches. The *Tbx2*-related branch includes mouse *Mm-Tbx2* and *Mm-Tbx3, Drosophila* optomotor-blind (*Dm-Omb*), and *Caenorhabditis elegans Ce-Tbx2* and *Ce-Tbx7* genes. From the localization of *Mm-Tbx2* to Chromosome (Chr) 11, we focused our search for the human homolog, Hs-TBX2, within a region of synteny conservation on Chr 17q. We used *Dm-Omb* polymerase chain reaction (PCR) primers to amplify a 137-basepair (bp) product from human genomic, Chr 17 monochromosome hybrid, and fetal kidney cDNA templates. The human PCR product showed 89% DNA sequence identity and 100% peptide sequence identity to the corresponding T-box segment of *Mm-Tbx2*. The putative Hs-TBX2 locus was isolated within a YAC contig that included three anonymous markers, *D17S792, D17S794,* and *D17S948,* located at Chr 17q21-22. Hybridization- and PCR-based screening of a 15-week fetal kidney cDNA library yielded several TBX2 clones. Sequence analysis of clone λcTBX2-1 confirmed homology to *Mm-Tbx2*—90% DNA sequence identity over 283 nt, and 96% peptide sequence identity over 94 amino acids. Similar analysis of Hs-TBX2 cosmid 154F11 confirmed the cDNA coding sequence and also identified a 1.7-kb intron located at the same relative position as in *Mm-Tbx2*. Phylogenetic analyses of the T-box domain sequences found in several vertebrate and invertebrate species further suggested that the putative human TBX2 and mouse *Tbx2* are true homologs. Northern blot analysis identified two major TBX2 transcripts of 3.5 and 2.8 kb, with high levels of TBX2 expression in fetal kidney and lung; and in adult kidney, lung, ovary, prostate, spleen, and testis. Reduced expression levels were seen in heart, white blood cells, small intestine, and thymus. These results suggest that Hs-TBX2 could play important roles in both developmental and postnatal gene regulation.

Introduction

The T-box gene family is characterized by the presence of a highly conserved, unique protein motif, the T-box domain, originally identified as a defined region of amino acid sequence homology between the mouse *T* locus (Brachyury; *Mm-T*) and the *Drosophila* optomotor-blind (*Dm-Omb*) genes (Pflugfelder et al. 1992a; Bollag et al. 1994). T-box (*Tbx*) genes have been identified across the animal kingdom, from nematodes to mammals (Agulnik et al. 1995). Phylogenetic analysis of T-box genes in *C. elegans, Drosophila, Xenopus,* zebrafish, and mouse supports the existence of at least three separate T-box gene lineages that arose in an ancestor common to invertebrates and vertebrates (Bollag et al. 1994; Agulnik et al. 1995). These studies also showed that sequence homology is confined to the T-box domain (ranging in size from 166 to 203 amino acids), suggesting that locus-specific functions are encoded within the non-conserved regions. *Dm-Omb* (Pflugfleder et al. 1992b) and *Mm-T* (Kispert and Hermann 1993) T-box domains are associated with DNA-binding activity, suggesting that T-box genes may act as transcriptional regulators. The T-box gene family is most thoroughly characterized in the mouse, where, apart from the *T* locus, there are at least six additional T-box genes, *Tbx1* to *Tbx6,* each showing unique temporal and spatial expression patterns during embryonic development and in certain adult tissues (Bollag et al. 1994; Agulnik et al. 1995; unpublished data). Expression patterns, potential DNA-binding activities, and widespread occurrence among invertebrate and vertebrate species suggest that T-box genes play essential roles in developmental gene regulation.

Three mouse T-box genes have been mapped with a C57Bl/6J × SPRET/Ei backcross. *Tbx1, Tbx2,* and *Tbx3* reside on Chr 5, 11, and 16 respectively (Bollag et al. 1994). The *T* locus is a classical gene mapped in many studies to mouse Chr 17. *Mm-Tbx2* is located near homeobox 2 (*Hox2*) and myeloperoxidase (*Mpo*), in a region of conserved synteny with human Chr 17q21-23 (Copeland et al. 1993; Lossie et al. 1994). We therefore confined our search for the human TBX2 homolog to Chr 17q, using a strategy based on T-box domain sequence conservation. In the current study, we describe genomic and cDNA cloning, localization to Chr 17q21-22, phylogenetic analysis, and RNA expression patterns of the human TBX2 gene.

Materials and methods

PCR primers and conditions. PCR primer pairs were designed from the T-box domains of the *Drosophila Omb* cDNA (Pflugfelder et al., 1992a; GenBank # 81796) and mouse *Tbx2* cDNA (Bollag et al. 1994; GenBank # U15566) with the program PRIMER (Lincoln et al. 1991). Primer pairs were: OMB1-F (5′-TCCTGAACTCGATGCACAAG-3′) with OMB1-R (5′-ATTTTGATATGCTGTGACGGC-3′); and *Tbx2*-S (5′-ACAACATTTCTGACAAGCATGG-3′) with *Tbx2*-AS (5′-GTAGGCAGTGACAGCGATGA-3′). Human TBX2 primers (F: 5′-ATCCACCCAGACAGCCCA-3′ and R: 5′-GTCTCCGGGAACACG-TAGGT-3′) were designed by use of the sequences reported here. Primers also were designed for *D17S948* (EMBL # Z24102) and *D17S794* (EMBL # Z16761) because the original Généthon primers (Gyapay et al. 1994) derived from these loci were not robust. The primer pairs were: D17S948-F (5′-ACTGGATGGAGTGTACACATTTG-3′) with D17S948-R (5′-AACTCCCTAAGGACAGAGACCC-3′) and D17S794-F (5′-AAAATAGCATAAATCGAGAATGGC-3′) with D17S794-R (5′-TGGGGGACAGAAGAAGTCC-3′). PCR cycling conditions were 35 cycles of 94°C, 30 s; 55°C, 30 s; and 72°C, 30 s, with Boehringer Mannheim *Taq* DNA polymerase and buffer. The PCR additive TaqExtender™ (Stratagene, LaJolla, California) improved primer specificity for many primer pairs. Templates consisted of 200 ng of genomic DNA, 1 ng of single-

Correspondence to: D.J. Law

stranded cDNAs (Clontech), or an aliquot of the fetal kidney cDNA library containing 5×10^5 pfu. The human monochromosome 17 hybrid DNA, NA10498, is contained in the NIGMS somatic cell hybrid mapping panel #2 (Coriell Research Institute, Camden, N.J.).

Yeast artificial yeast chromosome (YAC) and Chr 17 cosmid library screening and clone characterization. The Centre d'Etude du Polymorphisme Humain (CEPH; Albertsen et al. 1990) and Généthon (Bellanné-Chantelot et al. 1992) human YAC libraries, maintained as gridded arrays (Bentley et al., 1992), were screened by PCR analysis of DNA matrix pools (Green and Olson 1990; Kwiatowski et al., 1990). YAC clone sizes were estimated by contour-clamped homogeneous electric field (CHEF) electrophoresis (Vollrath and Davis 1987), and clones were assayed for chimerism using single-color fluorescence in situ hybridization (FISH) as previously described (Flejter et al. 1993). Additional YAC clone information, including size, chimerism, marker content, cross-hybridization and fingerprint data, was obtained from the CEPH/Genethon database (Cohen et al. 1993). The human Chr 17 cosmid library (Deaven 1990) was screened by hybridization to high-density filter replicates of gridded colony arrays, with random-primed (Feinberg and Vogelstein, 1984) *Mm-Tbx2* cDNA fragments. The *Mm-Tbx2* S/AS primers were used to amplify a fragment from a mouse *Tbx2* cDNA clone (Bollag et al., 1994) to avoid cross-hybridization with vector sequences.

Human fetal kidney cDNA library. A 15-week fetal kidney cDNA library was generously provided by Andrew Feinberg of Johns Hopkins University. The library contains both oligo(dT)- and random-primed reverse transcripts, and was prepared by Stratagene (LaJolla, CA) with RNA provided by Feinberg's laboratory. The vector is λZAP II (Short et al. 1988), the complexity is 7×10^6, and average insert size is >0.4 kb.

Northern blot analysis. Human fetal and adult multiple tissue Northern blots were obtained from Clontech (Palo Alto, CA) and hybridized according to the manufacturer's recommendations. Semi-quantitative estimates of TBX2 were based on band intensity after normalization with an actin probe to correct for unequal sample loading.

DNA sequencing. PCR products were isolated on agarose gels (NuSieve, FMC Bioproducts) and purified with β-agarase (New England Biolabs, Beverly, Massachusetts). Cosmid and cDNA clones were purified by use of the Qiagen (Chatsworth, CA) plasmid DNA chromatography system. Double-stranded sequencing was performed with the automated cycle sequencing protocol previously described (Kaiser et al. 1993) or the Sequenase 2.0 system (U.S. Biochemical/Amersham, Cleveland, Ohio). For cosmid templates, sequencing primers were used at a 10-fold higher concentration than is specified for plasmid templates.

Sequence comparisons and phylogenetic analyses. The human T-box DNA sequences and deduced polypeptide sequences were compared initially by use of the BLAST program (Altschul et al. 1990; NCBI BLAST network) and the FASTA (Pearson and Lipman 1988) and PILEUP (Feng and Doolittle 1987) programs of the Wisconsin Genetic Computer Group (GCG) package, as previously described (Bollag et al. 1994; Agulnik et al. 1995). A more robust phylogenetic analysis was performed by data bootstrapping with maximum likelihood and maximum parsimony algorithms (Felsenstein 1991; Maddison and Maddison 1992). Sequences were obtained from the following GenBank files: *Mm-Tbx2*–U15566; *Mm-Tbx3*–U15567; *Dm-Omb*–M81796; *Ce-Tbx2*–U11279; *Mm-T*–X51683.

Results

Drosophila Omb T-box domain primers amplify human Chr 17 DNA sequences. In an effort to isolate a human TBX2 homolog, we designed PCR primers within the highly conserved T-box domains of two members of the T-box 2 gene family, the prototypical mouse *Tbx2* and the *Drosophila Omb* genes (Bollag et al. 1994; Agulnik et al. 1995; Pflugfelder et al. 1992a). The *Mm-Tbx2* primers encompass 161 nucleotides (nt), and the *Dm-Omb* primers encompass 137 nt of their respective T-box coding domains. Each primer pair amplified sequences in human genomic, Chr 17 somatic cell hybrid and fetal kidney cDNA templates (data not shown), suggesting that a *Tbx2*-related gene is located on Chr 17 and is expressed in fetal kidney. To better define the chromosomal location and structure of the putative human TBX2 gene, we used the *Omb* and *Tbx2* sequences to screen human genomic and fetal kidney cDNA libraries.

Isolation of genomic and cDNA clones encompassing the TBX2 locus at 17q22. The *Tbx2* and *Omb* PCR primers were used to screen matrix pools of the CEPH and Généthon human YAC libraries and to perform sequence-tagged site (STS)-content analysis of overlapping clones. The YAC clone contig, shown in Fig. 1, spans at least 0.9 Mb and links TBX2 to three Genethon microsatellite markers. YAC clone 961F1 (890 kb) establishes direct physical linkage between TBX2 and markers *D17S794* and *D17S948*, located at 17q21-22 (Gyapay et al. 1994; Rigault and Pollier 1994). Four cosmid clones were isolated from a human Chr 17-specific cosmid library (Deaven 1990) by hybridization with a 161-bp *Mm-Tbx2* cDNA fragment. STS-content analysis with Hs-TBX2 primers showed that the cosmids overlap with each other and with the TBX2 YAC clones. The mouse *Tbx2* probe also was used to screen a 15-week fetal kidney cDNA library, yielding clone λcTBX2-1. The 1.5-kb *Eco*RI insert of λcTBX2-1 was hybridized to *Eco*RI-digested human genomic as well as TBX2 YAC and cosmid DNAs. Fragments of 6.7, 11.5, and 12.6 kb were identified in all DNA sources, confirming locus specificity and suggesting that at least two introns interrupt the 1.5-kb cDNA fragment. To confirm the identity of Hs-TBX2, we compared its DNA and deduced peptide sequences to those of other T-box genes.

DNA sequencing reveals a high degree of homology between Hs-TBX2 and Mm-Tbx2. Cosmid clone 154F11 and cDNA clone λcTBX2-1 were sequenced using TBX2 and vector primers. A 283-bp, T-box encoding segment and its deduced polypeptide sequence are shown in Fig. 2. The human TBX2 segment shares 90% DNA sequence identity and 96% peptide sequence identity with the corresponding regions of *Mm-Tbx2*. Genomic sequence and

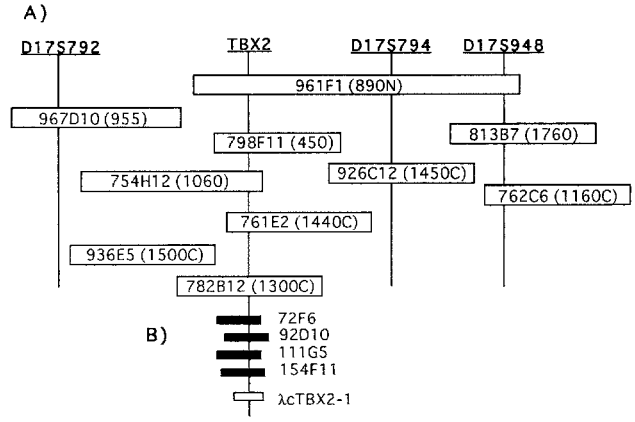

Fig. 1. Genomic and cDNA cloning of the human TBX2 gene, Hs-TBX2. **Panel A** shows a YAC clone contig encompassing TBX2 and anonymous markers *D17S792*, *D17S794*, and *D17S948* (Gyapay et al. 1994). Clones were isolated from the CEPH and Généthon libraries. Numbers in parentheses indicate clone size estimates; N = nonchimeric; C = chimeric; no suffix = not determined. Map is not to scale since extent of Chr 17 content is not known for most clones. **Panel B** shows TBX2 cosmids (solid), isolated from a Chr 17-specific library (Deaven 1990), and a cDNA clone (open) isolated from a 15-week fetal kidney cDNA library. Sequence confirmation of cosmid 154F11 and λcTBX2-1 is described below.

PCR analysis indicate that a 1.7-kb intron is located between codons 60 and 61 of the human sequence. This corresponds to a conserved intron site between residues 211 and 212 in the mouse gene. Consensus splice acceptor and donor sequences are present in the human intron.

Phylogenetic comparisons of TBX2 to other members of the T-box family. A phylogenetic analysis was performed on the complete T-box domain in all previously characterized genes of the TBX2 subfamily, four representative T locus genes, and other T-box genes characterized in the nematode *C. elegans*. This analysis was accomplished by data bootstrapping with maximum likelihood and maximum parsimony algorithms (Felsenstein 1991; Maddison and Maddison 1992). The consensus phylogenetic tree obtained through this analysis is shown in Fig. 3.

The most important results obtained for the purposes of this study are the phylogenetic association of the mouse *Tbx2* gene and the putative human TBX2 gene in 100% of all bootstrap runs. This

A) Comparison of Human and Mouse *TBX2* T-box Sequences.

```
            1  D  C  R  D  K  F  H  N  S  R  W  M  V  T  D  K  T  D  P  E   20
Hsa-TBX2:   1  GATTGCCGCGATAAGTTCCACAACTCACGCTGGATGGTGACGGACAAGACCGACCCTGAG  60
               ||  ||||||  |||||||  ||||||||  ||||||||||||  | |  |||  |||||||  |||
Mmu-Tbx2: 512  GACTGCCGCTATAAGTTTCACAACTCCCGCTGGATGGTGGCAGGCAAAGCCGACCCCGAG 571
          152  D  C  R  Y  K  F  H  N  S  R  W  M  V  A  G  K  A  D  P  E  171

           21  M  P  K  R  M  Y  I  H  P  D  S  P  A  T  G  E  Q  W  M  A   40
Hsa-TBX2:  61  ATGCCCAAACGCATGTACATCCACCCAGACAGCCCAGCCACGGGGGAGCAGTGGATGGCT 120
               |||||  ||||||||||||||||||||||  |||||  ||  |||||  |||||  ||||||||||
Mmu-Tbx2: 572  ATGCCTAAACGCATGTACATCCACCCGGACAGTCCGGCCACAGGGGAACAGTGGATGGCC 631
          172  M  P  K  R  M  Y  I  H  P  D  S  P  A  T  G  E  Q  W  M  A  191

           41  K  P  V  A  F  H  K  L  K  L  T  N  N  I  S  D  K  H  G  F   60
Hsa-TBX2: 121  AAGCCTGTGGCCTTCCACAAGCTGAAGCTGACCAACAACATCTCTGACAAGCACGGCTTC 180
               ||||||||||||||||||||  ||||||||||||||||||  ||||||||||  |||||
Mmu-Tbx2: 632  AAGCCTGTGGCCTTCCACAAACTGAAGCTGACCAACAACATTTCTGACAAGCATGGCTTT 691
          192  K  P  V  A  F  H  K  L  K  L  T  N  N  I  S  D  K  H  G  F  211

           61  T  I  L  N  S  M  H  K  Y  Q  P  R  F  H  I  V  R  A  N  D   80
Hsa-TBX2: 181  ACCATCCTAAACTCCATGCACAAGTACCAGCCGCGCTTCCACATAGTGCGAGCCAACGAC 240
               |||||||  ||||||||||||||||||||||||||  ||||||||  ||||||||||  |||
Mmu-Tbx2: 692  ACCATCCTGAACTCCATGCACAAGTACCAGCCGCGATTCCACATCGTGCGAGCCAATGAC 751
          212  T  I  L  N  S  M  H  K  Y  Q  P  R  F  H  I  V  R  A  N  D  231

           81  I  L  K  L  P  Y  S  T  F  R  T  Y  V  F    94
Hsa-TBX2: 241  ATCCTGAAGCTGCCTTACAGCACCTTCCGCACCTACGTCTTCC 283
               |||  |||||||  ||  ||||||||||||||||||  |||||||
Mmu-Tbx2: 752  ATCTTGAAGCTCCCATACAGCACCTTCCGCACCTATGTCTTCC 794
          232  I  L  K  L  P  Y  S  T  F  R  T  Y  V  F   245

                DNA Sequence Identities =          254/283 (90%)
                Peptide Sequence Identities =      90/94 (96%)
```

INTRON SEQUENCES FLANKING CODONS 60/61:

(60) GTGAGTGTTG GGGCAGGGTG// **~1.7KB** //GACCCCCACC CTCCCCGCAG (61)

B) Sequence similarity of *Hs-TBX2* to other *Tbx2* branch genes

Hs-TBX2 compared to:	Identity / Interval length: DNA (%/nt)	Peptide (%/aa)
Mm-Tbx2	90 / 283	96 / 94
Dm-Omb	76 / 281	89 / 94
Mm-Tbx3	81 / 208	94 / 69
Ce-Tbx2	66 / 103	73 / 92

Fig. 2. Comparison of human and mouse TBX2 DNA and polypeptide sequences. **(A)** Two-pass DNA sequence of cDNA clone cTBX2-1 and its deduced polypeptide sequence are compared with the corresponding T-box encoding segment of mouse *Tbx2*. Coding sequence was confirmed, and an intron, between codons 60 and 61, was identified by sequence analysis of cosmid 154F11. Consensus splice donor and acceptor sites are underlined. DNA sequence identity is indicated by vertical lines; unique human peptide residues are underlined. Mouse *Tbx2* sequences and residue numbers according to GenBank U15566. **(B)** Sequence comparisons of Hs-TBX2 and other Tbx2 subfamily T-box segments. Data obtained by BLAST analysis (Altschul et al. 1990).

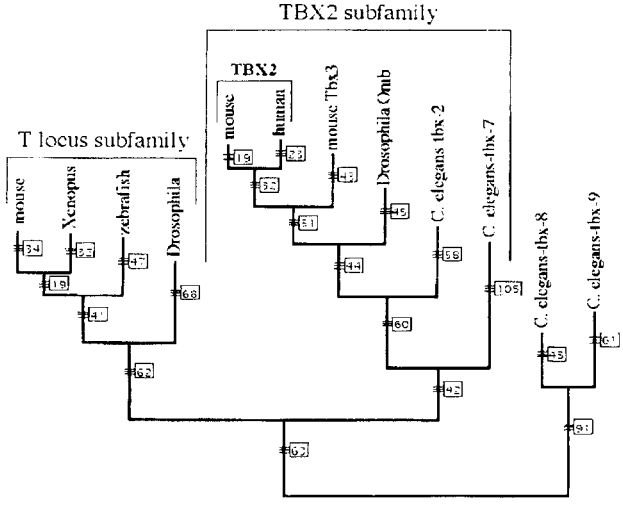

Fig. 3. Phylogenetic tree of the T-box genes. Phylogenetic analysis was performed on the T-box domain of the genes shown with two different approaches—maximum likelihood analysis and maximum parsimony analysis. Both analyses produced the same consensus tree with a confidence level of greater than 95% for all branches except those leading to *C. elegans*-tbx-7 and the tbx-8/tbx-9 gene pair. The relative order and location of these two ancient branchpoints cannot be resolved with the available data; one possible configuration is shown here. The numbers shown along each branch represent the most likely number of nucleotide changes as determined by the MacClade program (Maddison and Maddison 1992).

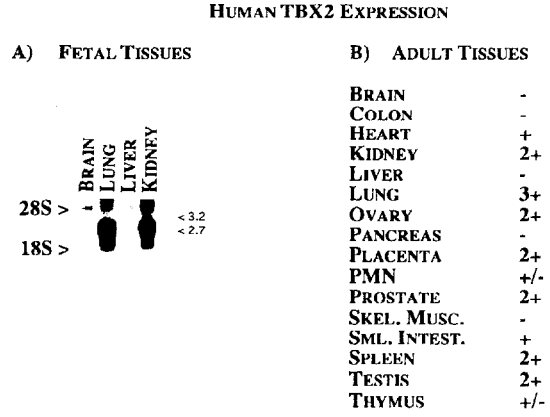

Fig. 4. Hs-TBX2 expression in fetal and adult tissues. Northern blot analysis of Hs-TBX2 expression in a variety of fetal (**A**) and adult (**B**) tissues. Hybridization with the labeled human TBX2-F/R PCR product showed major transcripts of 3.5 and 2.8 kb (arrows on right). Abundant expression is seen in fetal lung and kidney. Correction for unequal loading, as determined by hybridization with an actin probe (data not shown), suggests that expression in the two tissues is comparable since the fetal lung lane contains more RNA. Faint expression in fetal liver also was apparent on the original autoradiogram. Semi-quantitative analysis of TBX2 expression (actin-normalized) in several adult tissues is summarized in panel B.

phylogenetic result, together with the observed conservation of synteny in the mouse and human genomes, provides strong evidence in support of our contention that mouse *Tbx2* and human TBX2 are true homologs.

TBX2 expression in fetal and adult tissues. Hs-TBX2 RNA expression was assayed by hybridizing multiple tissue Northern blots with the labeled Hs-TBX2 PCR product. As shown in Fig. 4A, major transcripts of 3.5 and 2.8 kb are abundantly expressed in fetal kidney and lung, with low levels of expression in fetal liver. These transcripts are similar in size to the 3.2- and 2.7-kb *Tbx2* transcripts in the mouse (Bollag et al. 1994). Semi-quantitative analysis of TBX2 expression in several adult tissues is summarized in panel B. Moderate expression levels are observed in kidney, lung, ovary, placenta, prostate, spleen, and testis; with lower levels detected in heart, polymorphonucleocytes, small intestine, and thymus. The human TBX2 expression pattern shows general similarity to the mouse *Tbx2* pattern (Bollag et al. 1994), where direct comparison is possible. Two exceptions are spleen and testis, tissues that show no expression in the mouse.

Discussion

A comparative genetic approach was used to identify the human TBX2 gene, the first human member of the *Tbx2* subfamily (Agulnik et al. 1995). Close physical linkage to anonymous markers *D17S792, D17S494,* and *D17S948* localized Hs-TBX2 to 17q21-22, as predicted by conserved synteny with distal mouse Chr 11, site of the mouse *Tbx2* locus. Structural and functional analyses confirmed the high degree of relatedness between mouse and human TBX2 loci. T-box encoding segments shared 90% DNA and 96% amino acid sequence identity within the regions compared. The relative position of an intron was also conserved within the two mammalian genes. Transcript sizes and expression patterns in fetal and adult tissues were also very similar for the two genes. The expression levels of the two major human transcripts were always comparable within a given tissue type. Whether the two products are derived from a single TBX2 locus or multiple TBX2-related loci, their levels appear to be coordinately regulated.

The inclusion of human TBX2 in a phylogenetic analysis of T-box sequences provides further evidence for the existence of at least two separate subfamilies of T-box genes named according to the first gene uncovered in each subfamily—the T locus subfamily and the TBX2 subfamily. The presence of both mammalian and invertebrate sequences within both subfamilies strongly suggests that the primordial ancestor to all metazoan organisms had at least two T-box genes—one representative of each subfamily—in its genome.

Members of the mouse T-box gene family, including *Tbx2,* are expressed with unique spatial and temporal patterns in a variety of embryonic and adult tissues (Bollag et al. 1994), supporting their specific involvement in developmental signaling and gene regulation. By extrapolation, Hs-TBX2 becomes a candidate for regulating developmental gene expression and could be involved in certain developmental and adult-onset diseases. The TBX2 locus is near the site of an interstitial deletion, del (17) (q21.3-q23), that is associated with multiple congenital abnormalities, including brachycephaly, club feet, and developmental and growth delays (Park et al., 1992). TBX2 is also within a region of minimal interstitial chromosomal loss observed in a fraction of sporadic and familial ovarian tumors (Godwin et al., 1994), and ovarian tissue is one of the sites of Hs-TBX2 expression. Gain-of-function mutations in members of the PAX (Maulbecker and Gruss 1993a) and HOX (Maulbecker and Gruss 1993b) gene families can confer oncogenic potential, establishing a role for developmental gene mutations in oncogenesis. TBX2 also is localized within the candidate region (17q21-22) for Disinhibition-Dementia-Parkinsonism-Amyotrophy Complex (DDPAC; Wilhelmsen et al., 1994), an adult-onset, progressive neurological disease with a primary characteristic of frontal lobe dementia. Further analysis of TBX2 structure and function will help to elucidate its role in mediating gene expression in normal and pathological states.

Acknowledgments. We thank Dr. Miriam H. Meisler for useful discussions and for helping to establish the collaboration between Princeton University and the University of Michigan. We gratefully acknowledge Dr. Andrew P. Feinberg of Johns Hopkins University for providing an aliquot of the amplified fetal kidney cDNA library. Dr. Spencer W. Thomas and Jeffrey D. Aspinall provided excellent informatics support. Madhu A. Prasad, Rita

E. Fox, and Diane Miller provided technical support; and Louise Hallett and Susan H. Semelroth provided administrative assistance. This work was supported by the National Institutes of Health (NIH) Genome Center grant HG00209 (Miriam H. Meisler, Director) and by NIH grant HD20275 to L.M. Silver.

Note Added in Proof

During the review of this manuscript, an independent description of the human TBX2 gene was reported (Campbell et al., 1995).

References

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

Albertsen, M.H., Abderrahim, H., Cann, H.M., Dausset, J., LePaslier, D. Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid genome equivalents. Proc. Natl. Acad. Sci. USA 87, 4256–4260.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Bellanné-Chantelot, C., Lacroix, B., Ougen, P., Billault, A., Beaufils, S., Betrand, S., Georges, I., Gilbert, F., Gros, I., Lucotte, G., Susini, L., Codani, J.-J., Gesnouin, P., Pook, S., Vaysseix, G., Lu-Kuo, J., Ried, T., Ward, D., Chumakov, I., LePaslier, D., Barillot, E., Cohen, D. (1992). Mapping the whole human genome by fingerprinting yeast artificial chromosomes. Cell 70, 1059–1068.

Bentley, D.R., Todd, C., Collins, J., Holland, J., Dunham, I., Hassock, S., Bankier, A., Giannelli, F. (1992). The development and application of automated gridding for efficient screening of yeast and bacterial gridded libraries. Genomics 12, 534–541.

Bollag, R.J., Siegfried, Z., Cebra-Thomas, J.A., Garvey, N., Davison E.M., Silver, L.M. (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., Goodrich, K., Casey, G., Beatty, B. (1995). Cloning and mapping of a human gene (TBX2) sharing a highly conserved protein motif with the Drosophila *omb* gene. Genomics 28, 255–260.

Cohen, D., Chumakov, I., Weissenbach, J. (1993). A first-generation physical map of the human genome. Nature 366, 698–701.

Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J., Miller, J.C., Dietrich, W.F., Weaver, A., Lincoln, S.E., Steen, R.G., Stein, L.D., Nadueu, J.H., Lander, E.S. (1993). A genetic linkage map of the mouse: current applications and future prospects. Science 262, 57–66.

Deaven, L.L. (1990). Chromosome-specific gene libraries. In *The Encyclopedia of Human Biology*, R. Dulbecco, ed., (San Diego: Academic Press), Vol. II, pp. 455–456.

Feinberg, A.P., Vogelstein, B. (1984). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6–13.

Felsenstein, J. (1991). *PHYLIP (Phylogeny Inference Package)*. Univ. Washington, Seattle.

Feng, D.F., Doolittle, R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25, 351–360.

Flejter, W.L., Barcroft, C.L., Guo, S.-W., Lynch, E.D., Boehnke, M., Chandrasekharappa, S., Hayes, S., Collins, F.S., Weber, B.L., Glover, T.W. (1993). Multicolor FISH mapping with Alu-PCR-amplified YAC clone DNA determines the order of markers in the BRCA1 region on chromosome 17q12-q21. Genomics 17, 624–631.

Godwin, A.K., Vanderveer, L., Schultz, D.C., Lynch, H.T., Altomare, D.A., Buetow, K.H., Daly, M., Getts, L.A., Masny, A., Rosenblum, N., Hogan, M., Ozols, R.F., Hamilton, T.C. (1994). A common region of deletion on chromosome 17q in both sporadic and familial epithelial ovarian tumors distal to BRCA1. Am. J. Hum. Genet. 55, 666–677.

Green, E.D., Olson, M. (1990). Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA 87, 1213–1217.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Lathrop, M., Weissenbach, J. (1994). The 1993-4 Généthon human genetic linkage map. Nature Genet. 7, 246–339.

Kaiser, R., Hunkapiller, T., Heiner, C., Hood, L. (1993). Specific primer-directed DNA sequence analysis using automated fluorescence detection and labeled primers. Methods Enzymol. 218, 122–153.

Kispert, A., Hermann, B.G. (1993). The *Brachyury* gene encodes a novel DNA binding protein. EMBO J. 12, 3211–3220.

Kwiatkowski, T.J., Jr., Zoghbi, H.Y., Ledbetter, S.A., Ellison, K.A., Chinault, A.C. (1990). Rapid identification of yeast of artificial chromosome clones by matrix pooling and crude lysate PCR. Nucleic Acids Res. 18, 7191–7192.

Lincoln, S.E., Daly, M.J., Lander, E.S. (1991). PRIMER: a computer program for automatically selecting PCR primers. Version 0.5 MIT Center for Genome Research and Whitehead Institute for Biomedical Research, Cambridge, Mass.

Lossie, A.C., MacPhee, M., Buchberg, A.M., Camper, S.A. (1994). Mouse Chromosome 11. Mamm. Genome 5 (Suppl.), 164–180.

Maddison, W.P., Maddison, D.R. (1992). MacClade: analysis of phylogeny and character evolution. (Sunderland, Mass.: Sinauer Assoc.).

Maulbecker, C.C., Gruss, P. (1993a). The oncogenic potential of Pax genes. EMBO J. 12, 2361–2367.

Maulbecker, C.C., Gruss, P. (1993b). The oncogenic potential of deregulated homeobox genes. Cell Growth Differ. 4, 431–441.

Park, J.P., Moeschler, J.B., Berg, S.Z., Bauer, R.M., Wurster-Hill, D.H. (1992). A unique *de novo* interstitial deletion del(17) (q21.3q23) in a phenotypically abnormal infant. Clin. Genet. 41, 54–56.

Pearson, W.R., Lipman, D.J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444–2448.

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

Pflugfelder, G.O., Roth, H., Poeck, B. (1992b). A homology domain shared between *Drosophila optomotor-blind* and mouse *Brachyury* is involved in DNA binding. Biochem. Biophys. Res. Commun. 186, 918–925.

Rigault, P., Poullier, E. (1994). QUICKMAP: Compact database and navigation tool for integration of CEPH-Genethon mapping data. Abstract A206 in *Genome Sequencing and Mapping*. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Short, J.M., Fernandez, J.M., Sorge, J.A., Huse, W.D. (1988). Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. Nucleic Acids Res. 16, 7583–7600.

Vollrath, D., Davis, R.W. (1987). Resolution of DNA molecules greater than 5 megabases by contour-clamped homogeneous electric fields. Nucleic Acids Res. 15, 7865–7876.

Wilhelmsen, K.C., Lynch, T., Pavlou, E., Higgins, M. and Nygaard, T.G. (1994). Localization of disinhibition-dementia-Parkinsonism-amyotrophy complex to 17q21-22. Am. J. Hum. Genet. 55, 1159–1165.