***TCOF1* gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treacher Collins Syndrome throughout its coding region**

CAROL A. WISE^{*†}, LYDIA C. CHIANG[‡], WILLIAM A. PAZNEKAS[‡], MRIDULA SHARMA[‡], MAURICE M. MUSY^{*}, JENNIFER A. ASHLEY^{*}, MICHAEL LOVETT^{*}, AND ETHYLIN W. JABS^{‡§}

^{*}Departments of Otorhinolaryngology, Molecular Biology and Oncology, and the McDermott Center, University of Texas Southwestern Medical Center, Dallas, TX 75235; and [‡]Departments of Pediatrics, Medicine, and Surgery, Center for Medical Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD 21287-3914

Communicated by Victor A. McKusick, The Johns Hopkins University, Baltimore, MD, January 13, 1997 (received for review November 23, 1996)

ABSTRACT Treacher Collins Syndrome (TCS) is the most common of the human mandibulofacial dysostosis disorders. Recently, a partial *TCOF1* cDNA was identified and shown to contain mutations in TCS families. Here we present the entire exon/intron genomic structure and the complete coding sequence of *TCOF1*. *TCOF1* encodes a low complexity protein of 1,411 amino acids, whose predicted protein structure reveals repeated motifs that mirror the organization of its exons. These motifs are shared with nucleolar trafficking proteins in other species and are predicted to be highly phosphorylated by casein kinase. Consistent with this, the full-length *TCOF1* protein sequence also contains putative nuclear and nucleolar localization signals. Throughout the open reading frame, we detected an additional eight mutations in TCS families and several polymorphisms. We postulate that TCS results from defects in a nucleolar trafficking protein that is critically required during human craniofacial development.

Treacher Collins Syndrome (TCS) is the most common human mandibulofacial dysostosis disorder. It shows autosomal dominant inheritance and has an estimated incidence of 1 in 50,000 live births, with approximately 60% arising from new mutations (1, 2). Although thought to be fully penetrant, TCS shows wide variability, with some cases so mildly affected that they may go undiagnosed. The major diagnostic criteria of this disease include bilaterally symmetric midface hypoplasia, downward slant of palpebral fissures and colobomata of lower lids, micrognathia, microtia, and other deformities of the ear often leading to conductive hearing loss. From the structures affected and from studies in mice that were exposed to the teratogens cis- and trans-retinoic acid, it has been deduced that the disease is a result of interference in the development of the first and second branchial arches (1, 3–5).

The TCS gene, *TCOF1* (MIM no. 154500) was initially localized to chromosome 5q31–33.3 by genetic linkage analysis (6, 7). This was then refined to the region between the colony stimulating factor 1 receptor gene (*CSF1R*) (proximally) and osteonectin gene (*SPARC*) (distally) (8, 9), and subsequently further refined to between marker D5S519, which is distal to *CSF1R* and *SPARC* (10). Physical and transcription maps of these regions were constructed (11–13). A candidate gene for TCS was recently identified (14) which, confusingly, in light of previous studies (10, 12) lies proximal to D5S519, between it and *CSF1R*. Five mutations, presumably leading to premature termination of translation, were detected in three exons from affected individuals. Five more mutations in *TCOF1* recently

have been reported in TCS-affected individuals (15), bringing the total number of identified mutations to 10. Of these, all are either nonsense mutations, insertions, deletions, or splicing mutations that apparently lead to premature termination of translation. Moreover, all appear to be unique to each family. A partial cDNA encoded by the *TCOF1* gene was identified and shown to contain an open reading frame for a putative protein product called “Treacle.” How Treacle functions normally or in the etiology of TCS was not clear, since no obvious similarity to any known proteins was described in these studies.

We present the genomic structure and complete coding sequence of *TCOF1*. Our analysis of the structure of the predicted *TCOF1* protein has revealed conserved motifs comprising individual exons, which are organized as repeated domains. By alignments with related motifs in other organisms, we predict that *TCOF1* encodes a highly phosphorylated nucleolar protein. We have detected an additional eight mutations within this open reading frame in nine individuals (one insertion and seven deletions). Each of these appears to lead to premature termination of translation. These and other findings (14, 15) support a model of haploinsufficiency of a nucleolar phosphoprotein in a critical stage of human mandibulofacial development leading to the pathogenesis of TCS.

MATERIALS AND METHODS

***TCOF1* Clone Extension.** The *TCOF1* gene (14) was extended by 5' and 3' rapid amplification of cDNA ends (BRL) as suggested by the manufacturer. The 3' end was extended in placental cDNA using gene-specific primer (GSP) 1: 5'-GAG CCA GAA GAG GAG CTT CA-3'; the product was amplified with GSP2: 5'-(CUA)₄ CGG TTG AAG GTG GAG ATC AA AG-3' (primers derived from exon 23). The 5' end of *TCOF1* was also extended in placental cDNA using GSP1A: 5'-AGG AGT TGG TAG ATG CTA GT-3' in exon 4. The product was deoxyadenosine-tailed for second strand synthesis, which was amplified with GSP2A1: 5'-TTC GGC TTC TGC TTC TTC CT-3' in exon 3 and reamplified with GSP2A: 5'-(CUA)₄ TTA CGG GCT GAG CCA GGA A-3' in exon 2. All PCR amplifications included 35 cycles of 94°C for 30 sec, 50°C for 30 sec, and 72°C for 30 sec. The DNA was isolated by gel purification using the Wizard PCR Preps DNA Purification System (Promega). The sequences of these PCR products were obtained from both directions using an Applied Biosystems 377 automated DNA sequencer.

Abbreviations: TCS, Treacher Collins Syndrome; GSP, gene-specific primer; ASO, allele-specific oligonucleotide.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. U76366).

[†]Present address: Texas Scottish Rite Hospital for Children, Dallas, TX 75219.

[§]To whom reprint requests should be addressed. e-mail: ewjabs@welchlink.welch.jhu.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA
0027-8424/97/943110-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

Sequence Analysis. The repeating motif within *TCOF1* was initially detected by alignment of the gene with itself using the BLASTN and BLASTX programs (16). Putative phosphorylation, nuclear, and nucleolar localization sites were identified by visual inspection of the sequence. By the newer 2.6.2 versions of BLASTX and BLASTN programs, multiple alignments were generated to *TCOF1* with genes of similar structures due to their repetitive nature.

Definition of *TCOF1* Exon/Intron Boundaries. Two genomic cosmid clones were sequenced to define exon/intron boundaries. Cosmid 245H1 from the Los Alamos chromosome 5 cosmid library (17) was used as a template for exons that are 5' of exon 16, and cosmid 40A6 was used to sequence the 3' end of *TCOF1*. The cosmid DNA was isolated using the

Plasmid Midi kit (Qiagen, Chatsworth, CA) and sequenced by The Johns Hopkins Genetic Resources Core Facility using an Applied Biosystems 373 automated DNA sequencer. PCR primer pairs were generated from intron sequences to amplify each exon in 20–50 μ l reactions containing 100–500 ng genomic DNA, 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl₂, 50 mM KCl, 0.2 mM dNTPs, 0.5 μ M each primer, and 1–2 units *Taq* DNA polymerase (Boehringer Mannheim). The PCR included 35 cycles of 94°C for 60 sec, annealing specific temperature (Table 1) for 60 sec, and 72°C for 60 sec. Sequencing the PCR products in both directions confirmed the exon/intron sequence derived from the cosmids.

Patient Population. The probands and their relatives were clinically examined. Genomic DNA from blood samples and

Table 1. Exons of *TCOF1*

Exon	Position in cDNA, bp	Splice acceptor* 5' to 3'	Splice donor* 5' to 3'	PCR primers† 5' to 3'	PCR size, bp‡ T _m , °C	Ref. source
1	5' UTR–108	ggtcgcgggt ATG GCC GAG G§	G AGC GGC CAG gtaagcggtc	AGGCGGGGCGTGCAGGTA	205	This study
2	109–164	ctctctgcag AAG TGT TTC C	AC TGG CAA CA gtaagtgggtg	CCGCTGATCTCCACATCTTG	65	
				CCTCCCAAAGTGCTGAGA	222	This study
				GTGTCCTGCTCCCTACTCCA	59	
3	165–304	tgtcctgcag A ACC TCA GAG	GCC AAA GCC A gtaagagcct	ATTCTTGTTGAGTTGTTT	308	This study
				CCCCAGGGTCTTTTAGGT	50	
4	305–378	tttcttgcag CC CCA AGA CT	A AAA GCC AAG gtgagtggga	TCATCTGGCTCCTTTAGCAG	141	15
				TAGGCAATAGCTTGAAGGC	63	
5	379–565	ttctctgtag GCA GAG ACA G	GCC AAG CCT G gtaagaagtc	TTGGGTTGAGATGCAAGTGG	297	15
				AAGTCTGGGGACTAGGTTC	59	
6	566–639	cgatcctcag GG ATG GTG TC	A GAC GTG GAG gtaattgccca	TGGAAGGGAGTCCCTCAGT	167	15
				GTTCTGGAAGGGTTAGAGG	62	
7	640–852	ttttccaccg GTA AAG GCC T	C CTG CTT CAG gtgaggcctg	GTTTTCACAAGCAGGAGAGC	313	This study, 14
				AGAAGGCCTTCTGGGGATG	60	
8	853–1047	gtttctccag GCG AAG GCC T	G CCT GCT CAG gtgaggcaga	GTGTCTGTGTCTCTCTCAC	298	15
				TTTAGGCATGGGGCTACTCT	63	
9	1048–1257	ctcactccag GCG AAG CCT T	T GCA GCT CAG gtgaggctgg	ACCTTTGCCACATCCAGCTC	328	15
				TCTTTTGAGGCAGGGCAGC	65	
10	1258–1473	tgtctcccag GTG AAG CCC T	C CCG GCT CAG gtgaggcccc	ACTCCCTCCCTAATCTTGTC	278	This study, 14
				GAAAGAGCCTTACAGGAAGG	60	
11	1474–1662	ctcactccag GAA AAG TCC T	T GCA GCT CAG gtgaggcctg	TACCCTGGGCTCCCTCTC	295	This study
				CCGGGGGTGCTGACTGTG	64	
12	1663–1911	gtcccccacag GCA AAA CCA G	C GTG GGA CAG gtgaggcctg	GTGGGGCAGAACAGATGG	408	This study
				GGGATGACAAGGGGAAGA	62	
13	1912–2109	gtcatcccag GCA AAG TCT G	T CCA GCA CAG gtgaggccta	GGAGACACCTCTCTCCCTC	257	This study, 14
				GGATGGGCCTGCTCCTCTCA	58	
14	2110–2247	ctccactcag GTG AAA ACC T	T CCA TCC AAG gcaagtgggg	CAATCTCACCTTCTCCCTCCT	219	15
				AACCTCCACACCTCCTGTG	62	
15	2248–2427	tgcaattcag GTG AAG CCA C	G CTG GCT CAG gtgaggggga	GGGAGTGGGACCTGAAAGAA	277	15
				CCCATGTAGGGGATGATCTC	56	
16	2428–2628	ctccctccag GCG AAG CCT T	C TCT GCC CAG gtaagacttg	AGATCATCCCCTACATGG	360	This study
				CCCTATACCCCGTTCTG	55	
17	2629–2815	gtttttcaag GTG ATT AAA C	TTG ACT CCT G gtgagcgcag	CCCCATCACCTCCTTTCC	335	This study
				CCAGTGTCTGTCCCTTCTG	64	
18	2816–2952	tccatttcag GC ATC AGA AC	A GCC ACT CAG gtacctgggtg	GCACAGGCCGGTAAATTG	315	This study
				TTGCAGGCCATCCCATCA	65	
19	2953–3066	ccaccacacag GTG TCA AAG A	T TCA AGT GGG gtgagcttgg	CCCCAGCCAGACAGCATC	320	This study
				AGGGAGGCCAAACCAAGTG	64	
20	3067–3286	accgaattag GTT GAC AGT G	CAC ACG CTG G gtgaggggtgc	ACTTGCCCTAATTTTTC	320	This study
				CCACACAACACCTCTTC	55	
21	3287–3369	tctccagtag GT CCC ACC CC	G CCA TCC CAG gtaactgcaa	ATGGGGGTGAGGGACCTG	275	This study
				CTGAGGGATCGGGTAGAC	60	
22	3370–3550	gcttcttcag TCT CTC CTC T	CCT AAA ACA G gtaagttaag	CTCTGTGCTTGTGTGCC	387	This study
				CACCTGCCCTGTCCCTCTG	63	
23	3551–4111	ctctccatag GT GGA AAA GA	TCC GAC AAG A gtgagtgacc	ACTCCCTGCACCTCTTC	364	This study
				TGGTCTCCCGATAGCTTC	60	
				AGAAAGAAGGTGGTGGAC	466	
				TCTACATGGGAGGAATGA	50	
24	4112–4209	cttcccttag GA AAA AAA GA	G AAG AAA AAG gtagagagtt	ATTGACCCAGCACTTAG	248	This study
				GAAGGGGGCAGGAACCA	60	
25	4210–3' UTR, 4528	ctctccacag AAG AAG ACA G	ccaggcacag gtaagcttcc	GCAGTGGGTGGGGAAG	168	This study
				CCACAGGGGACACAGAG	60	
26	3' UTR, 4259–4741	cttcccttag ggatttcccta	gaggaagggtpolyA¶	ATGCCAGATTTTCATTTTC	406	This study
				CTGTGGAGCAAGGTGGTG	50	
				AGTGACCTCCTCTCCTTC	411	
				TAGTTTAGATGCCACCTC	52	

UTR, untranslated region.

*Uppercase letters are coding sequences. Lowercase letters are noncoding sequences. Spaces in sequences indicate exon/intron boundaries or reading frames.

†Primer pair sequences are noted for PCR amplification of each exon. Two pairs of PCR primers are required to amplify the entire length of exons 23 or 26.

‡The PCR product size and the annealing temperature used for PCR amplification are noted.

§This is the sequence prior to and including the translation start site. The A nucleotide of the ATG is numbered 1.

¶This is the sequence prior to where the poly(A) tract is added.

lymphoblast cultures were isolated from members of 55 unrelated TCS families and 60 controls.

Heteroduplex Analysis. Heteroduplex analysis was performed to scan a total of 13 *TCOF1* exons (4–16) for mutations. Five microliters of each PCR product was denatured at 98°C for 10 min and annealed at 68°C for 1 hr. The heteroduplexes were analyzed on a 1× Mutation Detection Enhancement gel (FMC) at 580 V for 16 hr. The gel was stained with ethidium bromide to visualize heteroduplexes. PCR products with mobility shifts on the Mutation Detection Enhancement gel were run on 2.0% NuSieve (FMC) agarose gels.

Mutation and Polymorphism Detection. Isolated DNA was sequenced in both directions. Sequence data were used to establish allele-specific oligonucleotide (ASO) hybridization and restriction enzyme-based screening methods that would distinguish between the mutant and normal sequence or between polymorphisms. For ASO hybridizations, PCR products were dotted onto Hybond-N+ (Amersham) filters, pre-hybridized for 30 min, and hybridized for 1 hr with Rapid-hyb buffer (Amersham). The ASO was ³²P-end-labeled with polynucleotide kinase. The filters were washed once at room temperature for 20 min in 5× SSC/0.1% SDS and then washed two times for 15 min at the specific ASO temperature wash (Table 2) in 0.2× SSC/0.1% SDS. The same filter was used for both the mutant and the corresponding normal ASO probes. Restriction enzyme digests were performed according to manufacturer's instructions (New England Biolabs), using 10 μl from the PCR reaction in a total volume of 20 μl.

TA Cloning. To obtain each allele with insertions or deletions, the PCR product was cloned into vector pCR2.1 using the TA cloning kit (Invitrogen) according to manufacturer's instructions. Clones containing the mutant or polymorphic alleles were detected by ASO hybridization or restriction

enzyme digestion. PCR products were prepared for automated sequencing as stated above.

RESULTS

Deriving the Full-Length Coding Region of the *TCOF1* Gene. The published partial cDNA sequence of *TCOF1* was extended in both the 5' and 3' directions by rapid amplification of cDNA ends. This extended the sequence 548 nucleotides 3' to the poly(A) tail, which lies 17 nucleotides downstream of a consensus polyadenylation signal and 505 nucleotides downstream of a translational stop site. 5' rapid amplification of cDNA ends extended the sequence 126 nucleotides, including an in-frame translational start site (ATG with a Kozak initiation consensus sequence) to an in-frame stop signal (TAA). This gave a composite cDNA of 4,840 nucleotides. The size of the *TCOF1* transcript is approximately 5.3 kb, as measured by our Northern blot analysis of mRNA from fetal brain, lung, liver, and kidney [CLONTECH, data not shown; previous estimates of transcript size are 5.8 and 6.3 kb (14)]. When the predicted poly(A) tract length is subtracted from this we estimate that at most there are a few hundred nucleotides remaining in the 5' untranslated region of the gene. Translation of the 4,233-nucleotide open reading frame predicts a 1,411 amino acid protein of approximately 155 kd. The cDNA sequence can be accessed in GenBank (accession no. U76366).

Sequence Analysis. The *TCOF1* predicted protein is of very low complexity. Alanine is the majority amino acid (14.7%), followed by serine (13.6%), lysine (11.1%), glutamic acid (9.1%), and proline (9.1%). Close visual inspection revealed a repeated motif that mirrors the exon/intron boundaries within the gene (Fig. 1). Within each of these repeated motifs are highly acidic clusters of amino acids, which contain the consensus casein kinase phosphorylation site (S/T-X-X-D/E)

Table 2. *TCOF1* mutations in TCS

Exon/nucleotide	Amino acid	Screening method	Case*
5 422insA	H141Q† STOP at 33 aa away	ASO‡ ATGCCACAACCCTGCCAC 62 ATGCCACACCCTGCCAC 60	1F
5 497delATAC	N166I† STOP at 48 aa away	ASO‡ TCAGCAATACGTTGGTCT 48 TCAGCAAATACTACGTTG 45	1U
10 1408delAG	S470Q† STOP at 1 aa away	ASO‡ CAGAGCAGTAGTGAGGAG 48 CAGAGAGCAGTAGTGAGGAG 53	2F
16 2490delC	PR830–831PG§ STOP at 42 aa away	ASO‡ CCAAGGAGTCCCCAGGAA 53 CCAAGGAGTCCCCAGGAA 55	1F
16 2526delAG	TG842–843TA§ STOP at 11 aa away	<i>Eco0109I</i> (nl)¶	1F
16 2552delA and 2561delA	K851S STOP at 13 aa away	ASO for 2552delA‡ CAGGCAGGGAGCAGGA 48 CAGGCAGGAAGCAGG 48 <i>Sfa</i> NI for 2561delA	1S
16 2565delAG	SG855–856SE§ STOP at 9 aa away	<i>DdeI</i> (nl)¶	1F, 1S

ins, insertion; del, deletion; STOP, termination codon; aa, amino acid.

*The number of familial (F) and sporadic (S) cases in which mutations occurred is designated. If it is unknown whether a case is familial or sporadic because relatives are unavailable, it is designated as U.

†The amino acid change indicates the first amino acid that is substituted. All subsequent amino acids are changed as a result of the frameshift.

‡ASO primer sequences in 5' to 3' orientation are listed with *T_m* °C used to wash membranes. The top oligonucleotide hybridizes to the mutant sequence and bottom oligonucleotide hybridizes to the normal sequence.

§The amino acid change indicates (i) the amino acid in which a frameshift occurred without substitution and (ii) the next amino acid that is substituted. All subsequent amino acids are changed as a result of the frameshift.

¶Restriction enzymes used to detect mutant sequences are listed. (nl) indicates that the enzyme digests the normal, not the mutant sequence.

||The amino acid change indicates the result of both mutations in the same gene.

FIG. 1. Repeated domains in *TCOF1*, polymorphisms, and mutations. Exons are aligned to show repeating protein structure of exons 7–16. The amino acids shared by all or the majority of exons are shown in boldface type. Amino acids that could be phosphorylated by casein kinase are shown in green, and nuclear and nucleolar localization signals are shown in blue. All polymorphic variations in pink and mutations in red, found to date, are shown. The red amino acids boxed in green indicate that these can be phosphorylated and have been shown to be mutated.

(18). Ten of the exons, 7–16, are most similar to each other with exons 8 and 16 having the highest protein identity of 78%. However, exons 3, 6, 17, 20, and 21, although less well conserved, also contain several potential casein kinase phosphorylation sites. Virtually every serine within the *TCOF1* repeats could be phosphorylated, and only a few exons, such as the first exon, are without obvious potential phosphorylation sites. The carboxyl terminus of *TCOF1* is extremely lysine rich (Fig. 1). This region contains several potential nuclear localization signals (K-K/R-X-R/K) (19, 20) at amino acid 1,285 (KKKKK), 1,314 (KRDK), 1,325 (KKGK), 1,362 (KKEK), 1,370 (KRKKDK), 1,377 (KKEKKKKAKK), and 1,398 (KKKKKKKK). An additional potential nuclear localization signal is contained at the amino-terminal end at position 74 (KKTR).

The entire *TCOF1* gene, as well as its individual repeats, were searched against the *Saccharomyces cerevisiae* and *Caenorhabditis elegans* genomes to identify potential homologs. A large number of similar regions were detected, most of which are predicted to encode serine- or lysine-rich regions, by BLAST searches (16). It is not yet clear which if any of these genomic regions might encode a homolog to *TCOF1*. However, a search of the literature revealed a number of proteins from various species that contain multiple casein kinase phosphorylation sites. One protein, Nopp140 (for nucleolar phosphoprotein) shows striking similarity to the *TCOF1* protein. This protein, identified in rat, has been shown to shuttle between the nucleolus and the cytoplasm (21). Nopp140, like *TCOF1*, is a low complexity protein (17.2% serine/16.1% lysine/14.3% alanine/9.7% glutamic acid/9.2% proline), which contains a 10-fold repeat of alternating acidic and basic regions, with each acidic domain containing numerous casein kinase phosphorylation sites. Nopp140 also contains seven potential nuclear localization signals. It is difficult to assign a numerical value to the degree of similarity between the Nopp140 and *TCOF1*

because multiple alignments can be generated due to the repetitive nature of each gene. A comparison of the predicted protein sequences of the two most similar regions within each (amino acids 891–1117 of *TCOF1*, amino acids 302–528 of Nopp140) reveals 21% identity and 35% similarity when conservative changes are included in the calculation.

Genomic Structure and Mutation Analysis. Locations of exon/intron boundaries were determined by comparing genomic sequence to the extended cDNA sequence. Genomic sequence was derived from cosmids that were identified from a chromosome 5-specific genomic cosmid library (17) and confirmed by sequence from PCR products of each exon amplified from intronic primer pairs. The gene is encoded by 26 exons (49–561 bp in size) with the termination codon in exon 25 (nucleotides 4234–4236). Individual exons plus flanking intron sequences have been deposited in GenBank (accession nos. U84640–U84665) (Table 1). The genomic size of the gene was determined to be greater than 20 kb by Southern blot analysis (data not shown). PCR primer sets were used to amplify 13 of the 26 *TCOF1* exons from affected individuals as well as unaffected controls. These PCR products were scanned for mutations by heteroduplex analysis, and mutations were confirmed by direct sequencing. ASOs were developed to the mutant and control sequences in five cases, and hybridization conditions were established that discriminated between the two. The three other mutations resulted in a gain or loss of a restriction enzyme recognition site (Fig. 2). In all, eight mutations were detected in 9 individuals including six familial cases, two sporadics, and one of unknown origin. In two instances, a mutation (deletion of AG) was shared by two unrelated families, one at nucleotide 1408 and the other at nucleotide 2565. In all other cases, mutations were only detected in a single family. Conversely, one individual was found to have two mutations in a single *TCOF1* allele, and neither parent was found to have either mutation (Fig. 2 and Table 2). Mutations segregated with the disease in familial cases and were

FIG. 2. Sequences, ASO hybridization, and restriction enzyme digestions of new *TCOF1* mutations. The order of the mutations are the same as in the Table 2. The top sequence shows the heterozygous mutation in a patient and compares it with the bottom sequence from a control. The pink asterisks denote the nucleotides involved in a mutation. When the precise nucleotide is not known because of the presence of repeats, the region where the change could have occurred is indicated by a bracket. Below the sequences are the family designation, one pedigree for each mutation, and results of either ASO or restriction enzyme digestion, confirming the presence of the mutation. A shaded box indicates a family member who was unavailable for DNA analysis, and the diamond indicates that the sex of the fetus was unknown. For the ASO experiments, the upper dot blot illustrates the detection of the mutant allele only in affected family members, whereas the lower shows the detection of the normal allele. For the electrophoretic experiments, the DNA marker in the right lane is ϕ X174 RF DNA/*Hae*III. The entire PCR product size for exon 16 is 360 bp. For family FTCSWK29-5.1, the fragment sizes from the *Eco*0109I digest are 200 and 160 bp for the normal allele and 360 bp for the mutant allele. With the *Sfa*NI digestion for family FTCSAW10m12, the normal product does not cut, but the mutant allele yields 220 and 139 bp fragments (the latter fragment is not shown). *Dde*I digests for family FTCSAP51.3, the normal allele to 235, 85, 21, and 19 bp fragments and the mutant allele to 318, 21, and 19 bp (the fragments less than 235 bp in size are not shown).

neither detected in unaffected members of the families nor in 60 controls. This study brings the total number of mutations to 18, scattered among 8 different exons and at potential phosphorylation sites in *TCOF1* (14, 15). In each case, the mutations are predicted to result in premature termination of translation.

Table 3. *TCOF1* polymorphisms

Exon/ nucleotide	Amino acid	Screening method	Frequency*
10	P/L439	<i>Eco</i> 0109 $\ddagger$ (P1)	1%
1316C/T $\dagger$			1/104
10	P/P449	<i>Sma</i> I $\ddagger$ (P2)	77%
1347T/C $\dagger$			92/120
11	V/I518	<i>Alu</i> NI $\ddagger$ (P1)	3%
1552G/A			6/198
11	S/S537	ASO $\S$	52%
1611G/A		AGGAGTCGTCGGACAGTG 53 AGGAGTCATCGGACAGTG 51	61/118
Intron 15	None	ASO $\S$	88%
2428-20ins		CCTCCAGGCTCTCTCATCC 58 CTCTC (P2) GGCTCTCTCTCTCATCC 53	151/172
16	A/V810	<i>Hph</i> I $\ddagger$ (P2)	82%
2429C/T			106/130
16	S/L845	ASO $\S$	1%
2534C/T		AGGGCCTTCGGCTGC 47 AGGGCCTTTGGCTGC 45	1/122

*Percentage and absolute number of P2 allele over total number of alleles studied in unrelated individuals, primarily Caucasians.

$\dagger$ These polymorphisms were originally reported in ref. 14.

$\ddagger$ Enzymes used to detect polymorphic sequences are listed. (P1) indicates that the enzyme cuts the first allele (e.g., 1316C), not the second allele (P2) (e.g., 1316T).

$\S$ ASO primer sequences in 5' to 3' orientation are listed with T_m °C used to wash membrane. The top primer hybridizes to P1 and the bottom primer hybridizes to P2.

An unexpectedly large number of polymorphisms were detected within *TCOF1* in this study: five different polymorphisms were detected while screening for mutations by heteroduplex analysis (Table 3). All of these changes were determined to be polymorphic because they are either silent mutations that do not alter the amino acid; they occur in noncoding regions but do not change the sequence to resemble splice sites; and they were detected in unaffected relatives and controls. Only one polymorphism, a 5-bp insertion in intron 15, occurred in a noncoding region. Three of the five remaining polymorphisms are predicted to produce amino acid changes: valine-518 to isoleucine, alanine-810 to valine, and serine-845 to leucine. The latter two changes occur within the context of a CpG dinucleotide that has an increased propensity for a C to T transition. Thus far, six polymorphisms have been detected (14) in three different exons and at potential phosphorylation sites of *TCOF1*.

DISCUSSION

The emerging picture is that TCS is caused by a wide spectrum of private mutations in *TCOF1*, which has immediate implications for genetic counseling. Since no common mutation has been found to date, mutation data will most likely need to be established for each individual family. This will be facilitated now that the complete genomic structure of the *TCOF1* coding region has been established. However, the possibility still exists that TCS can be caused by more than one gene, because recombination in affected individuals has been reported to preclude *TCOF1* as the disease-causing gene (10, 14). Although the majority of families appeared to be linked to this region of chromosome 5, heterogeneity would not be detected if two such disease-causing genes were near each other.

It is interesting that *TCOF1* has such a large number of detected polymorphisms; the expected rate is approximately one per 1–3 kb of coding sequence (22), and thus far seven

independent polymorphisms, one-half with coding changes, have been detected in only a portion of the coding region (exons 4–16, 2,325 bp). This frequency approaches that found in noncoding sequence (22). It may be that redundancy in the structure reflects redundancy in function of the repeated motifs in the protein. Therefore, these variations may be better tolerated, and the selection process against such changes may be less strict. However, this does not fully explain the fact that some 60% of TCS cases appear to arise *de novo*. The mechanism that drives such a high rate of independent *de novo* mutations is not clear at this point; one study suggested a slight increase in paternal age in sporadic TCS (2). It will be interesting to investigate if mutations in *TCOF1* can indeed be correlated with parental age or male versus female germ-line mosaicism, or if *TCOF1* represents a mutational hot spot.

The intriguing sequence and structure of *TCOF1* suggest its possible function. First, it contains several potential nuclear localization signals. Upon closer examination, *TCOF1* also contains two regions, at amino acids 1,362–1,365 and 1,370–1,385, which resemble previously identified nucleolar localization signals (19, 23–25). In these studies, it was demonstrated that two short basic regions separated by 1–4 amino acids, at least one of which was acidic, was sufficient to target a fusion protein to the nucleolus of COS7 cells. Second, *TCOF1* contains numerous casein kinase phosphorylation sites, suggesting a possible regulatory mechanism. Casein kinase is a serine/threonine kinase that requires acidic residues (“determinants”) on the carboxyl-terminal side of the phosphorylated residue; the activity is enhanced by acidic residues on the amino-terminal side as well. It prefers to phosphorylate serine more than threonine, and aspartate serves as a better determinant than glutamate (18). The +3 acidic position is especially critical. Phosphoserine can replace glutamate or aspartate as determinants. Thus, phosphorylation of serine can enhance the phosphorylation of an upstream serine and so on, leading to an overall cooperative mechanism of phosphorylation (26). Third, *TCOF1* bears a particularly striking resemblance to the Nopp (“nucleolar phosphoprotein”) family of proteins identified in human (27), rat (21), and *Xenopus* (28). The Nopp proteins, in turn, resemble other proteins including the yeast proteins NSR1 (29, 30) and NPI46 (31) and human nucleolin (32). Each of these proteins contains repeated domains which contain potential sites of casein kinase phosphorylation, and each has been localized to the nucleolus. Primarily by virtue of their localization these proteins have been proposed to function in some aspect of ribosome assembly. In the case of rat Nopp140, perhaps the best studied, the protein has been visualized shuttling between cytoplasm and the dense fibrillar component of the nucleolus along fibrillar tracks. Moreover, it binds nuclear localization signals contained within peptides *in vitro*. Thus, it has been proposed that this protein possibly functions as a chaperone, shuttling proteins, presumably ribosomal proteins or preribosomes, in and/or out of the nucleolus (21).

With the elucidation of the genomic structure of *TCOF1* it is clear that the repeated structure is a reflection of the exon structure of the gene, suggesting that it arose evolutionarily by a gene duplication mechanism. These repeated regions are rich in serine, lysine, alanine, glutamic acid, and proline; in rat Nopp140 these are also the most abundant amino acids and are organized into repeating acidic and basic regions. Serine residues found within these repeats in rat Nopp140 are heavily phosphorylated by casein kinase *in vivo*, and it is this form of the protein, and not the unphosphorylated form, which binds nuclear localization signals *in vitro*, implying that phosphorylation could be a means of regulating the activity of this protein. Perhaps the TCOF1 protein is subject to the same regulation.

The mutations we and others (14, 15) have detected in *TCOF1* would cause either decreased mRNA levels or protein truncations and presumably either haploinsufficiency or a functionless protein because of the loss of the nuclear localization signals that are required for its shuttling capacity. How could haploinsufficiency of a widely expressed protein, which perhaps functions in ribosome assembly, cause the specific features of TCS? One possibility is that *TCOF1* could function at a critical rate-limiting step during development when high levels of translational activity are essential. For example, *Minute* strains in *Drosophila melanogaster* harbor mutations in different components of the translational machinery. These strains are homozygous lethal, yet heterozygotes display a common structural phenotype of small bristles and small body size (33). Alternatively, deficiency of a *TCOF1*-shuttled protein involved specifically in craniofacial embryogenesis may lead to the restricted pattern of anomalies seen in TCS. Another question to be answered is why TCS displays such a variable phenotype. Perhaps other cellular proteins modify or partially replace the activity of TCOF1. Although a loss of function is likely, a dominant negative mechanism cannot be ruled out and could also explain the variability in expression (34). These speculations await cellular localization and functional studies of the TCOF1 protein.

We thank Roxann Ashworth, A. F. Scott, and T. D. Howard for generous assistance. This work was supported by National Institutes of Health Grants DE10180 and DE11131 (E.W.J.), Mental Research Center Grant HD24061, Outpatient General Clinical Research Center Grant RR00722, and Pediatric Clinical Research Center Grant RR00052.

- Gorlin, R. J., Cohen, M. M. & Levin, L. S. (1990) *Syndromes of the Head and Neck* (Oxford Univ. Press, Oxford).
- Jones, K. L., Smith, D. W., Harvey, M. A. S., Hall, B. D. & Quan, L. (1975) *J. Pediatr.* **86**, 84–88.
- Sulik, K. K., Johnston, M. C., Smiley, S. J., Speight, H. S. & Jarvis, B. E. (1987) *Am. J. Med. Genet.* **27**, 359–372.
- Poswillo, D. (1975) *Br. J. Oral Surg.* **13**, 1–26.
- Wiley, M. J., Cauwenbergs, P. & Taylor, I. M. (1983) *Acta Anat.* **116**, 180–192.
- Dixon, M. J., Read, A. P., Donnai, D., Colley, A., Dixon, J. & Williamson, R. (1991) *Am. J. Hum. Genet.* **49**, 17–22.
- Jabs, E. W., Li, X., Coss, C. A., Taylor, E. W., Meyers, D. A. & Weber, J. L. (1991) *Genomics* **11**, 193–198.
- Dixon, M. J., Dixon, J., Raskova, D., Le Beau, M. M., Williamson, R., Klinger, K. & Landes, G. M. (1992) *Hum. Mol. Genet.* **1**, 249–253.
- Jabs, E. W., Li, X., Lovett, M., Yamaoka, L. H., Taylor, E., Speer, M. C., Coss, C., Cadle, R., Hall, B., Brown, K., Kidd, K. K., Dolganov, G., Polymeropoulos, M. H. & Meyers, D. A. (1993) *Genomics* **18**, 7–13.
- Dixon, M. J., Dixon, J., Houseal, T., Bhatt, M., Ward, D. C., Klinger, K. & Landes, G. M. (1993) *Am. J. Hum. Genet.* **52**, 907–914.
- Li, X., Wise, C. A., Le Paslier, D., Hawkins, A. L., Griffin, C. A., Pittler, S. J., Lovett, M. & Jabs, E. W. (1994) *Genomics* **19**, 470–477.
- Dixon, J., Gladwin, A. J., Loftus, S. K., Riley, J., Perveen, R., Wasmuth, J. J., Anand, R. & Dixon, J. J. (1994) *Am. J. Hum. Genet.* **55**, 372–378.
- Loftus, S. K., Dixon, J., Koprivnikar, K., Dixon, J. J. & Wasmuth, J. J. (1996) *Genome Res.* **6**, 26–34.
- The Treacher Collins Syndrome Collaborative Group (1996) *Nat. Genet.* **12**, 130–136.
- Gladwin, A. J., Dixon, J., Loftus, S. K., Edwards, S., Wasmuth, J. J., Hennekam, R. C. M. & Dixon, M. J. (1996) *Hum. Mol. Genet.* **5**, 1533–1538.
- Althschul, S., Gish, W., Miller, W., Meyers, E. & Lipman, D. (1990) *J. Mol. Biol.* **215**, 403–410.
- Longmire, J. L., Brown, N. C., Meincke, L. J., Campbell, M. L., Albright, K. L., Fawcett, J. J., Campbell, E. W., Moyzis, R. K., Hildebrand, C. E., Evans, G. A. & Deaven, L. L. (1993) *Genet. Anal. Tech. Appl.* **10**, 69–76.
- Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. G. (1987) *J. Biol. Chem.* **262**, 9136–9140.
- Dang, C. V. & Lee, W. M. F. (1989) *J. Biol. Chem.* **264**, 18019–18023.
- Chelsky, D., Ralph, R. & Jonak, G. (1989) *Mol. Cell. Biol.* **9**, 2487–2492.
- Meier, U. T. & Blobel, G. (1992) *Cell* **70**, 127–138.
- Bowcock, A. M. & Cavalli-Sforza, L. L. (1991) *Genomics* **11**, 491–498.
- Garcia, J. A., Harrick, D., Pearson, L., Mitsuyasu, R. & Gaynor, R. B. (1988) *EMBO J.* **7**, 3143–3147.
- Dang, C. V. & Lee, W. M. F. (1988) *Mol. Cell. Biol.* **8**, 4048–4054.
- Hunt, C. & Morimoto, R. I. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6455–6459.
- Meggio, F. & Pinna, L. A. (1988) *Biochim. Biophys. Acta* **971**, 227–231.
- Pai, C.-Y., Chen, H.-K., Sheu, H.-L. & Yeh, N.-H. (1995) *J. Cell Sci.* **108**, 1911–1920.
- Cairns, C. & McStay, B. (1995) *J. Cell Sci.* **108**, 3339–3347.
- Lee, W.-C., Xue, Z. & Melese, T. (1991) *J. Cell Biol.* **113**, 1–12.
- Lee, W.-C., Zabetakis, D. & Melese, T. (1992) *Mol. Cell. Biol.* **12**, 3865–3871.
- Shan, X., Xue, Z. & Melese, T. (1994) *J. Cell Biol.* **126**, 853–862.
- Srivastava, M., McBride, O. W., Fleming, P. J., Pollard, H. B. & Burns, A. L. (1990) *J. Biol. Chem.* **265**, 14922–14931.
- Boring, L. F., Sinervo, B. & Shubiger, G. (1989) *Dev. Biol.* **132**, 343–354.
- Kern, S. E., Pietenpol, J. A., Thiagalingam, S., Seymour, A., Kinzler, K. W. & Vogelstein, B. (1992) *Science* **256**, 827–830.