A functional SNP in the promoter region of TCOF1 is associated with reduced gene expression and YY1 DNA–protein interaction

Cibele Masotti a, Lucia M. Armelin-Correa a, Alessandra Splendore b, Chin J. Lin c,d, Angela Barbosa c,e, Mari C. Sogayar f, Maria Rita Passos-Bueno a,*

a Instituto de Biociências, Universidade de São Paulo, Departamento de Biologia, Rua do Matão, 277/Sala 200, 05508-900 São Paulo/SP, Brazil
b Instituto Nacional do Câncer, Divisão de Genética, Rio de Janeiro, Brazil
c Laboratório de Hormônios e Genética Molecular LIM/42, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Departamento de Clínica Médica, São Paulo, Brazil
d Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Departamento de Patologia, São Paulo, Brazil
e Centro de Biotecnologia, Instituto Butantan, São Paulo, Brazil
f Instituto de Química, Universidade de São Paulo, Departamento de Bioquímica, São Paulo, Brazil

Received 20 April 2005; received in revised form 10 June 2005; accepted 22 June 2005
Available online 18 August 2005
Received by A.J. van Wijnen

Abstract

Treacher Collins syndrome (TCS) is an autosomal dominant craniofacial malformation caused by null mutations in the TCOF1 gene. High inter and intra familial clinical variability, ranging from mild malar hypoplasia to perinatal death due to airway collapse is observed, but, to date, no genotype–phenotype correlation has been reported. Considering haploinsufficiency as the molecular mechanism underlying the disease, we have hypothesized that mutations in the promoter region of the gene, which has never been previously characterized, in trans with a pathogenic mutation, could modulate the phenotype. Therefore, the aims of the present study were to determine the TCOF1 gene’s core promoter and to identify mutations in this region that could contribute to the phenotypic variation observed in this syndrome. We have delimitated the minimal promoter to a region of less than 150 bp, with 63% of identity among 5 different species. We screened 1.2 kbp of the TCOF15′ flanking sequence in the DNA obtained from 21 patients and 51 controls and identified four new single nucleotide polymorphisms (SNPs), one of which (−346C>T), was proved to be functional, as it decreased the promoter activity by 38%. Electrophoretic mobility shift assay (EMSA) analysis demonstrated that the −346T allele impairs DNA-binding to the YY1 transcription factor. This promoter variant represents a candidate allele to explain the clinical variability in patients bearing TCS.

© 2005 Elsevier B.V. All rights reserved.

Keywords: TCOF1 promoter; Clinical variability in Treacher Collins syndrome; Functional SNP; Modifier allele

1. Introduction

Treacher Collins syndrome (TCS; OMIM 154500) is a craniofacial disorder that affects the development of structures derived from the embryonic first and second branchial arches (Poswillo, 1975). TCS is inherited in an autosomal dominant pattern with 60% of the cases resulting from de novo mutations (Splendore et al., 2003). TCS has variable expressivity and high penetrance, since only one unequivocal case of non-penetration has been reported to date (Dixon et al., 1994). The syndrome
has an incidence of 1:50,000 live births and the common clinical features include down slanting palpebral fissures with lower eyelid coloboma, hypoplasia of the mandible and the zygomatic complex, malformed ears and conductive hearing loss due to atresia of the external ear canal (Gorlin et al., 2001).

The gene associated with the syndrome, TCOF1, contains 28 exons and several alternative splices (The Treacher Collins Collaborative Group, 1996; Edwards et al., 1997; So et al., 2004). The TCOF1 product, a nucleolar protein named treacle, plays a role in ribosome biogenesis and interacts with hNop56p in pre-ribosomal ribonucleoprotein complexes and also with the RNA polymerase I transcription factor UBF (Hayano et al., 2003; Valdez et al., 2004). More than 118 different pathogenic mutations were reported in the coding region of TCOF1, most of which are family-specific deletions or duplications causing a premature stop codon (Splendore et al., 2005). Absence of the truncated protein in fibroblasts from TCS patients suggests that null mutations in TCOF1 probably lead to nonsense mediated mRNA decay (Isaac et al., 2000). Therefore, haploinsufficiency has been proposed as the molecular mechanism underlying the disorder. Identification of families with no pathogenic mutation in the coding region of TCOF1 has suggested the possibility of genetic heterogeneity or the existence of different mechanisms leading to the syndrome (Isaac et al., 2000; Splendore et al., 2000, 2002, 2005).

TCS has a high intra and interfamilial phenotypic variation, ranging from perinatal death, due to compromised airways, to individuals that cannot be diagnosed on clinical grounds alone. Several attempts to evaluate the clinical variability in TCS have demonstrated no genotype–phenotype correlation in the syndrome (Edwards et al., 1997; Gladwin et al., 1996; Dixon, 1996). It has also been shown that clinical variability does not depend on the type or location of the mutation, sex or on whether the case is a sporadic or a familial one (Splendore et al., 2000; Teber et al., 2004). A recent study has demonstrated that the genetic background has a major effect on the phenotype of Tcof1+/− mice, suggesting that variations in other genes in the same or related pathways affect the function of the protein in different individuals (Dixon and Dixon, 2004).

We hypothesized that clinical variability can be determined by the degree of TCOF1 expression regulated by the wild-type allele. Several possible mechanisms could account for variations in treacle’s expression, including mutations in the promoter region of the TCOF1. In this regard, we have hypothesized that mutations in cis-acting elements of the wild-type allele could modulate the phenotype in TCS patients. In the present work, we aimed at delimitating the TCOF1 gene’s minimal promoter region, identifying functional SNPs in 1.2 kbp of the 5’ upstream region and evaluate any genotype–phenotype correlation in TCS patients.

2. Materials and methods

2.1. Patients and controls

Twenty-one Brazilian unrelated TCS patients with previously identified pathogenic mutation (Splendore et al., 2000, 2002, 2003, 2005) were screened for sequence variation in the promoter region of TCOF1. For each identified variation, at least 51 controls were screened. For the −346C>T SNP, 13 additional patients and 162 controls were included in the screening, since this SNP initially exhibited a higher frequency (>20%) in patients as compared to controls. We also screened three other patients (2 familial cases and 1 sporadic) for whom no pathogenic mutation in the coding region of the gene was detected through sequencing or Southern blot analysis (Splendore et al., 2005).

DNA was extracted according to standard techniques (Miller et al., 1988; Richards et al., 1993). The study protocols were previously approved by the ethical committee of our institution and informed consent was obtained from both patients and control subjects or from their legal tutors.

2.2. Mutation screening of the TCOF1 promoter region

The 5’ flanking sequence from the human TCOF1 gene was obtained from the Homo sapiens chromosome 5 genomic contig (GenBank accession no. NT_029289) after a BLASTn (http://www.ncbi.nlm.nih.gov/blast/) search, using the exon 1 sequence (NM_000356). Four overlapping segments (P1, P2, P3 and P4), corresponding to 1241 bp of the upstream sequence relative to the first methionine (NT_029289), were amplified with primers available upon request.

PCR products were analyzed through SSCP and dHPLC (Transgenicom). Samples showing mobility shifts or altered peaks were sequenced in both directions in an automated sequencer (MegaBACE 1000, GE Healthcare Bio-Sciences).

To verify whether the −1025G>C and −948G>A SNPs were in cis, PCR products were cloned with the TOPO TA™ Cloning Kit (Invitrogen Corporation). Once we observed that they were in linkage disequilibrium, we have screened only the −1025G>C transition in the control population, by PCR, followed by digestion with the HincII restriction enzyme (New England Biolabs). The −346C>T SNP was screened through Single Nucleotide Primer Extension (SNuPe™ Amersham Pharmacia Biotech) with the 5’-agatgagtaaaacgcagacc-3’ primer.

All the promoter mutations detected in this study were named relative to the start codon of the TCOF1 gene (http://www.genomic.unimelb.edu.au/mdi/mutnomen; Den Dunnen and Antonarakis, 2000). The +1 position is located 76 nucleotides downstream of the 5’ end of TCOF1 gene (position 109000247 in the reference sequence NT_029289,
TCOF1 from 10900171 to 10942791). Hardy-Weinberg equilibrium in patients and controls was verified by $\chi^2$ test at a 0.05 significance level.

2.3. Plasmid constructs

The luciferase reporter gene assays were performed using plasmid constructs bearing the TCOF1 gene 5’ flanking region: pGL3-1200wt (from −1241 to +56), pGL3-600wt (−571 to +56), pGL3-400wt (−369 to +56), pGL3-300wt (−258 to +56) and pGL3-200wt (−151 to +56) (Fig. 1). The 1200 bp (−1241 to +56 bp) and 600 bp (−571 to +56 bp) genomic DNA fragments were amplified from a DNA sample of a control subject with oligonucleotides containing 5’ terminal restriction sites to the enzymes BglII or HindIII. They were inserted into the pGL3-Basic vector (Promega), in either sense (pGL3-1200wt) or anti-sense orientations (pGL3-1200wtAS). The pGL3-1200wt plasmid was used to generate four 5’ deleted constructs either by digestion with SmaI (pGL3-300wt) or by PCR amplification using primers with a 5’ terminal restriction site for the HindIII enzyme (pGL3-600wt, pGL3-400wt and pGL3-200wt; primer sequences available upon request). The −346C>T variant was also chosen for functional studies (pGL3-600pol). The pGL3-600pol vector’s fragment was generated by PCR amplification of the same pGL3-600wt region, but using a patient’s DNA sample containing the T variant in position −346. All constructs were verified by direct sequencing.

2.4. Cell culture and transient transfection assays

Human immortalized hepatocytes (HepG2—HB8065—obtained from American Type Culture Collection) were grown in Dulbecco’s Modified Eagle’s Medium (DMEM) supplemented with 10% fetal bovine serum and 1% ampicillin/streptomycin, maintained in a 5% CO2 incubator at 37 °C. Twenty-four hours before transfection, cells were plated onto 12-well plates (1×10⁵/well) without antibiotics. Samples (1 μg) of each luciferase reporter plasmid were transfected, in duplicate wells, with Lipofectamine 2000 (Invitrogen), according to the manufacturer’s instructions. The pRL-TK plasmid (0.5 μg/well; Promega) was used to normalize the transfection efficiency. Cells were harvested 48 h after transfection, and the reporter gene expression was analyzed using the Dual-Luciferase Reporter Assay System (Promega) in a microplate luminometer (LB96V MicroLumat Plus—EG&Berthold). The TCOF1 gene promoter-fragments’ activities were normalized relative to the activity of the pGL3-1200wt vector and expressed as relative luciferase unit. Luciferase data were generated by pooling the results of at least three experiments performed in duplicate. Statistical significance (p-value) was determined using Kruskall-Wallis Test, followed by Dunn’s adjustment for comparing all vectors and Friedmann Test to compare the pGL3-600wt and 600 pol vectors. The significance level adopted was 0.05.

2.5. In silico analysis

We used the prediction web softwares TESS MASTER (“TESS: Transcription Element Search Software on the WWW”, http://www.cbil.upenn.edu/tess) and TFSEARCH (“TFSEARCH: Searching Transcription Factor Binding Sites”, http://www.rwcp.or.jp/papia/) to search for promoter elements. Alignment of the TCOF1 promoter sequences was performed using the UCSC (University of California-Santa Cruz Genome Bioinformatics, http://genome.cse.ucsc.edu/) Genome Browser with the following assemblies: human May 2004, chimpanzee Nov 2003, mouse May 2004, rat Jun 2003 and dog Jul 2004.

2.6. EMSA

Nuclear extracts from HepG2 cells were prepared as described (Read, 1998) and the protein concentration was determined using the Bradford’s method. Two 51 bp double-
stranded DNA probes, spanning from position −367 to −317, a wild-type and another one containing the −346C>T SNP, were prepared by annealing the desired sense and anti-sense oligonucleotides. Before annealing, the sense oligonucleotides were 5’-end-labeled with gamma-[32P] dATP. The protein binding reactions were carried out in 20 or 25 µL of buffer (20 mM Hepes pH 7.9, 80 mM KCl, 5 mM MgCl₂, 2% (v/v) Ficoll, 5% (v/v) Glycerol, 0.1 mM EDTA, 1 mM Dithiothreitol) containing 0.5–1 × 10⁶ cpm of labeled probe, 1 µg of double-stranded polyribonucleotide Poly(dI-dC), 50 µg of bovine serum albumin and 5 µg of HepG2 nuclear protein. As a non-specific competitor, we used a fragment derived from the Cytochrome P450 Family 17 (CYP17) gene promoter. The binding reaction mixture was pre-incubated with mouse monoclonal antibody to human Sp1, or rabbit polyclonal antibodies to human Sp3 or YY1 (Santa Cruz Biotechnology, Inc.; sc-420, sc-644 and sc-1703, respectively). The protein–DNA complexes were resolved by electrophoresis (6% non-denaturing polyacrylamide gel) in 0.5 TBE running buffer (44.58 mM Tris Base, 44.58 mM Boric Acid, 1.25 mM Na₂EDTA H₂O). The gels were exposed onto a phos-

### Table 1

Frequencies of the −346C>T, −1025G>C and −948G>A SNPs

<table>
<thead>
<tr>
<th>SNP</th>
<th>−346C&gt;T</th>
<th>−1025G&gt;C</th>
<th>−948G&gt;A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>CC</td>
<td>CT</td>
<td>TT</td>
</tr>
<tr>
<td>Patients</td>
<td>n f</td>
<td>n f</td>
<td>n f</td>
</tr>
<tr>
<td></td>
<td>28 0.824</td>
<td>6 0.176</td>
<td>0 0</td>
</tr>
<tr>
<td>Controls</td>
<td>186 0.873</td>
<td>25 0.117</td>
<td>2 0.009</td>
</tr>
<tr>
<td>Total</td>
<td>214 0.886</td>
<td>31 0.125</td>
<td>2 0.009</td>
</tr>
</tbody>
</table>

Fig. 2. A: Alignment of (A) the minimal promoter region (−152 to +56) and (B) the −346C>T SNP region (−367 to −317) of the TCOF1 gene: human (chromosome 5), chimpanzee (chromosome 4), mouse (chromosome 18), rat (chromosome 18) and dog (chromosome 4) candidate sequences. Coding sequence (part of exon 1) is represented by capital letters. The arrows mark the transcription start site (the 5’ end of mRNA from NM 00356 was considered the transcription start site), the conserved functional SNP and the YY1 consensus binding site (anti-sense strand). Asterisks represent conserved nucleotides among the five species.
phoscreen and the autoradiography captured using a STORM 860 scanner (GE Healthcare Bio-Sciences) and analyzed with the Image Quant® (5.2) software.

3. Results

3.1. Identification of SNPs in the TCOF1 gene promoter region

Screening of 1.2 kbp of the TCOF1 promoter sequence in 21 TCS patients and 51 controls revealed the presence of 5 SNPs. Two of these SNPs were identified only in the control group (−106A>C and −332T>C), with a frequency lower than 2%. The three others (−948G>A, −1025G>C and −346C>T), with a frequency higher than 10% and in Hardy–Weinberg equilibrium (p > 0.05; Table 1), were identified in both patients and controls and with no statistical difference between the groups (p > 0.05; Table 1). The −106A>C SNP was the only one previously described (dbSNP/GenBank accession number rs3806951). All individuals who displayed the −948G>A were also heterozygous for the −1025G>C SNP, suggesting that they are in linkage disequilibrium (LD). Indeed, we observed that both variations were in the same chromosome. On the other hand, the −346C>T variant was not found in LD with the −948G>A and −1025G>C SNPs.

The three patients with no detectable pathogenic mutation in the coding region of TCOF1 did revealed no alterations in the promoter region by direct sequencing.

3.2. Species-conserved SNPs in the proximal TCOF1 promoter

Computational analysis and visual inspection of the 1.2 kbp TCOF1’s promoter region revealed the absence of the TATAA-box element, and no initiator or downstream promoter element consensus sequence. In silico analysis also indicated more than 15 consensus sites for the ubiquitous Sp1 transcription factor and a high G/C content (75%) in the segment from position −200 bp to the first ATG. We also verified that the region from −152 to −1 bp shares 63% identity among the 5 species studied (Fig. 2). The −332T>C change is conserved only between human and chimpanzee, while the other four SNPs are located in conserved nucleotides in all species analyzed (human, mouse, rat, dog and chimpanzee; Fig. 2 and data not shown).

3.3. Functional analysis

All the vectors tested, except for the negative control pGL3-1200AS, showed a significantly higher (p<0.05) activity than the promoter-less pGL3-Basic vector (p<0.05;
Fig. 4. The \(-346C>T\) SNP and nuclear protein affinity to the \(TCOF1\) promoter. A: Gel shift evidencing the DNA–protein specific complexes 1 and 2 for wild-type (WT) and polymorphic (POL) probes: lanes 1 and 10, free labeled probes; 2 and 11, labeled probes and HepG2 nuclear extracts (NE); 3, 4 and 5, WT labeled probe, NE and specific competitor (WT non-labeled probe) 10\(\times\), 30\(\times\) and 100\(\times\), respectively; 6, NE and 100\(\times\) non-specific competitor; 7 and 8, WT labeled probe, NE and antibodies to Sp1 and Sp3, respectively; 9 and 12, WT and POL labeled probes, NE and antibody to YY1 (abrogation of Complex 2). B: Competition gel shift (WT > POL). Lanes 1 and 8, free labeled probes; 2 and 9, labeled probes and NE; 3 to 6 and 10 to 13, indicated specific competitor in increasing concentrations; 7 and 14, labeled probes, NE and 50\(\times\) non-specific competitor. Complex 2 generated by wild-type probe persisted even in the presence of 30\(\times\) molar excess of non-labeled polymorphic competitor. By comparison, as few as 3\(\times\) molar excess of wild-type competitor was able to inhibit the polymorphic probe-generated complex 2. C: quantification of complexes 1 and 2 (YY1) in each concentration of specific competitor. Probes average intensity was calculated based in three competition gel shifts and considered as the sum above background.
The shortest construct, pGL3-200wt, displayed an activity 150% higher than that of the pGL3-1200wt construct. Since no statistical difference was observed between the transcriptional activities of pGL3-200wt and 300wt vectors, the 200 bp fragment presumably contains the minimal promoter. The pGL3-200wt’s insert contains 207 bp, of which 76 bp are not transcribed, 75 bp belong to the 5’ untranslated region and 56 bp are part of exon 1 (Fig. 1).

Comparing all constructs with the pGL3-1200wt vector, pGL3-600wt is not significantly different, but pGL3-400wt, pGL3-300wt and pGL3-200wt present higher activity (p<0.05; Fig. 3A). These data suggest that the region between −1200 and −400 might comprise a repressor element. Comparison of the polymorphic 600 bp construct (pGL3-600pol) with its wild-type counterpart evidenced that the −346C>T SNP is able to significantly decrease the reporter gene transcription by 38% (p<0.05; Fig. 3B).

**3.4. Impact of −346C>T SNP in DNA–protein interaction**

EMSA evidenced the formation of two specific DNA–protein complexes (Fig. 4A) for both wild-type and polymorphic probes (Fig. 4A). A third complex was shown to be non-specific, since co-incubation with a non-specific probe in a molar excess of 100× abolished its formation (Fig. 4A, lane 6). Addition of antibody against YY1 led to abrogation of only complex 2 in both the wild-type and the polymorphic probes (Fig. 4A, lanes 9 and 12), suggesting that YY1 participates in complex 2 formation. Western blot analysis showed that HepG2 cells express the YY1 protein (data not shown). Sp1 and Sp3 do not seem to be involved in the formation of complexes 1 and 2 (Fig. 4A, lanes 7 and 8).

In the EMSA experiments, the complexes formed with the polymorphic probe were consistently less intense than the wild-type ones, suggesting that the −346C>T SNP affects nuclear protein binding. Through competition assays between the two probes (Fig. 4B) we observed that complexes 1 and 2 display less intense bands in the presence of both wild-type and polymorphic competitors (4B, lanes 3–6 and 10–13). In comparison with complex 1, we also observed that complex 2 requires higher concentrations of competitors to diminish its band intensity (Fig. 4C). These data indicate that the complexes contain sequence-specific DNA-binding proteins and that the −346C>T SNP resides in a functionally important nucleotide, perhaps more relevant for complex 2.

**3.5. −346C>T segregation in TCS Brazilian patients**

We were able to verify that the pathogenic mutation was in cis with the −346C>T SNP in one familial case (family II; Fig. 5) and in trans in two familial cases (families I and III, Fig. 5). We observed that in these three genealogies that the affected heterozygous mother always had a milder phenotype than her offsprings, who presented abnormal external ears and hearing impairment, besides hypoplasia of the mandible and the zygomatic complex. The −346T>C SNP was also present in six other cases with a de novo pathogenic mutation, for which we were unable to determine its origin (Splendore et al., 2003); consequently, we were not able to establish the phase of the SNP in relation of the pathogenic mutation. It is of note that all of these sporadic cases also present a severe phenotype.

**4. Discussion**

This report represents the first analysis of the promoter region of the TCOF1 gene, with identification of a novel functional SNP in the regulatory region. We have delimited the TCOF1 minimal promoter to a region of 151 bp, which contains more than 15 potential Sp1 binding sites and lacks the consensus TATAA-box element.
The absence of a pathogenic mutation in 1200 bp of the 5'flank sequence in three patients without mutations in the coding region of TCOF1 further supports genetic heterogeneity of TCS. However, we cannot rule out that the cause of the disease in these patients is due to mutations in distant regulatory regions, as has been shown for preaxial polydactyly and for lactase persistence (Lettice et al., 2003; Olds and Sibley, 2003). We have detected three SNPs in the promoter region of TCS and we have also demonstrated that the −346C>T variant is functional. The −346T allele decreased by 38% the reporter gene transcription, which might have important physiological effects (Yan et al., 2002). This SNP is interfering with YY1 and other transcription factors’ anchorage and stability. Although located 11 bp way from the YY1 consensus sequence, it is possible that the decrease in TCOF1 promoter activity caused by the −346T allele is caused by a YY1’s reduced DNA-binding affinity to this region. YY1 is a transcription factor with four zinc-finger motifs and functions as a repressor, activator or initiator of transcription in promoters lacking the consensus TATAAA-box element (OMIM *600013; Smale, 1997). Numerous in vitro and in vivo studies have suggested that YY1 controls expression of developmentally regulated genes, including those involved in early mouse craniofacial and limb morphogenesis (Donohoe et al., 1999; Tan et al., 2002). These data together with our in vitro results suggest that YY1 may play a major role in TCOF1 regulation and, in this case, the SNP −346C>T might have important implications in the transcription levels of this gene.

It is noteworthy that the individuals with the functional SNP in trans with the pathogenic mutation have a mild phenotype, although the sample is small to yield any genotype–phenotype correlation. We must consider that the phenotype variation might reflect the underlying complexity of haplotypic structure involved in clinical variability. In addition, we cannot exclude that gene expression can be modulated by stochastic or environmental effects. Retinoic acid, a well-known teratogen, had been demonstrated to generate a TCS-like phenotype in mice as well as other craniofacial anomalies (Poswillo, 1975; Sulik et al., 1987). It is also important to observe that all of our propositi present a typical and severe form of TCS, possibly due to ascertainment bias. Therefore, considering the rarity of the disease, the relatively low frequency of the −346C>T SNP and the ascertainment bias, other approaches will be necessary to assess the effect of this polymorphism in the TCS phenotype.

Variability of the transcriptional level through genetic and environmental factors of the wild allele in carriers of autosomal dominant mutations, particularly those leading to haploinsufficiency of the protein product, can be a general mechanism to explain phenotypic variation in Mendelian disorders. Identification of a functional SNP in cis-regulatory elements and delimitation of the minimal TCOF1 promoter constitute the first step in the elucidation of the complex mechanisms involved in the modulation of the TCS phenotype.

Acknowledgments

We would like to thank Dr. Mendonça for the use of her laboratory facilities, Dr. Novak and Dr. Bachega for Sp3 and Sp1 antibodies. We would also like to thank Oscar T. Suzuki and Sheila M. B. Winnischofer for helpful discussions, Constância Gotto for secretarial assistance. This work was supported by grants from Fundação de Amparo a Pesquisa de São Paulo (FAPESP) and Conselho Nacional de Pesquisa (CNPq).

References


