Transcriptional map of the Treacher Collins candidate gene region.

S K Loftus, J Dixon, K Koprivnikar, et al.

Genome Res. 1996 6: 26-34
Access the most recent version at doi:10.1101/gr.6.1.26

RESEARCH

Transcriptional Map of the Treacher Collins Candidate Gene Region

Stacie K. Loftus, Jill Dixon,[1] Kathryn Koprivnikar, Michael J. Dixon,[1] and John J. Wasmuth[2]

Department of Biological Chemistry, College of Medicine, University of California, Irvine, California 92717; [1]School of Biological Sciences and Departments of Dental Medicine and Surgery, University of Manchester, Manchester, M13 9PT, UK

Treacher Collins syndrome (TCOF1) is a dominant disorder of craniofacial development that has been linked previously to a region of chromosome 5q31.3–32. Identification of recombination events in affected individuals has reduced the candidate gene region to a 0.5-Mb area between the loci *RPS14* (proximal) and *ANX6* (distal). A transcriptional map of this candidate gene region, generated by analysis of exon amplification clones, has identified the genomic location of four genes, heparan sulfate-*N*-sulfotransferase-*N*-deacetylase, glutathione peroxidase, as well as two novel, previously uncharacterized genes. Each of these genes, based on their location, must be considered candidates for TCOF1 locus.

Treacher Collins syndrome (TCOF1) is an autosomal dominant disorder of craniofacial development that occurs with an incidence of ~1/50,000 live births. More than 50% of the cases are thought to be de novo mutations. Clinical features of TCOF1 are bilateral in nature and include (1) abnormalities of the external ears, atresia of external ear canals, and malformation of middle ear ossicles that results in bilateral conductive hearing loss, (2) lateral downward sloping of palpebral fissures, frequently with colobomota of the lower eyelids, (3) hypoplasia of mandible and zygomatic complex, and (4) cleft palate (Rovin et al. 1964; Frazen et al. 1967). Accurate diagnosis can be difficult, as there is a high degree of variable expressivity of the phenotype. Because the tissues affected in TCOF1 arise from the first and second branchial archs during early embryonic development, it has been suggested that these clinical features may be the result of a defect in neural crest cell migration or improper cellular differentiation during development (Poswillo 1975).

TCOF1 has been linked to several DNA markers in the region of chromosome 5q32–33.1 (Dixon et al. 1991; Jabs et al. 1991; Edery et al. 1994). Previously, the candidate gene region had been localized to an ~840-kb segment between the proximal flanking marker D5S519 and the distal flanking marker SPARC (Dixon et al. 1993).

[2]Corresponding author.
E-MAIL JJWASMUTH@uci.edu; FAX (714) 824-3403

We have analyzed families with new short tandem repeat polymorphisms (STRPs) isolated from the TCOF1 candidate region. Newly identified recombination events have refined the TCOF1 candidate gene region to be between *RPS14* (proximal) and a STRP in *ANX6* (distal). This region is estimated to span ~450 kb (Loftus et al. 1993; Dixon et al. 1994). These data exclude *ANX6* from the candidate gene region.

In this report we describe a map of expressed sequences within the ~450-kb TCOF1 candidate region, which have been identified using the technique of exon amplification (Buckler et al. 1991; Church et al. 1994). Yeast artificial chromosome (YAC) contigs that span the region have been reported previously (Dixon et al. 1994; Li et al. 1994) and were used to identify corresponding cosmids from the Los Alamos National Laboratory (LANL) flow-sorted chromosome 5 cosmid library. Cosmids from three contigs spanning 390 kb of the candidate gene region were screened using exon amplification to identify potential expressed sequences. These clones were then used as probes to screen several cDNA libraries. This approach has identified four cDNAs located within the candidate gene region, glutathione peroxidase 3 (*GPX3*) (Chu et al. 1992), heparan sulfate-*N*-deacetylase/*N*-sulfotransferase (*HSST*) (Dixon et al. 1995), and two previously uncharacterized genes (9D2 and 199G4) that were isolated from an embryonic craniofacial and fetal brain cDNA libraries, respectively.

6:26–34 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

RESULTS

Linkage Analysis: Reduction of the Candidate Gene Region

Previously, we have shown that the TCOF1 locus is located between D5S519 (proximal) and SPARC (distal), the recombination events defining the critical region occurring in a minimally affected and a severely affected individual, respectively (Dixon et al. 1993). The families in which these recombination events occur have been reexamined, and where possible, a second blood sample was obtained. On reexamination of the individual defining the proximal recombination boundary, two clinicians felt that the diagnosis was incorrect and the individual was unaffected. The proximal recombination boundary of the TCOF1 critical region has been redefined as lying at *RPS14* on the basis of a recombination event occurring in an unequivocally affected member of a small nuclear TCOF1 family, reported previously as family 8 (Dixon et al. 1991). Screening of the cosmid 2F, which contains the *ANX6* locus with poly[d(C-A)(G-T)] revealed the presence of a single $(CA)_n$ STR. Sequencing of this repeat identified a $(GT)_{14}$. Primers designed to sequence flanking the repeat were used to amplify genomic DNA from 50 Centre d'Etudes du Polymorphisme Humain (CEPH) families. The ANX6 STRP was found to exhibit six alleles that are inherited in a codominant Mendelian fashion with a heterozygosity value of 0.72. The allele sizes and frequencies are 135 bp (0.07), 133 bp (0.03), 131 bp (0.04), 129 bp 9 (0.35), 127 bp (0.12), and 119 bp (0.39). Genotyping of TCOF1 families, in which recombination events in the vicinity of the TCOF1 candidate region have been identified previously (Dixon et al. 1991, 1993, 1994; Loftus et al. 1993), revealed two recombination events, both of which occur in affected individuals. In both cases, the affected diagnosis has been confirmed by clinical evaluation and radiography. Both of these individuals were nonrecombinant at RPS14. Conversely, the affected individual from family 8 (Dixon et al. 1991), who was recombinant at RPS14 was nonrecombinant for the ANX6 STRP. These observations define the TCOF1 critical region as lying between RPS14 (proximal) and ANX6 (distal).

Identification of ANX6 and GPX3 exons

As diagramed in Figure 1, 17 cosmids from the region between *RPS14* and *ANX6* were isolated and placed into three separate contigs (I, II, and III). These cosmids, which span ~390 kb (85%) of the newly defined 450-kb candidate region for TCOF1, were used for exon amplification to identify additional TCOF1 candidate genes.

Orientation of clones from centromere to telomere in each contig was determined by one or more of the following methods, including radiation hybrid (RH) mapping, YAC end-clone mapping and two-color fluorescent in situ hybridization (FISH). For example, in the small central contig (II), cosmids 3G3-4A and 3G3-6A and 3.1 and 3.2 were identified using previously described sequence-tagged site (STS) primer sets for YAC end clones 3G3R and 1ED1R, respectively

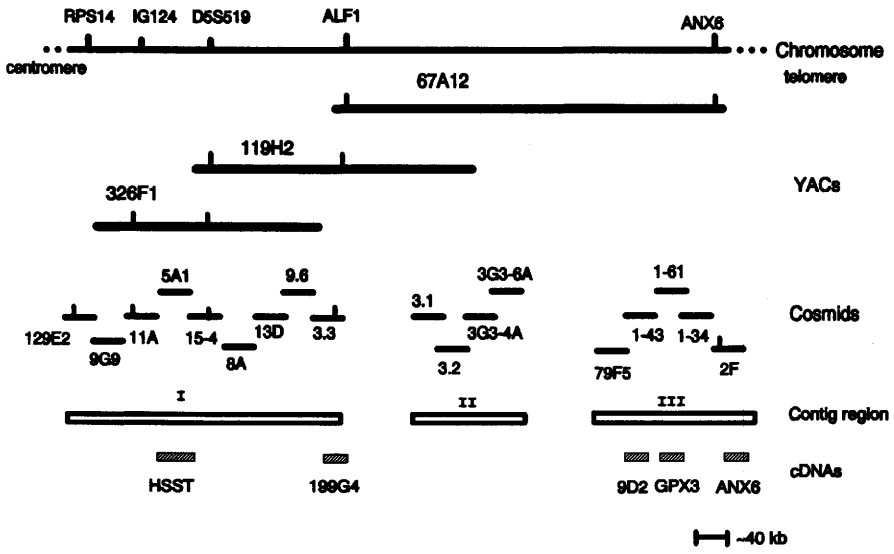

Figure 1 Physical and transcription map of the TCOF1 candidate region. Loci defined by PCR primers are shown above. YAC clones are indicated by heavy bars and are labeled corresponding to their coordinates from the CEPH I YAC library. Cosmids indicated were obtained from LANL flow-sorted chromosome 5 cosmid library. cDNA clones described in text are indicated by hatched bars. The vertical bars indicate the presence of loci *RPS14*, IG124, D5S519, ALF1, and *ANX6*, respectively. Overlap for YAC and cosmids clones was determined by PCR and Southern hybridization as detailed in Methods.

(Dixon et al. 1994). Orientation of the cosmid contig II was established as cosmids 3.1, 3.2, and 3G3-4A are found within YAC 119H2 by Southern hybridization, whereas 3G3-6A is not present in YAC 119H2. These results, placing cosmid 3.1 proximal to cosmid 3G3-6A, were also confirmed by FISH (data not shown).

From the 17 cosmid clones shown in Figure 1, a total of 120 exon clones were isolated and sequenced, 47 of which represented unique, single-copy sequences. Individual exon clones were confirmed to be contained within the relevant cosmids by Southern blot hybridization. After BLASTN and BLASTX computer searches, five clones were found to correspond to previously characterized genes, three of which (all from cosmid 2F) were identified as exons 14, 15, and 16 of *ANX6* (Smith et al. 1994). Cosmid 2F is the distal most cosmid of contig III and was known previously to contain the 3′-untranslated region of *ANX6* (Crompton et al. 1988). Identification of exons from *ANX6* provided a useful internal control for the exon amplification protocol. Two other exon clones (8F8; 4G3) showed 100% identity to segments of the previously characterized gene *GPX3*. Recently, *GPX3* was mapped to chromosome 5 (Chu 1994). *GPX3* exon clone 4G3, obtained from cosmid 1–61, contained both exons 2 and 3 spliced correctly. The second clone (8F8), which was isolated from the adjacent cosmid 1–43, corresponded to exon 4 of *GPX3*. Thus, we are able to place the *GPX3* gene ~30 kb proximal to the 3′ end of *ANX6*.

Identification of the *HSST* Gene

Results of BLASTN and BLASTX computer searches with the sequences of exon clones 3B1, 3A8, 1C6, and 1C9 revealed 70%–90% sequence identity to the rat and murine gene encoding *HSST* (Hashimoto et al. 1992; Eriksson et al. 1994). These exon clones were obtained from two overlapping cosmids in contig I, 5A1, and 15-4. These exon amplification clones were used as probes to isolate cDNA clones for the human *HSST* (Dixon et al. 1995). The orientation of *HSST* is centromere—5′ end/3′ end—telomere and the gene is located ~80 kb distal to *RPS14*.

Identification of Novel Transcripts 9D2 and 199G4

After sequence homology searches, 37/47 exon amplification clones did not show significant homology to sequences from the GenBank data base. Of these 37 clones without homology 15 exon amplification clones were selected, one clone from each cosmid, and primers designed. These primers were used to screen DNA pools from cDNA libraires. Two exon primer sets, 199G4 and 9D2, amplified predicted size fragments. Libraries were then plated and screened with corresponding radiolabeled exon fragment. Sequence information for the remaining exon amplification clones has been deposited in GenBank (accession nos. U4290, U4291, and U42570–U42579).

PCR primers for exon clone 199G4 amplified a fragment of 158 bp from a fetal brain cDNA library. Screening of this library yielded only one (2210 bp) clone that contained sequence identical to that of the 199G4 exon (Fig. 2). The cDNA clone identified with 199G4 appears to be a novel

```
   1  CCCAAAAAACATTTACCCAGAATACAAAAAAAGGACAGTATAGTGGTCAGATTCCCAGTG
  61  ACAATAAACAAATCCTCCTCTCTTCAATGTTTACTGTAAAGCCAGTTAACTGCCATTTAT
 121  TTGAAAGAAGAAAAGAAAAACAAGCCCCCAGATTGGTAAACACATTGGCAAGTTGCTAGA
 181  CTCCAAGCTGGCTACCCTTGTCAAATCTGTTCGAGTTTTTTCAAAATACTTGGGCTTTTT
 241  AACGCACCCAGCCTGCTGGGCACAGTTAAGTAAAAGGCATTTACTATAATCTAAGAATTC
 301  CCTGCTATTATCATTTTTAAATGTTTTCACATTTTTAAAACTGCCTTACCCTTTTGGATG
 361  TATTTGGATTCCATAAAGGTGAGTACAATCAACGAGAAACTGAAGTGGGAGATTCTACGA
 421  CTGCCCAGAGGGACACGCATGCATGGGGCTGGTTGTGGGTGCTGTGTGCAGTCCAGAGGA
                          M  G  L  V  V  G  A  V  C  S  P  E  E
 481  ATCATGGGAGAAAGGAAGTCAAGTCCTTAAACACCCTACAGAAACAAACACTGCAATCCA
       S  W  E  K  G  S  Q  V  L  K  H  P  T  E  T  N  T  A  I  Q
 541  GTATGGCTTATTCGGATGCATTTACCATGAAGCTACCAGAAGAAATTCAACCTACGAGGC
       Y  G  L  F  G  C  I  Y  H  E  A  T  R  R  N  S  T  Y  E  A
 600  TACTCTTAAAGTAACAGGATCGGCTATGTTGACGATGTACCCCTCCCAGGAATTCCTTC
       T  L  K  C  N  R  I  G  Y  V  D  D  V  P  L  P  G  I  P  S
 661  TAACTCTGATGGAACACGGCCAGTTGGCATGCTGGGGAAAGCCACATCTACCAGTGACAT
       N  S  D  G  T  R  P  V  G  M  L  G  K  A  T  S  T  S  D  M
 721  GCTGCTCAAACTGGCCCGGACCACACCCTACTACAAAAGGAATCGACCCCACATTTGCTC
       L  L  K  L  A  R  T  T  P  Y  Y  K  R  N  R  P  H  I  C  S
 781  CTTCTGGGTGAAAGGAGAGTGTAAGAGAGGAGAGGAATGTCCATACAGACATGAGAAGCC
       F  W  V  K  G  E  C  K  R  G  E  E  C  P  Y  R  H  E  K  P
 841  TACAGATCCAGATGACCCCCTTGCTGATCAGAATATTAAAGACCGTTATTACGGAATCAA
       T  D  P  D  D  P  L  A  D  Q  N  I  K  D  R  Y  Y  G  I  N
 901  TGATCCTGTAGCTGACAAGCTTCTAAAGCGGGCTTCAACAATGCCTCGGCTGGACCCACC
       D  P  V  A  D  K  L  L  K  R  A  S  T  M  P  R  L  D  P  P
 961  AGAGGATAAAACTATCACCACACTATATGTTGGTGGTCTAGGTGATACCATTACTGAGAC
       E  D  K  T  I  T  T  L  Y  V  G  G  L  G  D  T  I  T  E  T
1021  AGATTTAAGAAATCATTTCTACCAGTTCGGAGATCCGGACGATCACTGTTGTGCAGAG
       D  L  R  N  H  F  Y  Q  F  G  E  I  R  T  I  T  V  V  Q  R
1081  ACAGCAGTGTGCTTTCATCCAGTTTGCCACACGGCAGGCTGCAGAAGTGGCTGCTGAGAA
       Q  Q  C  A  F  I  Q  F  A  T  R  Q  A  A  E  V  A  A  E  K
1141  GTCCTTTAATAAGTTGATTGTAAATGGCCGCAGACTGAATGTGAAATGGGGAAGATCCCA
       S  F  N  K  L  I  V  N  G  R  R  L  N  V  K  W  G  R  S  Q
1201  GGCAGCCAGAGGAAAAGAAAAAGAGAAAGATGGAACTACAGACTCTGGGATCAAACTAGA
       A  A  R  G  K  E  K  E  K  D  G  T  T  D  S  G  I  K  L  E
1261  ACCTGTTCCAGGATTGCCAGGAGCTCTTCCTCCTCCTCCTGCAGCAGAAGAAGAAGCCTC
       P  V  P  G  L  P  G  A  L  P  P  P  P  A  A  E  E  E  A  S
1321  TGCCAACTACTTCAACTTGCCCCCAAGTGGTCCTCCAGCTGTGGTGAACATTGCTCTGCC
       A  N  Y  F  N  L  P  P  S  G  P  P  A  V  V  N  I  A  L  P
1380  ACCGCCCCCTGGCATTGCTCCACCCCCACCCCCAGGTTTTGGGCCACACATGTTCCACCC
       P  P  P  G  I  A  P  P  P  P  P  G  F  G  P  H  M  F  H  P
1441  AATGGGACCACCCCCTCCTTTCATGCGGGCTCCAGGACCAATCCACTATCCTTCTCAGGA
       M  G  P  P  P  P  F  M  R  A  P  G  P  I  H  Y  P  S  Q  D
1501  CCCTCAGAGGATGGGAGCTCATGCTGGAAAACACAGCAGCCCCTAGCACCTTGTCACCAC
       P  Q  R  M  G  A  H  A  G  K  H  S  S  P
1561  TCTGGGGCTCTGTGGAAGAAAGGGCACTTAAAACTCCCAGTAAATCTTGGAATAAATATA
1621  TTTTTCCTTCCCTTGTAGTTTCCATGGTAGCTGAATGTGCTCAGATGTGAGCAGTCAGAG
1681  ACTGACAGCCATGCTTTCCTATACTTGTTCAAAGGATCGATGGACCGTAAATAAGCTGCC
1741  ATTAACACATCTGGTTACTGCTGTAACATGACTAATAAAACCGAACGCCTGTTCCCCTTA
1801  CCCGTGTGGGGGACACGCAGATGAGTGAATTGGAATGTCCAGCAGAGTTACCCTCCCAAT
1861  TATATGTTCATTTTGTATATTTTTTGGTCGGGGGAAAAATTGACCTGCAGTAAAAAAACC
1921  TTTGACCATTTTTATGTCCATTGGATACTTTCCTTTTTATCATCTTAAAAAAAGATAACT
1980  AGTACTAATCATTGTAGTGGCCTAAGTGTGATTTAACTCTTGAAGTCACACCCTCCGAAA
2041  GATGAGTAGAAACCAGCACCAGCACAGCCCAGATCTTCTCTTTCCTCTCCTTTTCCTCAT
2101  TTATTCCTAAAGGAATCTGACCATTTTACGTCTCTACGGCCCAAAAAAAGACAAAATAA
2161  AAATTCCTTTTTATTCCTGTCAACTGGATGGAAACACAAATTTCATGGAG
```

Figure 2 DNA sequence of cDNA clone 199G4. The longest open reading frame of 334 amino acids is shown. The initial exon amplification clone is denoted in boldface type. The polyadenylation signal is underlined.

transcript that shows no homology to any known genes or proteins within the data base. It does, however, show homology to several expressed sequence tags (ESTs). GenBank accession numbers for these ESTs are T70273, R23976, R07690, F11315, R09682, R09681, and T27262. Isolation of these ESTs were from cDNA libraries derived from fetal liver/spleen, placenta, adult pancreatic islets, and infant brain. Northern blot analysis of fetal brain, lung, liver, and kidney RNA shows intense hybridization to a 2.4 kb transcript (Fig. 3A).

PCR primers specific for exon clone 9D2 amplified the predicted 124-bp fragment from a fetal craniofacial cDNA library. Hybridization of this library with the radiolabeled 9D2 exon fragment yielded three clones. The largest clone was 2146 bp and did not encode a poly(A) tail. Analysis of the 9D2 cDNA sequence revealed that it contained the sequence of three other exon clones, 9C2, 9A6, and 8H12 (Fig. 4).

After a search of the GenBank data base using BLASTN it was revealed that cDNA 9D2 contains a large region of homology with GenBank entry D30755. Sequence for clone D30755 was obtained by random sequencing of cDNA clones derived from a human myeloid cell line kG1. Clone D30755 appears to encode an alternately spliced form of the 9D2 gene. The two clones contain 100% sequence identity from nucleotide 615 to 1484 of cDNA 9D2. At this position, clone D30755 contains an additional 74 bp inserted within the sequence. After this 74-bp insertion, nucleotide identity between the two sequences is found for the remaining 661 bp of clone 9D2 with only 3 nucleotide differences (Fig. 4).

Results from Northern blot analysis of RNA from adult tissue shows that clone 9D2 hybridizes to a 3.2-kb transcript in all tissues examined (Fig. 3B). In addition, there are several larger transcripts in skeletal muscle that may be the result of additional alternate splice events.

DISCUSSION

Analysis of families with TCOF1 using newly identified STRPs has reduced the size of the candidate gene region by a factor of two. Previously, the region containing the gene associated with this disorder was located in a ~1-Mb region between D5S519 and *SPARC* at 5q 32–33.1. We can now define the candidate region as between the loci *RPS14* and *ANX6*. We have focused on this ~450-kb region to identify genes that, by ge-

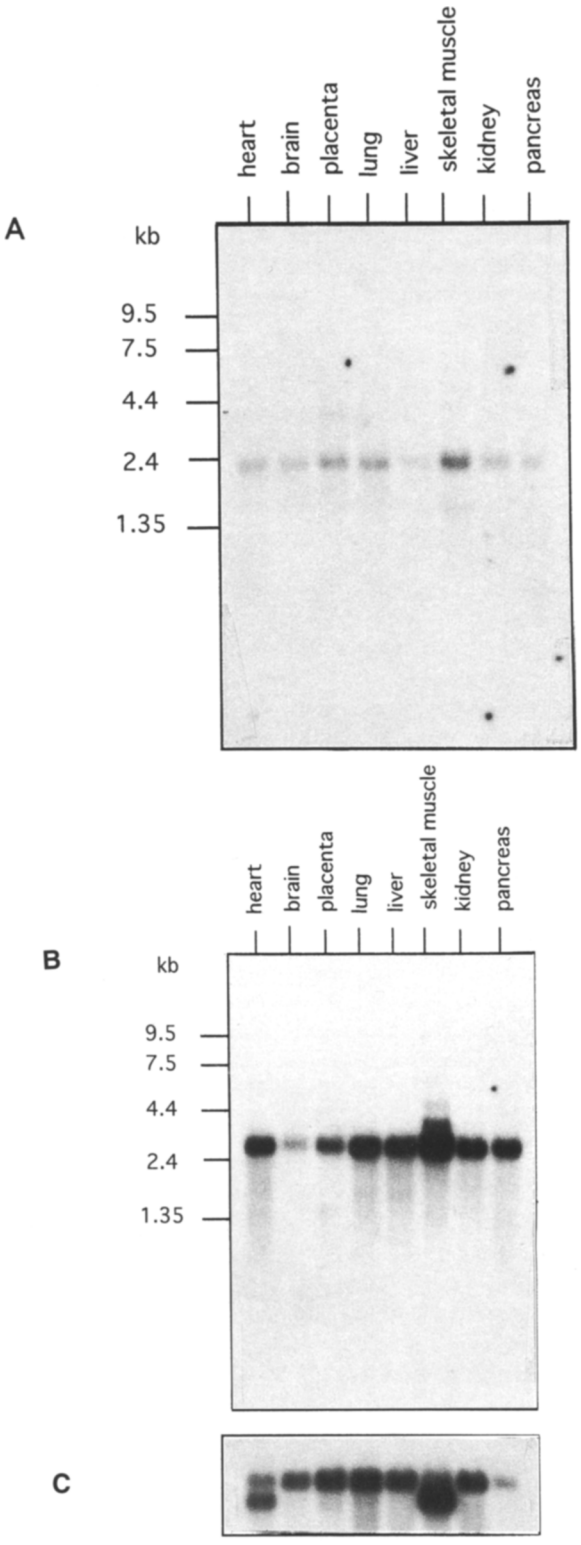

Figure 3 Northern blot analysis of cDNA clone 199G4 (*A*), cDNA clone 9D2 (*B*), and β-actin control (*C*). Filters containing RNA from the indicated adult tissues were purchased from Clontech and hybridized as described in Methods.

A

```
ACCAGAGAGCAGCAGCCATGCCAATGCGATGGCGCTGGGCCCCCTGCCCCGTGAGGACGG    60
CAACCTGATGCTGCACCTGCAGCGCCTGGAGACCACGCTGAGTGTGTGTGCCGAGGAGCC   120
GGACCACGGCCAGCTCTTCACCCACCTGGGCCGCATGGCCCTGGAGTTCAACCGACTGGC   180
ATCCAAGGTGCACAAGAATGAGCAGCGCACCTCCATTCTGCAGACCCTGTGTGAGCAGCT   240
TCGGAAGGAGAACGAGGCTCTGAAGGCCAAGTTGGATAAGGGCCTGGAACAGCGGGATCA   300
GGCTGCCGAGAGGCTGCGGGAGGAAAAATTTGGAGCTCAAGAAGTTGTTGATGAGCAATGG   360
CAACAAAGAGGGTGCGTCTGGGCGGCCAGGGTCAGCGAAGATGGAAGGGACAGGCAAGAA   420
GGCAGTCGGGTGGACAGCAGCAGGCTAGTGTGACGCGCAGTAAGGTCCCAGAGGTGGTGGC   480
CTTGGGCGCAGCCGAGAAGAAGGTGAAGATGCTGGAGCAGCAGCGCAGTGAGCTGCTGGA   540
AGTGAACAAGCAGTGGGACCAGCATTTCCGGTCCATGAAGCAGCAGTATGAGCAGAAGAT   600
CACTGAGCTGCGTCAGAAGCTGGCTGATTTGCAGAAGCAGGTGACTGACCTGGAGGCCGA   660
GCGGGAGCAGAAGCAGCGTGACTTTGACCGCAAGCTCCTCCTGGCCAAGTCCAAGATTGA   720
AATGGAGGAGACCGACAAGGAGCAGCTGACAGCAGAGGCCAAGGAGCTGCGCCAAAAGGT   780
CAAGTACCTGCAGGATCAGCTGAGCCCACTCACCCGACAGCGTGAGTACCAGGAAAAGGA   840
GATCCAGCGGCTCAACAAGGCCCTGGAGGAAGCACTGAGCATCCAAACCCCGCCATCATC   900
TCCACCAACAGCATTTGGGAGCCCAGAAGGAGCAGGGGCCCTCCTAAGGAAACAGGAGCT   960
GGTCACGCAGAATGAGTTGCTGAAACAGCAGGTGAAGATCTTCGAGGAGGACTTCCAGAG  1020
GGAGCGCAGTGATCGTGAGCGCATGAATGAGGAGAAGGAAGAGCTGAAGAAGCAAGTGGA  1080
GAAGCTGCAGGCCCAGGTCACCCTGTCAAATGCCCAGCTAAAAGCATTCAAAGATGAGGA  1140
GAAGGCAAGAGAAGCCCTCAGACAGCAGAAGAGGAAAGCAAAGGCCTCAGGAGAGCGTTA  1200
CCATGTGGAGCCCCACCCAGAACATCTCTGCGGGGCCTACCCCTACGCCTACCCGCCCAT  1260
GCCAGCCATGGTGCCACACCATGGCTTCGAGGACTGGTCCCAGATCCGCTACCCCCCTCC  1320
CCCCATGGCCATGGAGCACCCGCCCCCACTCCCCAACTCGCGCCTCTTCCATCTGCCGGA  1380
ATACACCTGGCGTCTACCCTGTGGAGGGGTTCGAAATCCAAATCAGAGCTCCCAAGTGAT  1440
GGACCCTCCCACAGCCAGGCCTACAGAACCAGAGCCAGCTGATCTCAGATTGCCAAGAAA  1500
CTAGAAGCCACTTGCACGGTGTGGCCAGAGCCTCAGCTGGATGAGAGGCTGAGATGGGTG  1560
GCCAGCTTGTATACCAGTCCCTGAACTGAGCTGTTTACAGGACTGGGGAGGCTCCACCCA  1620
GAAGGCTTTCATTTGTACTCTGCTGGGAGTGACTGGGAAAAACTCCTTCCCTGCTGCTGA  1680
GTGGAGAGAGGCCTCATCCGGCTTTGACCCACCATCCGTTGCAGAAGCCTCCAGGAGCAG  1740
CAATCCTAAGAGTGGGAGGCAGCCAAGACCCCCTTCCTTCAAAACCTCCCGGAAGTGGTT  1800
TCAGGCCCTCTAGTTGCCATGACCAATTTGTGTGTGTGTTTAATTTTTGCTTCAAGCTCT  1860
GTAGCAGGACCTGCCCCACGCACACCCCTACCCCTCTGTGAGGAGCTGTGGGAAGTGTGG  1920
GTTTGTCTCCAGAACAGAAGAGAATGATGGATATTCTGGCTCTGGGGCCCTCTCCACCAC  1980
CACTCACAGTAGCCTTGCTGAAGCCATCACAGATGGGAGAAGGCCATGCCAGCCACGTCC  2040
GCCGAGGGGCGCCAGCCTGAAGCTGCCAGGCCCTGAGGTTCAGACCCTGGACCCCATAGC  2100
TGGAGGCCTGTGGTGCCAGAAGCCCAGATTAGGGTGGCTGTCCATC                2146
```

B

```
TCTCCAAAAAATGACCGTGAGGGGCCTCAGTGAGACCAGATTGTGTCATTTGGCTCCAC
CTTCATCTTGCAGAG
```

Figure 4 (*A*) Sequence of cDNA clone 9D2. Individual exon amplification clones identified are as follows: exon 9D2, boxed; 9C2, reverse text; 9A6, underlined; 8H12, bold. The asterisk at nucleotide 676 denotes the start of homology between cDNA 9D2 and GenBank sequence D30755, which continues to nucleotide 2146 of cDNA 9D2. The arrow denotes the location of the 74-bp sequence insertion found in D30755 sequence. (*B*) Sequence of the 74-bp insertion denoted by arrow in *A*.

nomic location, are candidates for the TCOF1 locus. Toward this goal, we have generated a cosmid contig spanning ~390 kb. Potential exons were isolated from these cosmids using the technique of exon amplification. This procedure yielded 47 unique clones, a subset of which has been used to determine the precise location of four genes within the TCOF1 candidate gene region (Table 1). Previous analysis of YACs spanning this region identified a cluster of *Hpa*II tiny fragment (HTF) islands near the D5S519 locus, which may indicate the presence of other as yet uncharacterized genes (Dixon et al. 1993).

Data base searches showed that exons from two previously isolated and sequenced genes, *ANX6* and *GPX3,* were among the clones isolated in this study. Clones corresponding to exons 14, 15, and 16 of *ANX6* were isolated. Although *ANX6* can be excluded as a candidate for TCOF1 on the basis of recombination events occurring in two individuals affected by TCOF1, isolation of exons from this gene provided a useful control to show that the technique was working properly. Two clones were identified that represented exons 2–3 and exon 4 from *GPX3,* which had been assigned previously to chromosome 5 but had not been localized regionally (Chu 1994). The results presented here allowed us to determine the precise location of *GPX3* as being 30 kb proximal to ANX6. Based on the pattern of expression of *GPX3* and the function of the gene product, *GPX3* seems a very unlikely candidate for TCOF1 based on biological properties (Yoshimura et al. 1991).

Four different clones (3B1, 3A8, 1C9, and 1C6) were isolated that showed extensive (70%–90%) sequence homology to the rat *HSST* gene. The *HSST* gene product catalyzes the initial reaction in a series of sequential enzymatic modifications of heparan sulfate proteoglycan side chains. *HSST* is responsible for the reaction in which a portion of *N*-acetyl resides are removed and replaced by *N*-sulfate groups (Lindahl et al. 1986). Only regions of the polysaccaride that contain the *N*-sulfate group will be modified further by other enzymes in the pathway.

David et al. (1992) used monoclonal antibodies, one of which recognizes specifically the *N*-sulfate epitope, to examine the distribution of different modifications of the molecule during development in hamster embryos. This antibody intensely stains the outline of neural crest cells, specific regions of the brain, as well as mesenchymal tissues of the head and limbs (David et al. 1992). These results suggest that *HSST* is expressed in a temporal and spatial pattern during development that coincides with the early stages of craniofacial development, making this gene an interesting candidate for TCOF1. Even more in-

Table 1. Exon Clone Location and Homology

Cosmid	Exon	Size (bp)	Homology	Reference
2F	2H2	79	annexin VI exon 14	Smith et al. (1994)
	2G6	82	annexin VI exon 15	Smith et al. (1994)
	2G2	95	annexin VI exon 16	Smith et al. (1994)
1–43	8F8	100	*GPX3* exon 4	Chu et al. (1992)
	8H12	66	GenBank accession no. D30755	
1–61	4G3	258	*GPX3* exons 2 and 3	Chu et al. (1992)
	9A6	114	GenBank accession no. D30755	
	9C2	92	—	
	9D2	124	—	
3.3	199G4	158	—	
11A	3B1	120	rat *HSST*	Hashimoto et al. (1992)
5A1	3A8	90	rat *HSST*	Hashimoto et al. (1992)
	1C9	123	rat *HSST*	Hashimoto et al. (1992)
	1C6	87	rat *HSST*	Hashimoto et al. (1992)

teresting, heparin and heparan sulfate proteoglycans have been shown by several studies to play a major role in the interaction of fibroblast growth factors (FGF) with their cell surface receptors (FGFRs) (Yanon et al. 1991; Mckeehan and Kan 1994; Spivak-Kroizman et al. 1994; Aviezer et al. 1994). Binding of FGRs to the cognate FGFRs enhances receptor dimerization and activation of intracellular signal cascades. In the past year, mutations in FGFRs 1, 2, and 3 have been shown to cause eight different inherited disorders of craniofacial development and/or skeletal dysplasias (Mulvihill 1995; Tavormina et al. 1995). Therefore, it is very interesting to speculate that mutations in a gene (*HSST*) involved in the biosynthesis of heparan sulfate modifications could produce the phenotype of TCOF1, a disorder of craniofacial development.

Two additional transcripts from the TCOF1 region have been identified as possible candidates using exon amplification clones as probes. These clones, 9D2 and 199G4, have no homology to known genes within data bases. Northern analysis indicates that clone 9D2 hybridizes to a transcript of 3.2 kb. Currently, we are trying to obtain a full-length cDNA clone for 9D2. Additional coding sequence may provide insight into the function of the gene product. Without identifiable functions for the encoded proteins, it is difficult to speculate on how differences within either gene may cause TCOF1. Further analysis is required to determine both the function and the pattern of expression for each gene.

In summary, analysis of clones isolated using the exon amplification method has identified two genes that have been characterized previously, as well as two novel genes. Exon amplification clones have confirmed and refined the location for genes *ANX6* and *GPX3*. Exon clones also were used to isolate two novel genes, 9D2 and 199G4, of unknown function. Homology between four exons and the rat *HSST* gene sequence allowed for identification of the human *HSST* homolog. Currently, we are analyzing three genes, *HSST,* 9D2, and 199G4, in TCOF1-affected individuals and controls. By identifying the gene responsible for this developmental disorder, we hope to better understand how cellular migration or differentiation are controlled during craniofacial development.

METHODS

Linkage Analysis

Cosmid 2F, which contains the 3′-untranslated region of *ANX6,* was isolated as described below. The identification, optimization, and utilization of STRPs for genotyping has been described previously (Nelson et al. 1989). PCR primers 5′-GCATAGGCTGGAGAGAAAGA-3′ and 5′-AGGAAGGAATGTCATCGTGG-3′ were used to amplify the *ANX6* STRP.

Isolation of Cosmids

Cosmids were identified from the LANL chromosome 5

cosmid library using PCR primers specific to *ANX6, RPS14,* and D5S519 (Loftus et al. 1993). PCR-labeled products were used to screen the LANL chromosome 5-specific cosmid library by filter hybridization. Three approaches were used to identify cosmids corresponding to previously identified YACs that spanned the region. The cosmid library was screened by filter hybridization using each of the following probes: entire YACs; inter-Alu PCR fragments from YACs; and end-clone fragments from YACs (Dixon et al. 1992). Inter-Alu PCR products were obtained using Alu SJ1 primers (Nelson et al. 1989). Probes for directed chromosome walking and analysis of cosmid overlap were generated by asymmetric PCR labeling of cosmid end fragments using either a T3 or T7 sCOS-1 primer.

Identification of Exon Amplification Clones

Genomic DNA from 15 cosmids, spanning ~85% of the region of interest, was digested with *Pst*I or double-digested with *Bam*HI and *Bgl*II. The restriction fragments were ligated into corresponding sites of the pSPL3 vector (Church et al. 1994). The exon amplification protocol of Church et al. was followed with one modification. Amplified exon products were cloned using a deglycosolase-ligation independent cloning protocol (Rashtchian et al. 1992). PCR primers were modified such that uracil was substituted for thymidine on the 5′ end of the primer. Exon amplification products and pBluescript II were amplified with the modified PCR primers. After amplification, vector, "exon," and uracil deglycosolase (GIBCO-BRL) were incubated at 37°C for 30 min and immediately transformed into *Escherichia coli* DH5α strain. Exon amplification clones were sequenced using Sequenase version 2.0 (U.S. Biochemical) and analyzed by computer programs BLASTN and BLASTX to identify regions of homology at both the nucleic acid and protein level (Altschul et al. 1990). Cloned sequences were confirmed by Southern hybridization to be located within the TCOF1 candidate gene region.

cDNA Library Screening

DNA from several cDNA libraries was screened with exon-specific PCR primers (9D2–5′-GGAGGAAAATTTGGGCTCAAG/9D2–3′-GACAGGTCGGTGACGGAAGAA) and (199G4F–TCATTTCAGCCAGTTCGGAG/199G4R–ATTTCACATTCAGTCTGCGGC). DNA for each cDNA library was prepared from 10 plate lysates (>1×10^6 phage/plate) using Lambda Phage Kit (Quiagen). PCR reactions were performed with 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 200 μM each dNTP, 25 pmoles of each primer, 0.5 units of *Taq* (Boehringer Mannheim), and 5 ng of DNA. Reactions were done for 30 cycles under conditions of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec. Primers 9D2-5/9D2-3 amplified a fragment of predicted size, 124 bp from human craniofacial-specific cDNA from day 43 to 55 gestation-pooled samples from Jeff Murray, Michael Solursh, and Babu Padnanilam (University of Iowa, Iowa City). Primers 199G4F/199G4R amplified the expected 158-bp product from a a fetal brain library (Stratagene). Both libraries were plated on 20 plates, at 5×10^4 plaques/plate. DNA from plaques was transferred to duplicate filters and hybridized to radiolabeled PCR fragments. Filters were hybridized at 65°C overnight, and subsequently washed at 65°C in 2× SSC, 0.1% SDS, for 30 min; 1× SSC, 0.1% SDS, for 30 min; and 0.5× SSC, 0.1% SDS, for 30 min. Filters were exposed for 12–24 hr at 80°C to Kodak X-AR film. Positive primary plaques were isolated and purified by two more rounds of plaque screening.

Northern Blot Analysis

Northern blots of poly(A)$^+$ RNA from adult and fetal tissue (Clontech) were hybridized with radiolabeled cDNA probes from either exon clone 9D2 or exon clone 199G4. Filters were hybridized as per the Clontech protocol. Filters were subsequently washed at 50°C in 2× SSC, 0.1% SDS, for 30 min; 1× SSC, 0.1% SDS, for 30 min; and 0.5× SSC, 0.1% SDS, for 30 min. Filters were exposed for 12–24 hr at –80°C with an intensifying screen to Kodak X-AR film.

ACKNOWLEDGMENTS

We thank Ulla Bengtsson for technical assistance and Dr. Alan Buckler for the gift of the pSPL3 vector. S.K.L. was supported in part by National Institutes of Health (NIH) training grant GM07134. This work was supported by NIH grant AR42377-019 (J.J.W.), the Wellcome Trust 036797/Z/92/Z (M.J.D.) and 044684/Z/95/Z (M.J.D.), and by the Hearing Research Trust (M.J.D.). Dr. John Wasmuth passed away on December 29, 1995. We would like to dedicate this work to his memory.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Meyers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Aviezer, D.A., D. Hecht, M. Safran, M. Eisinger, G. David, and A. Yayan. 1994. Perlecan, basal lamina proteoglycan, promotes basic fibroblast growth factor-receptor binding, mitogenesis and angiogenesis. *Cell* **79:** 1005–1013.

Buckler, A., D.D. Chang, S.L. Graw, D. Brook, D.A. Haber, A.P. Sharp, and D.E. Housman. 1991. Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci.* **88:** 4005–4009.

Chu, F.F. 1994. The human glutathione peroxidase genes GPX2, GPX3, and GPX4 map to chromosomes 14, 5, and 19 respectively. *Cytogenet. Cell Genet.* **66:** 96–98.

Chu, F.F., R.S. Esworthy, J.H. Doroshow, K. Doan, and X.F. Liu. 1992. Expression of plasma glutathione peroxidase in human liver in addition to kidney, heart,

lung, and breast in humans and rodents. *Blood* **79:** 3233–3238.

Church, D.M., C.J. Stotler, J.L. Rutter, J.R. Murrel, J.A. Trofatter, and A.J. Buckler. 1994. Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6:** 98–105.

Crompton, M.R., R.J. Owens, N.F. Totty, S.E. Moss, M.D. Waterfield, and M.J. Crumpton. 1988. Primary structure of the human membrane associated CA2 + binding protein p68: A novel member of a protein family. *EMBO J.* **7:** 21–27.

David, G., X.M. Bai, B. Van Der Schueren, J.J. Cassiman, and H. Van Den Berghe. 1992. Developmental changes in heparan sulfate expression: In situ detection with mAbs. *J. Cell Biol.* **119:** 961–975.

Dixon, M.J., A.P. Read, D. Donnai, A. Colley, J. Dixon, and R. Willaimson. 1991. The gene for Treacher Collins syndrome maps to the long arm of chromosome 5. *Am. J. Hum. Genet.* **49:** 17–22.

Dixon, M.J., J. Dixon, D. Raskova, M.M. Le Beau, R. Williamson, K. Klinger, and G.M. Landes. 1992. Genetic and physical mapping of the Treacher Collins syndrome locus: Refinement of the localization to chromosome 5q32-33.2. *Hum. Mol. Genet.* **1:** 249–253.

Dixon, M.J., J. Dixon, T. Houseal, M. Bhatt, D.C. Ward, K. Klinger, and G.M. Landes. 1993. Narrowing the position of the Treacher Collins syndrome locus to a small interval between three new microsatellite markers at 5q32-33.1. *Am. J. Hum. Genet.* **52:** 907–914.

Dixon, J., A.J. Gladwin, S.K. Loftus, J.H. Ridley, R. Perveen, J.J. Wasmuth, R. Anand, and M.J. Dixon. 1994. A yeast artificial chromosome contig encompassing the Treacher Collins syndrome critical region at 5q31.3-32. *Am. J. Hum. Genet.* **55:** 372–378.

Dixon, J., A.J. Gladwin, S.K. Loftus, P.J. Scrambler, J.J. Wasmuth, and M.J. Dixon. 1995. Cloning of the human heparan sulfate N-deacetylase N-sulfotransferase gene from the Treacher Collins candidate gene region at 5q 32-33.1. *Genomics* **26:** 239–244.

Edery, P., Y. Manach, M. LeMerrer, M. Till, A. Vignal, S. Lyonnet, and A. Munnich. 1994. Apparent genetic homogeneity of the Treacher Collins-Franceschetti syndrome. *Am. J. Med. Genet.* **52:** 174–177.

Eriksson, I, P. Sandback, B. Ek, U. Lindahl, and L. Kjellen. 1994. cDNA cloning and sequencing of mouse mastocytoma glucosaminyl N-deacetylase/N-sulfotransferase, an enzyme involved in the biosynthesis of heparin. *J. Biol. Chem.* **269:** 10438–10443.

Frazen, L.E., J. Elmore, and H.L. Nadler. 1967. Mandibulofacial dysostosis (Treacher Collins syndrome). *Am. J. Dis. Child* **113:** 406–410.

Hashimoto, Y., A. Orellana, G. Gil, and C.B. Hirschberg. 1992. Molecular cloning and expression of rat liver N-heparin sulfate sulfotransferase. *J. Biol. Chem.* **268:** 20091–20095.

Jabs, E.W., X. Li, C.A. Coss, E.W. Taylor, D.A. Meyers, and J.L. Weber. 1991. Mapping of the Treacher Collins syndrome locus to 5q31.3-q33.3. *Genomics* **11:** 193–198.

Li, X., C.A. Wise, D. LePaslier, A.L. Hawkins, C.A. Griffen, S.J. Pittler, M. Lovett, and E.W. Jabs. 1994. A YAC contig of approximately 3 MB from human chromosome 5q 31-33. *Genomics* **19:** 470–477.

Lindahl, U., D.S. Feingold, and R. Lennart. 1986. Biosynthesis of heparin *Trends Biochem. Sci.* **11:** 221–225.

Loftus, S.K., S.J. Edwards, T. Scherpbier-Heddema, K.H. Buetow, J.J. Wasmuth, and M.J. Dixon. 1993. A combined genetic and radiation hybrid map surrounding the Treacher Collins syndrome locus on chromosome 5q. *Hum. Mol. Genet.* **2:** 1785–1792.

Mckeehan, W.L. and M. Kan. 1994. Heparan sulfate fibroblast growth factor receptor complex: Structure function relationships. *Mol. Reprod. Dev.* **39:** 69–81.

Mulvihill, J. 1995. Craniofacial syndromes: No such thing as a single gene disease. *Nature Genet.* **9:** 101–103.

Nelson, D.L., S. Ledbetter, L. Corbo, M.F. Victoria, R. Ramirez-Solis, T.D. Webster, D.H. Ledbetter, and C.T. Caskey. 1989. *Alu* polymerase chain reaction: A method for rapid isolation of human specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci.* **86:** 6686.

Poswillo, D. 1975. The oathogenesis of Treacher Collins syndrome (mandibulofacial dysostosis). *Br. J. Oral Surg.* **13:** 1–26.

Rashtchian, A., G.W. Buchman, D.M. Schuster, and M.S. Berninger. 1992. Uracil DNA glycosylase mediated cloning of polymerase chain reaction-amplified DNA: Application to genomic and cDNA cloning. *Anal. Biochem.* **206:** 91–97.

Rovin, S., S.F. Dachi, D.B. Borenstein, and W.B. Cotter. 1964. Mandibulofacial dysostosis, a familial study of five generations. *J. Pediatr.* **65:** 215–221.

Smith, P.D., A. Davies, M.J. Crumpton, and S.E. Moss. 1994. Structure of the human annexin VI gene. *Proc. Natl. Acad. Sci.* **91:** 2713–2717.

Spivak-Kroizman, T., M.A. Lemmon, I. Dikic, J.E. Ladbury, D. Pinchasi, J. Huang, M. Jaye, G. Crumley, J. Schlessinger, and I. Lax. 1994. Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation. *Cell* **79:** 1015–1024.

Tavormina, P.L., R. Shiang, L.M. Thompson, Y. Zhu, D.J. Wilkin, R.S. Lachman, W.R. Wilcox, D.L. Rimoin, D.H. Cohn, and J.J. Wasmuth. 1995. Thanatophoric dysplasia

(types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nature Genet.* **9:** 321–328.

Yanon, A., M. Klagsbrun, J.D. Esko, P. Leder, and D.M. Ornitz. 1991. Cell surface heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell* **64:** 841–848.

Yoshimura, S., H. Suemizu, Y. Taniguchi, K. Watanabe, Y. Nomoto, Y. Katsuoka, S. Arimori, and T. Moriuchi. 1991. Molecular cloning of human plasma glutathione peroxidase gene and its expression in the kidney. *Nucleic Acids Symp. Ser.* **25:** 163–164.

Received October 26, 1995; accepted in revised form December 18, 1995.