Cibele Masotti

Estudo de Mecanismos Regulatórios e Mapeamento de
Genes Associados a Malformações Craniofaciais

Instituto de Biociências
Universidade de São Paulo

São Paulo

2007

Cibele Masotti

Estudo de Mecanismos Regulatórios e Mapeamento de Genes Associados a Malformações Craniofaciais

Tese apresentada ao Instituto de
Biociências da Universidade de São
Paulo, para a obtenção de Título de
Doutor em Ciências, na Área de
Biologia/Genética.

Orientadora: Maria Rita dos Santos
Passos Bueno

São Paulo

2007

Ficha Catalográfica

Masotti, Cibeles
Estudo de Mecanismos Regulatórios e
Mapeamento de Genes Associados a
Malformações Craniofaciais
Número de páginas

Tese (Doutorado) - Instituto de
Biociências da Universidade de São Paulo.
Departamento de Biologia/Genética.

1. Síndrome de Treacher Collins 2.
Síndrome Aurículo-Condilar 3. *TCOF1* 1.
Universidade de São Paulo. Instituto de
Biociências. Departamento de Genética e
Biologia Evolutiva

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a). Maria Rita dos Santos Passo-Bueno
Orientador(a)

Aos meus pais e ao meu irmão,
Lezy, Leonardo e Gustavo.

Sócrates

Depois da leitura de um livro sobre a história da filosofia, o Senhor K. expressou-se desfavoravelmente sobre as tentativas dos filósofos de apresentar as coisas como fundamentalmente incognoscíveis. "Quando os sofistas afirmavam muito saber sem nada ter estudado", disse ele, "apresentou-se o sofista Sócrates com a arrogante assertiva de que sabia que nada sabia. Era de se esperar que acrescentasse a sua sentença: pois também eu não estudei nada. (Para saber algo, nós precisamos estudar). Mas ele parece que não seguiu falando e, provavelmente, os estrondosos aplausos que estalaram assim que pronunciou sua sentença e que duraram dois mil anos, teria afogado qualquer sentença que a seguisse.

Histórias do senhor Keuner
Bertolt Brecht

AGRADECIMENTOS

À minha orientadora Maria Rita Passos Bueno, que tanto admiro. Rita, obrigada pela confiança, pelo incentivo, pelo exemplo e pelo aprendizado. Sua paixão pela ciência é contagiante, seu pragmatismo é uma lição.

Aos colegas de laboratório que já passaram, que ainda passam, que já se vão: Dra. Andréia Sertié, Aline, Beatriz Burim, Camila, Dr. Carlos Maranduba, Cristian Corio, Dra. Daniela Bueno, Dra. Dinamar Gaspar, Dra. Eloísa Moreira, Érika Kague Neném, Érika Yeh, Dra. Fernanda Jehee, Dra. Flávia Errera, Gérson, Guilherme Orabonna, Gustavo de Alencastro, Karina de Oliveira, Kelly Bagatini, Mayara Ribeiro, Nélcio Oliveira e Roberto Fanganiello. Agradeço pela ajuda incondicional de todas as horas, pelo aprendizado, pelo carinho, pelos incentivos, pelas risadas dentro e fora da USP.

À Alessandra Splendore, quem felizmente me iniciou na biologia molecular e na *TCOF1*.

À Lucia Armelin-Corrêa e ao Oscar Suzuki em especial. Eles me ensinaram muito, participaram de quase todas as etapas do meu trabalho, ajudando na elaboração das hipóteses, na discussão dos dados e nas dificuldades técnicas. Estiveram presente nos momentos que mais precisei, da análise estatística às prateleiras pregadas na minha casa.

Aos colegas do Instituto de Biociências: laboratório da Dra. Mayana Zatz (em especial Constância, Toninha, Manuela, Agnes, Miguel e Éder), Centro de Estudos do Genoma Humano (em especial Kátia Rocha e Roberto Rivelino), laboratório da Dra. Cristina Miyaki (em especial Camila Ribas e Érika Sendra), laboratório da Dra. Célia Koiffmann (em especial Carla, Cris, Mônica, Rose e Ilana), laboratório da Dra. Regina Mingroni, laboratório do Dr. Luís Soares Netto, laboratório da Dra. Marie-Anne Van Sluys, Dr. Paulo Otto, Dr. Eduardo Gorab e Dra. Luciana Haddad.

Aos amigos e colaboradores Dra. Ângela Barbosa e Dr. Chin Jia Lin, que contribuíram imensamente para o meu trabalho. Obrigada pela amizade e pela ajuda de sempre.

À Dra. Mari Sogayar pela disponibilidade do laboratório e de seus alunos para ensinar a técnica de transfecção.

Ao Dr. Michael Lovett pela oportunidade do doutorado-sanduíche em seu laboratório na Washington University in Saint Louis, pelo aprendizado imenso que veio desta experiência. Agradeço ao seu grupo de trabalho, em especial Rose Veile.

À Dra. Anamaria Aranha Camargo, que me abriu as portas do seu laboratório. Agradeço imensamente pela ajuda nessa etapa final do doutorado, pelo interesse, pela confiança e pela oportunidade de trabalhar no seu grupo. Aos colegas do LMBG-Instituto Ludwig, meu segundo laboratório, muito obrigada pela acolhida tão calorosa.

À Fapesp pelo apoio financeiro e à CAPES pela bolsa PDEE de doutorado sanduíche.

Às amigas de Saint Louis, Kalinka Valar, Blima Fux e Maristela Hernandez, pelos momentos inesquecíveis e pelo suporte todo.

Aos amigos e colegas de UFRGS, que viram o meu interesse pela genética florescer e que muito me incentivaram para seguir pós-graduação na área, Renata Medina, Maurício Lehmann e Leonardo Karam.

Aos meus amigos longe dos olhos, mas perto do coração: Maíra Rocha, Clarice Madruga, Tiago Bernardon, Liu Yi Nan e Gislaine Marins.

À minha família “sul-paulistana”, que, mesmo indo e vindo de São Paulo, mesmo não entendendo nada de síndromes nem de moléculas, esteve presente em todas as etapas desse trabalho, na alegria e na tristeza da vida de dentro e de fora do doutorado: Andréia Zwetsch, Beatriz Santomauro, Cláudia Guzzo, Daniel Lorenzini, Fabiane Satler, Fabrício Brasiliense, Gerusa Mengarda, Janaína Augustin, Marcelo de Andrade, Renata Medina e Tatiana Cyro Costa.

À minha nova família paulistana, que tanto me apoiou nesta etapa final do trabalho, que tanto me preenche, dada a saudade da família no sul. Regina, Luiz e Aninha, obrigada pelo carinho.

À minha família de Porto Alegre, muito obrigada por tudo, principalmente pelo incentivo, pela compreensão e pelas recepções calorosas nas minhas idas “de volta pra casa”.

Aos meus pais e ao meu irmão, a quem dedico esta tese com todo o meu amor.

Ao Diogo, a quem dedico meu amor.

Índice

Capítulo 1:	
INTRODUÇÃO.....	3
1.1 Aspectos Embriológicos	
1.2 Síndromes de primeiro e segundo arcos faríngeos	
1.3 Síndrome Aurículo-Condilar	
1.4 Síndrome de Treacher Collins	
1.5 Questões não resolvidas	
1.6 Objetivos	
Capítulo 2:	
CASUÍSTICA E MÉTODOS	26
2.1 Pacientes e controles	
2.2 Considerações éticas	
2.3 Métodos	
Capítulo 3:	
Prefácio.....	47
3-I: TCOF1 MUTATION DATABASE: NOVEL MUTATION IN THE NEWLY DESCRIBED EXON 6A AND UPDATE IN MUTATION NOMENCLATURE.....	4
8	
3-II: NOVAS MUTAÇÕES NO GENE TCOF1 E CORRELAÇÃO GENÓTIPO-FENÓTIPO NA SÍNDROME DE TREACHER COLLINS.....	55
Capítulo 4:	
A FUNCTIONAL SNP IN THE PROMOTER REGION OF TCOF1 IS ASSOCIATED WITH REDUCED GENE EXPRESSION AND YY1 DNA-PROTEIN INTERACTION.....	83
Capítulo 5:	
ESTUDO DE SEQUÊNCIAS E PROTEÍNAS POTENCIALMENTE ENVOLVIDAS NA REGULAÇÃO DO GENE TCOF1.....	93
Capítulo 6:	
TCOF1 EXPRESSION LEVELS ARE DIFFERENT BETWEEN TCS PATIENTS AND CONTROLS, BUT ARE NOT ASSOCIATED TO PHENOTYPIC VARIATION AND CPG ISLAND METHYLATION STATUS.....	114
Capítulo 7:	
AURICULO-CONDYLAR SYNDROME: MAPPING OF A FIRST LOCUS AND GENETIC HETEROGENEITY.....	133

Capítulo 8:	
DISCUSSÃO GERAL E CONCLUSÕES.....	146

Referências Bibliográficas.....	153
--	------------

Anexos

- I. Termo de consentimento livre e esclarecido (maiores de 18 anos)
- II. Mapas dos vetores utilizados neste trabalho

Resumo

Neste trabalho, estudamos duas malformações craniofaciais mendelianas, decorrentes de um distúrbio do desenvolvimento dos primeiro e segundo arcos faríngeos: a síndrome de Treacher Collins (STC) e a síndrome Aurículo-condilar (SAC). A identificação de genes e de mecanismos moleculares associados a essas condições, além de contribuir para a compreensão do desenvolvimento das estruturas derivadas desses arcos faríngeos, é fundamental para o desenvolvimento do diagnóstico molecular, uma ferramenta importante para diagnóstico diferencial e aconselhamento genético.

Contribuímos para uma melhor caracterização clínica da SAC com a descrição de uma nova família com 11 afetados. Após excluirmos quatro genes/regiões candidatas para síndromes de 1º e 2º arcos faríngeos, realizamos estudos de ligação usando marcadores polimórficos ao longo do genoma. Mapeamos o primeiro locus associado à SAC, 1p21.1-q23.3 (*lod score*=3.0), e nossos dados sugerem que há heterogeneidade genética para essa patologia.

Com relação ao estudo da STC, realizamos uma extensa revisão da nomenclatura das mutações patogênicas descritas na literatura, além de investigar mecanismos mutacionais atípicos na STC, corroborando a hipótese de que mutações nos exons que sofrem *splicing* alternativo são capazes de gerar o fenótipo da síndrome. Demos continuidade à caracterização do espectro de mutações no gene *TCOF1* e investigamos a correlação genótipo-fenótipo numa amostra de 58 pacientes com STC. A análise dos dados de polimorfismos da região codificadora do gene permitiu que fizéssemos inferências sobre o regime de seleção ao qual o *TCOF1* está submetido, e os resultados sugeriram que o gene *TCOF1* está sob seleção purificadora, que atua sobre mutações fracamente deletérias. Também inferimos a fase para um conjunto de polimorfismos da região codificadora, verificando se havia associação de algum haplótipo à gravidade do quadro clínico ou à predisposição para a doença. Dada a observação de ausência de correlação haplótipo/genótipo-fenótipo, testamos a hipótese de que variações nos níveis de expressão do alelo normal poderiam ser responsáveis pela variabilidade clínica observada nos pacientes portadores da STC. Para tanto, iniciamos o estudo funcional das regiões regulatórias do gene *TCOF1*, inclusive, de regiões distantes do promotor mínimo, preditas como *enhancers*. Identificamos polimorfismos em sua região promotora, sendo um deles capaz de diminuir os níveis de transcrição e de afetar a ligação do fator de transcrição *YY1* ao promotor do *TCOF1*. Caracterizamos a ação de *YY1* como repressora *in vitro*. Testamos também a hipótese de haploinsuficiência como mecanismo associado à STC. Quantificamos os níveis de transcritos do *TCOF1* em indivíduos afetados e normais e observamos uma diferença significativa. Nosso trabalho corrobora a hipótese de haploinsuficiência e mostra pela primeira vez que pacientes têm degradação de transcritos. Também investigamos a possibilidade de os níveis de transcritos estarem correlacionados à variabilidade fenotípica. Comparamos os dados de expressão de cada paciente à frequência de nove sinais clínicos principais da STC, mas nenhuma correlação foi observada. Investigamos o padrão de metilação da ilha CpG do gene *TCOF1* em pacientes e em controles, com o intuito de verificar se diferentes níveis de metilação inter-individual estariam associados à grande variação na expressão gênica. Demonstramos que a metilação da ilha CpG não é o mecanismo regulatório por trás dessa ampla variação de expressão do *TCOF1*.

Summary

In the present study, we investigate two craniofacial mendelian disorders, resulting from abnormalities in the development of the first and second pharyngeal arches: the Treacher Collins Syndrome (TCS) and the auriculo condylar syndrome (ACS). The identification of genes and molecular mechanisms associated to these conditions, in addition to contributing to the understanding of the development of structures derived from these pharyngeal arches, is fundamental for the development of molecular diagnostics, an important tool for differential diagnosis and genetic counseling.

We contributed to a better clinical characterization of ACS with a description of a new family with 11 affected individuals. After excluding four candidate genes/regions for syndromes of the 1st and 2nd pharyngeal arches, we carried out linkage studies using polymorphic markers throughout the genome. We mapped the first locus associated to ACS, 1p21.1-q23.3 (*lod score*=3.0), and our data suggest genetic heterogeneity exists for this pathology.

With respect to the study of TCS, we carried out an extensive review of the nomenclature for the pathogenic mutations described in the literature, in addition to investigating atypical mutation mechanisms in TCS, corroborating the hypothesis that mutations in the exons that undergo alternative splicing can result in the TCS phenotype. We continued the characterization of the mutation spectrum for *TCOF1* and we investigated the genotype-phenotype correlation in a sample of 58 patients with TCS. The analysis of polymorphisms in the coding region allowed us to make inferences about the selective regime experienced by *TCOF1*, and our results suggested that *TCOF1* is under purifying selection, which acts upon weakly deleterious mutations. We also inferred the phase for a set of polymorphisms in the coding region, testing whether there was association between any haplotype and the severity of the phenotype or the susceptibility to the disease. Given the observation of no correlation between haplotype/genotype and phenotype, we tested the hypothesis that variation in the levels of expression of the normal allele could be responsible for the clinical variability observed in TCS patients. To do this, we started a functional study of the *TCOF1* regulatory regions, including regions distant from the minimal promoter, predicted to be enhancers. We identified polymorphisms in the promoter region, one of which reduced the levels of transcription and affected the binding of the YY1 transcription factor to the *TCOF1* promoter. Using an *in vitro* assay we characterized YY1 as a repressor. We also tested the hypothesis of haploinsufficiency as a mechanism associated to TCS. We quantified the levels of *TCOF1* transcripts in normal and affected individuals and found a significant difference. Our study corroborates the haploinsufficiency hypothesis and for the first time shows that patients have transcript degradation. We also investigated the possibility that transcripts levels are correlated to phenotypic variability. We compared the expression data for each patient with the frequency of nine clinical signs of TCS, but no correlation was found. We investigated the pattern of methylation on the CpG island of *TCOF1* in patients and controls, in order to test whether different levels of methylation among individuals were associated to the great variation in gene expression. We demonstrated that methylation of the CpG island is not the regulatory mechanisms underlying the broad variation in *TCOF1* expression.

INTRODUÇÃO

“O desenvolvimento facial envolve uma série complexa de eventos que, quando perturbados, resultam em uma ampla variedade de anomalias craniofaciais. Tais transtornos são extremamente preocupantes e estão entre as malformações congênitas mais comuns que afetam a humanidade. Existem numerosas anomalias craniofaciais e, cada uma delas, fornece uma oportunidade única para estudar um aspecto particular do desenvolvimento embrionário humano” (Gorlin *et al.*, 1990).

Este trabalho se refere ao estudo de duas malformações craniofaciais mendelianas, decorrentes de distúrbios do desenvolvimento embrionário, a síndrome de Treacher Collins e a síndrome Aurículo-Condilar (Figura 1). Primeiramente, descreveremos o desenvolvimento normal das estruturas embrionárias que darão origem às estruturas craniofaciais afetadas nessas patologias. Em seguida, apresentaremos uma breve descrição dos achados mais recentes da literatura e as questões que ainda permeiam o estudo dessas malformações.

1.1 Aspectos Embriológicos

À medida que o embrião se dobra em função do crescimento lateral e cranio-caudal, formam-se seis arcos faríngeos, estruturas revestidas pelo ectoderma e cujo tecido mesenquimal é formado principalmente por uma população de células que migram da crista neural. Num estágio anterior à formação completa desses arcos (hE20), na porção dorsal do embrião, ocorre um espessamento do ectoderma, formando a placa neural. A placa neural sofre um dobramento e as extremidades desse dobramento, chamadas cristas neurais, possuem uma população especial de células. Após o fechamento do tubo neural, as células da crista migram e passam a constituir o mesênquima dos arcos (Figura 2).

Existe uma subdivisão das células da crista neural, de acordo com sua origem cranio-caudal: tronco, circunfaríngeas e cranial. As células que povoarão os arcos faríngeos 1, 2 e 3 provêm da população cranial, partindo dos rombômeros 1 a 6 para os arcos (Figura 3a). Essas células, então, formarão as principais estruturas ósseas e epiteliais da face (Figura 3b).

Além das estruturas craniofaciais, as demais populações de células da crista neural darão origem a muitas outras estruturas do indivíduo adulto. Por isso, muitas vezes são chamadas de “quarto” folheto embrionário (Larsen, 1997; Carlson, 2005).

Diversas estruturas craniofaciais têm origem nos dois primeiros arcos faríngeos (Figura 4). A formação da orelha externa é um exemplo da interação precisa desses arcos. Por volta da quinta semana de gestação, massas nodulares de mesênquima se organizam próximo à primeira fenda faríngea. Esses nódulos crescem e se fusionam para formar a orelha externa. A orelha se localiza mais caudal em relação à posição definitiva. À medida que a mandíbula cresce, a orelha externa se desloca para uma posição mais cranial e vertical (Figura 5). Havendo qualquer perturbação no desenvolvimento desses dois arcos, é possível que observemos um reflexo no nas orelhas (Carlson, 2005; Embryo images, http://www.med.unc.edu/embryo_images/).

1.2 Síndromes de primeiro e segundo arcos faríngeos

Chamamos de síndromes de primeiro e segundo arcos faríngeos as malformações congênitas de estruturas derivadas desses arcos embrionários. A patogênese dessas doenças está associada a um defeito na formação dos arcos faríngeos, possivelmente envolvendo a migração e diferenciação das células da crista neural (Gorlin *et al.*, 2001). Há muitas síndromes descritas, e abordaremos somente alguma delas, que geralmente são consideradas para diagnóstico diferencial da síndrome de Treacher Collins e da síndrome Aurículo-Condilar.

A síndrome de Goldenhar, também conhecida como microsomia hemifacial, ou displasia óculo-aurículo-vertebral, ou espectro óculo-aurículo-vertebral (*OAV spectrum*), ocorre na frequência de 1:5.600 nascimentos. Os pacientes portadores desse espectro geralmente apresentam lipodermóide epibulbar, coloboma da pálpebra superior, apêndices pré-auriculares, micrognatia e anomalias vertebrais, sendo 65% dos casos unilaterais. A maioria dos casos é esporádica e existem raros casos familiares. Possivelmente há heterogeneidade genética e o risco empírico de recorrência é estimado de 2 a 3% (OMIM 164210; Gorlin *et al.*, 2001).

A síndrome de Nager, também conhecida como disostose acrofacial, apresenta hipoplasia maxilar e mandibular grave, com uma característica bastante importante que é anomalia de membros proximais. Essa síndrome é muito rara e possui padrão de herança

indefinido, uma vez que existem casos relatados com sugestão de herança autossômica dominante e recessiva. Possivelmente há heterogeneidade genética, pois vários locos (3p14, 9q32, 1q12-21.1) já foram associados à doença devido a anomalias cromossômicas (OMIM 154400; Gorlin *et al.*, 2001).

A síndrome de Miller também apresenta hipoplasia maxilar e mandibular, orelhas malformadas e anomalia de membros. Diferente da síndrome de Nager, as alterações de membros são pós-axiais, envolvendo agenesia ou hipoplasia do dedo mínimo. Apresenta possível herança autossômica recessiva e até hoje nenhum gene foi identificado (OMIM 263750; Gorlin *et al.*, 2001).

A Disostose maxilofacial (OMIM 155000) é uma condição caracterizada por fendas palpebrais em inclinação anti-mongolóide e encurtamento ântero-posterior da maxila com padrão de herança autossômica dominante. Há atraso na aquisição da linguagem e disartria, mas não há malformação dos pavilhões auriculares (Gorlin *et al.*, 2001).

A síndrome de Townes-Brocks (OMIM 107480), embora não seja uma patologia de primeiro e segundo arcos faríngeos, é considerada como diagnóstico diferencial para essas condições por possuir expressividade bastante variável, sendo que os casos mais leves podem apresentar somente alterações craniofaciais. A síndrome consiste de ânus atresico e mal posicionado, fístulas e fossetas pré-auriculares, surdez condutiva e neurosensorial, assimetria facial, micrognatia, macrostomia, orelhas e dedos malformados, anomalias renais e cardíacas. Johnson *et al.* (1996) descreveram uma família que havia sido previamente diagnosticada como portadora da síndrome de Goldenhar, mas o nascimento de um indivíduo com o quadro completo da síndrome de Townes-Brocks redirecionou o diagnóstico.

1.3 Síndrome Aurículo-condilar

A síndrome Aurículo-condilar (SAC, OMIM 602483) é uma síndrome de primeiro e segundo arcos faríngeos, tendo sido primeiramente relatada por Jampol *et al.* em 1998. Descrevendo uma família com vários afetados em cinco gerações, esses autores caracterizaram a condição pela presença de orelhas proeminentes e malformadas, hipoplasia mandibular e malformação da articulação têmporo-mandibular e dos côndilos mandibulares. A deformidade das orelhas é bastante característica: há uma fenda que divide a hélice superior do lóbulo, dando uma aparência de “ponto de interrogação” (Figura 1). Essa malformação, conhecida como “*question mark ear*”, já havia sido descrita por diversos autores como uma

deformidade distinta e restrita às orelhas em casos isolados e familiares (revisão em Guion-Almeida, 2002).

Os dois casos familiares descritos (Jampol *et al.* 1998 e Guion-Almeida 2002) sugerem um padrão de herança compatível com autossômico dominante. A penetrância é completa, mas há grande variabilidade fenotípica; esta é exemplificada por um propósito que possuía o côndilo mandibular mais gravemente afetado do que o de seu pai (Jampol *et al.*, 1998). Além disso, existem achados menos freqüentes na síndrome, como fenda palatina, ptose, língua hipoplásica, dentre outras alterações (Storm *et al.* 2005). Priolo *et al.* chegaram a sugerir que o caso relatado em 2000 pudesse ser uma entidade distinta, pois encontraram ptose, hipotonia e atraso no desenvolvimento. Da mesma forma, Erlich *et al.* (2000) descreveram um caso familiar de malformação das orelhas e hipoplasia mandibular como “complexo da disgnatia” (*dysgnathia*). No entanto, tendo em vista a semelhança clínica de todos esses casos relatados, Guion-Almeida *et al.* (2002) concluíram que eles pertencem à mesma condição, porém, com grau variável de expressividade. Casos isolados foram relatados por Guion-Almeida *et al.* (1999 e 2002) e Priolo *et al.* (2000), sugerindo que esses casos teriam ocorrido por mutações novas.

Os genes *ET1* (*endothelin-1*), *DLX2* (*distal-less homeobox 2*) e *HoxA2* (*homeobox A2*) foram sugeridos como candidatos, pois participam da morfogênese dos primeiros e segundos arcos faríngeos (Guion-Almeida *et al.* 2002 e Erlich *et al.* 2000). Esses genes, contudo, ainda não foram testados. Cabe ressaltar que nenhuma anomalia cromossômica ou região candidata foi associada à síndrome.

1.4 Síndrome de Treacher Collins

A síndrome de Treacher Collins (STC) (OMIM 154500) é uma anomalia congênita do desenvolvimento craniofacial, possui etiologia genética e padrão de herança autossômico dominante. É também conhecida como disostose mandibulofacial ou síndrome Franceschetti-Zwahlen-Klein. Embora tenha sido primeiramente descrita por Thomsom (1846), o nome associado à condição é do oftalmologista inglês Treacher Collins, que descreveu dois pacientes portadores de coloboma das pálpebras e hipoplasia malar décadas mais tarde (Treacher Collins, 1900). A incidência estimada é de 1:50.000 nascimentos vivos (Gorlin *et al.*, 2001), sendo que cerca de 40% dos casos diagnosticados têm história familiar e 60% parecem ser causados por mutações novas (Jones, *et al.* 1997).

Como as estruturas afetadas na STC são originárias do primeiro e do segundo arcos faríngeos, que têm uma grande contribuição da crista neural, tem sido proposto que a patologia possa resultar de um defeito na migração das células da crista neural ou de uma diferenciação celular inadequada durante o desenvolvimento (Poswillo 1975). No entanto, Dixon *et al.* (2006) demonstraram que a hipoplasia esquelética típica da STC é resultado direto de uma deficiência no número de células da crista neural, que ocorre por causa de defeitos na formação e diferenciação dessas células, descartando a hipótese de migração celular deficiente.

1.4.1 Características Clínicas

A STC apresenta anormalidades bilaterais e simétricas, sendo as características clínicas mais comuns (Figura 1): fendas palpebrais em inclinação antimongolóide (89%), coloboma no terço mais externo das pálpebras inferiores (69%), ausência de cílios medialmente a este defeito (53%), hipoplasia do osso zigomático e das bordas supra-orbitais (81%), micrognatia (78%), palato ocasionalmente fendido (28%), penugem capilar atípica na face (26%), malformação das orelhas (77%) e dos ossículos do ouvido médio, podendo resultar surdez condutiva (40%) (Gorlin *et al.*, 1990; Jones *et al.*, 1997). Os pacientes geralmente apresentam um espectro de achados clínicos. Raramente uma alteração isolada leva ao diagnóstico da síndrome.

Existe uma grande variação fenotípica, tanto inter como intra-familiar, num espectro que inclui casos em que o paciente é tão levemente afetado que clinicamente não pode ser diagnosticado, até casos de morte perinatal pelo colapso das vias respiratórias. Alguns indivíduos são diagnosticados somente após o nascimento de um filho gravemente afetado, o que torna o diagnóstico inequívoco e o aconselhamento genético uma tarefa bastante difícil (The Treacher Collins Syndrome Collaborative Group 1996; Hansen *et al.* 1996). A penetrância é quase completa, havendo somente um relato comprovado de não-penetrância, em que o indivíduo, mesmo sendo portador do alelo de risco e pai de uma criança gravemente afetada, não apresentou sinais clínicos e radiográficos da síndrome (Dixon *et al.*, 1994).

Alguns autores sugerem a possível existência de uma forma autossômica recessiva da síndrome de Treacher Collins, como os relatos de Lowry *et al.* (1985), que descreveram dois filhos afetados de casamento consanguíneo, e como os de Gollop, (1981), de Richieri-

Costa *et al.* (1993) e de Splendore *et al.* (2000), que relataram recorrência de afetados na prole de pais normais.

1.4.2 *TCOF1* e treacle

O gene responsável pela STC se localiza no braço longo do cromossomo 5 (5q32-33.1) e é denominado *TCOF1*. Por clonagem posicional, o gene foi identificado em 1996 (The Treacher Collins Collaborative Group, 1996) e teve a identificação completa de sua sequência codificadora e de sua organização genômica em 1997 (Dixon *et al.*, 1997b). Descobriu-se que o gene possui 26 exons, variando em tamanho de 49 a 561pb, sendo 25 traduzidos. O exon 1 contém, além do início da transcrição e da tradução, uma região 5' não traduzida de 93pb. Há também uma região 3' não traduzida, que se estende do final do exon 25 ao final do exon 26. Recentemente, So *et al.* (2004) identificaram mais dois novos exons no *TCOF1*, 6a e 16a, elevando o número de exons traduzidos para 28. No entanto, a isoforma mais abundante contém somente o exon 6a (AY460334).

O produto gênico do *TCOF1*, a proteína treacle, possui 1488 aminoácidos, 152kD e baixa semelhança com as proteínas conhecidas. Na sua porção N-terminal, a treacle possui dois domínios preditos por análise computacional: um sinal de exportação nuclear, localizado na região entre os aminoácidos 40 e 49, e o motivo LisH (*LIS1* homology motif), que pode estar associado a dimerização protéica e dinâmica de microtúbulos, localizado entre os aminoácidos 5 a 38. Não há evidências experimentais para nenhum desses domínios (Emes & Ponting, 2001; revisão em Dixon & Dixon, 2003).

A treacle também possui um domínio repetitivo na sua porção central que é alvo potencial de fosforilação pela enzima caseína quinase II (CKII). Esse domínio é codificado pelos exons 6 a 16 e é bastante semelhante ao domínio central da fosfoproteína nucleolar *Nopp140* do camundongo, uma ribonucleoproteína chaperona, que faz transporte de moléculas entre o núcleo e o citoplasma (Meier & Bobel, 1992; Dixon *et al.*, 1997b; Wise *et al.*, 1997). Consistentemente, Isaac *et al.* (2000) observaram que a treacle e a CKII co-imunoprecipitaram a partir de lisado de células HeLA. Além disso, a treacle possui uma série de sítios de localização nuclear e nucleolar preditos na sua porção C-terminal, que foram experimentalmente testados (Winokur & Shiang, 1998; Marsh *et al.*, 1998), demonstrando a localização subcelular da proteína no nucléolo.

A partir dessas observações, foi proposta a hipótese de que a treacle estaria envolvida em biogênese de ribossomos, já que as principais funções do nucléolo são transcrição e processamento do RNA ribossômico (RNAr), assim como sua posterior organização em subunidades ribossômicas (Winokur & Shiang, 1998; Marsh *et al.*, 1998). No entanto, foi somente em 2003 que se obteve a primeira evidência experimental dessa função: procurando caracterizar a fosfoproteína nucleolar humana NOP56, envolvida em processamento de pré-RNAr, Hayano *et al.* (2003) co-imunoprecipitaram a treacle com o complexo protéico associado à NOP56. Sugeriu-se então que a treacle teria semelhante função e que também estaria envolvida em processamento de pré-RNA ribossômico.

Em 2004, Valdez *et al.* demonstraram, por ensaio de duplo-híbrido e co-imunoprecipitação, que a treacle interage com o fator de transcrição da RNA Polimerase I, o UBF (*upstream binding factor*), e, por ensaio de RNA de interferência, que afeta os níveis de transcrição de DNA ribossômico e a proliferação celular. Recentemente, Gonzales *et al.* (2005a) demonstraram que a treacle também participa da metilação de pré-RNAr. Esse grupo observou que, embora a treacle co-localize com o UBF durante toda a mitose, ela também co-localiza com as proteínas NOP56 e fibrilarina, porém, somente durante a telófase, momento em que se inicia a metilação de RNAr. Esses autores então propuseram que a treacle participa da transcrição de DNA ribossômico por interagir com a RNA polimerase I, e também participa modificações pós-transcricionais do pré-RNAr por interagir com NOP56 (Figura 6).

1.4.3 Espectro de mutações do gene *TCOF1*

Já foram descritas mais de 100 mutações causadoras da doença no gene *TCOF1*, sendo que a maioria resulta em códons de parada prematuros. Já foram documentadas mutações em sítios de *splicing*, mutações do tipo *nonsense* e *missense*; porém, a maioria das alterações identificadas no gene *TCOF1* é dada pela deleção ou duplicação de 1 a 40 nucleotídeos na região codificadora, que geram uma mudança no quadro de leitura e um códon de parada prematuro adiante na sequência (The Treacher Collins Collaborative Group 1996; Gladwin, 1996; Wise *et al.* 1997; Edwards *et al.* 1997; Splendore *et al.* 2000, 2002; Ellis *et al.* 2002; Teber *et al.*, 2004; Horiuchi *et al.*, 2005).

As mutações identificadas são específicas para cada família, porém, existe uma principal exceção, a deleção de cinco pares de bases no exon 24, encontrada em mais de uma família independentemente (Edwards *et al.*, 1997). Uma vez que essa deleção no exon

24 é responsável por aproximadamente 18.6% dos casos, alguns autores sugeriram que essa mutação possa representar um *hot spot* mutacional dentro do gene. Dada a natureza repetitiva da região, que codifica longos trechos de adeninas, é possível que as mutações surjam por *replication slippage*. Além desse possível *hot spot*, há cinco exons (10, 15, 16, 23 e 24) com maior probabilidade de apresentarem mutações patogênicas; por outro lado, existem exons em que ainda nenhuma mutação foi encontrada (Splendore *et al.* 2000, 2002).

Muitas mutações do tipo *missense* já foram descritas no gene *TCOF1*, sendo a maioria polimórficas. Embora tenha sido demonstrada uma alta frequência de SNPs (polimorfismos de base única - *single nucleotide polymorphisms*) na região codificadora do gene, nenhum polimorfismo foi detectado nos exons 1 e 2 em 100 indivíduos normais. Além disso, a porção N-terminal da proteína é altamente conservada entre humano e camundongo, exibindo uma semelhança de 92,6% para essa região, em comparação com 54,8% para o restante da proteína. Essas informações sugerem que a região codificada pelos exons 1 e 2 seja extremamente importante para a função da proteína, e, portanto, tolere menos variações em sua sequência. Sendo assim, quatro mutações *missense*, encontradas em quatro pacientes com o fenótipo clássico, foram consideradas como causadoras da síndrome nesses indivíduos (Edwards *et al.*, 1997; Splendore *et al.*, 2002b; Ellis *et al.*, 2002; Dixon & Dixon, 2003).

Sugere-se que o mecanismo molecular associado à STC seja de haploinsuficiência de treacle durante o desenvolvimento. Dado que as mutações causadoras de códon de parada prematuro podem gerar uma redução dos níveis de RNA mensageiro pelo mecanismo de *nonsense mediated mRNA decay*, (NMD), supõe-se que os níveis da proteína codificada possam ser reduzidos em consequência disso (The Treacher Collins Collaborative Group 1996; Gladwin *et al.* 1996; Wise *et al.* 1997; Edwards *et al.* 1997; Dixon *et al.* 2000; Splendore *et al.* 2000, 2002). No entanto, Isaac *et al.* (2000) não observaram diferença nos níveis de treacle entre fibroblastos de pacientes e de indivíduos normais. Dessa maneira, esses autores consideraram a possibilidade de um efeito dominante-negativo de proteínas truncadas, mais curtas que a normal, no mecanismo da doença. Entretanto, essa hipótese foi refutada no mesmo estudo, que mostrou a ausência de proteína mutante no citoplasma de linfócitos e fibroblastos de indivíduos portadores da STC, tanto para as mutações em C, quanto em N-terminal. Esses autores então sugeriram a existência de um mecanismo de compensação de dose de treacle em adultos, capaz de manter os níveis apropriados da proteína na célula, mesmo com um único alelo funcional. Porém, esse suposto mecanismo estaria ausente

durante o desenvolvimento e, portanto, a dose de treacle estaria reduzida, causando o quadro clínico por haploinsuficiência.

1.4.4 Modelos animais

Antes da descoberta do gene associado à STC, fenocópias foram produzidas com teratógenos em ratos e camundongos, na tentativa de elucidar as bases do desenvolvimento relacionadas ao fenótipo. Poswillo (1975) administrou doses teratogênicas de vitamina A (palmitato) a ratos nos primórdios do desenvolvimento. Isso gerou morte celular localizada nas células da crista neural pré-óticas, antes/durante a sua migração a partir do tubo neural em fechamento. Esse autor, então, sugeriu que o fenótipo da STC seria gerado por uma destruição específica de células da crista neural destinadas à formação dos mesênquimas dos primeiro e segundo arcos faríngeos. Sulik *et al.* (1987) administraram doses teratogênicas de *13-cis retinoic acid* (composto ativo derivado da vitamina A) a camundongas com 9 a 9,5 dias pós-fertilização e obtiveram fenocópias das síndromes de Treacher Collins, Nager e Miller. Esses autores observaram apoptose excessiva nos placóides ectodérmicos e sugeriram que esse dano causou um dano secundário, o impedimento da migração adequada das células da crista neural. Assim, eles sugeriram que o dano tecidual localizado seria causador da migração anormal das células da crista neural, e não o dano nessas células especificamente.

Em 1997, Dixon *et al.* (1997a) isolaram o gene *Tcof1*, homólogo do *TCOF1* humano em camundongos, que apresenta 74,3% de semelhança ao nível de nucleotídeos. Em 2001, Haworth *et al.* (2001) isolaram o *Tcof1* em cães, que tem 80% de identidade com a sequência humana, revelando nove variantes polimórficas do *Tcof1* canino, uma delas associada à braquicefalia canina. Gonzáles *et al.* (2005b) identificaram o gene em *Xenopus laevis*: embora exiba baixa semelhança com os genes ortólogos de mamíferos, possui função bastante semelhante – a sua inibição em oócitos de *Xenopus*, mediada por RNA antisenso, inibe a transcrição de DNA ribossômico. Recentemente, um macaco rhesus foi identificado com o fenótipo semelhante à STC. O gene foi clonado e seqüenciado, apresentando uma semelhança de 91,6% ao nível de cDNA com a sequência humana. Nenhuma mutação patogênica foi detectada, porém, utilizando PCR quantitativo, esse grupo de pesquisadores observou que os níveis de RNAm de *Tcof1* apresentaram uma redução de 87% no baço do macaco afetado, em comparação com um macaco normal (Shows *et al.* 2006).

A identificação do *Tcof1* em camundongo (18q21; Dixon *et al.*, 1997a) possibilitou a análise da expressão do gene durante o desenvolvimento embrionário. Esses autores investigaram a expressão do *Tcof1* em embriões de camundongos normais e demonstraram que o gene é expresso em todo o embrião durante o seu desenvolvimento. No entanto, por volta do 9º e do 10º dia pós-fecundação, ocorre um pico de expressão do gene na região dos arcos faríngeos e das cristas neurais (Figura 7). Dessa maneira, esses autores sugeriram que a maior expressão do *Tcof1* ocorre durante o período mais crítico para formação das estruturas da face, e uma quantidade reduzida de treacle poderia explicar o quadro clínico da síndrome, reforçando a hipótese de haploinsuficiência.

Em 2000, Dixon *et al.* desenvolveram o camundongo nocaute para estudo da síndrome, substituindo o exon 1 do *Tcof1*, por meio de recombinação homóloga, por um cassete de resistência à neomicina, anulando a expressão do gene. Os camundongos heterozigotos apresentaram morte perinatal devido a graves anomalias craniofaciais, com sobreposição parcial de fenótipo - os fetos, além de malformação da maxila, do complexo nasal, da orelha média e externa, da hipoplasia mandibular, também apresentaram exencefalia, anoftalmia e agenesia nasal. Possivelmente, esses defeitos surgiram devido a um aumento massivo de células apoptóticas observadas nas cristas neurais pré-fusão (Figura 8a). Esses autores sugeriram que a dosagem correta de treacle é essencial para a sobrevivência das células das cristas neurais em camundongos. No entanto, o fenótipo dos camundongos *Tcof1*^{+/-} foi considerado muito mais grave do observado em humanos.

Esse mesmo grupo de pesquisadores levantou a hipótese de que a gravidade do fenótipo em camundongos poderia depender de alelos modificadores distribuídos no genoma (Dixon & Dixon, 2004). Dessa maneira, testaram o genótipo *Tcof1*^{+/-} em diferentes linhagens de camundongos. Os indivíduos heterozigotos exibiram fenótipos extremamente variáveis (de morte perinatal à aparente normalidade), todavia, linhagem-dependente (figura 8b). Esses autores então sugeriram que o *background* genético determina a variabilidade fenotípica. Gonzales *et al.*(2005) verificaram que as linhagens *Tcof1*^{+/-} que exibiram fenótipo letal também apresentaram uma redução significativa da metilação do nucleotídeo C₄₆₃ do RNAr 18S em relação aos indivíduos selvagens. Ao contrário, os indivíduos *Tcof1*^{+/-} de linhagens com aparente normalidade não exibiram diferença significativa na metilação de RNAr.

Em 2006, Dixon *et al.* reproduziram os achados anteriores (Dixon *et al.*, 1997a) e demonstraram que a treacle é um regulador espaço-temporal específico da biogênese de ribossomos. Por hibridização *in situ*, observaram que o gene é bastante expresso no

neuroepitélio de camundongos E8.5, além do tubo neural e mesênquima fronto-nasal, maxilar e mandibular de embriões E9.5. Por volta de E11.5-E12.5 a expressão do gene se torna difusa, praticamente indetectável.

Esses mesmos autores também observaram que a migração das células da crista neural, tanto em embriões de camundongos selvagens, quanto em *Tcof1*^{+/-}, é normal, ou seja, a haploinsuficiência de treacle não causa migração aberrante dessas células (Figura 9a). No entanto, demonstraram que há uma redução de aproximadamente 22,3% no número de células da crista neural em migração ao mesênquima dos arcos faríngeos nos embriões *Tcof1*^{+/-} (Figura 9b). Além disso, foi demonstrado que a haploinsuficiência de treacle leva as células do neuroepitélio à apoptose e a uma diminuição na sua capacidade proliferativa, sem, no entanto, comprometer a viabilidade daquelas que foram formadas. Por meio de imunohistoquímica com anticorpos para RNAr maduro (28S), esses autores também verificaram que o neuroepitélio e o mesênquima derivado das células da crista neural de camundongos *Tcof1*^{+/-} possuíam quantidade reduzida desse marcador (Dixon *et al.*, 2006).

1.5 Questões não respondidas

A grande variabilidade clínica encontrada nos portadores da STC é uma questão bastante intrigante. Indivíduos da mesma família com a mesma mutação apresentam diferença acentuada no fenótipo (Figura 9). Assim, a identificação da mutação em um indivíduo não pode prever a gravidade do quadro clínico de seus descendentes. Na tentativa de compreender esse fenômeno, diversos estudos confirmaram a ausência de correlação genótipo-fenótipo (Dixon, 1996; Gladwin *et al.*, 1996; Edwards *et al.*, 1997; Splendore *et al.*, 2000). Splendore *et al.* (2000) não encontraram correlação entre a gravidade do quadro clínico, o sexo do afetado e a origem da mutação, se herdada ou *de novo*. Da mesma maneira, outros grupos observaram ausência de correlação com o tipo e/ou a localização da mutação patogênica (Dixon, 1996; Edwards *et al.*, 1997). Teber *et al.* (2004) tentaram estabelecer critérios de gravidade para o quadro clínico, apresentando uma classificação fenotípica por pontuação de achados clínicos detalhados, com o objetivo de estabelecer correlação genótipo-fenótipo. Esses autores concluíram que os tipos de mutação (*splicing*, *nonsense*, duplicações e deleções), assim como a transmissão paterna ou materna, estavam igualmente presentes em casos leves e graves. Também observaram que a distribuição das mutações ao longo do gene não era diferente em leves e graves.

Outra questão intrigante são os pacientes com quadro clínico típico da STC sem mutação patogênica detectada na região codificadora, como os casos descritos por Splendore *et al.* (2000). Sugerimos a existência de um mecanismo mutacional atípico para esses casos, como deleções ou duplicações de exons inteiros, que não seriam detectadas pelos nossos métodos, ou mutações localizadas em regiões regulatórias. Em 2004, dois novos exons foram identificados no gene *TCOF1*, 6a e 16a, por So *et al.* (2004). Dado que o exon 6a está presente na isoforma mais abundante, consideramos possível que os pacientes sem mutação patogênica detectada apresentem alterações nesses exons.

Além disso, cabe ressaltar que a haploinsuficiência, mecanismo molecular proposto para a STC, não foi demonstrado em pacientes. Supõe-se que o mecanismo NMD degrade os transcritos portadores de códon de parada prematuro e que ocorra por consequência uma diminuição nos níveis da proteína. Embora os achados em modelos animais reforcem a hipótese de haploinsuficiência (Dixon *et al.*, 1997a, 2000, 2006), a análise quantitativa de treacle em humanos não revelou diferença na quantidade de produto gênico entre fibroblastos de pacientes e de indivíduos normais (Isaac *et al.*, 2000). Além disso, não há relatos na literatura de quantificação dos níveis de RNAm de *TCOF1* em pacientes que confirmem a degradação dos transcritos. Dado que o efeito dominante negativo foi excluído como mecanismo responsável pela STC (Isaac *et al.*, 2000), julgamos importante que mais evidências sejam coletadas para corroborar a hipótese de haploinsuficiência em humanos.

Considerando essas questões e a possibilidade do mecanismo de haploinsuficiência de treacle como base molecular para manifestação da síndrome, é possível que variações na expressão do *TCOF1* em pacientes estejam associadas à variação do fenótipo. Dado que os pacientes possuem um alelo não-funcional, sugerimos que os mecanismos que modulam os níveis de expressão, especificamente do alelo normal, sejam responsáveis pela variabilidade clínica.

Existem diversos mecanismos de regulação da expressão gênica, sendo divididos em pré e pós-transcricionais. Para a maioria dos genes, a sua expressão é modulada pelo controle do início da transcrição, já que é a maneira mais econômica para a célula. Em eucariotos, a regulação ao nível da transcrição é bastante complexa, pois é caracterizada pela ligação de elementos de ação *trans* (fatores de transcrição) às seqüências de nucleotídeos específicas da região promotora do gene (elementos de ação *cis*), que, juntamente com a RNA polimerase II, formam o complexo de início da transcrição. Dessa maneira, a integridade

da sequência genômica da região promotora é fundamental para que ocorra a síntese adequada de RNA mensageiro (Carey & Smale, 2000).

O promotor do gene *TCOF1* ainda não foi caracterizado funcionalmente. Possivelmente, mutações nessa região podem diminuir os níveis de expressão da proteína treacle na fase crítica do desenvolvimento das estruturas craniofaciais, causando o quadro clínico da síndrome ou interferindo na variabilidade fenotípica. Yan *et al.* (2002) afirmam que é esperado que as variações nos fenótipos das doenças sejam mais frequentemente causadas por mudanças nos níveis de expressão do que por alterações estruturais dos genes. Esses autores correlacionaram um menor nível de expressão do gene APC a uma maior predisposição a tumor coloretal hereditário. Joosten *et al.* (2001) identificaram cinco haplótipos diferentes para região promotora do gene *PDGF* (*platelet-derived growth factor alfa-receptor*), associado a defeitos de fechamento do tubo neural (NTD) em modelos animais. Foi verificado que em alguns indivíduos portadores de combinações específicas desses haplótipos havia níveis mais altos de transcrição de *PDGF*, o que estava associado a uma maior predisposição para ocorrência de NTD.

A caracterização da região promotora do gene *TCOF1* também poderá contribuir no entendimento do padrão de expressão do gene durante o desenvolvimento embrionário humano. Sabe-se que o *TCOF1* se expressa em uma gama variada de tecidos adultos, tais como fígado, pâncreas, coração, cérebro e pulmão (The Treacher Collins Collaborative Group, 1996). Entretanto, há um pico de expressão do gene no 9º e 10º dia de vida embrionária em camundongos (Dixon *et al.* 1997a, 2006). Essa variação temporal da expressão do gene nos primeiro e segundo arcos faríngeos e nas cristas neurais sugere que haja um mecanismo regulatório específico das células embrionárias dessas regiões, ou seja, provavelmente há um elemento ativador que está ausente nos demais tecidos. Entretanto, essa variação também pode ser regulada por mecanismos epigenéticos, como alívio da repressão por desmetilação de promotores envolvidos em proliferação celular (Robertson *et al.*, 2005).

Com relação à síndrome Aurículo-Condilar, as questões se concentram na caracterização fenotípica dos afetados bem como na identificação do gene responsável por essa condição, uma vez que nada se sabe sobre o mecanismo molecular ou genes associados a essa doença. Alguns autores sugeriram que as síndromes de Nager e Miller poderiam representar uma variação do espectro de variabilidade fenotípica da STC, isto é, diferentes manifestações fenotípicas de mutações num mesmo gene (Sulik *et al.*, 1987). No entanto, foi demonstrado que mutações no gene *TCOF1* geram um fenótipo bastante

específico (Splendore *et al.*, 2002b), excluindo os pacientes portadores das síndromes de Nager, Miller e Goldenhar de possuírem mutação patogênica nesse gene. Uma vez que a SAC parece pertencer a uma via de desenvolvimento semelhante/sobreposta a essas patologias, julgamos importante verificar se essas condições são alélicas. É importante ressaltar que essas síndromes não têm base genética ou mecanismo de expressividade variável conhecidos. Embora três regiões cromossômicas já tenham sido associadas à síndrome de Nager (3p14, 9q32, 1q12-21.1; revisão em Gorlin *et al.*, 2001), nenhum gene candidato foi descrito. Mais de 13 anomalias cromossômicas já foram associadas à síndrome de Goldenhar (revisão em Kobrinski *et al.*, 1993) e um estudo de ligação com duas famílias grandes excluiu a região candidata (14q32) em uma delas (Kelberman *et al.*, 2001), sugerindo a existência de heterogeneidade genética. Dessa maneira, o mapeamento e a identificação de mecanismos regulatórios de genes associados à STC e à SAC certamente trarão uma contribuição importante para a compreensão do desenvolvimento embrionário das estruturas envolvidas nessas condições, além de possibilitar o diagnóstico molecular para auxiliar no aconselhamento genético e no diagnóstico diferencial dessas patologias.

1.6 Objetivos

Nossos principais objetivos são:

- a. Investigar mecanismos mutacionais atípicos na STC e testar a hipótese de que mutações nos exons que sofrem *splicing* alternativo são capazes de gerar o fenótipo da síndrome;
- b. Caracterizar o espectro de mutações no gene *TCOF1* e investigar a correlação genótipo-fenótipo na amostra total de pacientes STC do Centro de Estudos do Genoma Humano (IB-USP);
- c. Testar a hipótese de que variações nos níveis de expressão do alelo normal podem ser responsáveis pela variabilidade clínica observada nos pacientes portadores da STC;
- d. Testar a hipótese de haploinsuficiência como mecanismo associado à STC, quantificando os níveis de transcritos do *TCOF1* em indivíduos afetados;
- e. Testar a hipótese de que a SAC é alélica a malformações de primeiro e segundo arcos faríngeos já conhecidas.
- f. Mapear o gene associado à SAC e testar a hipótese de heterogeneidade genética para essa patologia.

Síndrome Aurículo-Condilar

Síndrome de Treacher Collins

Figura 1: Pacientes típicos das síndromes Aurículo-Condilar e de Treacher Collins. Na síndrome Aurículo-Condilar: micrognatia grave e bochechas proeminentes são ilustradas; detalhe da malformação da orelha externa (*question mark ear*). Na síndrome de Treacher Collins, as fotos menores detalham os achados clínicos mais freqüentes: **a** pálpebras com inclinação anti-mongolóide, **b** hipoplasia malar, **c** micrognatia, **d** coloboma da pálpebra inferior, **e** malformação da orelha externa, **f** orelha com implantação baixa, **g** projeção de couro cabeludo na face.

Figura 2: Origem do mesênquima dos primeiro e segundo arcos faríngeos. **A** Foto de microscopia eletrônica de um embrião humano de 23º dia pós-fertilização. Nesse momento, ocorre o fechamento do tubo neural, estrutura que dará origem ao cérebro e à coluna espinhal. Por volta do vigésimo dia (E20), na porção dorsal do embrião (corte transversal), forma-se a placa neural. De E20 a E23, a placa neural sofre um dobramento, até que suas extremidades (cristas neurais) se encostem, fechando o tubo neural. **B** Microscopia eletrônica de embrião (E27), com primeiro e segundo arcos faríngeos apontados. Células da crista migram e passam a constituir o mesênquima dos arcos.

Figura 3: **A** Correntes migratórias das células da crista neural. As células que povoarão os arcos faríngeos 1 (rosa), 2 (verde) e 3 provêm da população cranial, partindo dos rombômeros 1 a 6. **B** Principais estruturas ósseas e epiteliais adultas formadas a partir das células da crista neural do mesênquima dos arcos faríngeos.

© Elsevier Ltd. Carlson: Human Embryology and Developmental Biology 3E www.studentconsult.com

Figura 4: Principais estruturas adultas derivadas dos componentes nervosos (b), musculares (c) e esqueléticos (d) dos primeiro (em verde) e segundo (em rosa) arcos no adulto.

© Elsevier Ltd. Carlson: Human Embryology and Developmental Biology 3E www.studentconsult.com

Primeiro arco faríngeo

Segundo arco faríngeo

http://www.med.unc.edu/embryo_images

Figura 5: Interação entre os primeiro (verde) e segundo (vermelho) arcos para formação da orelha externa. Microscopia eletrônica de um embrião com cinco semanas de gestação. Massas nodulares de mesênquima (*auricular hillocks*) se organizam próximo à primeira fenda faríngea. Esses nódulos crescem e se fusionam para formar a orelha externa.

Gonzales *et al.*, 2005.

Figura 6: Modelo de funcionamento da treacle, proposto por Gonzales *et al.* (2005). A treacle co-localiza e interage com proteínas envolvidas na transcrição e na metilação do pré-RNA ribossômico. No início da mitose, a ligação de treacle à maquinaria da RNA Polimerase1 pode bloquear a iniciação da transcrição. Ao interagir com UBF, tardiamente na telófase, a treacle possibilitaria a RNA Polimerase1 a prosseguir a transcrição. Simultaneamente com a ativação da RNA Polimerase1, a treacle recrutaria ribonucleoproteínas do complexo de metilação do pré-RNAr.

Dixon *et al.*, 1997.

Figura 7: Hibridização *in situ* de *TCOF1* em embriões de camundongo de 8,5, 9 e 9,5 dias (pós-fertilização; Dixon *et al.*, 1997^a). Ilustração do pico de expressão do gene nas cristas neurais, durante o fechamento do tubo neural (E8,5). Os arcos branquiais também possuem alta expressão em E9 e E9.5, que equivale ao fim da quarta semana do desenvolvimento humano.

Figura 8: **A** Corte transversal do tubo neural em fechamento. Número massivo de células apoptóticas nas cristas neurais pré-fusão. O corante azul marca as células em apoptose (Dixon *et al.*, 2000). **B** Diferentes linhagens de camundongos *Tcof1*^{+/-} (Dixon & Dixon, 2004). Camundongos com a mesma mutação exibiram fenótipo linhagem-dependente extremamente variável (morte perinatal à aparente normalidade).

A

Dixon *et al.*, 2006.

B

Dixon *et al.*, 2006.

Figura 9: Análise do número e do padrão de migração das células da crista neural de embriões de camundongo *Tcof1*^{+/-} (modificada de Dixon *et al.*, 2006). **A** Células da crista neural em migração aos arcos faríngeos 1 e 2 (ba1 e ba2), ao nível dos rombômeros 2 (r2) e 4 (r4), em embriões (E8,5) de camundongos selvagens (A, B, E, F) e *Tcof1*^{+/-} (C, D, G, H), coradas em vermelho. **B** Células da crista neural marcadas com GFP (*green fluorescent protein*) de um camundongo (E9,0) selvagem. Ao lado, um histograma demonstrando a proporção de células marcadas em embriões *Tcof1*^{+/-} (d, barra verde) e selvagens (c, barra amarela). As barras vermelha (b) e azul (a) representam controles positivo e negativo, respectivamente. Há uma redução de 22.3% no número de células dos indivíduos mutantes, em comparação ao selvagem. Haploinsuficiência de treacle não causa migração aberrante dessas células; porém, diminui o número de células viáveis.

Capítulo 2

CASUÍSTICA E MÉTODOS

2.1 Pacientes e Controles

Nossa amostra total de pacientes portadores da síndrome de Treacher Collins (STC) consistiu de 58 indivíduos, incluindo propósitos e familiares afetados com mutação patogênica identificada no gene *TCOF1*. Nessa amostra estão incluídos os pacientes portadores de mutação patogênica cujos diagnósticos moleculares estão descritos no Capítulo 3-II (Tabela 1). Também incluímos no estudo de haplótipos 30 familiares normais.

Com exceção dos casos novos descritos no Capítulo 3-II, a triagem de mutação no *TCOF1* (em pacientes e familiares normais) foi realizada por Splendore *et al.* (2000 e 2002a). A amostra total foi constituída no Centro de Estudos do Genoma Humano ao longo dos anos de aconselhamento genético, sendo que o estudo de grande parte dessa amostra gerou publicações prévias de nosso grupo em STC. Os pacientes analisados nos Capítulos 3-I, 4 e 6 pertencem à amostra de 58 indivíduos portadores da STC.

Os 30 pacientes estudados no Capítulo 3-II foram encaminhados para aconselhamento genético e diagnóstico molecular no Centro de Estudos do Genoma Humano (CEGH, IB-USP), ou tiveram suas amostras de DNA enviadas pelo correio, partindo dos seguintes locais e respectivos pesquisadores responsáveis: (1) Hospital de Clínicas de Porto Alegre (Dra. Têmis Félix); (2) Arcispedale “S. Maria Nuova”, Ambulatório di Genética Clinica, Divisione di Pediatria, Reggio Emilia, Itália (Dra. Livia Garavelli); (3) Vall d’Hebron Hospital Materno-Infantil, Unitat de Genética Clínica i Citogenètica, Barcelona, Espanha (Dr. E. Sarret); (4) Centre de Génétique Médicale, Lovreval, Bélgica (Dr. Lionel Van Maldergem); (5) Department of Clinical Genetics, Karolinska Hospital, Estocolmo, Suécia (Dra. Kristina Lagerstedt); (6) Department of Clinical Genetics, Oulu University Hospital, Oulu, Finlândia (Dr. Aki Mustonen e Dr. Jukka Moilanen); (7) Medical Genetics Center, Vrije Universiteit Brussel, Bruxelas, Bélgica (Dra. Sara Sêneca).

As amostras de 213 DNAs de controles utilizadas fazem parte de uma amostra de aproximadamente 1000 indivíduos, constituída em um projeto conjunto de alguns grupos de pesquisa do CEGH. A amostra é caracterizada por doadores do Hemocentro do Hospital das Clínicas de São Paulo, com idades e raças variáveis. Esse projeto de foi aprovado pela Comissão de Ética para análise de projetos de pesquisa do Hospital das Clínicas e da Faculdade de Medicina Universidade de São Paulo em 26/06/2001(protocolo 349/01).

A amostra de pacientes portadores da síndrome Aurículo-Condilar (SAC) consistiu de 30 indivíduos, sendo 14 da família 1 (9 afetados e 5 familiares normais) e 16 da família 2 (6 afetados e 10 familiares normais; Guion-Almeida *et al.*, 2002). Os pacientes da família 1 foram avaliados pela equipe multidisciplinar do Centro de Assistência a Indivíduos Portadores de Fissuras de Lábio e Palato (CAIF).

2.2 Considerações éticas

Os pacientes que foram atendidos pelo CEGH passaram por consulta com uma equipe multidisciplinar constituída por geneticistas clínicos, biólogos moleculares e aconselhadores genéticos. As dúvidas dos familiares dos consulentes foram esclarecidas e informações adicionais sobre a pesquisa foram oferecidas sob consulta. Os responsáveis pelos pacientes são informados sobre os testes genéticos disponíveis, a capacidade diagnóstica dos mesmos e sob a finalidade de pesquisa. A coleta de material (amostra de sangue periférico) para extração de DNA dos afetados aconteceu somente após consentimento do responsável do paciente. A divulgação científica e publicação dos dados, incluindo fotos, dos pacientes foram realizadas somente mediante autorização dos responsáveis. Todos os pacientes que foram submetidos à triagem de mutações no *TCOF1* receberam um relatório escrito com o diagnóstico molecular alcançado. Relatórios escritos com o resultado da pesquisa serão entregues a todos os pacientes mediante finalização dos estudos em andamento.

As amostras de DNA encaminhadas para diagnóstico molecular provenientes de outros centros foram obtidas de acordo com os protocolos de ética dos respectivos locais. O presente projeto foi aprovado pelo comitê de ética do Instituto de Biociências (Registro CONEP 8083, Parecer 1390/2003) e seguimos as devidas recomendações. Todos os pacientes assinaram o termo de consentimento permitindo a participação nesta pesquisa (Anexo I).

2.3 Métodos

2.3.1 Extração de DNA de linfócitos de sangue periférico, de esfregaço bucal e de cultura de células.

A extração do DNA genômico, a partir de linfócitos, foi realizada com amostras de 5-10 ml de sangue total coletado em tubos contendo 0,4 ml de EDTA 5 % com anticoagulante. A extração foi realizada de acordo com a metodologia descrita por Miller et al. (1988). As amostras de DNA de sangue periférico mais recentes foram extraídas pelo robô de extração de DNA (*AutopureLS*, *Gentra Systems*), instalado no Centro de Estudos do Genoma Humano.

A extração de DNA de esfregaço bucal consiste na coleta de células da mucosa bucal com uma pequena escova descartável, de acordo com o protocolo de Richards *et al.* (1993). Para a maioria das amostras, utilizamos o kit QIAamp DNA Mini Kit 250 (*QIAGEN*).

Para a extração de DNA de cultura de células, utilizamos o método de fenol e clorofórmio, usado como rotina no Instituto Ludwig. Nesse protocolo, as células com confluência de aproximadamente 95%, em garrafas de 25 cm³ ou de 50 cm³ são tripsinizadas e lavadas com PBS 1X. Após centrifugação, o pellet é lisado com tampão TES (Tris HCl pH 8.0 10mM, EDTA 50mM, SDS 0,5%) e proteinase K (para concentração final de 200 µg/mL), sendo incubado *overnight* a 55° C. O lisado de células é então purificado com fenol equilibrado (pH 8.0) e posteriormente com clorofórmio-álcool-isoamílico. Após, o DNA é precipitado com acetato de sódio 3M e etanol absoluto, a -70° C por 1h, centrifugado e lavado com etanol 70%. Após a precipitação, o pellet de DNA seco é ressuspendido em TE (Tris-HCl pH 8.0 10mM, EDTA 50mM) e incubado a 65° C por 30 minutos para inativar as DNases. A qualidade e na concentração do DNA é verificada no aparelho NanoDrop ND-1000 (espectrofotômetro da *peqlab biotechnologie GmbH*) e em gel de agarose 1%.

2.3.2 Extração de RNA de linfócitos de sangue periférico e de cultura de células

Extraímos o RNA de linfócitos de acordo com o método Trizol (Gibco), conforme o protocolo do fabricante. Para extração de RNA de linhagens celulares, utilizamos o mesmo protocolo de extração de RNA de linfócitos, porém com algumas modificações. As células com

confluência de aproximadamente 95%, em garrafas de 25 cm³, foram tripsinizadas e centrifugadas, e o pellet foi submetido ao método de extração de RNA total com Trizol (Gibco).

A qualidade e a concentração do RNA é verificada no aparelho NanoDrop ND-1000 (espectrofotômetro da *Peqlab biotechnologie GmbH*) e em gel de agarose 1,2% após desnaturação a 65° C com formamida.

2.3.3 Cultura de células

As linhagens celulares, obtidas do ATCC (*American Type Culture Collection - Manassas, VA, EUA*), foram cultivadas em meio de cultura DMEM (*Dulbecco's Modified Eagle's Medium, Invitrogen*) com 10% de soro fetal bovino (*Gibco*) e 1% de estreptomicina/ampicilina (*Sigma-Aldrich*), mantidos em estufa a 37°C (5% CO₂). A linhagem de células-tronco de polpa de dente foi mantida em DMEM, 1% estreptomicina/ampicilina e 10% soro fetal bovino (*Hyclone*).

As linhagens celulares imortalizadas HepG2 (*hepatocellular carcinoma*), HEK293 (*human epithelium from kidney*), HUVEC (*human umbilical vein endothelial cells*), HeLa (*cervical carcinoma*), SH-SY5y (*neuroblastoma*) e hFOB (*human fetal osteoblasts*) foram trazidas para o nosso laboratório devido a uma colaboração com o grupo da Dra. Mari Sogayar, e a linhagem SaOS-2 (*epithelial-like osteosarcoma*), por colaboração com a Dra. Anamaria A. Camargo. A linhagem de células-tronco de polpa de dente foi estabelecida a partir de um dente decíduo de uma criança normal, como parte de um projeto do nosso laboratório (Bueno, DF. 2007).

2.3.4 Triagem de mutações na região codificadora e promotora do gene *TCOF1*

Amplificação dos fragmentos de DNA

O DNA genômico de pacientes e controles foi amplificado com a reação da cadeia em polimerase (PCR - *Polymerase Chain Reaction*), nas seguintes condições: 4 minutos de desnaturação a 95°C, 35 ciclos de 30 segundos a 95°C, 40 segundos na temperatura de hibridação (*annealing* – especificamente calculada para cada par de *primers*) e 40 segundos a 72°C, seguidos de uma etapa de extensão de 6 minutos a 72°C. Para cada reação de 20µL geralmente utilizamos: 100ng de DNA genômico, tampão 1X (*Amersham Biosciences*), 1u Taq DNA Polimerase (*Amersham Biosciences*), 200 µM dNTPs (*Amersham Biosciences*) e 0,5 a 1µM

de cada *primer*. O sucesso da PCR foi verificado em gel de agarose 1,5%, submetido a eletroforese horizontal, corado com brometo de etídeo (0,5µg/mL) e visualizado sob luz ultravioleta.

Os *primers* (oligonucleotídeos iniciadores) utilizados para amplificação dos exons do gene *TCOF1* foram publicados por The Treacher Collins Collaborative Group (1996), Gladwin *et al.* (1996), Edwards *et al.* (1997), Wise *et al.* (1997). As seqüências desses *primers*, assim como a metodologia de triagem de mutações no *TCOF1*, estão descritas em Splendore, A. (2002). As seqüências (5'-3') do *primers* para amplificação dos dois novos exons (So *et al.*, 2004) são: 6aF- CCA TCC TCT TGT TCC AG, 6aR- GTG AGG CCT GGA GGA GGG CTG, 16aF- AGG TCA GGC CAG CAG GTA TC e 16aR- CTG GAA GCT GTG AGG TGC TG.

A seqüência 5' adjacente ao início da transcrição do gene *TCOF1* foi obtida pelo alinhamento da seqüência do exon 1 (NM_000356) com a seqüência do *contig* do cromossomo 5, usando o programa BLASTn (<http://www.ncbi.nlm.nih.gov/blast/>). Desenhamos quatro pares de *primers* (Prom1, 2, 3 e 4), que se sobrepõem numa região de 1241pb do promotor. Para tanto, utilizamos o programa Primer3 (<http://www-genom.wi.mit.edu/cgi-bin/primer/primer3.cgi>). As seqüências dos *primers* utilizados para triagem de mutação na região promotora do gene *TCOF1* estão na Tabela 1.

Tabela 1: Primers utilizados para triagem de mutações na região promotora do *TCOF1*.

Nome	Tamanho	Annealing	Forward	Reverse
Prom 1	351 pb	56°C	CAG CCG GAA GGA TTA AAA GA	TGG TGG TAG ATC AGG GGA AG
Prom 2	331 pb	56°C	GAG GGC AAC TGC CAT GTA TT	GGG ACT TGG ACG ACT CTC TTT
Prom 3	317 pb	59°C	GTT TAG GGT TCC CAG GCA AT	AAT AAG CAC CTA ACT TGG GCT A
Prom 4	343 pb	56°C	CCA TCC TTC CCT CAA CAA CT	AGA GGG CAC TGC ATG AGT CT

SSCP e DHPLC

Para triagem de mutações no promotor realizamos dois métodos: polimorfismo conformacional de fita simples, SSCP (*Single Strand Conformation Polymorphism* - Orita *et al.*, 1989), para Prom1, 2 e 4, ou cromatografia líquida denaturante de alta performance, ou DHPLC (*Denaturing High Performance Liquid Chromatography* – Oefner, 2000), para Prom2 e 3. Ambos os métodos são comparativos, ou seja, as amostras dos pacientes foram sempre analisadas em

relação à amostra de controles normais. As amostras que apresentaram um padrão diferente do normal foram posteriormente seqüenciadas. O gel de SSCP (*MDE™ Gel solution, Bio Whittaker Molecular Applications*) foi realizado em eletroforese vertical com adição de glicerol 2,5% (8W, 12-14h). Adicionamos às amostras de PCR o SSCP *loading buffer* (formamida 95%, EDTA 0,02M, xilenocianol 0,5%, bromofenol blue 0,5%), na proporção 1:1. Os produtos foram visualizados com a coloração do gel com nitrato de prata (Bassam *et al.*, 1991) e fotografados em filme Typopaque TR-DO 18 (*Typon, Promega*). Para DHPLC, utilizamos o aparelho *Wave^R Nucleic Acid Fragment Analysis System (Transgenomic Inc)* e o programa *Wavemaker*, de acordo com o protocolo do fabricante. As temperaturas de corrida (*melting*) de cada *amplicon* foram: 59°C (Prom1), 57°C e 60°C (Prom2), 56°C e 60°C (Prom3) e 56°C (Prom4).

SNUPE

Empregamos o método de SNUPe (*Single Nucleotide Primer Extension - Amersham Biosciences*) para triagem do SNP (*Single Nucleotide Polymorphism*) -346C>T em pacientes e controles, de acordo com o protocolo do fabricante. Após a amplificação de Prom2, o produto de PCR purificado é submetido à reação de SNUPe (o kit contém um pré-mix com tampão específico e dideoxynucleotídeos) com um primer único (5'-aga tga gta aaa cgc agac c-3'). O método se baseia no princípio de que dideoxynucleotídeos marcados com diferentes fluorocromos, quando incorporados, cessam a síntese da molécula, o que ocorre logo após a última base na posição 3' do *primer*, que corresponde a posição da base polimórfica. Este método foi realizado no seqüenciador automático (MegaBACE™ 1000 DNA Analysis Systems).

Clonagem TOPO

Utilizamos o *TOPO TA Cloning® kit (Invitrogen)* para confirmar um aparente desequilíbrio de ligação entre dois polimorfismos do promotor do *TCOF1*: c.-948G→A e -1025G→C. O produto de PCR foi misturado aos reagentes de ligação e a reação foi incubada à temperatura ambiente por 30 minutos. Após a incubação, 2 µL do produto de ligação foram eletroporados com 50 µL de bactérias *OneShot®TOP10 Eletrocompetent E.coli (Invitrogen)* e essas foram recuperadas em tubos de 15 mL com meio SOC, a 37°C, por uma hora, com agitação de 200 rpm. Logo após, 60 µL das bactérias recuperadas foram plaqueadas meio seletivo. O PCR das colônias positivas foi feito com os primers universais M13F/R, fornecidos no próprio kit de clonagem. O PCR de colônias foi analisado em gel de agarose 1,5%, corado com brometo de etídeo e visualizado sob

luz ultra-violeta. Os produtos de PCR de colônias com o inserto do tamanho esperado foram purificados e seqüenciados.

Triagem de polimorfismo por digestão com endonucleases

Uma vez que detectamos desequilíbrio de ligação para os polimorfismos -948G>A e -1025G>C, utilizamos endonucleases de restrição para triar somente a variante -1025G>C em 50 indivíduos normais. O produto de PCR (Prom4) foi digerido a 37°C por 1h, após 80°C por 15 min, com a endonuclease de restrição *HincII* (*New England Biolabs*). O produto digerido foi analisado em gel de acrilamida 8% com o uso do sistema de eletroforese vertical *Hoefer miniVE Electrophoresis Systems*[®] (*Amersham Pharmacia Biotech Inc*).

Sequenciamento

Para o seqüenciamento, os produtos de PCR (8µL) foram purificados por meio de adição de 10u da enzima Exonuclease I (EXO, *USB*) e de 5u da enzima *Shrimp Alkaline Phosphatase* (SAP, *Amersham Biosciences*), seguida de incubação a 37°C por 30 minutos e inativação da enzima a 80°C por 15 minutos. Também utilizamos os kits de purificação *Quiaquick*[®] *PCR Purification Kit* (*Qiagen*). O produto de PCR purificado foi quantificado por meio de eletroforese em gel de agarose 1,5%, com utilização do marcador de peso molecular *Low DNA Mass*[™] *Ladder* (*Invitrogen*). O seqüenciamento foi realizado com 30-100ng de produto de PCR, em ambas as direções segundo o protocolo do kit *DYEnamic IT Dye Terminator Cycle Sequencing for MegaBACE* (*Amersham Biosciences, GE Healthcare*). A reação foi realizada em 25 ciclos de desnaturação à 95°C por 10 segundos, *annealing* a 50°C por 10 segundos e extensão à 60°C por 1 minuto. A precipitação dos produtos foi realizada com acetato de amônia 1,5mM e etanol. Os produtos foram analisados no seqüenciador automático *MegaBACE*[™] *1000 DNA Analysis Systems* (*Amersham Biosciences, GE Healthcare*).

Nomenclatura de mutações

Todas as mutações detectadas neste trabalho foram nomeadas de acordo com a isoforma mais abundante do gene *TCOF1* (AY460334). Para as mutações do promotor, nomeamos as mutações em relação ao início da tradução (primeiro ATG), de acordo com Den Dunnen & Antonarakis (2000) (<http://www.genomic.unimelb.edu.au/mdi/mutnomen>). Consideramos como +1 a posição 109000247 da seqüência de referência NT_029289.

Estabelecimento das frequências dos polimorfismos

Verificamos se os polimorfismos da região promotora estavam em equilíbrio de Hardy-Weinberg, utilizando o teste qui-quadrado com valor crítico de 0,05.

2.3.5 Ferramentas de bio-informática

Abaixo descreveremos as ferramentas mais utilizadas, geralmente disponíveis gratuitamente na internet:

UCSC Genome Browser:

Este website foi fundamental para o desenvolvimento desse trabalho. Além de fornecer a sequência genômica de diversos organismos, oferece diversas ferramentas, como BLAT (alinhamento de seqüências), PCR *in silico*, identificação de regiões conservadas entre diferentes espécies, identificação de ilhas CpGs, identificação de isoformas e EST (*expression sequence tags*) depositadas, dentre outras.

Identificação de regiões conservadas

Um programa bastante útil para identificação de regiões conservadas é o *VISTAgenome browser* (<http://www-gsd.lbl.gov/vista>). É uma ferramenta de visualização da genômica comparativa, no qual é possível visualizar a seqüência inteira do cromossomo de interesse, alinhada às seqüências ortólogas de outras espécies, identificando-se as seqüências com diferentes níveis de conservação (Frazer *et al.*, 2004; Brudno *et al.*, 2003). Também utilizamos a ferramenta *VISTA Enhancer Browser* (<http://enhancer.lbl.gov/>), que contém dois tipos de dados: (1) um banco de dados experimental de seqüências humanas não-codificantes, que foram testadas em camundongos transgênicos e caracterizadas como *enhancers*; (2) banco de dados computacional de todas as seqüências humanas não-codificantes do genoma, que foram identificadas como potencialmente regulatórias pelo grau de semelhança entre humano, rato e camundongo (Prabhakar *et al.* 2006).

Predição de promotores e de sítios de fatores de transcrição

Todos são de uso livre e se baseiam em matrizes de seqüências consenso de sítios de ligação de fatores de transcrição. O programa analisa uma determinada seqüência, verificando o

grau de homologia com essas matrizes, e, em geral apresenta um escore de similaridade. Essas análises orientam os estudos funcionais de elementos regulatórios. Os mais utilizados foram:

- GENOMATIX: *Genomatix Software GmbH* (<http://www.genomatix.de/>).
- TESS (www.cbil.upenn.edu/cgi-bin/tess/tess33?_if=1&RQ=WELCOME)
- Tfsearch (www.cbrc.jp/research/db/TFSEARCH.html)

2.3.6 Estudo funcional da região promotora: ensaio da luciferase

Construção dos plasmídeos

Para realização do ensaio da luciferase, utilizamos o vetor *pGL3-Basic Vector* (Anexo II, *Promega*) que não possui promotor eucariótico e possibilita a clonagem de seqüências regulatórias adjacentes ao gene repórter *luc+*. O gene *luc+* expressa luciferase quando a seqüência regulatória em teste é ativada nas células transfectadas com esse plasmídeo. Fizemos 7 construções dos primeiros 1200 pb com os seguintes insertos no vetor pGL3-Basic:

- 1) pGL3-1200wt: 1200 pb com seqüência selvagem (-1241 a +56);
- 2) pGL3-1200wtAS: 1200 pb com seqüência selvagem invertida;
- 3) pGL3-600wt: 600 pb com seqüência selvagem (-571 a +56);
- 4) pGL3-600pol: 600 pb contendo o polimorfismo c.-346C→T (-571 a +56);
- 5) pGL3-400wt: 400 pb com seqüência selvagem (-369 a +56);
- 6) pGL3-300wt: 300 pb com seqüência selvagem (-258 a +56);
- 7) pGL3-200wt: 200 pb com seqüência selvagem (-151 a +56).

Os insertos dos vetores foram construídos amplificando-se a região de um indivíduo controle com *primers* com cauda polylinker (Tabela 2) para as enzimas de restrição *Bgl II* e *Hind III* (*New England Biolabs*). Os produtos de PCR purificado e 1µL do vetor pGL3-Basic foram digeridos com as enzimas de restrição *Bgl II* e *Hind III* (37° C por 15h). O vetor foi desfosforilado após a digestão e, juntamente com os insertos digeridos, foi aplicado em gel de agarose 0,8%, submetidos a eletroforese horizontal, extraídos do gel e purificados com kits *Quiaquick® PCR Gel Purification Kit* (*Qiagen*). Insertos e vetor digeridos/purificados foram ligados na proporção de 3:1.

Acrescentou-se 1u da enzima T4 ligase (*Invitrogen*) e 1X tampão (5x *Invitrogen*), num volume final de 10 µL. A reação de ligação (2 µL) foi eletroporada em bactérias eletrocompetentes (*Top10*, *Invitrogen*) e estas foram plaqueadas em meio seletivo. Realizamos PCR das colônias que cresceram com *primers* do próprio vetor pGL3-Basic. Os produtos de PCR de colônias foram purificados por EXO/SAP e seqüenciadas com os *primers* do vetor e com *primers* intermediários do próprio inserto para verificar a integridade da construção.

A construção pGL3-1200wt serviu de molde para outras construções, sendo digerida com a enzima *SmaI* (*New England Biolabs*) para pGL3-300wt, e sendo amplificada com primers com cauda *polylinker* para *HindIII* (*New England Biolabs*), para pGL3-400wt e pGL3-200wt.

Tabela 2: Primers utilizados para construções com o plasmídeos pGL3.

Nome	Tamanho	Seqüência das bases (5'-3')
pGL3-1200wt-F	1297pb	GAAGATCTGGACCTCCTAGGGTTAGGG
pGL3-1200wt-R	Reverso de ambos	CCCAAGCTTCAGATGGTGGTAGATCAGG
pGL3-600wt-F	627pb	GAAGATCTTTTGAATCCCCCTAGCC
pGL3-400wt-F	314pb	GAAGATCTCAGAGATGAGTAAAACGCAGAC
pGL3-200wt-F	207pb	GAAGATCTCGGAGAGGGGACTACGTTTC

Cultura de células e ensaio de transfecção e co-transfecção com lipofectamina

Realizamos nossos ensaios em hepatócitos humanos imortalizados (HepG2 – HB8065 – ATCC), cultivados em meio de cultura DMEM, 10% de soro fetal bovino (*Gibco*) e 1% de antibiótico, mantidos em estufa a 37°C (5% CO₂). As células foram privadas de soro e antibiótico 24h antes da transfecção, tendo sido plaqueadas em placas de 6 ou 12 poços, em quantidade de 1x10⁵ e 2,5x10⁵ por poço, respectivamente.

Para o ensaio de transfecção, utilizamos 1µg das construções com pGL3 e 0,5µg de pRL-TK (*Promega*), em duplicatas, em placas de 12 poços, com *Lipofectamine* 2000 (*Invitrogen*), de acordo com o protocolo do fabricante. Para o ensaio de co-transfecção, utilizamos somente o vetor de expressão de YY1 (CMV-YY1) e o vetor pGL3-1200wt. Utilizamos 0,5µg de pRL-TK, 2,5µg de pGL3-1200wt e 0,1 a 1µg de CMV-YY1 por poço, em uma placa de 6 poços.

O vetor CMV-YY1 foi cedido pelo grupo do Dr. Peter Baumeister (utilizado em Seto *et al.*, 1991) e foi sido seqüenciado para verificar sua integridade. Os *primers* utilizados foram desenhados (5'-3'): YY1F- gga gac cat cga gac cac ag e YY1R-gtc tgc acc agg atc acc tc.

O vetor pRL-TK possui o gene da *Renilla* luciferase, que é controlado pelo promotor constitutivo HSV-tk. Esse vetor é usado como um controle interno para normalizar a eficiência de transfecção. As células são lisadas 48h após a transfecção, e a expressão do gene repórter é analisada usando o Dual-Luciferase Reporter Assay System (*Promega*) num luminômetro de placa (LB96V *Microlumat Plus* - EG&Berthold). A atividade de cada vetor foi considerada com a média de cada duplicata e os valores relativos da atividade da luciferase foram calculados em relação à atividade do vetor vazio pGL3-Basic ou pGL3-1200wt. Os dados de luciferase foram gerados pela media dos resultados de ao menos três experimentos independentes.

2.3.7 Identificação de fatores de transcrição da região promotora

Eletrophoretic Mobility Shift Assay (EMSA)

Preparamos extrato nuclear de HepG2 de acordo com o protocolo de Read (1998) e a concentração de proteínas foi determinada usando o reagente Bradford (Bio-Rad). Duas sondas de 51pb, uma com a seqüência selvagem e outra com o polimorfismo c.-346C>T, correspondendo à seqüência da posição -367 a -316bp do promotor do *TCOF1*, foram preparadas anelando cada uma com sua respectiva sonda complementar (Tabela 3). Antes do anelamento, as sondas senso foram marcadas com [γ^{32} P]ATP e T4 polynucleotide kinase (*New England Biolabs*), seguido de purificação usando colunas Sephadex G50 (*Amersham Pharmacia Biotech*). As reações de ligação de proteínas foram realizadas em 20 ou 25 μ L de tampão (25mM Tris-HCl, pH 7,5, 50mM KCl, 1mM dithiothreitol, 5% glicerol), com 300-500 cpm de sonda marcada, 1 μ g de poly (dl:dC) e 5 μ g de extrato nuclear de HepG2. As misturas foram incubadas por 30 min a temperatura ambiente na presença ou na ausência de competidores específicos não-marcados. Para os ensaios de supershift, a reação de ligação foi pré-incubada com anticorpo monoclonal humano de Sp1, Sp3 e anticorpo policlonal humano YY1 (*Santa Cruz Biotechnology, Inc*). Os anticorpos Sp1 e Sp3 foram cedidos pela Dra. Bachega, Laboratório de Hormônios e Genética Molecular LIM/42, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo.

Os complexos proteína-DNA foram analisados em eletroforese vertical (gel de acrilamida 6% não-desnaturante) com o tampão de corrida 0,5X Tris-borate-EDTA. Para quantificar a afinidade proteína-DNA de cada genótipo, comparamos a intensidade das bandas dos géis de cada complexo com a sua banda de sonda livre, usando o STORM scanner (Amersham Pharmacia Biotech) e o programa *Image Quant*®.

Tabela 3: Oligonucleotídeos usados para o ensaio de EMSA

Nome das sondas	Sequências das sondas F (5'- 3')
Selvagem (Wild-type probe)	gagatgagtaaaacgcagacccagccatcagagaattaacaaggaagataa
Polimórfica (Polymorphic probe -346C>T)	gagatgagtaaaacgcagacccagccatcagagaattaacaaggaagataa

Triagem de biblioteca de cDNA com o ensaio de Mono-híbrido

O ensaio de mono-híbrido permite identificar as proteínas que interagem com seqüências regulatórias alvo *in vitro*, com a utilização de uma biblioteca de cDNAs (Vidal, M. e Legrain, P. 1999). A linhagem de leveduras Y187 foi transformada com dois vetores: um contendo a seqüência regulatória alvo (pHIS2) e outro, contendo os cDNAs da biblioteca (pGADT7-Rec2 ou pACT2). Os cDNAs são transcritos fusionados com o domínio de ativação da GAL4. Caso uma das proteínas da biblioteca interaja com a seqüência regulatória alvo, o domínio de ativação da GAL4 se ligará à sua seqüência de ligação ao DNA (GAL4 DNA-binding domain), presente no vetor pHIS2, o que será suficiente para recrutar a maquinaria de transcrição para a região promotora mínima *Pmin HIS3*, promovendo, então, a expressão do gene repórter *HIS3*. Assim, a interação DNA-proteína é observada em colônias que são capazes de sobreviver em meio seletivo para Histidina. Para confirmação de interação DNA-proteína, também utilizamos a linhagem AH109, disponível em nosso laboratório.

Utilizamos um kit para construção de biblioteca e triagem da mesma num único ensaio (*Matchmaker™ One-Hybrid Library Construction & Screening kit*, Clontech). Utilizamos a linhagem de *Saccharomyces Cerevisiae* Y187, auxotrófica para histidina, e as condições de cultivo, de preparação e de transformação de leveduras competentes foram realizadas de acordo com o protocolo do fabricante. Para sequenciamento de colônias Y187, purificamos os plasmídeos de leveduras (*Isolation of plasmid DNA from yeast using the QIAprep® Spin Miniprep Kit*, Qiagen) e

posteriormente os transformamos em bactérias eletrocompetentes (*ElectroMAX DH10B Cells*, *Invitrogen*). Purificamos os plasmídeos (*QIAprep® Spin Miniprep kit*, *Qiagen*) das colônias que cresceram em meio seletivo e prosseguimos com o sequenciamento direto desses plasmídeos (250ng/reação) no seqüenciador automático MegaBACE 1000 (GE-Healthcare Bio-Sciences).

Construção dos plasmídeos pHIS2:

Os *primers* para as regiões R1, R1 REST, R1 SNP, R1 PMIN, R2 e R3 foram desenhados com o programa *Primer3* (seqüências na Tabela 4), tendo sido adicionados às suas seqüências caudas 5' com sítios para digestão com as endonucleases de restrição *EcoRI* e *MluI* (*Roche Molecular Biochemicals*). Amplificamos o DNA genômico de um indivíduo controle e os produtos de PCR foram digeridos, purificados e clonados no vetor pHIS2. Todos os plasmídeos foram seqüenciados para confirmar sua integridade (*primers* do vetor pHIS2 – Tabela 4). Transformamos leveduras competentes com 700ng dos plasmídeos purificados, de acordo com protocolo do kit de transformação.

Tabela 4: *Primers* utilizados nas construções dos plasmídeos com gene repórter (pHIS2).

Nome do <i>primer</i>	Seqüência (5'-3')	Aplicação
pHIS2-R1_F	Cgg aat tcT CAT GTG CCC TCA TTT GTT C	Construção R1
pHIS2-R1_R	Cga cgc gtA CTT CCC TCG CGC CCT TA	Construção R1
pHIS2-R2_F	Cgg aat tcA AAA GGG CCT GAC TCC TGA C	Construção R2
pHIS2-R2_R	Cga cgc gtG TGG GGA AGG AAC TGG ATT G	Construção R2
pHIS2-R3_F	Cgg aat tcC ATT GAC TAG CCG GGT GAT T	Construção R3
pHIS2-R3_R	Cga cgc gtG AAA GTG TGG TGA TGC ACA AA	Construção R3
pHIS2-R1_PrMin-F	Cgg aat tcG CGG TAT TGT TGA TCA CGA GTC	Construção R1 PMIN
pHIS2-R1_SNP-F	Cgg aat tcT GGC AAT TCG TGT CTG GTT C	Construção R1 SNP
pHIS2-R1_SNP-R	Cga cgc gtC CAT TGG CGC GGA GCA T	Construção R1 SNP
pHIS2-R1_REST-F	Cgg aat tcT CAT GTG CCC TCA TTT GTT C	Construção R1 RESTO
pHIS2-R1_REST-R	Cga cgc gtG TTT ATT CAT TGA TTC ATT CGC	Construção R1 RESTO
pHIS2-F	AAGGCGATTAAGTTGGGTAACG	Sequenciamento dos insertos clonados em pHIS2
pHIS2-R	TAGGGCTTTCTGCTCTGTCATC	Sequenciamento dos insertos clonados em pHIS2

Padronizamos a quantidade de 3-AT (*3-amino-1,2,4-triazol*, *Sigma-Aldrich*) para cada uma das regiões (Capítulo 5 - Resultados). 3-AT é um inibidor competitivo da proteína do gene *HIS3*

(His3p), e é usado para suprimir o crescimento de colônias capazes de sobreviverem em meio seletivo (-His) apenas com a fraca expressão basal ("*leaky expression*"). A concentração ideal para realização do ensaio é a menor capaz de suprimir o *background* de colônias com expressão basal. A quantidade de 3-AT recomendada pelo fabricante para qualquer construção varia de 5 a 60 mM; no entanto, padronizamos o vetor controle p53HIS2 (mapa no Anexo2) em 130 mM, cuja expressão "leaky" geralmente é suprimida com 5-10 mM.

Transposição da biblioteca de transcritos de E26(*day 26 human embryo*):

Os cDNAs de E26 foram amplificados com *primers* (pAMP1-T7 e pAMP1-3'; Tabela 5) para as regiões de homologia entre ambos os vetores (pGADT7-Rec2 e pAMP1), à exceção da porção 3' dos transcritos. O vetor pGADT7-Rec2 provém do kit *Matchmaker Library Construction and Screening kit (Clontech)*. Para que os produtos amplificados da biblioteca de cDNAs fossem capazes de recombinar na sua porção 3' com o vetor pGADT7-Rec2, desenhamos um oligonucleotídeo "adaptador" (40pb), cuja seqüência molde e complementar foram aneladas. Esse oligo dupla-fita contém a região 3' adaptadora dos cDNAs da biblioteca e parte do CDSIII terminus do vetor pGADT7-Rec2. As leveduras competentes foram transformadas (método de Acetato de Lítio) com o produto de PCR dos cDNAs purificado (72ng), com o oligo "adaptador" (72nmol) e com o vetor pGADT7-Rec2 linearizado (1000ng). As leveduras então recombinaram as regiões de homologia e geraram plasmídeos pGART7-Rec2 contendo os cDNAs da biblioteca de E26. As leveduras transformadas foram plaqueadas em meio seletivo para leucina (SD agar/-Leu). As colônias positivas foram raspadas dessas placas, recuperadas em um único recipiente com meio seletivo líquido (SD/-Leu) por 30 min a 30°C, e congeladas em glicerol. Alguns clones positivos da transposição da biblioteca foram seqüenciados: utilizamos os *primers* do vetor pGADT7-Rec2, que flanqueiam a região de recombinação (Tabela 5) e o kit BigDye™ Terminator Cycle Sequencing (Applied Biosystems), sendo as amostras analisadas no seqüenciador automático ABI Prism™ 377 (Applied Biosystems).

Tabela 5: *Primers* utilizados para transposição da biblioteca E26.

Nome do <i>primer</i>	Seqüência (5'-3')	Utilização
pAMP1- T7 promoter	GCG CTA ATA CGA CTC ACT ATA G	Transposição E26
pAMP1- 3' Region	AAG CAG TGG TAA CAA CGC AG	
Oligo adaptador -F	CTG CGT TGT TAC CAC TGC TTC ATG TCG GCC GCC TCG GCC T	Para recombinação homóloga da porção 3'cDNAs e pGADT7 – Rec2
Oligo adaptador -R	AGG CCG AGG CGG CCG ACA TGA AGC AGT GGT AAC AAC GCA G	
pGADT7-Rec2_F	CACGGCTAGTAAATTGATGATGG	Para seqüenciar insertos clonados em pGADT7-Rec2
pGADT7-Rec2_R	TTACATAAAAGAAGGCCAAAACGATG	

Triagem da biblioteca de transcritos de testículos humanos:

Para realização do ensaio de mono-híbrido propriamente dito, utilizamos uma biblioteca comercial de cDNAs de testículos humanos (*Pre-transformed Matchmaker cDNA AD-library*, Clontech). A biblioteca foi re-estriada em meio seletivo, amplificada em aproximadamente 3×10^6 clones e posteriormente purificada (Qiaprep spin maxiprep kit, QIAGEN). Confirmamos a expressão do gene *TCOF1* e *YY1* nessa biblioteca pela amplificação com *primers* específicos (5'-3'): *TCOF1*_25F-gtt cga agg ctg tgg gaa ac, *TCOF1*_25R-CCA CAG GGG ACA CCA GAG, para *YY1* vide co-transfecção CMV-YY1). Procedemos com a transformação seqüencial: primeiramente transformamos cada construção pHIS2 (pHIS2-R1, R1 RESTO, R2 ou R3) e, posteriormente, cada construção com 10µg da biblioteca comercial. Os clones positivos, capazes de sobreviver em meio seletivo para histidina (SD/-His) foram re-plaqueados no mesmo meio seletivo e somente aqueles que repetiram o fenótipo foram seqüenciados, com *primers* do vetor pACT2 (F-5'-TTACATAAAAGAAGGCCAAAACGATG-3' e R-5'-ctc ggc tag taa aat tga tga tgg-3').

2.3.8 Quantificação de expressão gênica

Síntese de cDNA

A síntese de cDNA das 41 amostras de pacientes, controles e linhagens celulares foi realizada de acordo com o protocolo da enzima transcriptase reversa SuperScript II (SSII, Gibco BRL) com algumas modificações. Primeiramente, tratamos 1µg de RNA com DNase (Promega,

30min a 37° C) e uma alíquota desse RNA é separado para reação-controle com primers intrônicos (gene *MLH1*). Essa reação tem por objetivo confirmar a eliminação de qualquer vestígio de DNA após o tratamento com DNase, não podendo haver nenhuma amplificação. O RNA tratado com DNase é então misturado a 25ng de oligo (dT) 12-18, 0,5mM de dNTP e o volume acertado para 12µL. Essa mistura é aquecida a 65° C por 5 minutos rapidamente resfriada em gelo. Em seguida, a reação é processada com tampão da enzima SSII 1X, 10mM de DTT, 10u de SSII e 2u de RNase OUT (Invitrogen), a 42° C, durante 60 minutos, e inativada a 70°C por 15 minutos.

PCR em tempo real (RT-PCR)

A metodologia utilizada para a análise foi o SYBR Green (*Applied Biosystems*) (Morrison *et al.* 1998). O SYBR Green é um agente intercalante que, quando se liga às moléculas de fita dupla, emite uma fluorescência. Como ele é adicionado ao PCR, à medida que novas moléculas de fita dupla são sintetizadas ao passar dos ciclos da reação, a fluorescência aumenta. O aparelho *ABI Prism 7700 Sequence Detection System (Applied Biosystems)* detecta a fluorescência crescente: quanto mais tardio o ciclo em que a fluorescência é detectada, menor é o nível de expressão do gene analisado. O valor quantitativo de expressão, denominado C_T (*Cycle Threshold*), é obtido durante a reação: corresponde ao número do ciclo em que a curva de amplificação exponencial se encontra na metade. Esse parâmetro é analisado pelo programa do aparelho. Desenhamos os *primers* para esse ensaio no programa *Primer Express 2.0*, do próprio aparelho *ABI Prism 7500* (Tabela 6).

Fizemos as reações em placas de 96 poços (*MicroAmp Optical, Applied Biosystems*) em duplicatas, cada uma com volume final de 20 µL, contendo 10µL de *SYBR Green PCR Master Mix (Applied Biosystems)*, 50nM a 800nM de cada primer e 0,01 a 100ng de cDNA. As concentrações dos primers foram padronizadas considerando dois fatores: (1) utilização das menores concentrações, a fim de evitar a formação de dímeros de primers; (2) obtenção dos menores valores de C_T possíveis. Com relação à quantidade de cDNA, recomenda-se o suficiente para obter o menor C_T .

Tabela 6: Primers para RT-PCR: seqüências, concentrações e eficiências de amplificação.

Gene	Forward (5'-3')	Reverse (5'-3')	Concentração final	Eficiência de Amplificação
<i>TCOF1</i>	GCCCCTGGAAAAGTTGTCAC	GGTTTCTCACTGGTGGCTTCAC	200nM	1,83
<i>HMBS</i>	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC	200nM	1,86
<i>GAPDH</i>	ACCCACTCCTCCACCTTTGA	CTGTTGCTGTAGCCAAATTCGT	400nM	1,94
<i>HPRT1</i>	GAACGTCTTGCTCGAGATGTGA	TCCAGCAGGTCAGCAAAGAAT	400nM	1,86
<i>BCRP</i>	CCTTCGACGTCAATAACAAGGAT	CCTGCGATGGCGTTCAC	400nM	1,85

Quanto à ciclagem, utilizamos o padrão recomendado: 50° C por 2min, 95° C por 10min, 40 ciclos de 95° C por 15s e 60° C por 1min. Após a amplificação, as amostras são submetidas a um ciclo de dissociação dos primers, em que a temperatura se eleva gradualmente de 60° C para 95° C a 0,2° C por segundo, durante 20 min. Essa última etapa é fundamental para confirmar a especificidade da reação, dada pela curva de dissociação. Os produtos da padronização são também analisados em gel de acrilamida 8%.

As eficiências de amplificação (E), tanto para o gene alvo (*TCOF1*) quanto para os genes normalizadores (*HMBS*, *GAPDH*, *BCRP* e *HPRT1*), são calculadas de acordo com a equação $E = 10^{-(-1/\text{slope})}$. *Slope* corresponde à inclinação da reta gerada pelos pontos do gráfico em que, no eixo y, representamos a média dos C_T s e, no eixo x, o *log* das 5 diferentes quantidades de cDNA (0,01, 0,1, 1, 10 e 100ng). Essas quantidades de cDNA são obtidas pela diluição seriada da amostra original, que foi sintetizada a partir de 1µg de RNA total.

A expressão diferencial dos transcritos de interesse (*TCOF1*) foi determinada pelo método de quantificação relativa em relação aos genes normalizadores. Utilizamos como amostra referência o cDNA sintetizado a partir de um *pool* de RNAs de 12 indivíduos normais. Para o cálculo da expressão diferencial foi utilizado o modelo matemático proposto por Pfaffel (2001):

$$R = (E \text{ alvo})^{\Delta C_{T\text{alvo}}} / (E \text{ normalizador})^{\Delta C_{T\text{normalizador}}}$$

E alvo corresponde à eficiência de amplificação dos primers do gene alvo e *E normalizador*, do gene normalizador. ΔC_T corresponde à diferença entre o C_T obtido para a amostra em análise (pacientes, controles e linhagens celulares) e o C_T obtido para a amostra referência, tanto para o gene alvo, quanto para o controle endógeno.

Utilizamos quatro genes normalizadores para análise dos dados de expressão: *HMBS*, *HPRT1*, *GAPDH* e *BCRP*. Empregamos uma ferramenta desenvolvida para análise dos dados de RT-PCR no *Microsoft Excel*, o *geNorm* (Vandesompele *et al.*, 2002; <http://medgen.ugent.be/~jvdesomp/genorm/>). Esse aplicativo modificado determina quais são os genes normalizadores mais estáveis em um conjunto de amostras de cDNA e, além disso, calcula um fator de normalização (FN) da expressão gênica para cada amostra com base na média geométrica dos valores obtidos para os diferentes normalizadores.

A estabilidade dos genes normalizadores é determinada pelo princípio de que dois controles endógenos ideais possuem razões de expressão idênticas em todas as amostras de cDNA a serem analisadas, independente das condições experimentais ou tipos celulares. O *geNorm* então avalia a razão de expressão de cada gene normalizador em relação aos demais normalizadores, sempre aos pares, permitindo a exclusão dos controles endógenos menos estáveis. O valor de estabilidade de corte é $M < 1,5$ (*cut off*), ou seja, acima desse valor, o normalizador é instável. Os valores apresentados nos gráficos (Capítulo 6) são produtos da normalização dos $\Delta\Delta C_{Ts}$ de *TCOF1* pelos FN calculados a partir dos $\Delta\Delta C_{Ts}$ dos controles endógenos.

2.3.9 Sequenciamento de produtos de PCR tratado com bissulfito de sódio

Para verificar o padrão de metilação da ilha CpG do *TCOF1*, tratamos DNA genômico de controles e de pacientes com bissulfito de sódio para posterior sequenciamento. O bissulfito de sódio induz a conversão de todas as citosinas do genoma para uracilas; porém, aquelas que estão metiladas (5-metil-citosinas) permanecem inalteradas (Frommer *et al.*, 1992). O DNA tratado é amplificado com primers especialmente desenhados para bases convertidas, que não podem anelar em regiões com dinucleotídeos CGs. Os *primers* amplificam somente uma das fitas do DNA, pois não há anelamento na fita complementar. O sequenciamento da região possibilita a identificação das bases metiladas.

Utilizamos um protocolo de tratamento com bissulfito modificado de Goldenberg *et al.* (2004). As amostras de DNA foram quantificadas no aparelho NanoDrop e analisadas em gel de agarose 0,8% para confirmar integridade. 500 a 1000 ng de cada amostra de DNA foram tratados.

Dividimos a ilha CpG do *TCOF1* em dois *amplicons* (CpG1 e CpG2; Capítulo 6 - *Methods*). Realizamos a reação de PCR nas seguintes condições: para um volume final de 25µL, adicionamos 1X PCR Buffer (*Applied Biosystems*), 1,5 mM de MgCl₂ (*Applied Biosystems*), 200µM de dNTP (*Amersham*), 0,32µM de cada primer, 1U de *Taq GOLD* (*Applied Biosystems*) e 1µL de DNA tratado com bissulfito. A seguinte ciclagem foi aplicada para as duas reações, com as mesmas temperaturas de anelamento: 94°C por 12 min; 94°C por 3 min, 53°C por 3 min, 72°C por 2 min; 94°C por 3 min, 55°C por 3 min e 72°C por 2 min, 5 ciclos; 94°C por 1 min, 54°C por 1 min e 72°C por 1 min, 35 ciclos. Para a reação *nested*, utilizamos 1µL da diluição 1:10 da primeira reação de PCR. Na tabela 7 descrevemos os primers utilizados para PCR e sequenciamento. O produto de PCR foi visualizado em gel de acrilamida 8%, corado com nitrato de prata.

Os produtos de PCR foram clonados no vetor *pGEM® - T Easy* (*Promega*), utilizando o kit de clonagem do próprio plasmídeo. Bactérias eletrocompetentes (*ElectroMAX DH10B Cells*, *Invitrogen*) foram transformadas com 1,5 µL da reação de ligação e plaqueadas em meio seletivo. Realizamos PCR das colônias positivas com os *primers* universais do vetor (M13F e M13R; Tabela 7) e os produtos foram analisados em gel de agarose 1%.

Seqüenciamos produtos de PCR ou plasmídeos purificados (*QIAprep miniprep spin kit*, *QIAGEN*) das colônias positivas com o inserto do tamanho esperado para as ilhas CpG1 e 2. Para a análise da seqüência completa de ambas as ilhas utilizamos os primers M13F e M13R, além do CpG5R, que é específico para a ilha CpG1. As reações foram feitas com o kit de sequenciamento BigDye V.3.1 (*Applied Biosystems*) e aplicadas no seqüenciador automático ABI 3100 (*Applied Biosystems*).

Para cada indivíduo, seqüenciamos ao menos quatro clones de cada ilha, para que seja provável que tenhamos analisado os dois alelos de cada indivíduo. A porcentagem de metilação para cada amostra foi calculada como o número de dinucleotídeos CpGs metilados sobre o número total de CpGs analisados nas colônias positivas.

Tabela 7: Primers (5'-3') utilizados para amplificação e sequenciamento de DNAs tratados com bissulfito.

Reação inicial	Reação <i>NESTED</i>	Sequenciamento
CpG4_F: GTGTGAAATGGATAAATGAGTAGGT	Ilha CpG 1	Ilha CpG1
	CpG3_F: GGATTTAGAGAAAGATTTAGATTAGGG	M13_F: GTAAAACGACGGCCAGT M13_R: GGAAACAGCTATGACCATG
	CpG1_R: ACACACTTTCTAACCCAACCCACTAC	CpG5_R: CTCTACTCCTTCACTTCCC
CpG3_R: CTAAAAACCTAAATTCCAACCTCAACT	Ilha CpG 2	Ilha CpG2
	CpG2_F: TTTTTGGGTTTTTATTTTTTTTATT	M13_F: GTAAAACGACGGCCAGT M13_R: GGAAACAGCTATGACCATG
	CpG2_novo_F: GGGAGTTATTTTTTTTGATTATTA	
	CpG2_R: AACTATAAAATCCCTTTTCCAAAC	

2.3.10 Estudo de Ligação

Realizamos análise de segregação com marcadores polimórficos tipo microssatélite para excluir e mapear regiões candidatas para a síndrome Aurículo-Condilar. O DNA genômico foi amplificado por *PCR* com *primers* fluorescentes *ABI PRISM™ Linkage Mapping Set Version 2.0*, (*Perkin-Elmer, Applied Biosystems*), marcados com as fluorescência FAM, HEX ou NED. Os marcadores desse kit são de repetições de dinucleotídeos, em sua maioria com alto grau de informação (heterozigosidade), com uma distância média entre um e outro de aproximadamente 10cM.

A reação de *PCR* utilizou tampão 1x e 0,7u de Taq DNA polimerase (*10X Taq DNA Polimerase Buffer, Invitrogen*), MgCl₂ 1,5mM, 0,25mM dNTP (*Amersham Biosciences*), 0,3μM de *primers* fluorescentes e 50-100ng de DNA genômico. A reação obedeceu a seguinte ciclagem: 94°C por 4min, trinta ciclos de 94°C por 15s, 55°C por 15s, 72°C por 30s e, por fim, 72°C por 15min. Os produtos de *PCR* foram analisados no seqüenciador automático MegaBACE 1000 (*Amersham, GE Healthcare, Little Chalfont, United Kingdom*), de acordo com o protocolo do fabricante, e a análise dos dados foi feita com o programa Megabace Genetic Profiler® (*Amersham, GE Healthcare*). Ao todo, analisamos 370 marcadores.

Para o cálculo de *lod score* de dois pontos, utilizamos o programa MLINK (*FASTLINK Version 5.1*; <http://linkage.rockefeller.edu/soft/>). Os parâmetros para a análise consideraram padrão de herança autossômico dominante, penetrância completa e frequência do alelo de risco igual a 0,001. As frequências de recombinação foram consideradas como $1/N$, N representa o número de alelos diferentes no heredograma. Para cálculo de *lod score multipoint* (vários pontos), utilizamos o programa Merlin 1.0.1 (Abecasis et al, 2002). Para cálculo dos *lod scores* finais, estabelecemos as frequências dos alelos para os marcadores D1S2726, D1S252, D1S498, e D1S484 em 20 controles.

Capítulo 3

O Capítulo 3 foi dividido em duas partes:

3-I: Artigo publicado:

TCOF1 Mutation Database: Novel Mutation in the Alternatively Spliced Exon 6A and Update in Mutation Nomenclature

Alessandra Splendore, Roberto D. Fanganiello, Cibele Masotti, Lucas S.C. Morganti, and M. Rita Passos-Bueno. *HUMAN MUTATION* 25:429-434 (2005)

3-II: Artigo a ser submetido para publicação:

Triagem de mutações no gene TCOF1 e correlação genótipo-fenótipo

Cibele Masotti, Diogo Meyer, Alessandra Splendore e Maria Rita Passos-Bueno.

Observação:

Para os pacientes portadores da síndrome de Treacher Collins (STC) e seus familiares que concordaram em participar desse projeto de pesquisa oferecemos, além do aconselhamento genético, o diagnóstico molecular da síndrome, que consiste na triagem de mutações do gene *TCOF1*. A primeira parte do capítulo (3-I) é uma publicação conjunta do grupo de alunos com projetos relacionados à STC do nosso laboratório. Nesse trabalho, a minha contribuição foi identificar a mutação do exon 6A e participar da criação do banco de dados, além de mantê-lo atualizado. A segunda parte deste capítulo (3-II) se refere ao estudo das mutações encontradas nos indivíduos estudados durante a realização desse projeto, além da análise dos dados clínicos da amostra total de pacientes, grande parte da qual foi constituída nos projetos anteriores do nosso laboratório.

TCOF1 Mutation Database: Novel Mutation in the Alternatively Spliced Exon 6A and Update in Mutation Nomenclature

Alessandra Splendore, Roberto D. Fanganiello, Cibele Masotti, Lucas S.C. Morganti, and M. Rita Passos-Bueno.

Human Mutation 25:429-434 (2005)

Resumo

Recentemente foi descoberto um novo exon no *TCOF1* que, apesar de sofrer *splicing* alternativo, está presente na principal isoforma da proteína. Além disso, a maior parte das mutações publicadas para esse gene não estão adequadas às normas atuais para nomenclatura de mutação. Dadas essas observações, desenvolvemos um banco de dados *online* de mutações do *TCOF1*, no qual todas mutações descritas são renomeadas de acordo com recomendações padrão e com referência às sequências genômicas e de cDNA referência (www.genoma.ib.usp.br/TCOF1_database). Nós também apresentamos neste trabalho: 1) resultados da primeira triagem para grandes deleções no *TCOF1* usando *Southern blot* em pacientes sem mutação detectada por sequenciamento direto; 2) a identificação da primeira mutação patogênica no recém-descrito exon 6; e 3) análise estatística das mutações patogênicas e da distribuição do polimorfismo ao longo do gene.

DATABASES

TCOF1 Mutation Database: Novel Mutation in the Alternatively Spliced Exon 6A and Update in Mutation Nomenclature

Alessandra Splendore,² Roberto D. Fanganiello,¹ Cibele Masotti,¹ Lucas S.C. Morganti,¹ and M. Rita Passos-Bueno^{1*}

¹Universidade de São Paulo, Centro de Estudos do Genoma Humano, Instituto de Biociências, São Paulo, Brazil; ²Instituto Nacional de Câncer, Divisão de Genética, Rio de Janeiro, Brazil

Communicated by Marc Greenblatt

Recently, a novel exon was described in *TCOF1* that, although alternatively spliced, is included in the major protein isoform. In addition, most published mutations in this gene do not conform to current mutation nomenclature guidelines. Given these observations, we developed an online database of *TCOF1* mutations in which all the reported mutations are renamed according to standard recommendations and in reference to the genomic and novel cDNA reference sequences (www.genoma.ib.usp.br/TCOF1_database). We also report in this work: 1) results of the first screening for large deletions in *TCOF1* by Southern blot in patients without mutation detected by direct sequencing; 2) the identification of the first pathogenic mutation in the newly described exon 6A; and 3) statistical analysis of pathogenic mutations and polymorphism distribution throughout the gene. *Hum Mutat* 25:429–434, 2005. © 2005 Wiley-Liss, Inc.

KEY WORDS: *TCOF1*; Treacher Collins syndrome; database; mutation nomenclature

DATABASES:

TCOF1 – OMIM: 606847, 154500; GenBank: NM_000356, AY460334, NT_029289 (*Homo sapiens*), NM_011552 (*Mus musculus*), AJ296287 (*Canis familiaris*); UniGene: Hs.301266; HGMD: TCOF1 www.genoma.ib.usp.br/TCOF1_database (TCOF Mutation Database)

INTRODUCTION

Treacher Collins syndrome (TCS; MIM# 154500) is an autosomal dominant craniofacial disorder affecting the development of structures derived from the first and second branchial arches during early embryonic development [Poswillo, 1975]. The resulting clinical features comprise down-slanting palpebral fissures with lower eyelid coloboma, malar and maxillary hypoplasia, malformed ears, and conductive hearing loss due to atresia of the external ear canal [Franceschetti and Klein, 1949]. The estimated incidence of TCS is 1 in 50,000, with 60% of the cases resulting from sporadic mutations [Gorlin et al., 2001]. There is marked phenotypic variability among patients, ranging from perinatal death due to a compromised airway to those that go undetected by medical examination. Penetrance is considered to be very high, but the extent of clinical variability makes it sometimes difficult to differentiate a case of nonpenetrance from one with minimal expression of the syndrome. In such cases of suspected nonpenetrance, detailed radiological examination may reveal subclinical hypoplasia of the zygoma [Dixon et al., 1991a]. Only one case of nonpenetrance even after radiological examination has been reported to date [Dixon et al., 1994].

The gene responsible for this syndrome, *TCOF1* (MIM# 606847), was mapped to 5q31–q34 in 1991 and positionally cloned in 1996. [Dixon et al., 1991b; Jabs et al., 1991; The

Treacher Collins Syndrome Collaborative Group, 1996]. By that time, it was determined that *TCOF1* consisted of 26 exons, and exons 1 to 25 coded an open reading frame (ORF) of 4,233 bp. Gene expression was detected in various adult and fetal tissues.

Recently, by aligning the known version of the *TCOF1* cDNA (U40847) with the human expressed sequence tag (EST) data bank (dbEST), So et al. [2004] identified alternatively spliced isoforms of the *TCOF1* mRNA, some of which included two previously unidentified exons. The novel exons are exon 6A (231 bp between exons 6 and 7), included in the most common isoform, and exon 16A (108 bp between exons 16 and 17), included in a

Received 4 August 2004; accepted revised manuscript 18 November 2004.

*Correspondence to: Maria Rita Passos-Bueno, Instituto de Biociências, Universidade de São Paulo, Departamento de Biologia, Rua do Matao, 277, 05508-900, São Paulo, SP, Brazil.
E-mail: passos@ib.usp.br

Grant sponsors: Centros de Pesquisa, Inovação e Difusão/Fundação de Amparo à Pesquisa do Estado de São Paulo (CEPID/FAPESP); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

DOI 10.1002/humu.20159

Published online in Wiley InterScience (www.interscience.wiley.com).

minor isoform. They also reported isoforms lacking exon 19. Interestingly, alternative splicing of exon 19 had already been reported in the canine homolog of *TCOF1* [Haworth et al., 2001]. The new cDNA sequence corresponding to the major isoform (including exon 6A, but without 16A) was deposited in GenBank with accession number AY460334.

The protein coded by the *TCOF1* gene, treacle, has a three-domain structure, with nucleolar localization signals in the C-terminus, 11 repetitive motifs coded by exons 6A–16 that are subject to phosphorylation by casein kinase II, and an N-terminus with a LIS1 homology motif [Jones et al., 1999; Marsh et al., 1998; Winokur and Shiang, 1998; Emes and Ponting, 2001; So et al., 2004]. Treacle bears weak similarity to the mouse nucleolar phosphoprotein Nopp140, believed to shuttle between nucleolus and cytoplasm. It is becoming clear that the protein treacle plays a role in ribosome biogenesis, as it has been shown to interact with hNop56p in preribosomal ribonucleoprotein complexes [Hayano et al., 2003] and also with the RNA polymerase I transcription factor UBF [Valdez et al., 2004]. These observations suggest that treacle may serve as a link between rDNA gene transcription and pre-rRNA processing.

Pathogenic mutations in the *TCOF1* gene were reported throughout its coding region and, with the exception of a 5-bp deletion in exon 24, are usually family-specific. Recent studies proposed the existence of mutational hot spots in *TCOF1*, indicating that exons 23 and 24 are responsible for roughly 1 out of 3 of all known pathogenic mutations [Splendore et al., 2000, 2002]. The purpose of this article is to provide an updated database in which all the reported mutations are renamed according to standard recommendations and in reference to a novel cDNA reference sequence to accommodate the newly described exon 6A, which is present in the major mRNA transcript. Furthermore, we discuss the type and distribution of mutations in *TCOF1* reported to date, and evaluate their effect on the phenotype.

MUTATION NOMENCLATURE AND *TCOF1* DATABASE

Many original mutation descriptions in *TCOF1* were made before the first recommendation for mutation nomenclature was published [Antonarakis, 1998]. As a consequence, many authors have used different notation for describing mutations, and protein alterations were described without specifying the nucleotide change. In addition to this, all published mutations were described in relation to the DNA sequence published by Dixon et al. [1997], U40847, or by Wise et al. [1997], U76366, which were now found to correspond to a minor, shorter isoform of treacle. Therefore, an update in mutation nomenclature is required to accommodate the recently described exon 6A.

The current recommendations for mutation nomenclature suggest that we use the genomic sequence when available [den Dunnen and Antonarakis, 2000]. Therefore, all mutations were renamed according to the genomic reference (NT_029289) produced by the Human Genome Project, and the cDNA that corresponds to the major treacle isoform (AY460334).

Table 1 illustrates a few examples of how mutations and polymorphisms were originally described and the revised genomic, cDNA and protein nomenclature. A complete online mutation database can be found at www.genoma.ib.usp.br/TCOF1_database with all known pathogenic and polymorphic alterations in the coding region, and mutations of undetermined status in *TCOF1*, described as suggested in den Dunnen and Antonarakis [2000] and www.hgvs.org/mutnomen/. The database includes the

TABLE 1. Examples of the Mutation Nomenclature Update*

cDNA	Original description (U40847 or U76366)		Genomic		cDNA		Protein		Reference	
		Protein								
–	1347T/C	K749K	g.21371G>A		c.2478G>A		altered splicing		Gladwin et al. [1996]	
–	2526delAG	P/P449	g.17757T>C		c.1578T>C		p.Pro526Pro		Wise et al. [1997]	
–		TG842-843TA	g.21959-21960delAG		c.2757-2758delAG		p.Gly920fs		Wise et al. [1997]	
–		stop at 11 aa away								
–		W53R	g.3533T>C		c.157T>C		P.Trp53Arg		Edwards et al. [1997]	
–	nt4135del((GAAAAA)	–	g.40706-40710delGAAAA		c.4366-4370delGAAAA		p.Lys1457fs		Edwards et al. [1997]	
–	3987-3988insG	–	g.39047dupG		c.4218dupG		p.Ser1407fs		Splendore et al. [2000]	
–	nt4127-34del(A)	–	g.40705delA		c.4365delA		p.Glu1456fs		Ellis et al. [2002]	

*For the genomic nomenclature, nucleotide numbering is the same used in the database references file. For the cDNA sequence, nucleotide + 1 is the A of the ATG-translation initiation codon.

genomic and cDNA description, deduced change at the protein level, as well as the original description and reference for each mutation.

FREQUENCY AND DISTRIBUTION OF MUTATIONS

As of this report, there are 118 different pathogenic mutations described in 165 unrelated patients. The only frequently recurring mutation, a 5-bp deletion in exon 24, has a frequency of 17%. An insertion of a guanine residue in exon 23 was also reported in five unrelated patients in the Brazilian population. For both of these mutations, a founder effect is unlikely, as they have been associated with different haplotypes or, in some cases, have been shown to have occurred de novo [Edwards et al., 1997; Splendore et al., 2000; Splendore et al., 2003]. It is possible that some of these recurring mutations have different incidences in diverse populations.

The most common class of mutation in *TCOF1* is small deletions (1–41 bp) that result in a frameshift, detected in 67% of patients. According to recent changes in mutation nomenclature, all single-base insertions in *TCOF1* are now appropriately described as duplications, and probably result from polymerase slippage copying a nucleotide twice. A true insertion is found only associated with a deletion (indel). Small duplications are the second most common class of mutations in this gene. Nucleotide changes were divided in three functional categories: those that produce a premature stop codon, missense changes, and changes at splice sites (only one mutation occurring at a splice site—c.2478 G>A, first reported as K749K [Gladwin et al., 1996]—was confirmed experimentally to alter mRNA splicing, whereas for all the others the effect on splicing was inferred). The relative frequencies of each class of pathogenic mutation are presented in Figure 1.

It has been previously suggested that five exons: 10, 15, 16, 23, and 24, are more likely to bear pathogenic mutations, and thus should be regarded as hot spots [Splendore et al., 2000]. Identification of mutational hot spots may highlight mutational mechanisms and is also useful in setting up screening protocols. We compared the expected distribution of mutation per exon (as a function of exonic length) with the number of pathogenic mutations actually observed. The first and last three intronic bases adjacent to the examined exon (splice donor and acceptor sites) were also included in the analysis. Exon 24 was confirmed as a hot spot, with the observed number of mutations (35/165) greatly exceeding the expected (3.8/165; $\chi^2 = 177.31$; $P < 0.0001$). This observation probably reflects the highly repetitive nature of this exon, which is composed of 60% of adenines. Of its 32 coded amino acids, 18 are lysines (codons AAA and AAG), making it more prone to polymerase slippage during

DNA replication [Streisinger et al., 1966; Ellis et al., 2002]. Consistent with this hypothesis, practically all the mutations reported in exon 24 are deletions. On the other hand, none of the other putative hot spots was confirmed to display more pathogenic mutations than theoretically expected. Exon 23, which also has a great frequency of pathogenic mutations (21/165), is also the largest exon, and the number of observed mutations matched the expected (20.9/165; $\chi^2 = 1.094 \times 10^{-5}$; $P = 0.99$).

We then tested if pathogenic mutations are evenly distributed across functional domains in the protein: region 1 (exon 1 to 6), corresponding to the N-terminal portion; region 2 (exon 7 to 16), corresponding to the repetitive phosphorylated domains; region 3 (exons 17 to 22), “spacer region”; and region 4 (exons 23 to 25), corresponding to the C-terminal domain bearing the nucleolar localization signals. For regions 1 and 2, the observed number of pathogenic mutations was not different from the theoretically expected ($\chi^2 = 0.03$; $P = 0.88$ and $\chi^2 = 1.66$; $P = 0.19$, respectively). Region 4, as expected, had a significantly greater number of mutations, due to the inclusion of exon 24 ($\chi^2 = 35.303$; $P < 0.0001$). Surprisingly, region 3 showed fewer mutations than expected ($\chi^2 = 12.668$; $P = 0.0004$). The same test was performed to determine if polymorphic alterations in the coding region of the gene were evenly distributed across different domains. No specific domain differed significantly from expected values.

MISSENSE MUTATIONS IN *TCOF1*

There are 15 reported nucleotide changes that alter the coded amino acid in the treacle protein. A total of 12 of these mutations are polymorphic, being found in controls with frequencies that vary between 1 and 84% [Wise et al., 1997; Edwards et al., 1997; Splendore et al., 2000]. These polymorphic missense alterations are found in exons 7, 8, 10, 11, 13, 16, 21, and 23. On the other hand, there are some missense alterations that are thought to be pathogenic.

The first reported pathogenic missense mutation was a p.Trp53Arg in exon 2. This alteration was not detected in the patient's parents or in controls, but an effect in splicing could not be ruled out [Edwards et al., 1997]. In 2002, another de novo missense mutation, p.Tyr50Cys, was reported in the same exon. Comparing the portion of the protein coded by exons 1 and 2 with the murine and canine orthologs, the authors noted that both amino acids involved in the pathogenic mutations were conserved, and the whole N-terminus of treacle has a high level of identity between species, suggesting the presence of a functionally important domain that is less tolerant to amino acid changes [Splendore et al., 2002].

Two other missense mutations in exon 2 (p.Ala41Val and p.Trp53Cys) [Ellis et al., 2002] and one in exon 1 (p.Ile14Phe) [Dixon et al., 2004] were reported since. There is no information if these mutations were tested in the patient's parents or other affected family members or controls. As the p.Trp53Cys mutation affects the same codon as the mutation described by Edwards et al. [1997], which is a codon conserved in both canine and murine homologs, we considered this mutation to be pathogenic. For the remaining missense alterations, however, determination of pathogenicity is not straightforward: while the isoleucine at position 14 is conserved in both known homologs, threonine substitutes for alanine 41 in both mice and dogs. Thus, in the absence of additional evidence, these mutations were assigned an undetermined status.

FIGURE 1. Relative frequencies of different classes of pathogenic mutations in *TCOF1*.

PATIENTS WITHOUT PATHOGENIC MUTATIONS AND THE FIRST MUTATION IN EXON 6A

A previous report with the highest mutation detection rate for patients with typical TCS suggested that the lack of mutation in some patients, even after sequencing the whole coding region of *TCOF1*, could be due to large deletions or rearrangements that would remain undetected under the conditions employed [Splendore et al., 2000]. We used a combination of Southern blot and long-range PCR to screen for large deletions or rearrangements in *TCOF1* in four patients with TCS and three patients with atypical phenotype in whom no mutation was found after screening the whole coding region of the gene. Four probes, spanning exons 1 to 24 of the gene were developed for Southern blot analysis as detailed in Table 2 and Figure 2. No alteration was found, indicating that large deletions and rearrangements are unlikely to be a significant cause of TCS.

The four patients with typical TCS phenotype were then screened for mutations in exons 6A and 16A by direct sequencing (primer sequences available on request). A single-base deletion

was detected in exon 6A of a sporadic patient with typical phenotype (c.726delT; p.Lys244fs), which was not detected in 60 controls. On the other hand, screening of exon 16A did not reveal any changes in its coding region, but a G>A transition 19 nucleotides upstream of the beginning of the exon (c.2859+3314G>A) was identified in two affected siblings and their normal mother. Although this intronic alteration was not found in 40 controls, due to its localization and the fact that the phenotypically normal mother is also a carrier, we have not considered this to be a pathogenic alteration.

For two of the patients without an identified pathogenic mutation, the coding region of the gene (except exons 6A and 16A) was screened by SSCP only, so there still is a small possibility of mutations going undetected under our method (approximately 2.5%, based on our previous report) [Splendore et al., 2000]. Alternatively, the clinical features of these three patients may also result from mutations in another gene.

Considering the observation of a patient with a pathogenic mutation in exon 6A (described here for the first time) and another with a mutation in exon 19 (c.3337delC) [Edwards et al.,

FIGURE 2. Schematic representation of genomic *TCOF1* with *ApaI* and *PstI* restriction sites and the regions spanned by the four probes. Exons (boxes) and introns (lines) are represented on scale. The length of each exon is indicated inside the box (in base pairs), while intronic length is written above the representative lines. The distance between the restriction sites represents the sizes of the restriction fragments. Regions that were not properly covered by Southern blot were analyzed by long-range PCR spanning exons 1–4, 7–9, and 18–20.

TABLE 2 RT-PCR Primers Used to Synthesize the Probes for Southern Blot Analyses*

Region	Forward primer	Reverse primer
Exons 1–8	CTTCCCTGATCTACCA	TCCTCACTGTCCGATTCTC
Exons 8–14	CAGGAATCGGACAGTGAGGA	TCTTCCAGGGAGTACTGGA
Exons 14–20	CAGCACAGGTGAAAACCTCA	GGACTTGTTCAGGGAGTGT
Exons 20–24	CCCCTGAGAGCTCAGATGAC	CTGGGACGGTGACTCAGAAT

*Probes were constructed using RNA extracted from cranial periosteum fibroblast cultures, reverse-transcribed with specific primers and cloned into TOPO-TA (Invitrogen) vectors.

1997], both of which can be alternatively spliced, we suggest that the isoforms containing these exons are critical for normal craniofacial development. Despite different isoforms being apparently present in the same proportion in various tissues [So et al., 2004], we still do not know if they have different functions during embryonic neural crest development. When mutations lying in alternatively spliced exons are described in a greater number of patients, a tentative correlation may be attempted to verify if these cases have a milder phenotype, as only one isoform of the mutant allele would be affected.

CONCLUSION

As two alternatively spliced exons of *TCOF1* were recently described, one of which is present in the major protein isoform, the reference cDNA sequence of *TCOF1* became obsolete. We recommend that the major isoform (AY460334) be adopted as reference sequence for describing mutations from now on, and that the appropriate nomenclature be used when describing novel mutations or referring to previously reported ones. The correct naming of mutations prevents confusion and duplication of published mutations under different names [den Dunnen and Paalman, 2003].

For mutations not expected to cause a premature stop codon (nonsense and frameshift mutations), assignment of pathogenic status must be approached with great care. In particular, missense mutations should be tested in the parents and other family members to determine if it segregates with the phenotype or arose de novo, and should always be tested in a control sample. In most cases, comparison with the murine and canine protein orthologs can also be useful in verifying if the amino acid involved in missense mutations is evolutionarily conserved. Specially in the absence of data on protein structure, the use of evolutionary conservation between homolog proteins has proven to be a useful tool for discriminating potential pathogenic from neutral missense changes [Ng and Henikoff, 2001]. Although the prediction power of this approach increases when a greater number of sequences included in the analysis, comparison with as little as two orthologs (as is the case with treacle), may highlight important functional domains that are less tolerant to amino acid changes. This type of evolutionary approach for the analysis of missense mutations has already been successfully applied to a number of genes, recent examples including *BRCA1*, *TP53*, and *CDKN2A (INK4a)* [Fleming et al., 2003; Greenblatt et al., 2003].

This is the first time that patients with TCS were tested for large deletions or rearrangements in *TCOF1* by Southern blot. As we have demonstrated here, it is more likely that those reported patients with clinical TCS and without an identified mutation after sequencing have mutations in exons 6A or 16A, that were not screened up to now. Genetic heterogeneity for TCS remains a possibility, because some patients still have no identified mutation even after screening the new exons and excluding large deletions.

Analyzing the distribution of pathogenic mutations throughout the *TCOF1* gene, we observe that the only true hot spot is exon 24. When exons were grouped according to functional domains in the protein, the region corresponding to exons 17–22 ("spacer or linker domain") showed a significantly lower rate of pathogenic mutations. It is possible that this region is less susceptible to duplications or deletions, but further studies will be necessary to confirm these findings.

ACKNOWLEDGMENTS

We thank all the patients for their collaboration; C.M. Maranduba, H.S. Takagi, and the Genome Center sequencing service for technical assistance; Dr. E.W. Jabs for helpful discussions; Mrs. C.G. Urbani for secretarial help; and Drs. T.M. Felix and E.O. Silva for referring some of the patients.

REFERENCES

- Antonarakis SE. 1998. Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. Hum Mutat 11:1–3.
- den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12.
- den Dunnen JT, Paalman MH. 2003. Standardizing mutation nomenclature: why bother? Hum Mutat 22:181–182.
- Dixon MJ, Haan E, Baker E, David D, McKenzie N, Williamson R, Mulley J, Farrall M, Callen D. 1991a. Association of Treacher Collins syndrome and translocation 6p21.31/16p13.11: exclusion of the locus from these candidate gene regions. Am J Hum Genet 48:274–280.
- Dixon MJ, Read AP, Donnai D, Colley A, Dixon J, Williamson R. 1991b. The gene for Treacher Collins syndrome maps to the long arm of chromosome 5. Am J Hum Genet 49:17–22.
- Dixon MJ, Marres HAM, Edwards SJ, Dixon J, Cremers CWRJ. 1994. Treacher Collins syndrome: correlation between clinical and genetical linkage studies. Clin Dysmorphol 3:96–103.
- Dixon J, Edwards SJ, Anderson I, Brass A, Scambler PJ, Dixon MJ. 1997. Identification of the complete coding sequence and genomic organization of the Treacher Collins syndrome gene. Genome Res 7:223–234.
- Dixon J, Ellis I, Bottani A, Temple K, Dixon MJ. 2004. Identification of mutations in *TCOF1*: use of molecular analysis in the pre- and postnatal diagnosis of Treacher Collins syndrome. Am J Med Genet 127A:244–248.
- Edwards SJ, Gladwin AJ, Dixon MJ. 1997. The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon. Am J Hum Genet 60:515–524.
- Ellis PE, Dawson M, Dixon MJ. 2002. Mutation testing in Treacher Collins syndrome. J Orthod 29:293–297.
- Emes RD, Ponting CP. 2001. A new sequence motif linking lissencephaly, Treacher Collins and oral-facial-digital type 1 syndromes, microtubule dynamics and cell migration. Hum Mol Genet 10:2813–2820.
- Fleming MA, Potter JD, Ramirez JC, Ostrander GK, Ostrander EA. 2003. Understanding missense mutations in the *BRCA1* gene: an evolutionary approach. Proc Natl Acad Sci USA 100:1151–1156.
- Franceschetti A, Klein D. 1949. Mandibulo-facial dysostosis, new hereditary syndrome. Acta Ophthalmol (Copenh) 27:143–224.
- Gladwin AJ, Dixon J, Loftus SK, Edwards SJ, Wasmuth JJ, Hennekam RCM, Dixon MJ. 1996. Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene. Hum Mol Genet 5:1533–1538.
- Gorlin RJ, Cohen MM, Hennekam RC. 2001. Branchial Arch and oral-Acral Disorders. In: Syndromes of the Head and Neck, 4th edition. Oxford: Oxford University Press. p. 790–849.

- Greenblatt MS, Beaudet JG, Gump JR, Godin KS, Trombley L, Koh J, Bond JP. 2003. Detailed computational study of p53 and p16: using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene* 22:1150–1163.
- Haworth KE, Islam I, Breen M, Putt W, Makrinou E, Binns M, Hopkinson D, Edwards Y. 2001. Canine *TCOF1*: cloning, chromosome assignment and genetic analysis in dogs with different head types. *Mamm Genome* 12:622–629.
- Hayano T, Yanagida M, Yamauchi Y, Shinkawa T, Isobe T, Takahashi N. 2003. Proteomic analysis of human Nop56p-associated pre-ribosomal ribonucleoprotein complexes. Possible link between Nop56p and the nucleolar protein treacle responsible for Treacher Collins syndrome. *J Biol Chem* 278:34309–34319.
- Jabs EW, Li X, Coss CA, Taylor EW, Meyers DA, Weber JL. 1991. Mapping the Treacher Collins syndrome locus to 5q31.3–q33.3. *Genomics* 11:193–198.
- Jones NC, Farlie PG, Minichiello J, Newgreen DF. 1999. Detection of an appropriate kinase activity in branchial arches I and II that coincides with peak expression of the Treacher Collins syndrome gene product, treacle. *Hum Mol Genet* 8:2239–2245.
- Marsh KL, Dixon J, Dixon MJ. 1998. Mutations in the Treacher Collins syndrome gene lead to mislocalization of the nucleolar protein treacle. *Hum Mol Genet* 7:1795–1800.
- Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. *Genome Res* 11:863–874.
- Poswillo D. 1975. The pathogenesis of the Treacher Collins syndrome (mandibulofacial dysostosis). *Br J Oral Surg* 13:1–26.
- So RB, Gonzales B, Henning D, Dixon J, Dixon MJ, Valdez BC. 2004. Another face of the Treacher Collins syndrome (*TCOF1*) gene: identification of additional exons. *Gene* 328:49–57.
- Splendore A, Silva EO, Alonso LG, Richieri-Costa A, Alonso N, Rosa A, Carakushansky G, Cavalcanti DP, Brunoni D, Passos-Bueno MR. 2000. High mutation detection rate in *TCOF1* among Treacher Collins syndrome patients reveals clustering of mutations and 16 novel pathogenic changes. *Hum Mutat* 16:315–322.
- Splendore A, Jabs EW, Passos-Bueno MR. 2002. Screening of *TCOF1* in patients from different populations: confirmation of mutational hot spots and identification of a novel missense mutation that suggests an important functional domain in the protein treacle. *J Med Genet* 39:493–495.
- Splendore A, Jabs EW, Felix TM, Passos-Bueno MR. 2003. Parental origin of mutations in sporadic cases of Treacher Collins syndrome. *Eur J Hum Genet* 11:718–722.
- Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M. 1966. Frameshift mutations and the genetic code. *Cold Spring Harb Symp Quant Biol* 31:77–84.
- The Treacher Collins Collaborative Group. 1996. Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat Genet* 12:130–136.
- Valdez BC, Henning D, So RB, Dixon J, Dixon MJ. 2004. The Treacher Collins syndrome (*TCOF1*) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc Natl Acad Sci USA* 101:10709–10714.
- Winokur ST, Shiang R. 1998. The Treacher Collins syndrome (*TCOF1*) gene product, treacle, is targeted to the nucleolus by signals in its C-terminus. *Hum Mol Genet* 7:1947–1952.
- Wise CA, Chiang LC, Paznekas WA, Sharma M, Musy MM, Ashley JA, Lovett M, Jabs EW. 1997. *TCOF1* gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treacher Collins syndrome throughout its coding region. *Proc Natl Acad Sci USA* 94:3110–3115.

NOVAS MUTAÇÕES NO GENE *TCOF1* E CORRELAÇÃO GENÓTIPO-FENÓTIPO NA SÍNDROME DE TREACHER COLLINS

Cibele Masotti¹, Diogo Meyer¹, Alessandra Splendore², e Maria Rita Passos-Bueno¹.

¹Instituto de Biociências, Universidade de São Paulo, Brasil.

²Department of Genetics, Stanford University, Califórnia, EUA

RESUMO

A síndrome de Treacher Collins (STC) é uma anomalia do desenvolvimento craniofacial com padrão de herança autossômico dominante. Existe uma grande variação fenotípica, tanto inter como intra-familiar, e a penetrância é considerada completa. O gene associado à síndrome é o *TCOF1* e mais de 80% das mutações patogênicas nele descritas são deleções e duplicações que levam a um códon de parada prematuro. O mecanismo molecular associado à STC é haploinsuficiência de produto gênico (a proteína treacle) durante o desenvolvimento embrionário. Diversos estudos confirmam a ausência de correlação genótipo-fenótipo e o mecanismo molecular associado à variabilidade clínica é ainda desconhecido. Neste trabalho, descrevemos seis novas mutações no gene *TCOF1* e analisamos a correlação genótipo-fenótipo em uma maior amostra de pacientes STC, portadores de diagnóstico clínico e molecular. Além disso, comparando o *TCOF1* entre humanos e chimpanzés e analisando os polimorfismos em humanos, inferimos que a história desse gene é caracterizada por um regime de seleção purificadora, e que uma parte das mutações presentes em humanos são fracamente deletérias.

INTRODUÇÃO

A síndrome de Treacher Collins (STC) (OMIM 154500) é uma anomalia congênita do desenvolvimento craniofacial com padrão de herança autossômico dominante, sendo que cerca de 40% dos casos diagnosticados têm história familiar e 60% parecem ser causados por mutações novas. Os achados clínicos mais comuns são fendas palpebrais em inclinação anti-mongolóide, coloboma no terço mais externo das pálpebras inferiores e ausência de cílios medialmente a esse defeito, hipoplasia do osso zigomático e das bordas supra-orbitais, micrognatia, palato fendido, penugem capilar atípica na face, malformação das orelhas e dos ossículos do ouvido médio, podendo resultar surdez condutiva. Os pacientes geralmente apresentam um espectro dos achados clínicos, bilateral e simetricamente, e raramente uma alteração isolada leva ao diagnóstico da síndrome (Jones *et al.*, 1997; Gorlin *et al.*, 2001).

Existe uma grande variação fenotípica, tanto inter como intra-familiar, num espectro que varia de casos em que o paciente é tão levemente afetado que clinicamente não pode ser diagnosticado, até casos de morte perinatal pelo colapso das vias respiratórias (The Treacher Collins Syndrome Collaborative Group 1996; Hansen *et al.* 1996). A penetrância é quase completa, havendo somente um relato comprovado de não-penetrância (Dixon *et al.* 1994).

O gene associado à síndrome, *TCOF1*, possui 28 exons, que variam de 49 a 561pb em tamanho, sendo 27 traduzidos (The Treacher Collins Collaborative Group, 1996; So *et al.*, 2004). O produto gênico do *TCOF1*, a treacle, é uma fosfoproteína nucleolar que participa da biogênese de ribossomos, associando-se ao complexo protéico de iniciação da transcrição e do processamento de rRNA (Gonzalez *et al.*, 2005).

Mais de 80% das mutações patogênicas descritas no *TCOF1* são deleções e duplicações que levam a um códon de parada prematuro (revisão em Splendore *et al.*, 2005), e têm como consequência predita uma redução nos níveis de RNA mensageiro pelo mecanismo de *nonsense mediated mRNA decay* (NMD). A via de NMD degrada transcritos portadores de códons de parada prematuros, com o objetivo de prevenir a tradução de RNAm aberrantes ou desnecessários (Khajavi *et al.*, 2006). Consequentemente, a quantidade de proteína funcional é reduzida, e, portanto, sugere-se que a STC é causada pela haploinsuficiência de treacle durante o desenvolvimento embrionário (The Treacher Collins Collaborative Group 1996; Gladwin *et al.* 1996; Wise *et al.* 1997; Edwards *et al.* 1997; Dixon *et al.* 2000; Splendore *et al.* 2000, 2002). Essa

hipótese foi corroborada por meio do estudo de camundongos com deficiência parcial de *treacle*, nos quais se observou que a hipoplasia esquelética típica da STC é resultado de uma deficiência no número de células da crista neural que formam o mesênquima do primeiro e do segundo arcos faríngeos, elementos embrionários dos quais se originarão as estruturas da face (Dixon *et al.*, 2006).

Apesar das várias evidências de que a STC é causada por haploinsuficiência de *treacle*, ainda não há mecanismo molecular para explicar a variabilidade clínica. Diversos estudos confirmaram a ausência de correlação genótipo-fenótipo (Dixon, 1996; Gladwin *et al.*, 1996; Edwards *et al.*, 1997; Splendore *et al.*, 2000). No entanto, os estudos de modelos animais para a STC sugerem que a variabilidade fenotípica depende de fatores genéticos (Dixon e Dixon, 2004). Uma hipótese sugerida pelo nosso grupo, é a de que os níveis de expressão do alelo selvagem podem contribuir para a variabilidade do fenótipo da síndrome (Masotti *et al.*, 2005).

Neste trabalho, descrevemos seis novas mutações no gene *TCOF1* e analisamos a correlação genótipo-fenótipo numa amostra significativamente maior do que a anterior (Splendore *et al.*, 2000; 2002a), contendo 58 indivíduos portadores da STC. Também analisamos o efeito do sexo, da origem parental e posição da mutação patogênica, além do haplótipo do alelo selvagem no fenótipo dos afetados. A análise comparativa com a sequência predita do *Tcof1* de chimpanzé e do banco de dados de mutações em humanos (Splendore *et al.*, 2005) nos permitiu investigar o mecanismo de seleção natural associado ao *TCOF1*. Nossos resultados indicam que a história desse gene é caracterizada por um regime de seleção purificadora, e que parte das mutações não-sinônimas são fracamente deletérias.

SUJEITOS E MÉTODOS

Casuística

Os 30 indivíduos descritos nas Tabelas 1 e 2 foram encaminhados para aconselhamento genético e diagnóstico molecular no Centro de Estudos do Genoma Humano (CEGH), ou tiveram suas amostras de DNA enviadas pelo correio, partindo dos seguintes locais e respectivos pesquisadores responsáveis: (1) Hospital de Clínicas de Porto Alegre (Dra. Têmis Félix); (2) Arcispedale “S. Maria Nuova”, Ambulatório di Genética Clinica, Divisione di Pediatria, Reggio

Emilia, Itália (Dra. Livia Garavelli); (3) Vall d'Hebron Hospital Materno-Infantil, Unitat de Genética Clínica i Citogenética, Barcelona, Espanha (Dr. E.Sarret); (4) Centre de Génétique Médicale, Loverval, Bélgica (Dr. Lionel Van Maldergem); (5) Department of Clinical Genetics, Karolinska Hospital, Estocolmo, Suécia (Dra. Kristina Lagerstdt); (6) Department of Clinical Genetics, Oulu University Hospital, Oulu, Finlândia (Dr. Aki Mustonen e Dr. Jukka Moilanen); (7) Medical Genetics Center, Vrije Universiteit Brussel, Bruxelas, Bélgica (Dra. Sara Sêneca). No total, somam 21 propósitos e 9 familiares. Os diagnósticos moleculares dos demais pacientes (Tabelas 3, 4, 5, 6, 7 e 8) foram realizados por Splendore *et al.* (2000 e 2002a).

Para análise de dados fenotípicos, estudamos 44 propósitos e 14 familiares afetados. Também incluímos o estudo de haplótipos de 30 familiares normais desses indivíduos. A amostra total foi constituída no Centro de Estudos do Genoma Humano ao longo dos anos de aconselhamento genético e diagnóstico molecular, sendo que o estudo de grande parte dessa amostra gerou publicações prévias de nosso grupo em STC.

A extração do DNA genômico a partir de linfócitos foi realizada de acordo com a metodologia descrita por Miller *et al.* (1988) e, a partir de esfregaço bucal, de acordo com Richards *et al.* (1993). Este trabalho foi aprovado pelo comitê de ética de nossa instituição e os termos de consentimento livre e informado foram obtidos de pacientes e familiares, ou de seus responsáveis legais.

Para correlação genótipo-fenótipo, empregamos a metodologia descrita por Gollop (1981). Os nove dados clínicos mais comuns da nossa amostra (hipoplasia malar, hipoplasia mandibular, fissura palatina, displasia do pavilhão auditivo externo, perda de audição, fendas palpebrais com inclinação anti-mongolóide, coloboma das pálpebras inferiores, ausência de cílios e cabelos pré-auriculares atípicos) foram classificados em presente (P), ausente (A) e não-referido (N). Um sinal clínico pode ser classificado como não-referido pelo fato de o geneticista clínico ter ignorado a avaliação do sinal em questão, ou pelo fato de que o paciente não apresentava esse sinal e o avaliador não o mencionou como ausente. Julgamos adequado, então, calcular a frequência dos achados como igual a $100P/(P+A)$, que corresponde à hipótese de que a presença e a ausência do sinal clínico, nos casos em que o sinal não foi referido (N), estão na mesma proporção em que no resto da amostra.

Triagem de mutações

Para triagem de mutações no gene *TCOF1* empregamos a mesma metodologia descrita em Splendore *et al.*, 2000 e 2002.

Análise de haplótipos da região codificadora

Uma amostra de 73 indivíduos (43 afetados e 30 familiares normais) foi submetida à triagem de mutações patogênicas nos estudos de Splendore *et al.* (2000 e 2002). Todos os exons do *TCOF1* foram analisados e 10 polimorfismos de base única (SNPs) foram identificados nesses indivíduos: c.1229 C>T (p.Ser410Leu), c.1578 T>C (p.Pro516Pro), c.1761 G>T (Gly587Gly), c.1842 G>A (p.Ser614Ser), c.2660 C>T (p.Ala887Val), c.2765 C>T (p.Ser992Leu), c.3183+66C>T, c.3527 C>G (p.Pro1176Arg), c.4050 G>A (p.Ser1350Ser) e c.4169 C>T (p.Ala1390Val). As frequências desses polimorfismos em pacientes e controles já foram descritas por Wise *et al.* (1997), Edwards *et al.* (1997) e Splendore *et al.* (2000).

Para estabelecer a fase dos diferentes SNPs, utilizamos o programa PHASE v2.1.1 (Stephens and Donnelly, 2003), obtido a partir de <http://www.stat.washington.edu/stephens/software.html>. Esse programa estima os haplótipos mais prováveis com base na frequência dos alelos na amostra estudada.

Teste de seleção natural

Testamos a hipótese de haver seleção natural no gene *TCOF1* usando duas abordagens. Primeiro, estimamos a razão entre as taxas sinônima (dN) e não-sinônima (dS) para o gene *TCOF1*. Valores de dN/dS > 1 indicam seleção positiva e valores de dN/dS < 1 indicam seleção purificadora (Nielsen, 2005). A seleção positiva ocorre quando mutações favoráveis são fixadas numa espécie. No caso de regiões codificadoras, tal regime de seleção leva a um aumento na proporção de substituições não-sinônimas, que são aquelas que alteram características da proteína. Para estimar as taxas de substituição sinônima e não-sinônima acumuladas na divergência entre humano e chimpanzé, alinhamos a sequência predita para o RNAm do *TCOF1* do chimpanzé (XM_527074, baseada na sequência genômica NW_001235420) com a sequência da isoforma principal do *TCOF1* humano (AY460334). Estimativas de dN e dS foram feitas usando o programa DNAsp (Rozas *et al.*, 2003).

A segunda abordagem utilizou os dados de polimorfismo para o gene *TCOF1* em humanos. Sob a hipótese de neutralidade, a proporção entre o número de sítios polimórficos sinônimos (p_n) e não-sinônimos (p_s) é igual a de diferenças sinônimas e não-sinônimas entre humanos e chimpanzés (ou seja, $p_n/p_s = dN/dS$; McDonald e Kreitman, 1991). Para testar a hipótese de que *TCOF1* está evoluindo de modo neutro ou sob seleção natural, comparamos o número de polimorfismos sinônimos e não sinônimos obtidos a partir do banco de dados de mutações do *TCOF1* (www.genoma.ib.usp.br; Splendore *et al.* 2005), com os valores de divergência sinônima e não sinônima entre humanos e chimpanzés.

RESULTADOS

Triagem de mutações no TCOF1

A triagem de mutação no gene *TCOF1* em 21 propósitos revelou a presença de mutação patogênica em 11 deles (Tabela 1). Foram também analisados nove familiares, dos quais cinco eram clinicamente normais. Os afetados apresentaram a mesma mutação do propósito, contudo, foi identificada a mesma mutação patogênica do propósito em dois dos indivíduos clinicamente normais (F1761-2 e F1785-2), mães de 2 propósitos.

As mutações patogênicas detectadas são famílias-específicas e apenas quatro já haviam sido descritas na literatura (c.1637_1640delAGAG, c.4218dupG, c.1446dupA, c.1095_1096delAG- (Gladwin *et al.*, 1996; Splendore *et al.*, 2000 e 2002a; Dixon *et al.*, 2004). Dentre as restantes, três (c.639+1G>A, c.1609C>T (p.Gln537X), c.2033C>A (p.Ser678X)), são trocas de base única: a primeira está localizada na primeira posição do íntron seis e possivelmente afeta o *splicing* desse íntron; e as duas outras, localizadas nos exons 10 e 12, substituem um códon normal por um de parada. Outras três mutações [c.431delC (p.Thr144fs), c.2499delC (p.Gln835fs) e c.3853delC (p.Gln1285fs)] são deleções de apenas um nucleotídeo nos exons 10, 15 e 23 que alteram o quadro de leitura e devem gerar códons de parada prematuros. Cabe ressaltar que essas mutações não foram encontradas em 50 controles normais da população brasileira (Splendore *et al.*, 2000). Assim sendo, considerando-se a natureza das mutações e a ausência destas em 100 cromossomos normais, elas foram classificadas como patogênicas.

A mutação c.4375_4377delAAG (p.Lys1459del) no exon 24 está presente no caso familiar F1785, em que os dois irmãos com fenótipo típico da STC (F1785-1 e -4) e a mãe aparentemente normal (F1785-2) são portadores da mutação. Essa deleção de três nucleotídeos elimina um resíduo de lisina inteiro, sem alterar a fase de leitura.

Nenhuma mutação patogênica foi detectada nos 10 propósitos restantes após análise dos 28 exons do *TCOF1* (Tabela 2). Investigamos apenas os exons 6A e 16A nos dois indivíduos F1704-1 e 2, uma vez que esses pacientes já haviam feito o diagnóstico molecular em seu país de origem, mas nenhuma mutação patogênica foi detectada.

Correlações genótipo-fenótipo

Casos isolados x casos familiares:

Analizamos a frequência dos achados fenotípicos de uma amostra de 58 pacientes portadores de mutação patogênica (Tabela 3). A amostra foi dividida conforme a ocorrência do caso, ou seja, se isolado (CI) ou familiar. Os afetados dos casos familiares foram divididos em dois grupos: CFT (total de casos familiares, incluindo propósitos e parentes afetados) e CFP (incluindo somente os propósitos dos casos familiares). A comparação dos achados clínicos (Tabela 4) de CI x CFT mostra que há uma tendência de as frequências serem diferentes ($p < 0,05$; teste do sinal). Após comparação das frequências de cada um dos nove sinais, observamos uma diferença significativa ($p < 0,05$, Teste Exato de Fisher com Tabelas de contingência 2x2) entre CI e CFT para displasia do pavilhão auditivo externo, perda de audição e ausência de cílios. Isso sugere que esses achados estão mais presentes nos casos isolados. Realizamos, em seguida, uma comparação das frequências dos achados clínicos dos propósitos de CI com somente os propósitos dos casos familiares (CFP) e não mais observamos essas diferenças ($p > 0,05$, Teste do Sinal; Teste Exato de Fisher com Tabelas de contingência 2x2; Tabela 3).

Posição da mutação no gene:

Comparamos a frequência dos dados clínicos dos 58 pacientes de acordo com a posição da mutação patogênica na região codificadora do gene, que foi dividida conforme os domínios protéicos por ela codificados: região 1 (porção N-terminal, exons 1 a 6), região 2 (domínios repetitivos e fosforilados, exons 6A a 16), região 3 (domínio espaçador, exons 17 a 22) e região 4

(porção C-terminal, exons 23 a 25) (Tabela 5). Não observamos diferença significativa ($p>0,05$, Teste do Sinal) das frequências clínicas nas regiões 2, 3 e 4; contudo, observamos uma maior frequência dos sinais clínicos nos pacientes portadores de mutação na região 1 ($p<0,05$, Teste do Sinal). Dado que Teber *et al.* (2004) encontraram menor frequência de surdez condutiva em pacientes com mutações na porção 3' do gene, analisamos a frequência de perda auditiva nos afetados com mutações na porção N-terminal da proteína, somando os indivíduos das regiões 1, 2 e 3 e comparando com os da região 4. Nenhuma diferença significativa foi observada ($p>0,05$; Teste Exato de Fisher, Tabela de contingência 2x2).

Sexo e origem parental da mutação:

Comparamos as frequências dos sinais clínicos de 35 pacientes do sexo feminino e de 23 do sexo masculino (Tabela 6) e não encontramos diferença significativa entre os sexos ($p=0,453$, Teste do Sinal), o que mostra que não há correlação da gravidade com o sexo do afetado. A proporção de homens ($n=23$) e mulheres ($n=35$) afetados não difere do esperado ($p>0,26$; qui-quadrado=1.26).

Também comparamos os achados clínicos dos indivíduos com mutação patogênica de origem materna com os daqueles de origem paterna (Tabela 7) e nenhuma diferença observada foi significativa ($p>0.05$ para Teste do Sinal e Teste Exato de Fisher para todos os achados clínicos).

Haplótipos dos SNPs mais frequentes do TCOF1

Dado que o *TCOF1* apresenta muitos sítios polimórficos, levantamos a hipótese de que exista um ou mais haplótipos do alelo normal (em *trans* com a mutação patogênica) que esteja associado à gravidade do quadro clínico. Para tanto, inferimos a fase de 10 SNPs do gene *TCOF1* que foram genotipados em 73 indivíduos. Dos 58 pacientes analisados nas tabelas anteriores, somente 43 afetados foram completamente genotipados para os 10 polimorfismos. A análise dos genótipos identificou 17 haplótipos com maior probabilidade de ocorrência nos 73 indivíduos analisados, sendo que os haplótipos 3, 4, 5 e 10 são os mais frequentes (Tabela 8).

Considerando que o alelo portador da mutação patogênica é degradado, procuramos identificar o haplótipo do alelo selvagem. Foi possível determinar a origem parental da mutação

patogênica em 20 indivíduos da amostra estudada (12 maternas e 8 paternas), sendo que sete pacientes pertencem aos casos isolados que tiveram sua origem parental definida no trabalho de Splendore *et al.* (2003). Utilizando a informação de origem parental da mutação e observando os indivíduos homozigotos para um haplótipo, identificamos os haplótipos expressos (selvagens) em 28 afetados. Dividimos os 28 pacientes em grupos, de acordo com o haplótipo expresso, e comparamos a frequência dos achados clínicos (Tabela 9). Três indivíduos foram desconsiderados por serem representantes únicos dos haplótipos 6, 16 e 17. Não observamos nenhuma evidência de associação do quadro clínico a qualquer um dos quatro haplótipos analisados, 3, 4, 5 e 10 ($p > 0,05$, Teste do Sinal).

Analizamos também a distribuição dos haplótipos em *cis* com a mutação patogênica, procurando verificar se algum deles estava preferencialmente ligado à doença. Comparamos a frequência do haplótipo 3 (o mais comum) em relação à frequência dos haplótipos 4, 5, 7, 10, 13, 14, 16 e 17 agrupados. Não observamos associação de um dos haplótipos à mutação patogênica ($p > 0,05$; Tabelas de contingência 2x2, Teste Exato de Fisher).

Análise de seleção natural

Analizamos a proporção de substituições sinônimas e não-sinônimas (taxa dN/dS) entre humanos e chimpanzés num total de 4.137 nucleotídeos (aproximadamente 1379 códons) alinhados entre as duas espécies. Obtivemos uma taxa dN/dS = 0,24.

Para investigar o modelo de seleção natural que atua sobre o *TCOF1*, comparamos a proporção de polimorfismos não-sinônimos e sinônimos em humanos ($p_n/p_s = 12/6$) com a proporção de substituições não-sinônimas e sinônimas entre humanos e chimpanzés (dN/dS=20/28). O *TCOF1* tem um excesso de polimorfismos não-sinônimos em populações humanas (p_n), em relação ao esperado com base na divergência entre humanos e chimpanzés (dN). As razões são marginalmente significativamente diferentes ($p = 0,06$, Teste Exato de Fisher para Tabela de contingência 2x2).

Usando os dados do *TCOF1 mutation database*, observamos que a maior parte (75%) dos alelos que resultam em mudanças não-sinônimas são raros (isto é, possuem frequências próximas

a 1%). Já os alelos que não alteram os aminoácidos da sequência codificadora estão igualmente distribuídos em frequências raras e comuns (três comuns e dois raros).

DISCUSSÃO

O presente trabalho representa uma continuidade da análise do espectro de mutações dos pacientes que foram encaminhados ao nosso centro para diagnóstico molecular. Analisamos 21 propósitos e detectamos mutação patogênica em 11 deles, mesmo após a análise dos exons 6a e 16a (recentemente descritos por So *et al.*, 2004). Cinco familiares clinicamente normais solicitaram o diagnóstico molecular e, após a triagem de mutações, dois deles (F1761-2 e F1785-2) foram diagnosticados como portadores da mutação patogênica presente em seus respectivos filhos, F1761-1 e F1785-1 e -4. Não podemos afirmar que esses casos se tratam de não-penetrância, uma vez que não realizamos análise radiográfica da face desses indivíduos. Para os dois propósitos sem mutação identificada que apresentam quadro clínico típico da síndrome, é possível que o sequenciamento completo revele alguma alteração. No entanto, não podemos excluir a possibilidade de um mecanismo mutacional atípico, ou mesmo de heterogeneidade genética para a STC. Por outro lado, os outros oito propósitos sem mutação detectada possuem achados clínicos atípicos (Tabela 2), geralmente não relacionados à STC, confirmando nossos achados anteriores que sugerem que mutações no gene *TCOF1* estão associadas a um fenótipo específico (Splendore *et al.*, 2002b). A solicitação elevada do teste genético mostra que este é considerado uma ferramenta de diagnóstico diferencial por alguns geneticistas clínicos, que, mesmo identificando alterações atípicas da STC de seus pacientes, optam por encaminhar as amostras de DNA para triagem de mutações no *TCOF1*. Isso mostra que a variabilidade clínica das síndromes malformativas craniofaciais ainda representa um desafio para o aconselhamento genético.

Descrevemos seis novas mutações patogênicas no gene *TCOF1*, sendo uma do tipo *missense*, que afeta o processamento normal do íntron 6, duas do tipo *nonsense*, e três do tipo *frameshift*. A identificação de uma deleção no exon 24 (c.4375_4377delAAG; p.Lys1459del) que não altera o quadro de leitura em um caso familiar da STC nos sugere que a eliminação de um resíduo de lisina no sinal de localização nucleolar seja capaz de causar o fenótipo da síndrome. De acordo com Winokur & Shiang (1998), há um sinal de localização nucleolar (NoLS) conservado

entre camundongos e humanos na porção C-terminal da treacle. Segundo esses autores, as construções da treacle fusionada com GFP (*green fluorescent protein*) que não continham o NoLS não co-localizavam com o nucléolo. Dado que a proteína exerce sua função no nucléolo, é possível que uma mutação que altere ou diminua a sua eficiência de localização também gere o quadro clínico por haploinsuficiência, uma vez que não há dose suficiente de treacle no nucléolo para síntese de ribossomos. Interessantemente, Teber *et al.* (2004) descreveram uma mutação semelhante: a duplicação de uma lisina do exon 24 (p.Lys1460dup), também junto ao NoSL, em um caso isolado de STC típico, sem qualquer outra mutação patogênica detectada. Duas deleções de resíduos de aminoácidos inteiros também foram identificadas por Ellis *et al.* (2002), p.Lys1366del e p.Lys1444del, contudo, ambas estão localizadas no exon 23, fora do sinal de localização nucleolar, e foram detectadas em dois pacientes com as síndromes de Goldenhar e de Nager, respectivamente, e possivelmente não são patogênicas. Cabe ressaltar que nenhum desses autores realizou estudo funcional para essas alterações.

Com a ampliação da amostra de pacientes com diagnóstico clínico confirmado pela triagem de mutações no *TCOF1*, re-analisamos o fenótipo de 58 indivíduos afetados, para testar se há correlação entre o fenótipo do indivíduo e sexo, ocorrência, posição da mutação no gene e origem parental da mutação.

A comparação entre o fenótipo da STC de casos isolados e familiares da literatura foi bastante documentada por Gollop (1981). Esse autor comparou a frequência de 19 dados clínicos entre os casos familiares com herança claramente autossômica dominante e casos familiares com sugestão de herança autossômica recessiva, ou seja, com recorrência de afetados na prole de pais normais. Uma diferença significativa foi observada entre as amostras, havendo maior frequência de alterações auriculares nos casos recessivos, levando o autor a propôr a existência de uma forma autossômica recessiva para a síndrome, capaz de levar a um fenótipo mais grave. A mesma diferença foi observada quando o autor comparou casos isolados e casos familiares dominantes, e a conclusão a partir desses dados foi que dentro dos casos isolados existiriam casos recessivos que estariam desviando a frequência dos achados clínicos em direção à severidade dos casos recessivos. No entanto, nenhum dos casos documentados nesse trabalho foi submetido à triagem de mutações no *TCOF1*, uma vez que a identificação do gene foi posterior a este trabalho. Diante disso, não podemos saber se os casos considerados como recessivos são exemplos de baixa expressividade da síndrome em pais portadores da mutação. Somente um

caso com fenótipo típico e sem mutação no *TCOF1* e com sugestão de herança recessiva foi documentado em nossa amostra (Splendore *et al.*, 2002a). No presente trabalho, também observamos uma diferença na frequência dos achados clínicos de casos isolados em relação aos familiares (CFT). Contudo, restringindo a comparação dos casos isolados com somente propósitos dos casos familiares (CFP), verificamos que as diferenças se tornaram não-significantes. Nossa conclusão é que, ao contrário do que Gollop (1981) supunha, há um viés de averiguação, pois os casos que chegam ao atendimento são os mais graves. Dessa maneira, o propósito do caso familiar tem uma gravidade semelhante ao do caso isolado, porém os parentes do propósito do caso familiar tendem a ter um quadro clínico mais leve e a inclusão dos dados clínicos desses familiares na análise mascara a gravidade do fenótipo dos propósitos dos casos familiares. Por isso, a significância observada está principalmente relacionada à malformação das orelhas e à perda auditiva, defeitos limitadores de função e esteticamente conspícuos quando moderados ou severos. Cabe ressaltar que viés de averiguação é comum em famílias com afetados por STC (Marres *et al.*, 1995; Gladwin *et al.*, 1996; Edwards *et al.*, 1997; Splendore *et al.*, 2000).

Os trabalhos de Dixon (1996), Edwards *et al.* (1997) e Splendore *et al.* (2000) não observaram qualquer associação da gravidade do fenótipo ao tipo ou à posição da mutação no gene. Nossos resultados confirmam esses achados de não-correlação. Embora seja possível que ocorra um fenótipo mais grave nos portadores de mutação na região 1, acreditamos que esse resultado esteja relacionado ao pequeno número amostral (apenas dois indivíduos possuem mutação nessa região). No entanto, esses resultados estão de acordo com o modelo de haploinsuficiência, em que mutações patogênicas em diferentes posições no gene geram semelhantes quadros clínicos. Teber *et al.* (2004) foram os únicos a constatar que os pacientes que apresentavam mutações próximas à porção 3' do gene tinham significativamente menos perda auditiva do que os demais; contudo, nossos resultados não confirmam esse achado. Além disso, não estabelecemos nenhuma correlação entre o sexo do afetado e a origem parental da mutação com a gravidade do fenótipo, o que está de acordo com a maioria dos estudos anteriores (Gollop, 1981; Splendore *et al.*, 2000; Teber *et al.*, 2004).

Analisamos a frequência dos haplótipos dos polimorfismos do *TCOF1* em pacientes e controles com o intuito de elucidar duas hipóteses. (1) Estando em *trans* com a mutação patogênica, um determinado haplótipo da região codificadora poderia predispor à gravidade do fenótipo nos pacientes por modular a expressão do alelo normal. Isso poderia resultar do fato dos

haplótipos da região codificadora estarem em desequilíbrio de ligação com mutações em regiões regulatórias capazes influenciar a expressão gênica. Além disso, combinações de SNPs da região codificadora podem ter um efeito sobre a função da proteína. (2) Estando em *cis* com a mutação patogênica, um haplótipo mais freqüente pode estar associado a um aumento da ocorrência de mutações.

Nenhuma correlação haplótipo-fenótipo foi observada e nossos achados sugerem, portanto, que os polimorfismos mais freqüentes da região codificadora do *TCOF1* do alelo selvagem, em *trans* com a mutação patogênica, não influenciam o fenótipo dos afetados. Além disso, nosso trabalho também mostra que não há maior probabilidade de ocorrer mutação patogênica num haplótipo em relação a outro.

Este trabalho permitiu também fazer inferências sobre o regime de seleção ao qual o *TCOF1* está submetido. Comparando humanos e chimpanzés, obtivemos uma taxa $dN/dS = 0,24$, típica para seqüências codificantes sob seleção purificadora. Comparando os números de polimorfismos sinônimos e não-sinônimos em humanos (pn/ps) com a divergência entre humanos e chimpanzés (dn/ds), identificamos um excesso de polimorfismos não-sinônimos (pn/ps > dn/ds; $p=0.06$). Uma interpretação evolutiva para esse achado é que as mutações não-sinônimas que ocorrem em humanos incluem algumas que são fracamente deletérias. Essas mutações persistem transientemente na população como polimorfismos e raramente são fixadas, mas resultam em diferenças entre as duas espécies. É possível que os polimorfismos não-sinônimos do *TCOF1* tenham um efeito clinicamente imperceptível sobre o fenótipo, entretanto, são capazes de afetar o valor adaptativo dos indivíduos. Nossos resultados não estão de acordo com o trabalho de Bustamante *et al.* (2005), que, analisando 11.000 genes do genoma humano, incluíram o gene *TCOF1* no grupo de genes sujeitos a seleção positiva (regime que favorece um alelo específico, cuja freqüência aumenta sob seleção natural). Sugerimos então que o gene *TCOF1* não foi selecionado positivamente ao longo da divergência entre as duas espécies, que ocorreu há aproximadamente sete milhões de anos atrás.

A não associação do haplótipo do alelo selvagem com o fenótipo dos pacientes sugere que as mutações fracamente deletérias não interferem significativamente no fenótipo da STC. No entanto, não podemos excluir a possibilidade desses haplótipos predisporer a doenças complexas ou a qualquer outra condição multifatorial, como o câncer, dado que *TCOF1* é um gene

envolvido em proliferação celular. Assim sendo, os nossos resultados sugerem que a expressividade variável da síndrome esteja associada a outros mecanismos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG. (2005) Natural selection on protein-coding genes in the human genome. *Nature*. Oct 20;437(7062):1153-7
2. Dixon, M.J.; Marres, H.A.M.; Edwards, S.J.; Dixon, J; Cremers, C.W.R.J. (1994) Treacher Collins syndrome: correlation between clinical and genetic linkage studies. *Clin Dysmorphol*, 3: 96-103.
3. Dixon J, Jones NC, Sandell LL, Jayasinghe SM, Crane J, Rey JP, Dixon MJ, Trainor PA. (2006) Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc Natl Acad Sci U S A*. Sep 5;103(36):13403-8. Epub 2006 Aug 28.
4. Edwards, S.J.; Gladwin, A.J.; Dixon, M.J. (1997) The Mutational Spectrum in Treacher Collins Syndrome Reveals a Predominance of Mutations That Create a Premature-Termination Codon. *Am J Hum Genet* 60:515-524
5. Gladwin, A.J.; Dixon, J.; Loftus, S.K.; Edwards, S.; Wasmuth, J.J.; Hennekam, R.C.M.; Dixon, M.J. (1996) Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene. *Hum Mol Gen* 5:1533-1538
6. Gollop, TR. (1981) Estudo Genético-clínico das Disostoses Mandibulofacial e Frontofacionasal. Tese de Doutorado. Instituto de Biociências, Universidade de São Paulo, SP, Brasil.
7. Gorlin, R.J.; Cohen, M.M.; Hennekam, R.C. (2001) Branchial Arch and oral-Acral Disorders. In: *Syndromes of the head and neck*, 4th edition. Oxford: Oxford University Press. p.790-849.
8. Hansen, M.; Lucarelli, M.J.; Whiteman, D.A.H.; Mulliken, J.B. (1996) Treacher Collins Syndrome: Phenotypic variability in a family including an infant with arhinia and uveal colobomas. *Am J Hum Genet* 61:71-74

9. Jones, K.L. (1997) *Smith's Recognizable Patterns of Human Malformations* 5th ed. Pennsylvania: Saunders.
10. McDonald, J. H., and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351: 652-654.
11. Marres, H.A.M.; Cremers, C.W.R.J.; Dixon, M.J.; Huygen, P.L.M.; Joosten, B.M. (1995) The Treacher Collins syndrome: A clinical radiological and genetic linkage study on two pedigrees. *Arch Otolaryngol Head Neck Surg* 121:509-514.
12. Miller, S.A.; Dykes, D.D.; Polesky, H.F. (1988) A simple testing out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
13. Nielsen, R. (2005) Molecular Signatures of Natural Selection. *Annual Review of Genetics*, 39: 197-218.
14. Richards, B.; Skoletsky, J.; Shuber, A.P.; Balfour, R.; Stern, R.C.; Dorkin, H.L.; Parad, R.B.; Witt, D.; Klinger, K.W. (1993) Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum Mol Genet* 2:159-163
15. Rogers, BO. (1964) Berry-Treacher-Collins syndrome: a review of 200 cases. *Br J Plast Surg*, 17:109.
16. Rogers, BO. (1976) The surgical treatment of mandibulofacial dysostosis. *Clin Plast Surg*, 3:653.
17. Rozas, J, Sánchez-DelBarrio, J.C, Messeguer, X, e Rozas, R. (2003) *Bioinformatics* 19: 2496-2497
18. Rovin, S; Dachi, SF, Borestein, DB, Cotter, WR. (1964) Mandibulofacial dysostosis, a familial study of five generations. *J Pediatr*, 64:215.
19. Szlajak, J. (1953) Treacher-Collins syndrome. *Canad Med Assoc J*, 69:274.
20. So RB, Gonzales B, Henning D, Dixon J, Dixon MJ, Valdez BC: Another face of the Treacher Collins syndrome (*TCOF1*) gene: identification of additional exons. *Gene* 2004 328: 49-57.
21. Splendore, A.; Silva, E.O.; Alonso, L.G.; Richieri-Costa, A.; Alonso, N.; Rosa, A.; Carakushansky G.; Cavalcanti, D.P.; Brunoni, D.; Passos-Bueno, M.R. (2000) High mutation

detection rate in *TCOF1* among Treacher Collins syndrome patients reveals clustering of mutations in 16 novel pathogenic changes. *Hum Mutat* 16:315-322

22. Splendore, A.; Jabs, E. W.; Passos-Bueno, M.R. (2002a) Screening of *TCOF1* in patients from different populations: confirmation of mutational hot spots and identification of a novel missense mutation that suggests an important functional domain in the protein treacle. *J Med Genet* 39:493-495
23. Splendore, A.; Jabs, E. W.; Maldergem, L.V.; Wulfsberg, E.; Passos-Bueno, M.R. (2002b) *TCOF1* mutations excluded from a role in other first and second branchial arch related disorders. *Am J Med Genet* 111:324-327
24. Splendore, A.; Jabs, E. W.; Félix, T.M.; Passos-Bueno, M.R. (2003) Parental origin of mutations in sporadic cases of Treacher Collins syndrome. *Eur J Hum Genet* Sep 11(9): 718-22
25. Splendore, A.; Fanganiello, R.D.; Masotti, C.; Morganti, L.S.C.; Passos-Bueno, M.R. (2005) *TCOF1* mutation database: novel mutation in the alternatively spliced exon 6A and update in mutation nomenclature. *Hum Mut*, 25: 429-434.
26. The Treacher Collins Collaborative Group (1996) Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat Gen* 12:130-136
27. Wise, C.A.; Chiang, L.C.; Paznekas, W.A.; Sharma, M.; Musy, M.M.; Ashley, J.A.; Lovett, M. Jabs, E.W. (1997) *TCOF1* gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treaches Collins Syndrome throughout its coding region. *Proc Natl Acad Sci USA* 94:3110-3115

Tabela 1: Pacientes encaminhados para triagem de mutação no gene *TCOF1* – com mutação patogênica detectada.

Registro	Ocorrência	Diagnóstico clínico e parentesco	Sinais Clínicos									Mutação patogênica (Ref Seq AY460334)
			Hipoplasia malar	Microgantia	Fenda palatina	Displasia pavilhão auditivo	Surdez	Fendas anti-mongolóides	Coloboma	Ausência de cílios	Cabelo na face	
F1342-1	Familiar	Propósito afetado	P	P	A	P	P	P	P	P	c.1637_1640delAGAG	
F1342-2	Familiar	Irmão afetado	P	P	A	P	P	P	A	P	c.1637_1640delAGAG	
F1379-1	Isolado	Propósito afetado	P	P	P	P	P	P	P	P	c.4218dupG	
F1380-1	Isolado	Propósito afetado	P	P	A	P	P	P	P	A	c.2033C>A	
F1491-1	Isolado	Propósito afetado	P	P	P	P	P	P	P	A	c.1446dupA	
F1491-2	Isolado	Mãe normal	-	-	-	-	-	-	-	-	Normal	
F1564-1	Isolado	Propósito afetado	P	P	A	A	P	P	P	A	c.1609C>T	
F1565-1	N	Propósito afetado	N	N	N	N	N	N	N	N	c.3853delC	
F1595-1	Familiar	Propósito afetado	P	P	P	P	P	P	P	P	c.431delC	
F1679-1	Familiar	Propósito afetado	P	P	A	P	P	P	A	P	c.639+1G>A	
F1757-1	Isolado	Propósito afetado	P	P	P	P	P	P	A	N	c.2499delC	
F1761-1	Familiar	Propósito afetado	P	P	A	P	P	P	P	P	c.1095_1096delAG	
F1761-2	Familiar	Mãe normal	A	A	A	A	A	A	A	A	c.1095_1096delAG	
F1785-1	Familiar	Propósito afetado	P	P	A	P	A	A	A	A	c.4375_4377delAAG	
F1785-2	Familiar	Mãe normal	A	A	A	A	A	A	A	A	c.4375_4377delAAG	
F1785-3	Familiar	Pai normal	-	-	-	-	-	-	-	-	Normal	
F1785-4	Familiar	Irmão afetado	P	P	A	P	P	P	P	P	c.4375_4377delAAG	

P, presente; A, ausente; N, não referido.

Tabela 2: Pacientes encaminhados para triagem de mutações no gene *TCOF1* – ausência de mutação patogênica.

Registro	Diagnóstico clínico e parentesco	Sinais Clínicos										Outros achados
		Hipoplasia Malar	Microgantia	Fenda Palatina	Displasia Pavilhão Auditivo	Surdez	Fendas Anti-mongolóides	Coloboma	Ausência de cílios	Cabelo na face		
F1343-1	Propósito afetado	P	N	P	P	P	P	N	N	N	Epilepsia, atraso no desenvolvimento, fistula traqueo-esofágica, microcefalia, escoliose, alterações cardíacas e de membros.	
F1458-1	Propósito afetado	N	N	N	N	N	N	N	N	N	Caso isolado	
F1476-1	Propósito afetado	N	N	N	N	N	N	N	N	N	Fenótipo típico, retardo mental moderado.	
F1521-1	Propósito afetado	N	N	N	N	N	N	N	N	N	N	
F1547-1	Propósito afetado	P	P	P	P	P	P	A	P	N	Úvula bifida, anomalias vertebrais cervicais, retardo mental.	
F1547-2	Mãe normal	P	P	N	P	P	P	A	A	N	N	
F1557-1	Propósito afetado	N	N	N	N	N	N	N	N	N	Caso atípico de STC	
F1569-1	Propósito afetado	P	P	A	A	P	P	P	P	P	N	
F1569-2	Mãe afetada	P	P	P	P	P	P	P	P	P	N	
F1570-1	Propósito afetado	P	P	A	P	N	P	A	N	P	História familiar de 'queixo pequeno'	
F1670-1	Propósito afetado	N	N	P	P	P	P	P	A	A	Assimetria facial, microsomia, macrostomia, fissura labial, atresia da coana, ptose, criptoquidismo, clinodactília, microcefalia, craniossinostose, alterações renais e vertebrais.	
F1704-1	Propósito afetado	P	P	N	P	N	P	P	P	P	Assimetria facial e alteração de orelha isolada na família.	
F1704-2	Mãe afetada	P	P	A	A	A	P	A	A	A	Assimetria facial e alteração de orelha isolada na família.	

Tabela 3: Frequência de sinais clínicos da amostra total de 58 pacientes STC.

Sinais Clínicos	P	A	N	Frequência (%)
Hipoplasia malar	51	5	2	91
Hipoplasia mandibular	52	6	0	90
Fissura palatina	7	51	0	12
Displasia do pavilhão auricular	41	17	0	71
Perda de audição	35	14	9	71
Fendas palpebrais antimongolóides	54	3	1	95
Coloboma de pápebras	37	16	5	70
Ausência de cílios	37	13	8	74
Cabelos pré-auriculares	16	29	13	36

P = presente

A = ausente

N = não-referido

Freq = Frequência igual a $100P/(P+A)$. Corresponde à hipótese de que a presença e a ausência do sinal clínico, nos casos em que o sinal não foi referido (N), estão na mesma proporção em que no resto da amostra.

Tabela 4: Comparação das frequências de sinais clínicos da diferentes grupos da amostra total de pacientes STC.

Sinais Clínicos	Casos isolados CI (n=25)				Casos familiares (total) CFT (n=33)				Casos familiares (propósitos) CFP (n=19)			
	P	A	N	Freq	P	A	N	Freq	P	A	N	Freq
									CI x CFT		CI x CFP	
Hipoplasia malar	24	1	0	96	27	4	2	87	18	0	1	100
Hipoplasia mandibular	24	1	0	96	28	5	0	85	19	0	0	100
Fissura palatina	5	20	0	20	2	31	0	6	2	17	0	11
Displasia do pavilhão auricular	22	3	0	88	19	14	0	58	17	2	0	89
Perda de audição	20	2	3	91	15	12	6	56	13	1	5	93
Fendas palpebrais antimongolóides	24	1	0	96	30	2	1	94	19	0	0	100
Coloboma de pápebras	20	5	0	80	17	11	5	61	13	3	3	81
Ausência de cílios	22	2	1	92	15	11	7	58	11	4	4	73
Cabelos pré-auriculares	7	16	2	30	9	13	11	41	7	6	6	54
										Teste do sinal		
										0,039		0,180

P = presente

A = ausente

N = não-referido

Freq = Frequência igual a $100P/(P+A)$.

NS = não-significante ($p>0,05$)

Teste exato de Fisher, bicaudal, em tabelas de contigência 2X2, uma para cada sinal clínico.

Tabela 5: Comparação das frequências de sinais clínicos dos pacientes TCS com mutação patogênica em diferentes regiões do gene.

Sinais Clínicos	Região 1 Exons 1 a 6				Região 2 Exons 6A a 16				Região 3 Exons 17 a 22				Região 4 Exons 23 a 25			
	P	A	N	Freq	P	A	N	Freq	P	A	N	Freq	P	A	N	Freq
Hipoplasia malar	2	0	0	100	27	2	1	93	6	2	0	75	15	1	1	94
Hipoplasia mandibular	2	0	0	100	29	2	0	94	6	2	0	75	15	2	0	88
Fissura palatina	1	1	0	50	5	26	9	16	0	8	0	0	1	16	0	6
Displasia do pavilhão auricular	2	0	0	100	22	9	0	71	6	2	0	75	11	6	0	65
Perda de audição	2	0	0	100	20	9	2	69	4	2	2	67	9	3	5	75
Fendas palpebrais antimongolóides	2	0	0	100	29	1	1	97	8	0	0	100	15	2	0	88
Coloboma de pápebras	1	1	0	50	21	7	3	75	6	2	0	75	8	6	2	57
Ausência de cílios	2	0	0	100	21	4	6	84	6	2	0	75	9	6	2	60
Cabelos pré-auriculares	2	0	0	100	8	14	9	36	1	7	0	13	5	8	4	38
Total de indivíduos	2				31				8				17			

Tabela 6: Comparação dos sinais clínicos de pacientes TCS do sexo feminino e masculino.

Sinais Clínicos	Feminino (n=35)				Masculino (n=23)			
	P	A	N	Freq (%)	P	A	N	Freq (%)
Hipoplasia malar	28	5	2	85	23	0	0	100
Hipoplasia mandibular	30	5	0	86	22	1	0	96
Fissura palatina	6	29	0	17	1	22	0	4
Displasia do pavilhão auricular	25	10	0	71	16	7	0	70
Perda de audição	20	10	5	67	15	4	4	79
Fendas palpebrais antimongolóides	32	2	1	94	22	1	0	96
Coloboma de pápebras	22	10	3	69	15	6	2	71
Ausência de cílios	22	10	3	69	15	3	5	83
Cabelos pré-auriculares	8	19	8	30	8	10	5	44
								NS

Tabela 7: Comparação das frequências de sinais clínicos entre pacientes STC com mutação patogênica de origem materna ou paterna.

Sinais Clínicos	Origem materna (n=12)				Origem paterna (n=8)			
	P	A	N	Freq (%)	P	A	N	Freq (%)
Hipoplasia malar	12	0	0	100	6	1	1	86
Hipoplasia mandibular	11	1	0	92	8	0	0	100
Fissura palatina	0	12	0	0	2	6	0	25
Displasia do pavilhão auricular	9	3	0	75	6	2	0	75
Perda de audição	7	3	2	70	2	3	3	40
Fendas palpebrais antimongolóides	12	0	0	100	7	0	1	100
Coloboma de pápebras	9	1	2	90	4	3	1	57
Ausência de cílios	7	3	2	70	5	1	2	83
Cabelos pré-auriculares	3	7	2	30	3	2	3	60
								NS

Tabela 8: Haplótipos dos 10 SNPs do gene *TCOF1* em 43 indivíduos portadores da mutação patogênica.

Paciente TCS	Haplótipos*	Cromossomo 1	Cromossomo 2	Ocorrência	Herança	Mutação Patogênica
CR17-2	11;15	CCTGCCCCGC	TTGATTCCAC	F		(EXON 15) c.2585_2594delCAGGGCCAGA
CR23-1	4;7	CTGATCCCCGT	CTGATCTCGC	F	materna	(EXON 12) c.2098_2099delGA
CR54-1	4;11	CTGATCCCCGT	CCTGCCCCGC	I		(EXON 14)c.2478G>A
CR60-1	3;3	CTGATCCCCGC	CTGATCCCCGC	F	paterna	(EXON 15) c.2586_2587delAG
CR71-1	3;11	CTGATCCCCGC	CCTGCCCCGC	I		(EXON 9) c.1397dupT
F239-1	4;5	CTGATCCCCGT	CTGATCCGGC	I	materna	(EXON 24) c.4366_4370delGAAAA
F279-1	3;10	CTGATCCCCGC	CCGGCCCCGC	I		(EXON 21) c.3544delC
F339-1	4;14	CTGATCCCCGT	TTGATCCCCGC	I	materna	(EXON 24) c.4366_4370delGAAAA
F340-1	3;16	CTGATCCCCGC	CCGACCCCCGC	I	paterna	(EXON 10) c.1639_1640delAG
F341-1	10;10	CCGGCCCCGC	CCGGCCCCGC	I		(EXON 20) c.3331delA
F352-1	3;3	CTGATCCCCGC	CTGATCCCCGC	F	materna	(EXON 15) c.2617G>T
F356-1	3;3	CTGATCCCCGC	CTGATCCCCGC	F	materna	(EXON 20) c.3317_3323delCACTCCC
F360-1	3;3	CTGATCCCCGC	CTGATCCCCGC	F	materna	(EXON 18) c.3053_3054delGA
F466-1	4;10	CTGATCCCCGT	CCGGCCCCGC	I	paterna	(EXON 24) c.4366_4370delGAAAA
F471-1	5;10	CTGATCCGGC	CCGGCCCCGC	I		(EXON 10) c.1518dupG
F502-1	3;10	CTGATCCCCGC	CCGGCCCCGC	I	paterna	(EXON 10) c.1637_1640delAGAG

F535-1	3;4	CTGATCCCGC	CTGATCCCGT	F		(EXON 24) c.4366_4370delGAAAA
11033	3;4	CTGATCCCGC	CTGATCCCGT	I		(EXON 24) c.4366_4370delGAAAA
11035	3;10	CTGATCCCGC	CCGGCCCCCGC	F	materna	(EXON 23c) c.4218dupG
F562-1	5;13	CTGATCCGGC	CCTGTCTCGC	I	paterna	(EXON 13) c.2249_2256delCAGTCACCC
F572-1	3;3	CTGATCCCGC	CTGATCCCGC	I	materna	(EXON 24) c.4366_4370delGAAAA
F703-1	3;4	CTGATCCCGC	CTGATCCCGT	F	materna	(EXON 15) c.2585_2594delCAGGGGCCAGA
F737-1	4;14	CTGATCCCGT	TTGATCCCGC	F	paterna	(EXON 23c) c.4218dupG
F761-1	3;5	CTGATCCCGC	CTGATCCGGC	F	materna	(EXON 12) c.2097_2104delAGATAGTG
F762-1	3;4	CTGATCCCGC	CTGATCCCGT	I		(EXON 23a) c.3942_3943delIAG
F770-1	4;5	CTGATCCCGT	CTGATCCGGC	I		(EXON 8) c.1246G>T
F800-1	3;3	CTGATCCCGC	CTGATCCCGC	I		(INTRON 17) c.3046+1delG
F927-1	3;3	CTGATCCCGC	CTGATCCCGC	I		(EXON 9) c.1446dupA
F948-5	4;7	CTGATCCCGT	CTGATCTCGC	F	materna	(EXON 24) c.4366_4370delGAAAA
F977-1	1;2	CTGACCCCGC	CTGACCCGGC	I		(EXON 23b) c.4061delC
F995-1	3;3	CTGATCCCGC	CTGATCCCGC	F		(EXON 24) c.4361_4365delAAAAA
F1253-1	4;4	CTGATCCCGT	CTGATCCCGT	F		(EXON 24) [c.4366dupG,c.4387_4388delAA]
F1255-1	3;4	CTGATCCCGC	CTGATCCCGT	I		(EXON 13) c.2257C>T
CR23-2	3;7	CTGATCCCGC	CTGATCTCGC	F	paterna	(EXON 12) c.2098_2099delGA

CR60-2	3;5	CTGATCCCGC	CTGATCCGGC	F		(EXON 15) c.2586_2587delAG
F352-2	3;10	CTGATCCCGC	CCGGCCCCGC	F	paterna	(EXON 15) c.2617G>T
F352-3	3;5	CTGATCCCGC	CTGATCCGGC	F	materna	(EXON 15) c.2617G>T
F356-2	3;17	CTGATCCCGC	CCTGCCCGGT	F		(EXON 20) c.3317_3323delCACTCCC
F360-2	3;3	CTGATCCCGC	CTGATCCCGC	F		(EXON 18) c.3053_3054delGA
11034	3;5	CTGATCCCGC	CTGATCCGGC	F		(EXON 23c) c.4218dupG
F737-3	5;14	CTGATCCGGC	TTGATCCCGC	F		(EXON 23c) c.4218dupG
F761-2	3;5	CTGATCCCGC	CTGATCCGGC	F		(EXON 12) c.2097_2104delAGATAGTG
F948-7	4;7	CTGATCCCGT	CTGATCTCGC	F		(EXON 24) c.4366_4370delGAAAA

*Haplótipos expressos em cinza (haplótipos em trans com a mutação patogênica).

Tabela 9: Frequência dos sinais clínicos de pacientes STC que expressam os haplótipos mais comuns de SNPs da região codificadora do *TCOF1*.

Sinais Clínicos	Haplótipo 3 CTGATCCCGC				Haplótipo 4 CTGATCCCGT				Haplótipo 5 CTGATCCGGC				Haplótipo 10 CCGGCCCCGC			
	P	A	N	Freq	P	A	N	Freq	P	A	N	Freq	P	A	N	Freq
Hipoplasia malar	8	2	0	80	4	0	1	100	5	0	0	100	4	0	1	100
Hipoplasia mandibular	9	1	0	90	5	0	0	100	4	1	0	80	5	0	0	100
Fissura palatina	2	8	0	20	0	5	0	0	1	4	0	20	0	5	0	0
Displasia do pavilhão auricular	6	4	0	60	4	1	0	80	1	4	0	20	4	1	0	80
Perda de audição	5	3	2	63	2	0	3	100	0	3	2	0	3	1	1	75
Fendas palpebrais antimonilóides	10	0	0	100	5	0	0	100	5	0	0	100	4	0	1	100
Coloboma de pápebras	5	3	2	63	3	1	1	75	3	1	1	75	3	1	1	75
Ausência de cílios	5	2	3	71	2	2	1	50	3	0	2	100	4	0	1	100
Cabelos pré-auriculares	1	5	4	17	1	2	2	33	2	3	0	40	3	1	1	75
Total de indivíduos	10				5				5				5			

A FUNCTIONAL SNP IN THE PROMOTER REGION OF *TCOF1* IS ASSOCIATED WITH REDUCED GENE EXPRESSION AND YY1 DNA–PROTEIN INTERACTION

Cibele Masotti¹, Lucia M. Armelin-Correa¹, Alessandra Splendore², Chin J. Lin³, Angela Barbosa⁴, Mari C. Sogayar⁵, Maria Rita Passos-Bueno¹.

¹Instituto de Biociências, Universidade de São Paulo, Departamento de Genética e Biologia Evolutiva, São Paulo, Brasil

²Instituto Nacional do Câncer, Divisão de Genética, Rio de Janeiro, Brasil

³Laboratório 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, Brasil

³Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Departamento de Patologia, São Paulo, Brasil

⁴Centro de Biotecnologia, Instituto Butantan, São Paulo, Brasil

⁵Instituto de Química, Universidade de São Paulo, Departamento de Bioquímica, São Paulo, Brasil

RESUMO

A síndrome de Treacher Collins (STC) é uma malformação craniofacial autossômica dominante, causada por mutações nulas no gene *TCOF1*. Há alta variabilidade intra e inter-familiar, compreendendo desde hipoplasia malar leve até morte perinatal devido ao colapso de via aéreas. Porém, até o momento, nenhuma associação genótipo-fenótipo foi descrita. Considerando a haploinsuficiência como sendo o mecanismo molecular subjacente à doença, nós hipotetizamos que mutações na região promotora desse gene, que até agora não foi caracterizada, quando em *trans* com a mutação patogênica, poderiam modular o fenótipo. Portanto, os objetivos do presente estudo foram determinar o promotor mínimo do *TCOF1* e identificar mutações nessa região que podem contribuir para a variabilidade fenotípica observada na síndrome. Nós delimitamos o promotor mínimo a uma região de menos de 150 pb, com 63% de identidade entre cinco espécies diferentes. Nós triamos 1.2kbp da região 5' flanqueadora do gene *TCOF1* no DNA de 21 pacientes e 51 controles e identificamos quatro novos polimorfismos de sítio único, um dos quais (-346C>T) mostrou-se funcional, uma vez que reduziu a atividade do promotor em 38%. Ensaio de *elektrophoretic mobility shift* (EMSA) demonstrou que o alelo -346T impede o fator de transcrição YY1 de se ligar ao DNA. Essa variante do promotor representa um alelo candidato para explicar a variabilidade clínica em pacientes com STC.

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 delimited 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 *TCOF1* 5' 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-penetrance has been reported to date (Dixon et al., 1994). The syndrome

Abbreviations: *TCOF1*, Treacher Collins–Franceschetti syndrome gene; YY1, Ying Yang 1 transcription factor; EMSA, Electrophoretic Mobility Shift Assay; SNP, Single Nucleotide Polymorphisms; SSCP, Single Strand Conformation Polymorphism; dHPLC, Denaturing High-Performance Liquid Chromatography; (prime), denotes a truncated gene at the indicated side; UBF, Upstream Binding Factor; bp, basepairs; kb, 1000 basepairs; cpm, counts per minute.

* Corresponding author. Tel.: +55 11 3091 7740; fax: +55 11 3091 7419.

E-mail address: passos@ib.usp.br (M.R. Passos-Bueno).

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 (Transgenomic). 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 χ^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 *Bgl*II or *Hind*III. 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 *Sma*I (pGL3-300wt) or by PCR amplification using primers with a 5' terminal restriction site for the *Hind*III 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 *Microlumet 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 Kruskal–Wallis Test, followed by Dunn's adjustment for comparing all vectors and Friedman 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-

Fig. 1. Schematic representation of the different *TCOF1* promoter fragments inserted in the pGL3-Basic vectors. The –600 bp pol fragment differs from its wild-type version (–600 bp wt) only by the –346C>T nucleotide change. The anti-sense construct is a negative control (pGL3-1200wtAS).

Table 1
Frequencies of the -346C>T, -1025G>C and -948G>A SNPs

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.; 1C6/sc-420 $\times$, d-20/sc-644 and sc-1703 $\times$, 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-

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. 3. Functional activity of *TCOF1* promoter-reporter constructs. Plasmids containing serial deletions of the *TCOF1* promoter (A) and the wild-type and the polymorphic (–346C>T variant) 600 bp constructs (B) were transiently transfected into HepG2 cells. Data are means of at least two-independent transfections, each performed in triplicate, normalized to the pGL3-1200wt construct's activity (relative means $\pm$ SD).

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 $\times$ 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.

Fig. 3A). 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 variant 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 (1 and 2) 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.

Fig. 5. TCS familial pedigrees and the –346C>T SNP segregation. The genotype for the SNP is represented by CC and TT for homozygous and CT for heterozygous individuals. Familial cases are informative to determine whether the pathogenic mutation is in *cis* or in *trans* with the (T) allele. The patients who had inherited the pathogenic mutation in the coding region of *TCOF1* and the T variant from the affected parent probably have both alterations in *cis* (family II). In families I and III, the (T) allele seems to be in *trans* with the pathogenic mutation.

The absence of a pathogenic mutation in 1200 bp of the 5' upstream 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 away 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 TATAA-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 too 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 probands 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 à Pesquisa de São Paulo (FAPESP) and Conselho Nacional de Pesquisa (CNPq).

References

- Den Dunnen, J.T., Antonarakis, S.E., 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Human Mutat.* 15 (1), 7–12.
- Dixon, M.J., 1996. Treacher Collins syndrome. *Hum. Mol. Genet.* 5, 1391–1396.
- Dixon, J., Dixon, M.J., 2004. Genetic background has a major effect on the penetrance and severity of craniofacial defects in mice heterozygous for the gene encoding the nucleolar protein treacle. *Dev. Dyn.* 229, 907–914.
- Dixon, M.J., Marres, H.A.M., Edwards, S.J., Dixon, J., Cremers, C.W.R.J., 1994. Treacher Collins syndrome: correlation between clinical and genetic linkage studies. *Clin. dysmorph.* 3, 96–103.
- Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., Shi, Y., 1999. Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell. Biol.* 19, 7237–7244.
- Edwards, S.J., Gladwin, A.J., Dixon, M.J., 1997. The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon. *Am. J. Hum. Genet.* 60, 515–524.
- Gladwin, A.J., et al., 1996. Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene. *Hum. Mol. Genet.* 5, 1533–1538.
- Gorlin, R.J., Cohen Jr., M.M., Hennekam, R.C.M., 2001. *Syndromes of the Head and Neck*, 4th ed. Oxford University Press, Oxford, pp. 799–802.
- Hayano, T., Yanagida, M., Yamauchi, Y., Shinkawa, T., Isobe, T., Takahashi, N., 2003. Proteomic analysis of human Nop56p-associated pre-ribosomal ribonucleoprotein complexes. Possible link between Nop56p and the nucleolar protein treacle responsible for Treacher Collins syndrome. *J. Biol. Chem.* 278 (36), 34309–34319.
- Isaac, C., et al., 2000. Characterization of the nucleolar gene product, treacle, in Treacher Collins syndrome. *Mol. Biol. Cell.* 11, 188–192.
- Lettice, L.A., et al., 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12 (14), 1725–1735.
- Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16, 1215.
- Olds, L.C., Sibley, E., 2003. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a *cis* regulatory element. *Hum. Mol. Genet.* 12 (18), 2333–2340.
- Poswillo, D., 1975. The pathogenesis of Treacher Collins syndrome (mandibulofacial dysostosis). *Br. J. Oral Surg.* 13, 1–26.
- Read, M., 1998. Electrophoretic mobility shift assay (EMSA). In: Docherty, K. (Ed.), *Essential Techniques*. Wiley & Sons, New York, US.

- Richards, B., et al., 1993. Multiplex PCR amplification from the *CFTR* gene using DNA prepared from buccal brushes/swabs. *Hum. Mol. Genet.* 2, 159–163.
- Smale, S.T., 1997. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta* 1351 (1–2), 73–88.
- So, R.B., Gonzales, B., Henning, D., Dixon, J., Dixon, M.J., Valdez, B.C., 2004. Another face of the Treacher Collins syndrome (*TCOF1*) gene: identification of additional exons. *Gene* 328, 49–57.
- Splendore, A., et al., 2000. High mutation detection rate in *TCOF1* among Treacher Collins syndrome patients reveals clustering of mutations in 16 novel pathogenic changes. *Human Mutat.* 16, 315–322.
- Splendore, A., Jabs, E.W., Passos-Bueno, M.R., 2002. Screening of *TCOF1* in patients from different populations: confirmation of mutational hot spots and identification of a novel missense mutation that suggests an important functional domain in the protein treacle. *J. Med. Genet.* 39, 493–495.
- Splendore, A., Jabs, E.W., Félix, T.M., Passos-Bueno, M.R., 2003. Parental origin of mutations in sporadic cases of Treacher Collins syndrome. *Eur. J. Hum. Genet.* 11 (9), 718–722.
- Splendore, A., Fanganiello, R.D., Masotti, C., Morganti, L.S., Passos-Bueno, M.R., 2005. *TCOF1* mutation database: novel mutation in the alternatively spliced exon 6A and update in mutation nomenclature. *Hum. Mutat.* 25 (5), 429–434.
- Sulik, K.K., Johnston, M.C., Smiley, S.J., Speight, H.S., Jarvis, B.E., 1987. Mandibulofacial dysostosis (Treacher Collins syndrome): a new proposal for its pathogenesis. *A.m. J. Med. Genet.* 27 (2), 359–372.
- Tan, D.P., et al., 2002. YY1 activates *Msx2* gene independent of bone morphogenetic proteins signaling. *Nucleic Acids Res.* 30 (5), 1213–1223.
- Teber, O.A., et al., 2004. Genotyping in 46 patients with tentative diagnosis of Treacher Collins syndrome revealed unexpected phenotypic variation. *Eur. J. Hum. Genet.* 12 (11), 879–890.
- The Treacher Collins Collaborative Group, 1996. Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat. Gen.* 12, 130–136.
- Valdez, B.C., Henning, D., So, R.B., Dixon, J., Dixon, M.J., 2004. The Treacher Collins syndrome (*TCOF1*) gene is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc. Natl. Acad. Sci.* 101 (29), 10709–10714.
- Yan, H., Weishi, Y., Velculescu, V.E., Vogelstein, B., Kinzler, K.W., 2002. Allelic variation in human gene expression. *Science* 297 (16), 1143.

ESTUDO DE SEQUÊNCIAS E PROTEÍNAS POTENCIALMENTE ENVOLVIDAS NA REGULAÇÃO DO GENE TCOF1

Cibele Masotti¹, Michael Lovett² e Maria Rita Passos-Bueno¹.

¹Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil.

²Human Genetics Division, Washington University in Saint Louis, Missouri, EUA.

RESUMO

A síndrome de Treacher Collins é uma malformação craniofacial que afeta estruturas derivadas de primeiro e segundo arcos faríngeos do embrião. O padrão de herança é autossômico dominante, a penetrância é completa e há extrema variabilidade fenotípica inter e intra-familiar. Mutações que alteram o quadro de leitura do gene *TCOF1* são responsáveis pela maioria dos casos de STC; entretanto, não há correlação genótipo-fenótipo. O *TCOF1* se expressa em uma gama variada de tecidos adultos; porém, foi observado um pico de expressão do gene nas células da crista neural, primeiro e segundo arcos faríngeos de camundongos de E8.5 a E9.5. A variação espaço-temporal da expressão do gene sugere a existência de uma regulação específica das células embrionárias dessas regiões. Dessa maneira, julgamos importante a caracterização de mecanismos regulatórios do gene *TCOF1*. No presente trabalho, demos continuidade à caracterização do promotor do *TCOF1*, iniciada em Masotti *et al.* (2005), e identificamos diversas regiões conservadas na sequência 5' do gene, com mais de 75% de semelhança entre humanos e camundongos. Caracterizamos como repressora a ação do fator de transcrição YY1 numa região de 1200pb do promotor do *TCOF1*. Também iniciamos a identificação de fatores de transcrição associados a duas regiões conservadas do promotor por triagem de mono-híbrido, que nos permitiu a elaboração de um modelo para regulação do *TCOF1*.

INTRODUÇÃO

A síndrome de Treacher Collins (STC; OMIM 154500), possui expressividade variável e tem como principais características clínicas anormalidades craniofaciais bilaterais e simétricas, sendo elas: fendas palpebrais em inclinação antimongolóide, coloboma das pálpebras inferiores, ausência de cílios, hipoplasia do osso zigomático, micrognatia, palato ocasionalmente fendido, penugem capilar atípica na face, malformação das orelhas e surdez condutiva (Gorlin *et al.*, 2001). A síndrome apresenta padrão de herança autossômico dominante e penetrância completa (Jones *et al.*, 1997). Como as estruturas afetadas na STC são originárias do primeiro e do segundo arcos faríngeos, cujo mesênquima tem uma grande contribuição da crista neural, sugere-se que a patologia resulte de um defeito na migração das células da crista neural durante o desenvolvimento (Poswillo, 1975; Dixon *et al.*, 2006).

A variabilidade clínica encontrada nos portadores da síndrome é uma questão ainda não respondida. Em algumas famílias, as alterações são tão próximas à normalidade que seguem não diagnosticadas ao longo das gerações, até o nascimento de um descendente mais gravemente afetado. Diversos grupos documentaram a ausência de correlação genótipo-fenótipo na STC (Gladwin *et al.*, 1996; Edwards *et al.* 1997; Splendore *et al.*, 2000; Teber *et al.*, 2004).

Mutações no gene *TCOF1* (5q32-33.1) são responsáveis pela maioria dos casos de STC. Já foram descritas mais de 118 diferentes mutações patogênicas no gene *TCOF1*, sendo a maioria delas causadoras de códons de parada prematura (Splendore *et al.*, 2005). Transcritos portadores de códons de parada prematuros geralmente são degradados pelo mecanismo de *nonsense mediated mRNA decay*, resultando na diminuição de produto gênico. Por isso, sugere-se que a haploinsuficiência da proteína codificada por este gene durante uma fase específica do desenvolvimento embrionário seja responsável pelo quadro clínico da STC (The Treacher Collins Collaborative Group 1996; Gladwin *et al.* 1996; Wise *et al.* 1997; Edwards *et al.* 1997; Dixon *et al.*, 2000; Splendore *et al.* 2000, 2002). O produto gênico, a treacle, é uma proteína nucleolar que participa da biogênese de ribossomos. Pouco se sabe sobre sua função e o modelo mais recente sugere que esteja envolvida na transcrição de rDNA-mediada por UBF1 (*upstream binding factor*) e RNA polimerase I e também no complexo protéico de modificações pós-transcricionais do pré-RNA (Gonzales *et al.*, 2005a).

O *TCOF1* se expressa em uma gama variada de tecidos adultos, tais como fígado, pâncreas, coração, cérebro e pulmão (The Treacher Collins Collaborative Group, 1996). No entanto, há um pico de expressão do gene nas células da crista neural e nos primeiro e segundo arcos faríngeos de camundongos de E8.5 a E9.5 (Dixon *et al.*, 1997a). Essa variação espaço-temporal da expressão do gene sugere que haja um mecanismo regulatório específico das células embrionárias dessas regiões.

Considerando o mecanismo de haploinsuficiência da proteína treacle como base molecular para manifestação da síndrome bem como a existência de uma regulação positiva durante a expressão, julgamos importante a caracterização dos mecanismos regulatórios do gene *TCOF1*. Em Masotti *et al.* (2005), caracterizamos uma região promotora mínima de 152pb sem TATAA-box, rica em GC e que possui mais de 15 sítios consenso para o fator de transcrição ubíquo Sp1. Além disso, identificamos um polimorfismo de base única capaz de diminuir os níveis da transcrição *in vitro* e de interferir na ligação do fator de transcrição YY1. No presente trabalho, demos continuidade à caracterização de seqüências regulatórias do promotor do *TCOF1* e identificamos diversas regiões conservadas (aquelas com mais de 75% de semelhança entre humanos e camundongos) na seqüência/região 5' do gene. Caracterizamos o mecanismo de ação do fator de transcrição YY1 em 1200pb do promotor do *TCOF1* como repressor. Também iniciamos a identificação de fatores de transcrição associados a duas regiões conservadas do promotor por triagem de mono-híbrido, que nos permitiu a elaboração de um modelo para regulação do *TCOF1*.

MATERIAL E MÉTODOS

Análise in silico

Para identificação das regiões conservadas na seqüência promotora humana, utilizamos o UCSC Genome Browser (<http://www.genome.ucsc.edu/>) e o VISTA genome browser (<http://www-gsd.lbl.gov/vista>). O VISTA genome browser é uma ferramenta de visualização da genômica comparativa, que permite visualizar a seqüência inteira do cromossomo de interesse, alinhada às seqüências ortólogas de outras espécies, identificando-se as seqüências com diferentes níveis de conservação (Frazer *et al.*, 2004; Brudno *et al.*, 2003). Também utilizamos a ferramenta VISTA Enhancer Browser (<http://enhancer.lbl.gov/>), que contém dois tipos de dados: (1) um banco de

dados experimental de seqüências humanas não-codificantes, que foram testadas em camundongos transgênicos e caracterizadas como enhancers; (2) um banco de dados computacional de todas as seqüências humanas não-codificantes do genoma, que foram identificadas como potencialmente regulatórias pelo grau de semelhança entre humano, rato e camundongo (Prabhakar *et al.* 2006). Para identificação de sítios de ligação de fatores de transcrição, analisamos o gene *TCOF1* com o programa *Genomatix Software GmbH* (<http://www.genomatix.de/>, Eldorado e Gene2Promoter).

Ensaio de co-transfecção

Co-transfectamos o vetor de expressão de YY1 (CMV-YY1) e o vetor portador da maior região do promotor do *TCOF1* clonada no vetor da luciferase (pGL3-1200Selv, Masotti *et al.*, 2005). O vetor CMV-YY1 foi cedido pelo grupo do Dr. Peter Baumeister (utilizado em Seto *et al.*, 1991), tendo sido seqüenciado para verificar sua integridade. Transfectamos hepatócitos humanos (HepG2, $2,5 \times 10^5$ células/well) com 3 a 4 µg de DNA total por poço (0,5 µg de pRL-TK, 2,5 µg de pGL3-1200Selv e 0,1 a 1 µg de CMV-YY1) em uma placa de 6 poços, com *Lipofectamine2000* (Invitrogen). O vetor pRL-TK (renilla luciferase) foi utilizado para normalizar a eficiência de transfecção. A atividade de cada vetor foi considerada com a média de cada duplicata e os valores relativos da atividade da luciferase foram calculados em relação à atividade do vetor vazio pGL3-basic (Promega).

Ensaio de mono-híbrido

Para identificação de fatores de transcrição associados às regiões conservadas do promotor do *TCOF1*, optamos por utilizar o ensaio de mono-híbrido (Vidal, M. e Legrain, P. 1999). A descrição da metodologia utilizada está dividida em quatro partes: (a) considerações gerais, (b) construção dos plasmídeos pHIS2, (c) transposição da biblioteca de transcritos de E26 e (d) triagem da biblioteca de transcritos de testículos humanos.

(a) Considerações gerais

Utilizamos um kit para construção de biblioteca e triagem da mesma num único ensaio (*Matchmaker™ One-Hybrid Library Construction & Screening kit*, Clontech). As seqüências dos primers estão descritas no Capítulo 2. Utilizamos a linhagem de *Saccharomyces Cerevisiae* Y187, auxotrófica para histidina, e as condições de cultivo, de preparação e de transformação de

leveduras competentes foram realizadas de acordo com o protocolo do fabricante. Para sequenciamento de colônias Y187, purificamos os plasmídeos de leveduras e posteriormente os transformamos em bactérias DH10B eletrocompetentes. Purificamos os plasmídeos (*Qiaprep miniprep kit*, QIAGEN) das colônias que cresceram em meio seletivo e prosseguimos com o sequenciamento direto desses plasmídeos no seqüenciador automático MegaBACE 1000 (GE-Healthcare Bio-Sciences).

(b) Construção dos plasmídeos pHIS2

As seqüências de DNA das regiões conservadas escolhidas para este estudo foram obtidas a partir do *UCSC Genome Browser*. Os primers para as regiões R1, R1 REST, R1 SNP, R1 PMIN, R2 e R3 foram desenhados com o programa *Primer3*, tendo sido adicionados às suas seqüências caudas 5' com sítios para digestão com as endonucleases de restrição EcoRI e MluI. Amplificamos o DNA genômico de um indivíduo controle e os produtos de PCR foram digeridos, purificados e clonados no vetor pHIS2. Todos os plasmídeos foram seqüenciados para confirmar sua integridade. Transformamos leveduras competentes com 700ng dos plasmídeos purificados, de acordo com o protocolo do fabricante.

Padronizamos a quantidade de 3-AT (3-amino-1,2,4-triazol, Sigma) para cada uma das regiões (Figuras 3 e 4). 3-AT é um inibidor competitivo da proteína do gene *HIS3* (His3p), e é usado para suprimir o crescimento de colônias capazes de sobreviverem em meio seletivo (-His) apenas com a fraca expressão basal. A quantidade de 3-AT recomendada pelo fabricante para qualquer construção varia de 5 a 60 mM; no entanto, padronizamos o vetor controle p53HIS2 (Mapa no Anexo2) em 130 mM, cuja expressão "leaky" geralmente é suprimida com 5-10 mM.

(c) Transposição da biblioteca de transcritos de E26

Cai *et al.* 2005 microdissecaram estruturas embrionárias de fetos humanos normais e construíram 12 bibliotecas de cDNAs para SAGE (*serial analysis of gene expression*) e para análise por microarray (Affymetrix). Como grande parte desse trabalho foi realizada pela equipe do Dr. Michael Lovett, essas bibliotecas estavam disponíveis em seu laboratório na ocasião do doutorado-sanduíche. Dessa maneira, dentre essas bibliotecas, optamos por utilizar a biblioteca de E26 (*day 26 human embryo*) para o ensaio de mono-híbrido, uma vez que é a biblioteca do período do desenvolvimento humano que mais se aproxima do período do desenvolvimento embrionário de camundongo (E8 a E9), no qual foi observado um pico de expressão do gene

TCOF1 (Dixon *et al.* 1997). Confirmamos a expressão do gene *TCOF1* e *YY1* nessa biblioteca pela amplificação com primers específicos.

A biblioteca de E26 está clonada no vetor pAMP1, o qual não serve para o estudo de mono-híbrido. Para a realização desse ensaio, iniciamos a transposição dessa biblioteca para o vetor adequado: pGADT7-Rec2. Esse vetor permite que os cDNAs sejam transcritos pelo promotor *ADH1* de leveduras e sejam fusionados com o domínio de ativação da GAL4. A transposição foi feita por meio de recombinação homóloga dos cDNAs amplificados da biblioteca com o vetor linearizado pGADT7-Rec2 (Figura 5).

Os cDNAs de E26 foram amplificados com primers para as regiões de homologia entre ambos os vetores (pGADT7-Rec2 e pAMP1), à exceção da porção 3' dos transcritos. Para que os produtos amplificados da biblioteca de cDNAs fossem capazes de recombinar na sua porção 3' com o vetor pGADT7-Rec2, desenhamos um oligonucleotídeo "adaptador" (40pb), cuja sequência molde e complementar foram aneladas. Esse oligo dupla-fita contém a região 3' adaptadora dos cDNAs da biblioteca e parte do CDSIII terminus do vetor pGADT7-Rec2. As leveduras competentes foram transformadas (método de acetato de lítio) com o produto de PCR dos cDNAs purificado (72ng), com o oligo "adaptador" (72nmol) e com o vetor pGADT7-Rec2 linearizado (1000ng). As leveduras então recombinaram as regiões de homologia e geraram plasmídeos pGADT7-Rec2 contendo os cDNAs da biblioteca de E26. As leveduras transformadas foram plaqueadas em meio seletivo para leucina (SD agar/-Leu). As colônias positivas foram raspadas dessas placas, recuperadas em um único recipiente com meio seletivo líquido (SD/-Leu) por 30 min a 30°C, e congeladas em glicerol. Para obter a representatividade dos transcritos humanos de E26, os criadores da biblioteca (Cai *et al.*, 2005) estimaram um número mínimo de 10⁶ clones positivos para uma transposição adequada.

(d) Triagem da biblioteca de transcritos de testículos humanos.

Para realização do ensaio de mono-híbrido propriamente dito, utilizamos uma biblioteca comercial de cDNAs de testículo humano (*Pre-transformed Matchmaker cDNA AD-library*, Clontech). A biblioteca foi re-estriada, amplificada em aproximadamente 3x10⁶ clones e posteriormente purificada (Qiaprep spin maxiprep kit, QIAGEN). Confirmamos a expressão do gene *TCOF1* e *YY1* nessa biblioteca pela amplificação com primers específicos. Procedemos com a transformação sequencial: primeiramente transformamos cada construção pHIS2 (pHIS2-R1, R1

RESTO, R2 ou R3) e, posteriormente, cada construção com 10µg da biblioteca comercial. Os clones positivos, capazes de sobreviver em meio seletivo para histidina (SD/-His), foram re-plaqueados no mesmo meio seletivo e somente aqueles que repetiram o fenótipo foram seqüenciados.

RESULTADOS

Efeito de YY1 sobre TCOF1

Dado que o fator de transcrição YY1 já foi associado à regulação do gene *TCOF1* (Masotti *et al.*, 2005), investigamos o seu possível papel no promotor do *TCOF1*. Para tanto, realizamos um ensaio de co-transfecção, adicionando quantidades crescentes do vetor de expressão de YY1 (CMV-YY1) a uma quantidade fixa do vetor pGL3-1200Selv (Masotti *et al.*, 2005), e medimos os níveis de luciferase emitidos pela construção contendo o promotor do *TCOF1* e o gene repórter.

Observamos que YY1 claramente exerce um papel de repressor na região de 1200pb do promotor de maneira dose-dependente em HepG2 (Figura 1). À medida que a concentração de CMV-YY1 aumenta, os níveis de luciferase diminuem. A co-transfecção de 1µg de CMV-YY1 com o vetor controle pGL3-basic não alterou significativamente os valores de expressão da luciferase em relação aos valores de pGL3-basic sozinho ($p>0,05$, Teste t).

Identificação de fatores de transcrição candidatos

Procuramos identificar novos sítios de ligação de YY1 na seqüência promotora do *TCOF1*. Analisamos a região de -55.192pb ao início da transcrição (NM_000356) com o programa Genomatix e identificamos 40 sítios preditos, sendo 13 sítios ativadores e 23 repressores. Para verificar a ligação de YY1 aos sítios preditos, assim como identificar novos fatores de transcrição associados ao promotor do *TCOF1*, optamos pelo ensaio de mono-híbrido.

Para realização desse ensaio, escolhemos apenas três regiões potencialmente regulatórias do *TCOF1*, segundo o grau de conservação entre as espécies e proximidade do início da transcrição. Alinhamos a seqüência promotora do *TCOF1* (do início da transcrição do gene vizinho ao início do mRNA NM_000356 do *TCOF1*) no *VISTA genome browser* e identificamos numerosas regiões com mais de 75% de conservação entre humanos e camundongos. Dessas

numerosas regiões, 18 foram identificadas como candidatas a *enhancers* (Figura 2, numeradas de 1 a 18). Dos 23 sítios repressores de YY1, oito estavam localizados em cinco desses enhancers. Por fim, iniciamos o estudo com três regiões (Figura 2): Região 1 (R1, -870 a +17 pb do início da tradução), Região 2 (R2, -4953 a -4273 pb) e Região 3 (R3, -7937 a -7193 pb). R1 contém 860pb da região do vetor pGL3-1200Selv. R2 e R3 contêm sítios de YY1 preditos em sua sequência e são duas das 18 regiões preditas como enhancers.

R1 apresentou crescimento de colônias em até 400 mM de 3-AT (Figura 3). Suspeitamos que a expressão basal aumentada estivesse sendo provocada pelo *core promoter* ou pela região do SNP funcional (c. -346C>T, Masotti *et al.*, 2005). Desmembramos essa região em três novas construções (R1 Promotor Mínimo, R1 SNP Funcional e R1 Resto) e novamente padronizamos a quantidade de 3-AT (Figura 4), porém isto não foi possível para R1 promotor mínimo. Portanto, decidimos realizar o ensaio de mono-híbrido somente com as regiões R1 SNP, R1 Resto, R2 e R3.

Para transposição da biblioteca de E26 (Figura 5), obtivemos uma eficiência de transformação altamente variável e, na grande maioria das vezes, extremamente baixa. Diante dessas dificuldades técnicas com a biblioteca de E26, optamos por realizar o ensaio com uma biblioteca comercial de testículo humano (*Pre-transformed Matchmaker cDNA AD-library* – no vetor pACT2).

Os clones positivos foram re-estriados em meio seletivo (SD/-Trp-Leu-His + 3-AT) e aqueles que repetiram o fenótipo foram posteriormente seqüenciados. Os clones que possuíam fragmentos de cDNA fora de fase de leitura com o domínio de ativação da GAL4 ou cujas seqüências traduzidas não alinharam com qualquer proteína conhecida, foram caracterizados como falso-positivos (Tabela 1). Dado o número de clones positivos obtidos, mesmo com uma alta dose de 3-AT, optamos por descartar a construção R1 SNP. Para R1 Resto somente duas colônias continham a seqüência traduzida em fase correta; observamos, contudo, que não se tratavam de proteínas envolvidas na regulação da transcrição.

Para a R2, obtivemos dois clones em fase, *PFAH1B1* (*platelet-activation factor acetylhydrolase-Ib*, *LIS1* ou *Miller-Diecker Lissencephaly protein*) e *RPS6* (*phosphoprotein NP33*). Como há evidências de que *PFAH1B1* esteja envolvida em regulação da transcrição, optamos por confirmar primeiramente essa interação DNA-proteína.

Para tanto, transformamos novamente leveduras Y187 com os vetores pHIS2-R2 e pACT2-LIS1, no entanto, na proporção 1:1. Realizamos duas transformações-controles para esses vetores: pHIS2-R2 + pACT2 (vazio) e pHIS2 (vazio) + pACT2-LIS1. Obtivemos crescimento de clones positivos em SD/-His em todas as transformações, incluindo os controles negativos.

Acreditando que esse resultado pudesse ser consequência de um problema na linhagem Y187, que poderia estar expressando seu gene repórter além dos níveis basais, optamos por transformar outra linhagem (AH109, disponível em nosso laboratório) com as mesmas construções/controles. No entanto, obtivemos o mesmo crescimento anômalo dos controles negativos em SD/-His.

DISCUSSÃO

YY1 é um fator de transcrição expresso durante o desenvolvimento e regula genes envolvidos na morfogênese craniofacial (Tan *et al.*, 2002; Gordon *et al.*, 2006). Além de funcionar como iniciador da transcrição em promotores sem *TATAA-box*, YY1 também ativa e reprime a transcrição, dependendo do contexto da sequência em que se liga e das proteínas com que interage (revisão em Gordon *et al.*, 2006). Interessantemente, vários sítios de ligação para este fator foram identificados por análise *in silico* na região promotora do *TCOF1*, nos quais YY1 pode atuar como repressor ou indutor da transcrição. No presente trabalho, mostramos que YY1 exerce uma função repressora do *TCOF1*, pois observamos níveis de luciferase reduzidos quando analisamos a sequência 1200pb da região promotora. No entanto, não é possível generalizar o papel desse fator de transcrição como repressor, uma vez que, como mencionado acima, a ação de YY1 é contexto-específica e estamos investigando somente 1200pb dessa sequência promotora. Além disso, é possível que em outros tecidos haja expressão de proteínas que inibam essa repressão.

Cai *et al.* (2005) analisaram o padrão de expressão de aproximadamente 12600 genes (Affymetrix HG-U95Av2) em 25 estruturas craniofaciais de embriões humanos. Utilizando os dados obtidos por estes autores, realizamos uma análise comparativa dos níveis de expressão de *TCOF1* e do fator de transcrição YY1 na proeminência fronto-nasal, primeiro, segundo, terceiro e quarto arcos faríngeos e rombômeros posteriores (quinto ao oitavo) de embriões com quatro semanas de desenvolvimento (Figura 6). Observamos que YY1 exerce uma ação antagônica ao

TCOF1, pois, nas estruturas dissecadas em que *TCOF1* apresenta aumento de expressão, há diminuição de expressão de *YY1*, e vice-versa, com exceção dos rombômeros posteriores. Esses dados indicam que *YY1* é uma proteína importante para a regulação da expressão do *TCOF1* não só em tecidos adultos, conforme apresentado neste trabalho, como também durante o desenvolvimento das estruturas da face. É interessante observar que em ambas as situações o *YY1* parece estar relacionado com a inibição de *TCOF1*.

Identificamos duas proteínas candidatas à regulação do *TCOF1* ligadas à R2 (-5.000pb): *PFAH1B1* (*platelet-activation factor acetylhydrolase-Ib*, *LIS1* ou *Miller-Diecker Lissencephaly protein*, OMIM 601545) e *RPS6* (*phosphoprotein NP33*, OMIM 180460). No entanto, devido às dificuldades técnicas do ensaio, essas interações DNA-proteína não foram confirmadas. Como há evidências de que *PFAH1B1* esteja envolvida em regulação da transcrição através de seu domínio WD-40 e, além disso, essa proteína possui o domínio LiSH, envolvido em dinâmica de microtúbulos, que também está presente na porção N-terminal da treacle, julgamos importante futuramente investigar essa possível interação. Da mesma maneira, não descartamos a possibilidade de interação da ribonucleoproteína S6 com R2, pois se trata de um componente da subunidade ribossômica 40S, altamente fosforilado mediante estímulo mitogênico, e pouco se sabe a respeito de sua função (Roux *et al.*, 2007).

A expressão desenfreada do gene repórter (histidina) dada pelas construções R1 PMÍN e R1 SNP demonstrou que essas regiões promotoras do *TCOF1* estavam recrutando fatores de transcrição da levedura e funcionando, portanto, como regiões regulatórias nesse organismo. Dessa maneira, essas regiões do promotor *TCOF1* parecem ser bastante “promíscuas”, ou seja, não há muita especificidade para a ativação da transcrição. Os dados obtidos com o ensaio da luciferase para as regiões do SNP funcional e do promotor mínimo (clonadas nos três vetores pGL3-200Sel, pGL3-300Sel e pGL3-400Sel; Masotti *et al.* 2005) mostram que essas promovem uma maior transcrição do gene repórter em comparação aos fragmentos maiores (pGL3-1200Sel e pGL3-600Sel; R1 possui 323 pb a menos que pGL3-1200Sel). Esses dados nos levaram a elaborar uma nova hipótese para o funcionamento do promotor do *TCOF1*: o promotor mínimo parece estar sempre ativo e os fatores de transcrição necessários para a sua ativação parecem ser pouco específicos. Dessa maneira, a regulação do *TCOF1* parece ocorrer por repressão do seu promotor mínimo. Assim, acreditamos que os elementos repressores, possivelmente, estão localizados nas regiões conservadas que se distribuem ao longo da sequência 5' do gene.

Segundo Nobrega e Pennacchio (2004), regiões não-codificantes com alto grau de conservação entre diferentes espécies compreendem seqüências regulatórias importantes, e, pela distância em relação ao início da transcrição, possivelmente, funcionam como enhancers do promotor mínimo.

Nossos achados sugerem que *YY1* seja um fator de transcrição importante na regulação do *TCOF1*. Se esse fator de transcrição atuar como repressor durante o desenvolvimento, é possível que o pico de expressão de *Tcof1*, observado em camundongos de E8.5 a E9.5 (Dixon *et al.*, 1997a), ocorra por um alívio da repressão por *YY1*. Considerando a hipótese de repressão, além dos demais achados, julgamos bastante importante dar continuidade ao presente estudo.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglu S; Glocal alignment: finding rearrangements during alignment. Special Issue on the Proceedings of the ISMB 2003, Bioinformatics, 19: 54i-62i.
2. Cai J, Ash D, Kotch LE, Jabs EW, Attie-Bitach T, Auge J, Mattei G, Etchevers H, Vekemans M, Korshunova Y, Tidwell R, Messina DN, Winston JB, Lovett M. Gene expression in pharyngeal arch 1 during human embryonic development. *Hum Mol Genet*. 2005 Apr 1;14(7):903-12.
3. Dixon, J.; Hovanes, K.; Shiang, R.; Dixon, M.J. (1997a) Sequence analysis, identification of evolutionary conserved motifs and expression analysis of murine *Tcof1* provide further evidence for a potential function for the gene and its human homologue, *TCOF1*. *Hum Mol Genet* 6: 727-737
4. Dixon J, Jones NC, Sandell LL, Jayasinghe SM, Crane J, Rey JP, Dixon MJ, Trainor PA. (2006) *Tcof1/Treacle* is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc Natl Acad Sci U S A*. Sep 5;103(36):13403-8.
5. Dixon, J.; Edwards, S.J.; Anderson, I.; Brass, A.; Scambler, P.J.; Dixon, M.J. (1997b) Identification of the Complete Coding Sequence and Genomic Organization of the Treacher Collins Syndrome Gene. *Genome Research* 7:223-234
6. Edwards, S.J.; Gladwin, A.J.; Dixon, M.J. (1997) The Mutational Spectrum in Treacher Collins Syndrome Reveals a Predominance of Mutations That Create a Premature-Termination Codon. *Am J Hum Genet* 60:515-524
7. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004 Jul 1;32 (Web Server issue):W273-9
8. Gladwin, A.J.; Dixon, J.; Loftus, S.K.; Edwards, S.; Wasmuth, J.J.; Hennekam, R.C.M.; Dixon, M.J. (1996) Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene. *Hum Mol Gen* 5:1533-1538

9. Gonzales B, Henning D, So RB, Dixon J, Dixon MJ, Valdez BC. (2005a) The Treacher Collins syndrome (TCOF1) gene product is involved in pre-rRNA methylation. *Hum Mol Gene* Jul 15;14(14):2035-43.
10. Gordon S, Akopyan G, Garban H, Bonavida B. (2006) Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*. Feb 23;25(8):1125-42.
11. Gorlin, R.J.; Cohen, M.M.; Hennekam, R.C. (2001) Branchial Arch and oral-Acral Disorders. In: Syndromes of the head and neck, 4th edition. Oxford: Oxford University Press. p.790-849.
12. Jones, K.L. (1997) Smith's Recognizable Patterns of Human Malformations 5th ed. Pennsylvania: Saunders.
13. Masotti C, Armelin-Correa LM, Splendore A, Lin CJ, Barbosa A, Sogayar MC, Passos-Bueno MR. (2005) A functional SNP in the promoter region of TCOF1 is associated with reduced gene expression and YY1 DNA-protein interaction. *Gene*. Oct 10;359:44-52.
14. Nobrega MA, Pennacchio LA. (2004) Comparative genomic analysis as a tool for biological discovery. *J Physiol*. Jan 1;554(Pt 1):31-9. Review.
15. Online Mendelian Inheritance in Man (OMIM) Johns Hopkins University, Baltimore, M.D. <http://www.ncbi.nlm.nih.gov/omim>
16. Poswillo, D. (1975) The pathogenesis of Treacher Collins syndrome (mandibulofacial dysostosis). *Br J Oral Surg* 13:1-26
17. Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, Pennacchio LA. (2006) Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res*. 16:855-63
18. Roux PP, Shahbazian D, Vu H, Holz MK, Cohen MS, Taunton J, Sonenberg N, Blenis J. (2007) RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. *J Biol Chem*. Mar 14; [Epub ahead of print]
19. Smale ST. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim Biophys Acta*. 1997 Mar 20;1351(1-2):73-88.
20. Seto, E; Shi, Y; Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription *in vivo*. *Nature* Nov 21; 354 (6350): 241-5.

21. Splendore, A.; Silva, E.O.; Alonso, L.G.; Richieri-Costa, A.; Alonso, N.; Rosa, A.; Carakushansky G.; Cavalcanti, D.P.; Brunoni, D.; Passos-Bueno, M.R. (2000) High mutation detection rate in *TCOF1* among Treacher Collins syndrome patients reveals clustering of mutations in 16 novel pathogenic changes. *Hum Mutat* 16:315-322
22. Splendore, A.; Fanganiello, R.D.; Masotti, C.; Morganti, L.S.C.; Passos-Bueno, M.R. (2005) *TCOF1* mutation database: novel mutation in the alternatively spliced exon 6A and update in mutation nomenclature. *Hum Mut*, 25: 429-434.
23. Tan DP, Nonaka K, Nuckolls GH, Liu YH, Maxson RE, Slavkin HC, Shum L. YY1 activates Msx2 gene independent of bone morphogenetic protein signaling. *Nucleic Acids Res.* 2002 Mar 1;30(5):1213-23.
24. Teber OA, Gillesen-Kaesbach G, Fischer S, Bohringer S, Albrecht B, Albert A, Arslan-Kirchner M, Haan E, Hagedorn-Greiwe M, Hammans C, Henn W, Hinkel GK, Konig R, Kunstmann E, Kunze J, Neumann LM, Prott EC, Rauch A, Rott HD, Seidel H, Spranger S, Sprengel M, Zoll B, Lohmann DR, Wieczorek D. (2004) Genotyping in 46 patients with tentative diagnosis of Treacher Collins syndrome revealed unexpected phenotypic variation. *Eur J Hum Genet.* 12(11):879-90.
25. The Treacher Collins Collaborative Group (1996) Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat Gen* 12:130-136
26. Vidal M, Legrain P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* 1999 Feb 15;27(4):919-29. Review.
27. Wise, C.A.; Chiang, L.C.; Paznekas, W.A.; Sharma, M.; Musy, M.M.; Ashley, J.A.; Lovett, M. Jabs, E.W. (1997) *TCOF1* gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treacher Collins Syndrome throughout its coding region. *Proc Natl Acad Sci USA* 94:3110-3115

Figura 1: Ensaio de co-transfecção com YY1 e pGL3-1200Selv. Atividade relativa da luciferase do vetor pGL3-1200Selv (Masotti *et al.*, 2005) (2,5µg) na ausência ou na presença de CMV-YY1 (0,1 a 1ug) em HepG2. As médias das duplicatas de cada vetor foram normalizadas pela média da duplicata do vetor pGL3-Basic (ausente no gráfico).

Chr5:149,656,428-149,760,063

Base genome: Human Mar. 2006

Figura 2: Alinhamento da seqüência genômica humana (entre o início do gene *ARS1* e o final do gene *TCOF1*) com a seqüência ortóloga de camundongo (janela superior) e de chimpanzé (janela inferior). Regiões em rosa são conservadas e não-codificadoras; regiões em azul são seqüências codificantes conservadas (exons); regiões em azul claro são seqüências transcritas e não-traduzidas (5' e 3' UTR). Em verde, posição dos sítios de YY1 preditos nas regiões conservadas não-codificadoras. Em vermelho, posição do sítio de YY1 estudado em Masotti *et al.*(2005). As regiões enquadradas foram estudadas pelo ensaio de mono-híbrido: R3(14), R2(15) e R1. Enumeramos de 1 a 18 as regiões conservadas preditas como enhancers pelo *VISTA Enhancer database*.

Construções	pHIS2-R1	pHIS2-R2	pHIS2-R3	pHIS2-R1 P _{min}	pHIS2-R1 SNP	pHIS2-R1Resto
Quantidade de 3-AT	>400 mM	60 mM	200 mM	>300 mM	>200 mM	100 mM

Figura 3: Padronização de 3-AT em meio seletivo (SD/-Trp-His) seleciona as células que incorporaram o vetor pHIS2, com crescentes quantidades do inibidor. A concentração ideal para realização do ensaio é a menor capaz de suprimir o "background" de colônias com expressão basal. Tabela com as concentrações de 3-AT padronizadas para cada construção. Algumas linhagens de leveduras transformadas com o vetor pHIS2 exibem uma expressão basal fraca de *HIS3*, o gene repórter para interação DNA-proteína. No entanto, essa expressão basal, conhecida como "leaky expression", é inibida pela presença de 3-AT no meio seletivo. 3-AT (3-amino-1,2,4-triazol) é um inibidor competitivo da proteína do gene *HIS3* (His3p), e é usado para suprimir o crescimento de colônias capazes de sobreviverem em meio seletivo (-His) apenas com a fraca expressão basal. Além disso, diferentes construções (região regulatória + pHIS2) têm diferentes capacidades de promover a expressão basal de *HIS3*. Por isso, é importante padronizar a quantidade de inibidor (3-AT) no meio seletivo para cada uma das construções, a fim de evitar o aparecimento de colônias positivas provenientes de "leaky expression" e não de verdadeiras interações DNA-proteína.

Figura 4: Região 1 (-870 a +17) reconstruída. RPMÍN (Promotor mínimo restringido pelo ensaio da luciferase), R1 SNP (Região que contém o SNP funcional, indicado com um retângulo vermelho porém, nessa construção o alelo é selvagem; -346C) e R1 RESTO. Ao lado, as padronizações de 3-AT definidas para as novas regiões.

Figura 5: Esquema da transposição dos vetores da biblioteca de E26. Os cDNAs da biblioteca de E26 são amplificados com primers com homologia para as regiões de clonagem do vetor pGADT7-Rec2. Após a preparação de células competentes, essas são transformadas (método de Acetato de Lítio) com o produto de PCR dos cDNAs, com o oligonucleotídeo dupla-fita "adaptador" e com o vetor pGADT7-Rec2 linearizado. As leveduras então recombina as regiões de homologia e geram plasmídeos contendo os cDNAs de interesse. As leveduras transformadas são plaqueadas em meio seletivo para leucina. Para obter a representatividade dos transcritos humanos de E26, é necessário o número de 10⁶ clones positivos.

Figura 6: Expressão normalizada de *TCOF1* e *YY1* nas seguintes estruturas embrionárias humanas com quatro semanas de desenvolvimento (E28): ProFroNa (proeminência fronto-nasal), 1° ArFar (primeiro arco faríngeo), 2° ArFar (segundo arco faríngeo), 3° /4° ArFar (terceiro e quarto arcos faríngeos), Romb Post (rombômeros posteriores – 5° a 8°). Dados extraídos dos ensaios de microarrays de Cai *et al.* (2005), disponibilizados pelo Dr. Michael Lovett.

Tabela 1: Resultados da triagem da biblioteca de testículos humanos.

Construções	R1 SNP		R1 Resto		R2			R3		
	1º ensaio		1º ensaio	2º ensaio	1º ensaio	2º ensaio	3º ensaio	1º ensaio	2º ensaio	3º ensaio
Ensaios										
Eficiência de transformação (colônias/ μg pACT2)	$2,6 \times 10^6$		$2,6 \times 10^6$	1×10^4	5×10^5	1×10^3	1×10^4	3×10^4	1×10^3	1×10^4
Número de clones triados	$2,6 \times 10^6$		$2,6 \times 10^6$	1×10^4	5×10^5	ED	1×10^4	3×10^4	ED	1×10^4
Número de colônias positivas	> 100	41	41	4	34	-	12	0	-	4
Colônias re-estriadas em meio seletivo que repetiram o fenótipo	50	41	41	4	34	-	12	0	-	4
Colônias sequenciadas	12	37	37	-	30	-	-	0	-	4
Falsos-positivos	12	37	37	-	30	-	-	0	-	4
Resultado	Construção descartada	2 clones em fase	Priorizamos 1º ensaio	2 clones em fase	-	-	Priorizamos 1º ensaio	Nenhum positivo	-	Nenhum clone em fase

ED, Experimento Desconsiderado.

Capítulo 6

TCOF1 EXPRESSION LEVELS ARE DIFFERENT BETWEEN TCS PATIENTS AND CONTROLS, BUT ARE NOT ASSOCIATED TO PHENOTYPIC VARIATION AND CPG ISLAND METHYLATION STATUS

Cibele Masotti¹, Ricardo Moura², Felícia Cavalher², Têmis Félix³, Nivaldo Alonso⁴, Anamaria Aranha Camargo² e Maria Rita Passos-Bueno¹.

¹Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil

²Ludwig Institute for Cancer Research, São Paulo Branch, Hospital Alemão Oswaldo Cruz, São Paulo, SP, Brazil

³Departamento de Genética Médica, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

⁴Departamento de Cirurgia Plástica, Hospital das Clínicas da Faculdade de Medicina, Universidade de São Paulo, SP, Brazil

RESUMO

A síndrome de Treacher Collins (STC) é uma malformação craniofacial autossômica dominante causada por mutações no *TCOF1*. Mutações STC são deleções ou duplicações que levam a mudanças no quadro de leitura, resultando em códons de parada prematuros que levam à degradação do RNA mensageiro por *nonsense mediated mRNA decay (NMD)*. Haploinsuficiência do produto gênico (treacle) durante o desenvolvimento embrionário é o mecanismo molecular subjacente à STC. Nenhuma correlação genótipo-fenótipo foi observada até o momento e nós hipotetizamos que a variabilidade clínica pode estar associada à variação nos níveis de expressão gênica no alelo selvagem dos pacientes STC. No presente trabalho nós investigamos se o níveis de transcritos de *TCOF1* em linfócitos está associado à variação fenotípica na STC. Nós encontramos uma grande variação na expressão de *TCOF1* tanto entre indivíduos normais como afetados, mas nenhuma associação foi observada entre níveis de transcrição de *TCOF1* e a gravidade do fenótipo STC. Nós também observamos uma diferença significativa nos níveis de RNA mensageiro de *TCOF1* entre os dois grupos, indicando pela primeira vez que as células maduras de pacientes não possuem a mesma quantidade de transcritos de *TCOF1* que indivíduos normais. Além disso, nós investigamos a condição de metilação das ilhas CpG do promotor do *TCOF1* em diferentes tipos de célula de pacientes STC e controles. A ausência de citosina metiladas sugeriu que esse mecanismos de regulação não é responsável pela variação individual nos níveis de expressão de *TCOF1* em pacientes e controles, ou em células diferenciadas ou indiferenciadas.

Capítulo 6

***TCOF1* EXPRESSION LEVELS ARE DIFFERENT BETWEEN TCS PATIENTS AND CONTROLS, BUT ARE NOT ASSOCIATED TO PHENOTYPIC VARIATION AND CPG ISLAND METHYLATION STATUS**

Cibele Masotti¹, Ricardo Moura², Felícia Cavalher², Têmis Félix³, Nivaldo Alonso⁴, Anamaria Aranha Camargo² e Maria Rita Passos-Bueno¹.

¹Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil

²Ludwig Institute for Cancer Research, São Paulo Branch, Hospital Alemão Oswaldo Cruz, São Paulo, SP, Brazil

³Departamento de Genética Médica, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

⁴Departamento de Cirurgia Plástica, Hospital das Clínicas da Faculdade de Medicina, Universidade de São Paulo, SP, Brazil

ABSTRACT

Treacher Collins syndrome is an autosomal dominant craniofacial disorder caused by mutations in *TCOF1*. TCS mutations are frameshift deletions or duplications that cause premature termination codons that lead to mRNA degradation by nonsense mediated mRNA decay (NMD). Haploinsufficiency of the gene product (treacle) during embryonic development is the molecular mechanism underlying TCS. No genotype-phenotype correlation has been observed so far and we have hypothesized that clinical variability could be associated to variation in gene expression levels of the wild-type allele in TCS patients. In the present report we investigated if *TCOF1* transcript levels in lymphocytes were associated to phenotypic variation in TCS. We detected a wide variation of *TCOF1* expression among normal and affected individuals, but no association was observed between *TCOF1* transcription levels and severity of the TCS phenotype. We also observed a significant difference in *TCOF1* mRNA levels between the two groups, demonstrating for the first time that patients' mature cells do not have the same amount of *TCOF1* transcripts as normal individuals. In addition, we investigated the methylation status of *TCOF1* promoter's CpG island in different cell types from normal or TCS patients. The absence of methylated cytosines suggested that this regulatory mechanism is not responsible for individual *TCOF1* expression variation in patients and controls, or in differentiated and undifferentiated cell lines.

INTRODUCTION

Treacher Collins syndrome (TCS; OMIM 154500) is an autosomal dominant craniofacial disorder caused by mutations in *TCOF1* (OMIM 606847; The Treacher Collins Collaborative Group, 1996). The gene encodes a nucleolar protein called treacle, which has recently been described to be involved in rRNA transcription by interacting with UBF1 and RNA polymerase I, and also to participate in pre-rRNA post-transcriptional modification by interacting with hNop56 from pre-rRNA methylation complex (Gonzalez *et al.*, 2005).

More than 80% of TCS mutations are frameshift deletions or duplications of 1-41bp in *TCOF1* coding region that cause premature termination codons (PTC). Mutations are usually family-specific, with the exception of a recurring 5bp deletion in exon 24 that is responsible for 17% of the cases (reviewed in Splendore *et al.*, 2005). Truncated proteins are not detected in TCS patients' fibroblast and lymphocyte cytoplasm, suggesting that mRNA bearing PTCs are being degraded by nonsense mediated mRNA decay (NMD; Isaac *et al.*, 2000).

TCOF1 is expressed in various adult and embryonic tissues (The Treacher Collins Collaborative Group, 1996), and haploinsufficiency of treacle during embryonic development is the molecular mechanism underlying TCS. In situ hybridization studies of the *Tcof1* orthologue demonstrated that there is a peak of expression of the gene in E8.5-9.5 mice embryos, especially in the first and second pharyngeal arch. These authors suggested that abnormal levels of treacle during this specific embryonic period result in the craniofacial abnormalities observed in TCS (Dixon, J *et al.*, 1997a). The strongest evidence came with *Tcof1*^{+/-} mice embryos that presented high levels of apoptosis in their migrating neural crest cells (Dixon, J *et al.*, 2000), demonstrating that critical dosage of *Tcof1* gene product is necessary to normal development. Interestingly, Isaac *et al.* (2000) identified the same amounts of treacle in fibroblasts and lymphocytes derived either from adult patients or controls, and these authors suggested that a dosage compensation mechanism could occur in adulthood to compensate the null allele.

TCS patients present bilateral and symmetrical malformations, which frequently include hypoplasia of the mandible and zygomatic complex, down-slanting palpebral fissures, coloboma of the lower eyelid and absence of eyelashes medial to this defect, external and middle ear malformation, and conductive hearing loss (Gorlin, 2001). The penetrance is considered to be complete (Dixon *et al.*, 1994), but there is a high inter and intra-familial phenotypic variation,

ranging from cases with perinatal death due to airway obstruction by severe orofacial malformations to those that are not clinically diagnosed.

No genotype-phenotype correlation has been observed in TCS and there is also no evidence of association between the disease severity and parental origin or type of the pathogenic mutation, male or female sex, sporadic or familial cases (Gladwin *et al.*, 1996; Dixon, 1996; Edwards *et al.*, 1997; Splendore *et al.*, 2000; Splendore *et al.*, 2005; Masotti *et al.*, 2007). By testing *Tcof1*^{+/-} mice mutation in different genetic backgrounds, Dixon and Dixon (2004) observed a high phenotypic variation, similar to the observed in TCS patients. These authors proposed that genetic background was responsible for disease expressivity. The methylation level of pre-rRNA in mutant mice was also correlated to phenotype severity (Gonzales *et al.*, 2005).

Considering that *TCOF1* dosage is critical for craniofacial development, we have hypothesized that clinical variability is associated to gene expression variation levels of the wild-type allele, as the mutant is degraded by NMD. We identified a functional single nucleotide polymorphism (SNP) that decreases *in vitro* transcription and interferes in YY1 transcription factor DNA binding in HepG2 cell line. Although these results suggest that *TCOF1* expression levels are at least partially dependent on genetic factors, we were unable to verify if this SNP could play an important role in the modulation of the TCS phenotype (Masotti *et al.*, 2005).

In the present report we investigate if *TCOF1* transcript levels are associated to phenotypic variation in TCS. First, we detected a wide variation of *TCOF1* expression in lymphocytes among both normal and affected individuals, but no association was observed between *TCOF1* transcription levels and severity of the TCS phenotype. Second, comparing patients with controls, we demonstrated a significant difference in *TCOF1* mRNA levels between the two groups. This result is in accordance with the haploinsufficiency mechanism and shows for the first time that TCS mature cells do not have the same amount of *TCOF1* transcripts as normal individuals. We also investigated a large CpG island predicted in 5' region of *TCOF1* in order to verify if TCS patients have abnormal methylation status in comparison to controls as well as if methylation is involved in the regulation of *TCOF1* transcription levels. The analysis did not reveal any considerable methylation and we show that promoter methylation is not the mechanism responsible for individual *TCOF1* expression variation in patients and controls, and in differentiated and undifferentiated cell lines.

SUBJECT AND METHODS

Patients, controls and cell lines

We analyzed 22 TCS patients (with 3 to 41 years old) referred to our center for genetic counseling and 15 controls (with 20 to 50 years old). The study was approved by the ethical committee of our Institution and informed consent was obtained from both patients and control subjects or from their legal tutors. The human immortalized cell lines HepG2 (hepatocellular carcinoma), HEK293 (human epithelium from kidney), HUVEC (human umbilical vein endothelial cells), HeLa (cervical carcinoma), SH-SY5y (neuroblastoma) and hFOB (human fetal osteoblasts) were kindly donated by Prof. Dr. M.C. Sogayar, and SaOS-2 (epithelial-like osteosarcoma) was purchased from American Type Culture Collection (Manassas, VA, USA). Dental pulp stem cells were obtained from a deciduous tooth of a normal child and culture was established by Bueno, DF 2007.

For *TCOF1* expression and phenotype correlation, we considered nine major clinical signs of TCS: zygomatic complex hypoplasia, micrognathia, cleft palate, abnormal external ears, hearing impairment, downward slanting palpebral fissures, lower eyelid coloboma, absence of eyelashes, and hair projecting onto face. Clinical signs were classified as *P* (present), *A* (absent), *N* (not reported), and the frequency of clinical signs per patient was calculated according to formula $\text{frequency} = P/(P+A)$ (Gollop, 1981).

RNA isolation and cDNA synthesis

Total RNA from lymphocytes and cell lines was isolated using Trizol® (Gibco BRL), analyzed in 1.2% agarose gel after denaturation and quantified by measuring UV absorbance at 260 nm in the spectrophotometer NanoDrop ND-1000 (*Peqlab biotechnologie GmbH*). Aliquots of total RNA were treated with *DNAse* (Promega) and reverse transcription reactions were performed with 1µg of DNA-free RNA using *Superscript II Reverse Transcriptase* (Gibco BRL), according to manufacturer's protocol.

Real Time PCR

Real time quantitative PCR (RT-PCR) reactions were performed in duplicates with a final volume of 20µl, using 12,5 ng of cDNA, 1X *SYBR Green PCR Master Mix* (Applied Biosystems),

and 200 or 400nM of each primer. We used *ABI Prism 7700 Sequence Detection System* (Applied Biosystems) with standard temperature protocol. Primers were designed with Primer Express software V.2.0 (Applied Biosystems; sequences available on request) and the amplification efficiency (E) of each primer was calculated according to the equation $E = 10^{(-1/\text{slope})}$. The expression data of *TCOF1* transcripts was determined by relative quantification in comparison to four endogenous controls (*GAPDH*, *BCRP*, *HPRT1*, and *HMBS*), according to the methodology proposed by Pfaffel (2001). To calculate ΔC_T s we used as reference sample a cDNA synthesized with a pool of RNAs from 12 controls. We verified the gene expression stability of endogenous controls through *geNorm* VBA applet designed for Microsoft Excel. This tool calculates the most stable reference genes from a set of tested candidate reference genes in a given sample panel, and calculates the gene expression normalization factor (NF) for each target sample based on the geometric mean of a user-defined number of housekeeping genes (Vandesompele *et al.*, 2002; <http://medgen.ugent.be/~jvdesomp/genorm/>). We calculated the NF for each sample based on the four endogenous controls. The expression data in Table 1 and Graphic 1 is given by the ratio between *TCOF1* $\Delta\Delta C_T(E^{\Delta C_T})$ and NF.

In silico CpG island identification

We used the *UCSC genome browser* (University of California-Santa Cruz Genome Bioinformatics, <http://genome.cse.ucsc.edu>) to identify a CpG island in the *TCOF1* sequence. The interspecies conservation degree was observed through alignment of human, mouse, rat, chimpanzee and dog assemblies in this browser.

Analysis of DNA methylation by sequencing of sodium bisulfite-treated DNA

Genomic DNA of patients and controls was extracted from peripheral blood samples, according to Miller *et al.* (1998), and DNA from cultured cell lines was obtained by digestion with proteinase K (Gibco, BRL) followed by phenol/chloroform extraction. The genomic DNA was subjected to sodium bisulfite treatment to modify unmethylated cytosine to uracil using standardized protocol (Goldenberg *et al.*, 2004). The CpG island of *TCOF1* was divided in two amplicons, CpG1 (743pb) and CpG2 (255pb), to facilitate cloning and sequencing. We amplified bisulfite-treated DNA with a nested-PCR protocol using the following primers (5'-3'): CpG_F-GTG TGA AAT GGA TAA ATG AGT AGG T and CpG_R-CTA AAA ACC TAA ATT CCA ACT CAA CT for the first reaction; CpG1_F-GGA TTT AGA GAA AGA TTT AGA TTA GGG, CpG1_R-ACA CAC

TTT CTA ACC CAA CCC ACT AC, CpG2_F-GGG AGT TAT TTT TTT TGA TTT ATT A, and CpG2_R-AAC TAT AAA ATC CCT TTT CCA AAC for the second reaction. PCR was performed in a volume of 25µl containing 1XPCR buffer (Applied Biosystems); 1.5mM of MgCl₂ (Applied Biosystems); 200µM of dNTPs (Amersham); 0.32µM of each primer, and 1U of *Taq* GOLD (Applied Biosystems). The PCR conditions were: 94°C for 12min, 94 °C for 3min, 53 °C for 3min, 72 °C for 2 min; five cycles of 94 °C for 3min, 55 °C for 3min, and 72 °C for 2min; and 35 cycles of 94 °C for 1min, 57 °C for 1min, 72 °C for 1min, for first and second (nested) reaction. Amplified products were cloned into a TA cloning vector (pGEM-T-Easy vector, Promega) and at least four positive clones were sequenced for each amplicon (CpG1 and CpG2) from 8 controls, 8 patients and 6 cell lines. We sequenced using the vector's universal primers set (M13F and M13R) and an internal primer for CpG1 (CpG1int_R-TAA TAA ATC AAA AAA AAT AAC TCC C), using BigDye V.3.1 and an automated sequencer ABI 3100 (*Applied Biosystems*). The methylation percentage for each sample was calculated as the number of methylated CpG dinucleotides from the total number of CpGs analyzed in all positive clones.

RESULTS

Expression data in TCS patients and controls

We first verified the range of *TCOF1* transcription levels in patients (n=20), controls (n=12) and 8 different cell lines through relative quantification RT-PCR assay. Besides *TCOF1*, we also analyzed the expression of four endogenous controls (*GAPDH*, *BCRP*, *HPRT1*, and *HMBS*), which were highly stable in all 40 samples, and the normalization factor (NF) was calculated based on the four reference genes.

We observed a wide expression variation in *TCOF1* transcript levels among both patients and controls, with a standard deviation of approximately 1.5. Comparing individual expression variation in the two groups independently, the standard deviation is 1.2 for patients and 1.4 for controls (Table 1), and variances are not significantly different ($p=0.5828$, F test). In patients, the expression variation is not correlated to age ($p=0.39$; Two-tailed Pearson correlation regression).

Comparing patients (MEAN $\pm$ SEM = 6.846 ± 0.3278) and controls (MEAN $\pm$ SEM = 8.093 ± 0.3587), we observed that *TCOF1* expression levels are significantly different in these two

groups ($p=0,0164$; Two-tailed unpaired t test with Welch's correction; Figure 1). TCS patients have approximately 18% less *TCOF1* transcript levels than normal individuals.

We examined if *TCOF1* expression varies among different tissues by comparing the expression levels in six different cell lines: HUVEC, HEK293, SH-SY5y, hFOB, HepG2, HeLA, and SaOS-2. There is a wide variation in expression values among cell lines (Table 1), HUVEC cells having the lower and hFOB the higher expression. It is important to note that hFOB and SH-SY5y have similar expression values to lymphocytes (Figure 2).

We also examined whether *TCOF1* expression levels are correlated to phenotypic variation in TCS patients. The comparison of normalized expression data and frequency of clinical signs (Table 1) revealed no significant correlation between the values ($p=0.414$, Two-tailed Pearson correlation regression). Individuals with similar expression values have different clinical severity degree (Figure 3).

TCOF1 promoter CpG island identification and methylation analysis

In silico analysis revealed a large predicted CpG island of 753 bp for *TCOF1*, spanning from -232bp in promoter to +521bp in intron 1 (Figure 4), with 59.4% of GC content and 69 CpGs (ratio observed CpGs/expected CpGs for the segment = 1.06). The alignment of orthologous sequences from chimpanzee, mouse, rat, dog and rabbit showed high conservation for this region. Large CpG islands are usually associated to CpG methylation and promoter silencing, so we considered it important to analyze the methylation status of *TCOF1*'s promoter island.

We analyzed a total of 445 dinucleotide CpGs within a region of 946bp, including the *TCOF1* CpG island and its flanking regions (Figure 4). The analysis of eight patients and eight controls revealed that less than 1% of cytosines were methylated in either groups (Table 2), with methylated bases equally distributed among individuals, and with no preferentially methylated positions (*data not shown*).

We next decided to investigate if *TCOF1* CpG island methylation occurs in different tissues, and we sequenced bisulfite treated genomic DNA from five immortalized cell lines: SH-SY5y, HeLA, HepG2, HEK293, and Sa-OS2. We included in our analysis genomic DNA from stem cells obtained from dental pulp of a normal individual, in order to verify if pluripotent cells present a different epigenetic signature than immortalized ones. Bisulfite genomic sequencing of the same

region studied in TCS patients and controls revealed a similar number of methylated cytosines, with less than 0,22% of the total CpGs methylated in all analyzed cell lines (Table 2).

DISCUSSION

We present here for the first time the quantification of *TCOF1* transcripts in lymphocytes from TCS patients and controls and also in different cell lines. We observed a wide range of expression levels and the variance was very similar independently of whether the tissue harbored a non-functional allele due to a *TCOF1* pathogenic mutation. The expression levels did show dependence on age in TCS patients.

We detected a significant difference in *TCOF1* transcript levels between patients and controls. Our results demonstrate that the dosage compensation mechanism, proposed by Isaac *et al.* (2000), does not occur in adult and juvenile lymphocytes, at least at the transcriptional regulation level. We cannot rule out the existence of post-translational regulatory mechanisms that assure the same amounts of treacle, but, before investigating this mechanism, we first suggest replicating the finding of Isaac *et al.* (2000) in a larger sample, as only four normal and four affected individuals were tested by these authors. In addition, the reduction in *TCOF1* expression in patients as compared to controls was of 18%, which is in accordance to the observed 5-25% reduction of the normal transcripts levels caused by NMD in the presence of premature stop codon mutations in mRNA molecules (Kuzmiak and Maquat, 2006). It is also possible that the methods adopted by Isaac *et al.* (2000) for treacle quantification, southern blot and immunofluorescence intensity measurement, were not able to detect an 18% reduction in protein levels.

Our finding of lower levels of *TCOF1* expression in patients is in accordance with the current hypothesis of haploinsufficiency for TCS, which predicts that the allele bearing a premature stop codon is degraded by nonsense mediated mRNA decay (NMD), and consequently patients have less gene product than normal individuals. Further, lower *TCOF1* levels in mature cells of TCS patients imply that treacle deficiency may lead to other clinical findings in adulthood not necessarily associated to craniofacial development, and therefore it will be important to better investigate other possible associate medical conditions with TCS, as increased susceptibility to some cancer types.

We observed that neuroblastoma and fetal osteoblasts derived cell lines had the highest expression levels among the different tissues tested, with values comparable to the median level for lymphocytes (Figure 2). Considering that each cell line was derived from different individuals with normal phenotype, we cannot yet know if this variation is due to inter-individual or to tissue-specific *TCOF1* expression levels. Anyhow, our results suggest that lower levels of *TCOF1* in the absence of a pathogenic mutation do not necessarily imply in disease. In this regard, it would be important to revisit the status of the macaque with a Treacher-Collins syndrome like phenotype in which a significant reduction in *Tcof1* transcript levels was observed but no pathogenic mutation within the coding region of *Tcof1* was detected (Shows *et al.*, 2006).

In addition, we observed that *TCOF1* expression levels were not correlated with presence or absence of nine TCS clinical signs, thus suggesting that the transcript levels in adult and juvenile lymphocytes are not associated to disease severity. Although we cannot guarantee that our findings are representative to the transcription profile in embryos, if *TCOF1* expression levels are genetically controlled, we can speculate that individuals with higher *TCOF1* levels in adulthood would have relatively higher levels during embryogenesis as compared to those individuals with lower levels. If this assumption is correct, our results suggest that expression levels of *TCOF1* are not the main factor to explain TCS clinical variability. Interestingly, Cui e Dimario (2007) evaluated the phenotypic variation of flies knocked-down for two isoforms of Nopp140, the closest homologue to treacle in *Drosophila*, and observed a correlation between transcript levels and severity of malformations. These authors observed that a 30% reduction in transcripts was able to produce malformations, but individuals reached adulthood. More than 50% of normal transcript levels resulted in fly lethality. Considering that we did not find similar correlation between transcript levels and phenotype, we suggest that these *Minute*-like flies are not a good model to understand human expressivity in TCS.

Bisulfite sequencing of *TCOF1* CpG island in patients and controls revealed very few methylated cytosines, with an average of 1 methylated CpG for each 160-200 CpGs analyzed (Table 2), representing less than 0,6% of total CpGs. Although it is known that CpG islands are associated to approximately 60% of human genes and the great majority are unmethylated in all stages of development and different tissue types (review in Bird, 2002), we cannot exclude the possibility of other as yet unknown methylated *TCOF1* regulatory regions. It is also important to mention that gene silencing is usually associated to a threshold of methylated CpG sites. Costa *et*

al. (2004) observed the ADAM23 gene (*A Desintegrin And Metalloprotease Domain 23*) silencing with 40-60% of CpG sites modified. Further, it was shown that genes with unmethylated CpG islands can have their expression modified by methylated 'neighborhood', by a mechanism called long-range epigenetic gene silencing (Frigola *et al.*, 2006).

It is noteworthy that *TCOF1* was considered a candidate transcript for paternal imprinting by Nikaido *et al.* (2003). These authors used large-scale cDNA microarray to detect differential expression dependent upon chromosomal parent of origin of E9.5 mice embryos, and created a predicted human imprinted transcript map based on syntenic regions (<http://fantom2.gsc.riken.go.jp/imprinting/>). Interestingly, three authors (Szczak, 1953; Rogers, 1964 and 1976; Rovin *et al.*, 1964; review in Gollop, 1981) mentioned that TCS affected women transmitted the condition to descendents more frequently than affected men. Although coherent with paternal imprinting, these observations were discontinued in literature, and parental origin of pathogenic mutation does not seem to be associated to TCS severity (review in Gollop, 1981; Teber *et al.*, 2004; Dixon, 2003; Masotti *et al.*, 2007).

Absence of methylation of the *TCOF1*'s CpG island was also observed in five tumoral cell lines (HEK 293, HepG2, HeLa, SH-SY5y and Sa-OS2) and in cells derived from dental pulp stem cells, similarly to our findings with lymphocytes. Although the undifferentiated cells used in our assays are the best match to the conditions of the embryonic cellular environment, we cannot exclude the possibility that *TCOF1* methylation occurs during the first and second pharyngeal arch development, as germ cells and early embryos methylation patterns are reprogrammed in the whole genome (Reik *et al.*, 2001; Bibikova *et al.*, 2006).

In conclusion, our results also suggest that the pathogenic mutations in *TCOF1* associated to TCS are associated to NMD and that even mature cells have decreased levels of *TCOF1*. It seems that *TCOF1* expression level varies according to different cell type and that it is not a major contribution to phenotypic variability. Finally, the *TCOF1* expression variability is not controlled by methylation of the CpG island at least in non-fetal cells.

REFERENCES

1. Bibikova M, Chudin E, Wu B, Zhou L, Garcia EW, Liu Y, Shin S, Plaia TW, Auerbach JM, Arking DE, Gonzalez R, Crook J, Davidson B, Schulz TC, Robins A, Khanna A, Sartipy P, Hyllner J, Vanguri P, Savant-Bhonsale S, Smith AK, Chakravarti A, Maitra A, Rao M, Barker DL, Loring JF, Fan JB. (2006) Human embryonic stem cells have a unique epigenetic signature. *Genome Res.* 16(9):1075-83. Epub 2006 Aug 9.
2. Bueno, DF. (2007). Uso de Células Tronco Adultas para Estudo da Etiopatogenia das Fissuras Lábio Palatinas e bioengenharia de Tecidos. Instituto de Biociências, Universidade de São Paulo, SP, Brasil.
3. Costa, FF *et al.* (2004) *Epigenetic silencing of the adhesion molecule ADAM23 is highly frequent in breast tumors.* *Oncogene*, Vol 23 (7):1481-8.
4. Cui, Z & Dimario, PJ. (2007) RNAi Knockdown of Nopp140 Induces Minute-like Phenotypes in *Drosophila*. *Mol Biol Cell*. Mar 28; [Epub ahead of print]
5. Dixon, J and Dixon MJ. (2004) Genetic background has a major effect on the penetrance and severity of craniofacial defects in mice heterozygous for the gene encoding the nucleolar protein Treacle. 229(4):907-14
6. Dixon, J.; Hovanes, K.; Shiang, R.; Dixon, M.J. (1997a) Sequence analysis, identification of evolutionary conserved motifs and expression analysis of murine *Tcof1* provide further evidence for a potential function for the gene and its human homologue, *TCOF1*. *Hum Mol Genet* 6: 727-737
7. Dixon, J.; Brakebusch, C.; Fassler, R.; Dixon, M. (2000) Increased levels of apoptosis in the prefusion neural folds underlie the craniofacial disorder, Treacher Collins. *Hum Mol Gen* 9:1473-1480
8. Dixon, M.J.; Marres, H.A.M.; Edwards, S.J.; Dixon, J; Cremers, C.W.R.J. (1994) Treacher Collins syndrome: correlation between clinical and genetic linkage studies. *Clin Dysmorphol*, 3: 96-103.
9. Edwards, S.J.; Gladwin, A.J.; Dixon, M.J. (1997) The Mutational Spectrum in Treacher Collins Syndrome Reveals a Predominance of Mutations That Create a Premature-Termination Codon. *Am J Hum Genet* 60:515-524

10. Frigola, J; Song, J; Stirzaker, C; Hinshelwood, R.A.; Peinado, M.A.; Clarck, S.J. (2006) Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat Genet* 38, 540-549.
11. Gladwin, A.J.; Dixon, J.; Loftus, S.K.; Edwards, S.; Wasmuth, J.J.; Hennekam, R.C.M.; Dixon, M.J. (1996) Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene. *Hum Mol Gen* 5:1533-1538
12. Goldenberg, D; Harden, S.; Masayeva, B.G.; Benoit, N.; Westra, W.H.; Koch, W.M.; Sidransky, D.; Califano, J.A. (2004) Intraoperative molecular margin analysis in head and neck cancer. *Arch Otolaryngol Head Neck Surg.* 130(1), 39-44.
13. Gollop, TR. (1981) Estudo Genético-clínico das Disostoses Mandibulo-facial e Frontofacionas. PhD dissertation. Instituto de Biociências, Universidade de São Paulo, SP, Brazil.
14. Gonzales B, Henning D, So RB, Dixon J, Dixon MJ, Valdez BC. (2005a) The Treacher Collins syndrome (TCOF1) gene product is involved in pre-rRNA methylation. *Hum Mol Gene* Jul 15;14(14):2035-43.
15. Gorlin, R.J.; Cohen, M.M.; Hennekam, R.C. (2001) Branchial Arch and oral-Acral Disorders. In: *Syndromes of the head and neck*, 4th edition. Oxford: Oxford University Press. p.790-849.
16. Isaac, C.; Marsh, K.L.; Panzekas, W.A.; Dixon, J.; Dixon, M.J.; Jabs, E.W.; Meier, U.T. (2000) Characterization of the nucleolar gene product, treacle, in Treacher Collins syndrome. *Mol Biol Cell* 11: 188-192.
17. Kuzmiak, HA and Maquat LE. (2006) Applying nonsense-mediated mRNA decay research to the clinic: progress and challenges. *TRENDS Mol Med.* 12(7): 306-316.
18. Miller, S.A.; Dykes, D.D.; Polesky, H.F. (1988) A simple testing out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
19. Online Mendelian Inheritance in Man (OMIM) Johns Hopkins University, Baltimore, M.D.
<http://www.ncbi.nlm.nih.gov/omim>
20. Pfaffle, M.W. (2001). *A new mathematical model for quantification in real-time-RT-PCR*. *Nucleic Acids Research*, Vol 29 (9): 2002-7.

21. Reik, W; Dean, W; Walter, J. (2001) Epigenetic reprogramming in mammalian development. *Science*, 293: 1089-1093.
22. Rogers, BO. (1964) Berry-Treacher-Collins syndrome: a review of 200 cases. *Br J Plast Surg*, 17:109.
23. Rogers, BO. (1976) The surgical treatment of mandibulofacial dysostosis. *Clin Plast Surg*, 3:653.
24. Rovin, S; Dachi, SF, Borestein, DB, Cotter, WR. (1964) Mandibulofacial dysostosis, a familial study of five generations. *J Pediatr*, 64:215.
25. Teber OA, Gillessen-Kaesbach G, Fischer S, Bohringer S, Albrecht B, Albert A, Arslan-Kirchner M, Haan E, Hagedorn-Greiwe M, Hammans C, Henn W, Hinkel GK, Konig R, Kunstmann E, Kunze J, Neumann LM, Prott EC, Rauch A, Rott HD, Seidel H, Spranger S, Sprengel M, Zoll B, Lohmann DR, Wieczorek D. (2004) Genotyping in 46 patients with tentative diagnosis of Treacher Collins syndrome revealed unexpected phenotypic variation. *Eur J Hum Genet*. 12(11):879-90.
26. Shows KH, Ward C, Summers L, Li L, Ziegler GR, Hendrickx AG, Shiang R. (2006) Reduced TCOF1 mRNA level in a rhesus macaque with Treacher Collins-like syndrome: further evidence for haploinsufficiency of treacle as the cause of disease. *Mamma Genome* Feb;17(2):168-77. Epub 2006 Feb 7.
27. Splendore, A.; Silva, E.O.; Alonso, L.G.; Richieri-Costa, A.; Alonso, N.; Rosa, A.; Carakushansky G.; Cavalcanti, D.P.; Brunoni, D.; Passos-Bueno, M.R. (2000) High mutation detection rate in TCOF1 among Treacher Collins syndrome patients reveals clustering of mutations in 16 novel pathogenic changes. *Hum Mutat* 16:315-322
28. Splendore, A.; Fanganiello, R.D.; Masotti, C.; Morganti, L.S.C.; Passos-Bueno, M.R. (2005) TCOF1 mutation database: novel mutation in the alternatively spliced exon 6A and update in mutation nomenclature. *Hum Mut*, 25: 429-434.
29. Szlazak, J. (1953) Treacher-Collins syndrome. *Canad Med Assoc J*, 69:274.
30. The Treacher Collins Collaborative Group (1996) Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat Gen* 12:130-136

31. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. (2002)
Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple
internal control genes. *Genome Biol.* Jun 18;3(7)

Figure 1: Normalized *TCOF1* expression levels in TCS patients (circles) and controls (squares). The wider horizontal lines are the mean expression value and the two outer horizontal lines are standard error of the mean. Patients (MEAN $\pm$ SEM = 6.846 ± 0.3278) and controls (MEAN $\pm$ SEM = 8.093 ± 0.3587) have significant different *TCOF1* expression means ($p=0,0164$; Two-tailed unpaired t test with Welch's correction).

Figure 2: Normalized *TCOF1* expression levels in six different cell lines: HUVEC, HEK293, SH-SY5y, hFOB, HepG2, HeLA, and SaOS-2. **SUGESTÃO:** As a references, we plotted in this graphic the median value (Table 1; Control D) of *TCOF1* expression in lymphocytes (grey circle). The horizontal line divides expression values above and under the median value.

Figure 3: Frequency of the presence of nine clinical signs and normalized *TCOF1* expression level in 18 TCS patients. Median expression value (Table 1; Patient 10) is marked by a dashed vertical line. Four TCS patients bearing a pathogenic mutation are shown below the graph, and their expression values are indicated by labels (A, B, C, and D). Patient A is B's daughter.. Individuals with similar expression values have different clinical scores and individuals with different expression values have similar clinical scores.

Figure 4: Predicted CpG island of *TCOF1* (-232 to +521 from translation initiation site). The analyzed region by bisulfite sequencing spans from promoter to intron 1 and was divided in two amplicons, CpG1 (743pb) and CpG2 (255pb). A total of 445 CpGs dinucleotides (represented by circles) were analyzed.

Table 1: Expression data from relative quantification analysis of *TCOF1* in TCS patients, controls and cell lines.

Sample	$\Delta\Delta CT$	NF	Normalized expression	Mean	Standard Deviation	Clinical Data %
Patients				6,85	1,46	P/(P+A)
1	10,43	1,52	6,84			78
2	11,35	1,23	9,15			78
3	8,72	1,39	6,24			67
4	8,51	1,56	5,42			75
5	5,19	1,18	4,39			No clinical data
6	9,02	1,24	7,23			75
7	9,78	1,54	6,33			11
8	9,84	1,17	8,39			75
9	7,23	0,82	8,75			86
10	9,04	1,40	6,43			89
11	8,94	1,39	6,39			0
12	9,46	1,39	6,78			67
13	8,41	1,46	5,73			89
14	10,62	1,01	10,48			78
15	10,15	1,65	6,12			100
16	9,73	1,17	8,29			83
17	8,51	1,39	6,12			56
18	6,87	1,16	5,89			57
19	3,84	0,67	5,72			No clinical data
20	4,62	0,75	6,11			86
Controls				8,09	1,24	-
A	7,61	0,89	8,47			-
B	4,38	0,42	10,21			-
C	1,88	0,32	5,88			-
D	7,75	0,92	8,36			-
E	5,56	0,63	8,80			-
F	6,03	0,71	8,43			-
G	6,68	0,71	9,34			-
H	10,08	1,33	7,53			-
I	7,21	0,79	9,02			-
J	1,30	0,19	6,52			-
K	1,44	0,20	7,02			-
L	1,48	0,19	7,47			-
Median expression (patients and controls)				7,34	1,56	-
Cell lines						-
Huvec	3,19	0,81	3,91			-
HEK293	18,57	3,00	6,18			-
SH-SY5y	24,23	2,93	8,24			-
hFOB	25,20	2,85	8,83			-
HepG2	14,80	2,91	5,07			-
HeLA	33,08	4,54	7,27			-
SaOS-2	6,12	1,05	5,78			-

Table 2: Methylation profile of TCOF1 CpG islands from control, patients, and cell lines.

<i>Individuals</i>	<i>CpG1 clones</i>	<i>CpG2 clones</i>	<i>Total of CpGs analyzed</i>	<i>Methylated cytosines</i>	<i>% Methylation</i>
Controls (n=8)	32	35	2.893	14	0,5
Patients (n=8)	39	40	3.486	22	0,6
<i>Tissues/Cell lineages</i>	<i>CpG1 clones</i>	<i>CpG2 clones</i>	<i>CpGs analyzed</i>	<i>Methylated cytosines</i>	<i>% Methylation</i>
SH-SY5y	5	5	445	0	0
HeLA	5	5	445	0	0
Dental pulp stem cells	5	5	445	1	0,22
HepG2	5	5	445	1	0,22
HEK293	5	5	445	1	0,22
Sa-OS2	5	5	445	0	0

% Methylation is given by the ratio of the number of methylated cytosines and total CpGs analyzed.

AURICULO-CONDYLAR SYNDROME: MAPPING OF A FIRST LOCUS AND GENETIC HETEROGENEITY

Cibele Masotti¹, Karina G. Oliveira¹, Alessandra Splendore², Fabiana Poerner³, Josiane Souza³, Renato Freitas³, Roseli Zechi-Ceide⁴, Maria Leine Guion-Almeida⁴ e Maria Rita Passos-Bueno¹

¹Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo, Brazil

²Department of Genetics, Stanford University, California, USA

³Integral Attending Center to the Cleft Lip and Palate Affected (CAIF), Curitiba, Parana, Brazil

⁴Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, Bauru, São Paulo, Brazil

RESUMO

A síndrome aurículo-condilar (SAC) é uma patologia autossômica dominante de primeiro e segundo arcos faríngeos. Pacientes geralmente apresentam orelhas proeminentes e mal-formadas ('orelhas em ponto de interrogação'), microstomia, anomalias na articulação temporomandibular e hipoplasia do côndilo mandibular condyle. Apesar da penetrância ser aparentemente completa, há alta variação fenotípica intra e inter-familiar. No presente trabalho descrevemos um novo caso familiar, com 11 indivíduos afetados em quatro gerações. Nós apresentamos assimetria facial nessa família, um achado clínico atípico. Uma vez que a síndrome de Treacher Collins, Espectro Oculo-aurículo-vertebral e a síndrome de Townes-Brocks são consideradas diagnósticos diferenciais em relação à SAC, usando a análise de segregação nós excluimos as regiões cromossômicas envolvidas com essas doenças como sendo responsáveis pelo fenótipo SAC nessa família recém-descrita, assim como naquela descrita por Guion-Aleida *et al.*(2002). A exclusão dessas quatro regiões nos levou a realizar uma busca genômica. Mapeamos o primeiro locus associado à SAC (1p21.1-q23.3) e também detectamos heterogeneidade genética.

AURICULO-CONDYLAR SYNDROME: MAPPING OF A FIRST LOCUS AND GENETIC HETEROGENEITY

Cibele Masotti¹, Karina G. Oliveira¹, Alessandra Splendore², Fabiana Poerner³, Josiane Souza³, Renato Freitas³, Roseli Zechi-Ceide⁴, Maria Leine Guion-Almeida⁴ e Maria Rita Passos-Bueno¹

¹Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo, Brazil

²Department of Genetics, Stanford University, California, USA

³Integral Attending Center to the Cleft Lip and Palate Affected (CAIF), Curitiba, Parana, Brazil

⁴Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, Bauru, São Paulo, Brazil

ABSTRACT

Auriculo-condylar syndrome (ACS) is an autosomal dominant disorder of first and second embryonic pharyngeal arches. Patients usually present prominent and malformed ears ('question mark ears'), microstomia, abnormal temporomandibular joint, and mandibular condyle hypoplasia. Although the penetrance seems to be complete, there is high inter and intra-familial phenotypic variation. We herein describe a new familial case, with 11 affected individuals in four generations. We report facial asymmetry in this family, an atypical clinical finding. As Treacher Collins syndrome, Oculo-auriculo-vertebral spectrum and Townes-Brocks syndrome are considered for differential diagnosis in relation to ACS, through segregation analysis, we excluded the chromosomal regions involved in these disorders as responsible for ACS phenotype in this newly described family, as well as the one reported by Guion-Almeida *et al.* (2002). The exclusion of these four regions led us to perform a wide genome search. We mapped the first locus associated to ACS (1p21.1-q23.3) and also detected genetic heterogeneity.

INTRODUCTION

Auriculo-condylar syndrome (ACS, OMIM 602483), an autosomal dominant disorder, was first described by Jampol et al. (1998), who reported relatives from five generations of a family with prominent and malformed ears, microstomia, abnormal temporomandibular joint, and mandibular condyle hypoplasia. This syndrome is also known as question mark ear due to peculiar shape of the ear, which characteristically presents marked constriction at the junction between the lower and middle thirds of the pinna, separating the lobule from the helix and giving the ear an appearance of a 'question mark' (reviewed in Guion-Almeida et al., 2002).

There are only 5 ACS familial cases reported in the literature. Although the penetrance seems to be complete in these cases, there is high inter and intra-familial phenotypic variation (Jampol et al., 1998; Erlich et al., 2000; Guion-Almeida et al., 2002; Storm et al., 2005). The phenotype variation is not only defined by the severity of the malformations, but also by the presence/absence of less frequent clinical findings such as cleft palate, hypoplastic tongue, hypotonia, and developmental delay (Erlich et al., 2000; Priolo et al., 2000; Storm et al. 2005). Based on the major clinical features, ACS can be classified as a disease of the first and second pharyngeal arches.

Treacher Collins syndrome (TCS; OMIM 154500), oculo-auriculo-vertebral spectrum (OAV; OMIM 164210) and Townes-Brocks syndrome (OMIM 107480), typical disorders of first and second pharyngeal arches, have several overlapping clinical signs with ACS and they should be considered for differential diagnosis. Mapping and identification of the ACS gene will certainly bring important contributions to understand the development of embryonic structures derived from these pharyngeal arches, as well as to perform differential diagnosis of the above mentioned syndromes. As currently there is no gene or chromosomal region associated to ACS, we first proposed to investigate if ACS is linked to the chromosomal regions 5q31-32, associated to TCS gene (*TCOF1*), 14q32, candidate region to hemifacial microsomia (Kelberman et al., 2001), 16q12.1 and 14q11, associated to *SALL1* and *SALL2* genes, known to be involved in Townes-Brocks syndrome, through segregation analysis in two large ACS Brazilian families. The exclusion of these four regions led us to perform a wide genome search. One of the families had already been reported (Guion-Almeida et al., 2002), while the other one is herein described for the first time. This newly diagnosed family also allowed us to test ACS for genetic heterogeneity.

SUBJECT AND METHODS

Family data

Family 1 (F1489)

The family was seen by clinicians from the Integral Attending Center to the Cleft Lip and Palate Affected (CAIF), Curitiba, Paraná, Brazil. Clinical evaluation included general examination with particular concern to the face and mouth, and complementary evaluation included skull and facial X-rays. Audiological evaluation included puretone audiometry, speech audiometry, tympanometry, stapedius reflex measurement, and brainstem evoked response.

Family 2 (F1700)

This family was originally reported by Guion-Almeida et al. (2002). Since its description no additional affected child was born in the family.

DNA analysis

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 patients or from their legal tutors.

Linkage analysis

Segregation analysis was performed using highly informative dinucleotide repeats microsatellite markers, with an average inter-marker distance of 10cM across the genome, from ABI PRISM® Linkage Mapping Set version 2.0 (*Perkin-Elmer, Applied Biosystems*). The PCR products amplified by fluorescently-labeled primers were loaded onto the automated sequencer Megabace 1000 (*Amersham, GE Healthcare, Little Chalfont, United Kingdom*) according to manufacturer's protocol and analyzed with Megabace Genetic Profiler® software (*Amersham, GE Healthcare*). We analyzed four candidate regions (5q31-32, 14q32, 16q12.1 and 14q11) associated to craniofacial malformations in both families using 18 markers. The genome-wide scan was performed only in family 2 and 370 markers were analyzed.

Two-point lod scores were calculated using the MLINK program from FASTLINK package version 5.1 (<http://linkage.rockefeller.edu/soft/>). The disease was analyzed as an autosomal dominant trait, with a penetrance equal to 100% and disease gene frequency of 0.001. Recombination frequencies were assumed to be equal in males and females and allele frequencies were considered to be 1/N (N= number of different alleles observed on the pedigree). For multipoint lod score we used Merlin 1.0.1 software (Abecasis et al, 2002). We also screened 20 control individuals to calculate the allele frequencies for the markers D1S2726, D1S252, D1S498, and D1S484 which showed evidence of linkage with the disease. These allelic frequencies were used for final Lod score calculations, both two point as well as multipoint linkage analysis.

RESULTS

Clinical report of family 1

The family history revealed 11 affected individuals with 'question mark ears' in at least four successive generations and a highly variable phenotype (pedigree, Figure 1a and Figure 2). In addition to malformed ears, the major features of these patients were prominent cheeks, round facial appearance, micrognathia, microstomia, malocclusion, and hearing loss. After radiological examination in four individuals, temporomandibular joint abnormalities were identified, leading us to the diagnosis of Auriculo-condylar syndrome.

The penetrance seems to be complete, but there is clear variable expressivity. The family comprises more affected women in the genealogy (8 affected women to 3 affected men), but no sex segregation deviation was observed according to parental transmission ($p=0,33$; Fisher's Exact Test). The most severe cases were observed in the fourth generation (individuals IV-3 and IV-6), including one (IV-7) who died in early childhood.

The ear malformations are highly variable (Figure 2), as we can observe in individual IV-3, who has auricular clefts, cupped pinna, underdevelopment and overfolding of the superior helices, pre and post-auricular cartilaginous tags, and her mother (III-10), who has none of these features. Audiometric exams were performed in six individuals, and all of them had at least some degree of conductive hearing loss, including individuals II-8 e III-5 who did not have stenotic ear canals. Individual II-8 had conductive hearing loss of the right and neurosensorial of the left ear.

Most of the patients had malocclusion and limitations to open their mouths, but only individuals II-8, IV-6 and IV-7 had mastication and speech articulation difficulties. These three individuals also presented high arched palate, but only IV-7 had a cleft.

Besides the typical clinical signs of ACS in family 1 (summarized in table 1), we also identified facial asymmetry, that is clearly evident in patients II-4, III-10, III-13, III-14 and IV-3, but mildly in IV-6 and IV-7 (figure 2). Intelligence was normal and no developmental delay was observed in any of the affected individuals. Clinical data are summarized in Table 1.

Linkage study

Initially we tested family 1 and 2 for candidate regions involved in craniofacial anomalies (5q31-32, 14q32, 16q12.1 and 14q11), but all were excluded. We thereafter conducted a genome scan in family 2.

A genome search with 250 markers revealed evidence of linkage to D1S252 marker on chromosome 1 in family 2 (Figure 1b), with a two-point lod score of 3.0 ($Z_{\max} = 3.0$, at $\theta=0$). Analyzing the haplotypes, we observed recombination in individuals II-3 and II-9: one recombination between D1S206 and D1S2726, positioning the gene distal to D1S206, and another, also in II-9, between D1S484 and D1S2878 markers, which positioned the gene proximal to D1S2878. These results suggested that the disease locus lies between markers D1S206 and D1S2878. The multipoint lod score was positive for all the markers of these regions, that is from D1S206 to D1S2878, with the maximum score (3.01) also for marker D1S252. The region comprised by the markers D1S206, D1S2726, D1S252, D1S498, D1S484 and D1S2878 corresponds to 1p21.1-q23.3. This region has about 43cM of genetic distance and corresponds to approximately 60 Mbp of genomic sequence (Figure 3).

In family 1, the 1p21.1-q23.3 region was excluded because ACS affected individuals III-10 and IV-3 do not share the same haplotype as the other affected individuals of the genealogy. In addition, individual III-8 is normal and also has the supposed disease haplotype (Figure 1a). Negative two-point lod scores for all markers at this region were observed for this family (for D1S252 and D1S498: $Z = -\infty$ at $\theta=0$; -1.6 at $\theta=0.1$; -0.8 at $\theta=0.2$)

DISCUSSION

The great intra and inter-familial phenotypic variation in ACS has previously been noted and is discussed in several clinical reports (Jampol et al., 1998; Guion-Almeida et al., 1999 and 2002; Storm et al, 2005). We further confirm a wide spectrum of clinical variability in the new family (1) herein reported, in which the mildest cases presented only 'question mark ears', whereas the most severe ones present the full phenotype, as was the case for individual IV-7, who died in early childhood due to airway obstruction by severe malformations. Besides the malformed ears, the most frequent clinical findings in family 1 were round facial appearance (90%), puffy cheeks (80%), micrognathia (60%), microstomia (80%), malocclusion (88%), abnormal mandibular condyle (75%), and temporomandibular joint abnormality (100%).

Family 2, originally described by Guion-Almeida et al. (2002), shares almost all of the phenotypic characters described in family 1, the major exceptions being facial asymmetry, present only in family 1, and upslanted palpebral fissures, present in the proband from family 2. In addition to these few atypical findings, both families have the same facial 'gestalt' and also cleft palate and hearing loss segregating through the generations. The frequency of clinical signs from family 1 was also compared with the frequency of the signs observed in previous reports (Table 2; modified from Storm et al., 2005). Of the sixteen clinical signs listed as the major ones for ACS by these authors, only 7 have similar frequency in the family here described. Prominent cheeks, microstomia, hearing loss and feeding difficulties seem to be more represented in family 1, but abnormal palate, macrocephaly, ptosis, mild developmental delay, microglossia, and hypotonia are in lower frequency or were not even diagnosed in family 1.

Facial asymmetry is an unusual clinical signal for ACS. Surprisingly, however, it was observed in 60% of the affected individuals from family 1. This is an important finding because facial asymmetry is usually considered for differential diagnosis of first and second pharyngeal arch disorders, such as OAV spectrum, Treacher Collins, Nager (OMIM 154400), and Miller syndromes (OMIM 263750). By studying animal models of hemifacial microsomia, Poswillo (1973) proposed that the pathogenesis of invariably asymmetrical malformations is different from that of symmetrical malformations. This author suggested that hemifacial microsomia could be explained by embryonic hemorrhage and Treacher Collins syndrome (TCS) by a mechanism acting uniformly over a precise temporal period in the developing embryo, because, even inducing hemorrhage in both

sides of mice embryo, symmetrical disposition of the craniofacial defects was never observed, as is the case in TCS patients. No definitive causal agent for hemifacial microsomia has been discovered so far (Vendramini et al, 2007), but the gene for TCS was identified (The Treacher Collins Collaborative Group, 1996) and it is known that 98% of TCS patients bearing the pathogenic mutation have bilateral and usually symmetrical malformations (Splendore et al., 2005).

We excluded four regions involved in craniofacial disorders of the first and second pharyngeal arches. This result implies that ACS is not allelic to TCS or any of the known loci associated to OAV spectrum and Townes-Brock syndrome. We identified linkage to the disease locus in family 2 for the region 1p21.1-q23.3, which represents the first chromosomal region associated to the ACS phenotype. Given the size of the candidate region (60Mbp) and the number of genes between 1p21.1-q23.3 (around 250), it is very important to detect additional families to narrow down the mapped region.

Interestingly, this candidate region was excluded in family 1. We observed that patients III-10 and her affected daughter did not share any of the alleles at the putative at-risk chromosomal 1 region. Although we cannot rule out that these two patients can have another disease, we observed that they share many of the similar clinical features of the other affected relatives, including facial asymmetry. Therefore, it is very likely that this genealogy represents another condition, representing genetic heterogeneity for the ACS phenotype.

The clinical data together with the linkage data here reported suggest that ACS is a very heterogeneous clinical entity, which includes many different clinical signs besides 'question mark ears', such as hypotonia, ptosis, microglossia, developmental delay, and facial asymmetry. Like OAV spectrum, ACS could be classified as a 'question mark ears' spectrum, in which we should include ACS variations and dysgnathia complex. We also observed that the phenotype in both families became more severe through the generations; although this can be due to an ascertainment bias, it also suggests clinical anticipation. It will thus be important to evaluate this hypothesis in additional families. The identification of this disease gene will certainly contribute to our knowledge in the development of the first and second pharyngeal arch structures as well as to the understanding of OAV conditions, as some of the isolated OAV cases might represent subtypes of ACS.

REFERENCES

1. Erlich, M.S.; Cunningham, M.L.; Hudgins, L. (2000) Transmission of the Dysgnathia Complex From Mother to Daughter. *Am J Med Genet* 95:269-274
2. Guion-Almeida, M.L.; Kokitsu-Nakata, N.M.; Zechi-Ceide, R.M.; Vendramini, S. (1999) Auriculo-Condilar syndrome: Further Evidence for a New Disorder. *Am J Med Genet* 86: 130-133
3. Guion-Almeida, M.L.; Zechi-Ceide, R.M.; Vendramini, S.; Kokitsu-Nakata, N.M. (2002) Auriculo-Condilar syndrome: additional patients. *Am J Med Genet* 112:209-214
4. Jampol, M. *et al.* (1998) Prominent, Constricted Ears With Malformed Condyle of the Mandible. *Am J Med Genet* 75:499-452
5. Miller, S.A.; Dykes, D.D.; Polesky, H.F. (1988) A simple testing out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
6. Online Mendelian Inheritance in Man (OMIM) Johns Hopkins University, Baltimore, M.D. 602483 [2007] <http://www.ncbi.nlm.nih.gov/omim>
7. Poswillo, D. (1973) The pathogenesis of the first and second branchial arch syndrome. *Oral Surg Oral Med Oral Pathol.* Mar;35(3):302-28.
8. Priolo, M. *et al.* (2000) Question mark eras, temporo-mandibula joint malformation and hipotonia: auriculo-condylar síndrome or a distinct entity? *Clin Dysmorph* 9: 277-280
9. Richards, B.; Skoletsky, J.; Shuber, A.P.; Balfour, R.; Stern, R.C.; Dorkin, H.L.; Parad, R.B.; Witt, D.; Klinger, K.W. (1993) Multiplex PCR amplification from the *CFTR* gene using DNA prepared from buccal brushes/swabs. *Hum Mol Genet* 2:159-163
10. Splendore, A; Fanganiello, RD; Masotti, C; Morganti,LSC; Passos-Bueno, MR. (2005) *TCOF1* Mutation Database: Novel Mutation in the Alternatively Spliced Exon 6A and Update in Mutation Nomenclature. *Hum Mut* 25:429-434.
11. Storm AL, Johnson JM, Lammer E, Green GE, Cuniff C. (2005) Auriculo-condylar syndrome is associated with highly variable ear and mandibular defects in multiple kindreds. *Am J Med Genet* Oct 1;138(2):141-5
12. Vendramini, S; Richieri-Costa, A; Guion-Almeida, ML. (2007) Oculoauriculovertebral spectrum with radial defects: a new syndrome or an extension of the oculoauriculovertebral spectrum? Report of fourteen Brazilian cases and review of the literature. *Eur J Hum Genet* 15, 411–421

Table 1: Clinical signs of ACS patients from family 1 (P: present; A: absent; N: not referred; Frequency: Present/(present+absent))

Clinical Signs (OMIM)	Patients											
	II-4	II-6	II-8	III-5	III-10	III-13	III-14	IV-3	IV-6	IV-7	P/(P+A)	Freq
Macrocephaly	N	N	A	A	A	A	A	A	A	A	0	0
Micrognathia	A	P	P	A	P	A	A	P	P	P	6/10	60%
Round Facial appearance	P	P	P	P	P	P	A	P	P	P	9/10	90%
Prominent cheeks	P	A	P	P	P	P	A	P	P	P	8/10	80%
Malformed ears	P	P	P	P	P	P	P	P	P	P	10/10	100%
Auricular clefts	A	A	P	P	A	A	P	P	P	P	6/10	60%
Cleft at the junction of the lobule and helix	P	A	P	P	A	P	P	P	P	P	8/10	80%
Lobule separate from the rest of external ear	A	A	P	A	A	A	A	P	A	A	2/10	20%
Cleft at the superior portion of the pinna	A	A	A	A	A	A	A	P	A	P	2/10	20%
Cupped pinna	A	A	N	A	A	A	A	P	A	A	1/9	11%
Overfolding of the superior helices	P	A	P	P	A	P	P	P	P	P	8/10	80%
Underdeveloped superior helices	A	P	P	P	A	P	P	P	P	A	7/10	70%
Pre and post-auricular skin or cartilaginous tags	A	A	A	A	A	A	A	P	A	P	2/10	20%
Low-set ears	A	A	A	A	A	A	A	A	A	P	1/10	10%
Posteriorly rotated ears	P	A	P	A	A	A	A	A	A	A	2/10	20%
Stenotic ears canals	N	N	A	A	N	A	A	P	N	P	2/6	33%
Ear constriction	A	N	A	A	N	A	A	A	N	N	0	0
Hearing loss	N	N	P	P	P	N	N	P	P	P	6/6	100%
Microstomia	P	P	P	P	P	A	A	P	P	P	8/10	80%
Glossoptosis	A	N	P	A	A	A	A	P	P	P	4/9	44%
Abnormal palate	A	A	P	A	A	A	A	A	P	P	3/10	30%
Cleft palate	A	A	A	A	A	A	A	A	A	P	1/10	10%
Mastication difficulties	A	A	P	A	A	A	A	A	P	P	3/10	30%
Speech articulation difficulties	A	A	P	A	A	A	A	A	A	P	2/10	20%
Crowded teeth	A	N	P	A	A	P	N	P	P	N	4/7	57%
Malocclusion	P	P	N	P	A	P	P	P	P	N	7/8	88%
Open anterior bite	P	P	N	P	A	A	P	P	P	N	6/8	75%
Posterior crossbite	P	P	N	A	P	P	P	P	P	N	7/8	88%
Respiratory difficulties de to orofacial malformations	A	A	P	A	A	A	A	P	P	P	4/10	40%
Facial asymmetry	P	A	A	A	A	P	P	P	P	P	6/10	60%
Apnea	A	A	P	A	A	A	A	P	P	P	4/10	40%
Temporomandibular joint abnormality	P	N	P	N	N	N	N	P	P	N	4/4	100%
Abnormalities of the mandibular condyle	A	N	P	N	N	N	N	P	P	N	3/4	75%

Table 2: Comparison of clinical signs from patients with ACS from previous reports and current study.

Clinical signs	Previous reports**		Guion-Almeida, 2002		Current report	
	Cases	Frequency	Cases	Frequency	Cases	Frequency
Temporomandibular joint/condyle abnormality	21/21	100%	1/1	100%	4/4	100%
Ear constriction (malformed?)	30/31	96.8%	6/6	100%	10/10	100%
Micrognathia	22/31	71.0%	5/5	100%	6/10	60%
Abnormal Palate	10/16	62.5%	1/3	33%	3/10	30%
Prominent cheeks	16/28	57.1%	5/5	100%	8/10	80%
Microstomia	14/27	51.9%	4/5	80%	8/10	80%
Glossoptosis	5/11	45.5%	2/2	100%	4/9	44%
Respiratory distress	8/22	36.4%	2/2	100%	4/10	40%
Stenotic ear canals	3/10	30.0%	1/1	100%	2/6	33%
Ptosis	3/11	27.3%	0	0	0	0
Macrocephaly	3/12	25.0%	0	0	0	0
Mild developmental delay	3/13	23.1%	0	0	0	0
Feeding difficulties	4/19	21.1%	1/1	100%	3/10	30%
Hearing loss	4/19	21.1%	2/2	100%	6/6	100%
Hypotonia	2/11	18.2%	0	0	0	0
Microglossia	2/11	18.2%	0	0	0	0

**Articles used for the previously reported column: Usupa [1978], Jampol et al. [1998], Guion-Almeida et al. [1999], Erlich et al. [2000], Priolo et al. [2000], Guion-Almeida et al. [2002] and Storm et al. [2005].

Figure 1: Families studied; **A** Family 1; **B** Family 2, previously reported by Guion-Almeida *et al.* (2002).

Figure 2: ACS affected patients from family 1. The numbers correspond to family 1's genealogy number.

Figure 3: Mapped region 1p21.2-1q23.3 for ACS in family 2. The candidate region is signaled in chromosome 1 ideogram and is expanded below. Each small square represents a reference sequence from GenBank localized in this region (approximately 250 genes). The black line signals the region with lod score 3.0, that includes centromeric region.

DISCUSSÃO GERAL E CONCLUSÕES

Estudamos duas malformações craniofaciais mendelianas de herança autossômica dominante, decorrentes de um distúrbio do desenvolvimento dos primeiro e segundo arcos faríngeos: a síndrome de Treacher Collins (STC) e a síndrome Aurículo-condilar (SAC). As duas condições possuem expressividade variável e sobreposição fenotípica com outras síndromes, também consideradas de primeiro e segundo arcos faríngeos, como as de Nager, Miller e Goldenhar. A identificação de genes e de mecanismos moleculares associados a essas condições, além de contribuir para a compreensão do desenvolvimento das estruturas derivadas desses arcos faríngeos, é fundamental para o desenvolvimento do diagnóstico molecular, uma ferramenta importante para diagnóstico diferencial e aconselhamento genético dessas patologias.

Em relação à SAC, o presente trabalho contribuiu para uma melhor caracterização clínica da síndrome com a descrição de uma nova família com 11 afetados. Observamos que, além dos sinais clínicos característicos desta síndrome, os afetados desta genealogia apresentam assimetria facial, presente em 60% dos afetados. Após excluirmos quatro genes/regiões candidatas para síndromes de 1º e 2º arcos faríngeos, realizamos estudos de ligação usando marcadores ao longo do genoma em uma das famílias e mapeamos o primeiro locus associado à SAC, 1p21.1-q23.3 (*lod score*=3.0). Nossos dados sugerem que há heterogeneidade genética para SAC, uma vez que essa região foi excluída na família portadora de assimetria facial. No entanto, não podemos deixar de considerar que a família cuja região candidata foi excluída tenha outra doença segregando em sua genealogia. Dada a ampla variabilidade clínica dos afetados na SAC, sugerimos que a SAC seja classificada como um espectro de malformações associadas à malformação típica das orelhas em forma de ponto de interrogação, “question mark ears spectrum”.

Onze anos após a identificação do gene *TCOF1* (The Treacher Collins Collaborative Group, 1996) como responsável pela síndrome de Treacher Collins (STC) e o estabelecimento do

diagnóstico molecular, a impossibilidade de prever a gravidade do quadro clínico nos descendentes de afetados ainda é a maior limitação para o aconselhamento genético. Neste trabalho, demos continuidade à caracterização do espectro de mutações em pacientes, procurando identificar uma correlação genótipo-fenótipo, além de caracterizar mecanismos mutacionais e regulatórios envolvidos na patogênese da síndrome.

Com a descrição de dois novos exons no *TCOF1* (So *et al.*, 2004), foi necessário adequar a nomenclatura de mutações à nova e mais abundante isoforma do gene (AY460334). Realizamos um extenso trabalho de revisão da nomenclatura das mutações patogênicas descritas na literatura, criando um banco de dados disponibilizado na internet (www.genoma.ib.usp.br/TCOF1_database). A análise do conjunto de 118 mutações, descritas em mais de 165 pacientes, trouxe algumas conclusões importantes quanto ao espectro de mutações no gene. Primeiramente, ao contrário do proposto por Splendore *et al.* (2000, 2002a), os exons 10, 15, 16 e 23 não se revelaram pontos quentes de mutações, somente o exon 24, dada a sua natureza repetitiva. Segundo, a identificação de uma mutação patogênica no exon 6A, que sofre splicing alternativo, mostra que a isoforma que contém esse exon é fundamental para o desenvolvimento craniofacial, e, além disso, que pacientes sem mutação na região codificadora e com fenótipo típico podem ter mutações em exons ainda não preditos. Terceiro, a investigação de deleção ou duplicação de exons inteiros pela técnica de *Southern blotting* demonstrou a ausência desse mecanismo mutacional no *TCOF1*. Cabe ressaltar que os exons 6a e 16a foram incluídos na rotina de triagem de mutações para STC.

Analisamos 21 propósitos encaminhados para diagnóstico molecular e aconselhamento genético. Detectamos mutação patogênica em 11 deles e em seus familiares afetados. Nossa análise confirma a grande variabilidade clínica na síndrome, uma vez que dois pais clinicamente normais foram diagnosticados como portadores da mutação após o diagnóstico molecular. A baixa taxa de detecção de mutações patogênicas observada neste trabalho (52%) não está associada à sensibilidade do método, uma vez que a metodologia de triagem de mutações foi a mesma descrita por Splendore *et al.* (2000), que obteve 97% de eficiência. Na realidade, nossos achados confirmam dados anteriores, que mostraram que mutações no *TCOF1* geram um fenótipo muito específico (Splendore *et al.*, 2002b). Com exceção de dois casos, os propósitos sem mutação detectada possuíam achados fenotípicos atípicos, não relacionados à STC. Dado que a maioria desses pacientes foi encaminhada por médicos de outros centros, a solicitação elevada do teste

genético mostra que esse é considerado uma ferramenta de diagnóstico diferencial por alguns geneticistas clínicos, que, mesmo identificando alterações atípicas da STC de seus pacientes, optam por encaminhar as amostras de DNA para triagem de mutações no *TCOF1*.

Analizamos a correlação genótipo-fenótipo numa amostra de 58 indivíduos, sendo este o maior número amostral já estudado de pacientes da STC portadores de diagnóstico clínico e molecular. Nossa análise confirma achados anteriores de ausência de correlação entre a severidade da doença e sexo do afetado, ocorrência, posição da mutação no gene e origem parental da mutação (Dixon, 1996; Gladwin *et al.*, 1996; Edwards *et al.*, 1997; Splendore *et al.*, 2000). Interessantemente, observamos uma diferença na frequência dos achados clínicos de casos isolados em relação aos familiares. Contudo, nossos achados demonstraram um viés de averiguação, já observado na STC (Marres *et al.*, 1995; Gladwin *et al.*, 1996; Edwards *et al.*, 1997; Splendore *et al.*, 2000).

Caracterizamos 17 haplótipos que envolvem nove polimorfismos da região codificadora do *TCOF1*. Pela primeira vez foi investigada a possibilidade de associação entre haplótipos da região codificadora, variabilidade clínica e predisposição à mutação patogênica. Nenhuma correlação haplótipo-fenótipo foi observada e nossos achados demonstram que os polimorfismos mais frequentes da região codificadora do *TCOF1*, estando em *trans* com a mutação patogênica (presentes no alelo normal), não influenciam o fenótipo dos afetados. Além disso, nosso trabalho também mostra que não há maior probabilidade de ocorrer mutação patogênica num haplótipo em relação a outro.

A análise dos dados de polimorfismos da região codificadora do gene permitiu que fizéssemos inferências sobre o regime de seleção ao qual o *TCOF1* está submetido. Para tanto, comparamos os números de polimorfismos sinônimos e não-sinônimos identificados nas diferentes populações humanas com os números obtidos com base na divergência entre humanos e chimpanzés. A identificação de uma baixa frequência dos polimorfismos não-sinônimos na população e o excesso desses em relação ao esperado com base na divergência de humanos e chimpanzés sugerem que o gene *TCOF1* está sob seleção purificadora, que atua sobre mutações fracamente deletérias. Essas mutações persistem transientemente na população como polimorfismos e raramente são fixadas, mas resultam em diferenças entre as duas espécies. É possível que os polimorfismos não-sinônimos do *TCOF1* tenham um efeito clinicamente

imperceptível sobre o fenótipo, entretanto, são capazes de afetar o valor adaptativo dos indivíduos. Dessa maneira, esses alelos tendem a desaparecer ao longo das futuras gerações.

A não-associação dos haplótipos da região codificadora com a variabilidade clínica sugere que a expressividade variável da síndrome esteja associada a outros mecanismos regulatórios, independentes da região codificadora, como a modulação da expressão do alelo selvagem por regulação ao nível da transcrição. Além disso, o modelo de seleção purificadora sugere que, embora as mutações não-sinônimas não tenham efeito evidente sobre o fenótipo da STC, os haplótipos podem predispor o indivíduo a doenças complexas, como o câncer.

Dessa maneira, investigamos a possibilidade de mutações na região promotora do alelo normal do *TCOF1* modularem o fenótipo da STC. Identificamos cinco polimorfismos de base única (SNPs), com frequências de 2 a 10% na população, numa região de -1200pb ao início da tradução. Identificamos uma variante funcional: mostramos que o SNP -346C>T foi capaz de reduzir os níveis de transcrição *in vitro* e de afetar a ligação do fator de transcrição YY1 à sua sequência consenso. No entanto, dado o pequeno número de pacientes portadores do SNP na nossa amostra, e a dificuldade em inferir a fase do SNP com a mutação patogênica, não pudemos estabelecer correlação genótipo-fenótipo.

A análise *in silico* da região promotora do *TCOF1* revelou algumas características importantes de sua arquitetura. Procuramos pelos elementos principais do *core promoter* (Carey & Smale, 2000) e observamos ausência da sequência iniciadora consenso TATAA-box, ausência do elemento iniciador, e também não encontramos o *downstream promoter element* (DPE). Identificamos mais de 15 sítios do fator de transcrição ubíquo Sp1 numa região rica em conteúdo GC (75%), com aproximadamente 63% de conservação entre cinco diferentes espécies. Segundo Smale (1994), essa arquitetura de *core promoter* está geralmente presente em genes do tipo *housekeeping*: Sp1 direciona a formação do complexo de pré-iniciação da transcrição para uma região 40 a 100pb abaixo (3') de seus sítios de ligação. Interessantemente, Azizkhan *et al.* (1993) observaram que a regulação da transcrição de muitos genes sem a sequência consenso TATAA-box (*TATA-less*) ocorre em resposta a estímulos de proliferação celular.

Identificamos oito sítios de YY1 em quatro regiões do promotor do *TCOF1* com alto grau de conservação entre as espécies, computacionalmente preditas como enhancers. No entanto, só foi possível analisar o papel de YY1 na regulação da transcrição exercida pela região de -1200 pb

ao início da tradução do *TCOF1*. Com o ensaio de co-transfecção, mostramos que *YY1* possui claramente um papel de repressor nessa região de maneira dose-dependente em hepatócitos humanos. Além disso, analisamos os dados de expressão de diferentes tecidos embrionários humanos (Cai *et al.*, 2005), incluindo primeiro e segundo arco faríngeo, e também observamos uma ação antagonista de *YY1* em relação ao *TCOF1* durante a quarta semana do desenvolvimento. Diante dessas observações e dos dados experimentais obtidos com transformação das regiões regulatórias do *TCOF1* em leveduras, elaboramos uma hipótese para o funcionamento do promotor: o promotor mínimo parece estar sempre ativo e os fatores de transcrição necessários para a sua ativação parecem ser pouco específicos. Dessa maneira, a regulação do *TCOF1* parece que ocorre por repressão do seu promotor mínimo.

Considerando o mecanismo molecular da STC como haploinsuficiência, consideramos importante investigar se os níveis de transcritos de *TCOF1* estariam associados à patologia. Pela primeira vez, os níveis de transcritos de *TCOF1* em linfócitos de pacientes e controles e em diferentes linhagens celulares foram quantificados. Observamos uma grande variação inter-individual da expressão gênica nas amostras estudadas, independente do tecido e da presença da mutação patogênica. Comparamos os níveis de expressão de pacientes e controles, e observamos uma redução significativa, de aproximadamente 18%, na expressão de *TCOF1* de pacientes. Nossos resultados mostraram que o mecanismo de compensação de dose, proposto por Isaac *et al.* (2000) não ocorre em linfócitos de pacientes adultos, ao menos ao nível de regulação da transcrição. Nossos achados estão de acordo com o modelo de haploinsuficiência para a STC: os transcritos com códon de parada prematuro são degradados pelo mecanismo de *nonsense mediated mRNA decay* (NMD) e, consequentemente, os indivíduos portadores de um alelo não-funcional, possuem menor dose de *TCOF1*.

Dada a grande variabilidade inter-individual observada, investigamos a possibilidade de os níveis de transcritos estarem correlacionados à variabilidade fenotípica. Comparamos os dados de expressão de cada paciente à frequência de nove sinais clínicos principais da STC, mas nenhuma correlação foi observada. Embora não possamos afirmar que os níveis de transcritos observados representem o padrão de expressão no embrião, caso a variação nos níveis de transcritos seja regulada por mecanismos genéticos, os indivíduos que possuem baixos níveis de expressão nas células adultas também devem ter baixos níveis de expressão durante o desenvolvimento embrionário. Se essa suposição estiver correta, nossos achados sugerem que

variações nos níveis de transcritos do *TCOF1* não são fundamentais para a ocorrência de variabilidade fenotípica. Estes resultados corroboram os nossos dados de não-associação entre haplótipos do alelo selvagem e variabilidade clínica.

Identificamos no gene *TCOF1* uma ilha CpG com aproximadamente 69 dinucleotídeos CGs. Investigamos o padrão de metilação dessa ilha em pacientes e em controles, como intuito de verificar se diferentes níveis de metilação inter-individual estariam associados à grande variação na expressão gênica detectada nos indivíduos. Detectamos menos de 0,6% de CpGs metilados em pacientes e controles, mostrando que a metilação da ilha CpG não é o mecanismo regulatório por trás dessa ampla variação de expressão do *TCOF1*.

Também avaliamos o padrão de metilação em diferentes linhagens celulares tumorais e em uma linhagem de células-tronco mesenquimais obtidas a partir de polpa de dente. Da mesma maneira que em linfócitos de pacientes e controles, não detectamos metilação significativa. Embora as células-tronco simulem um ambiente semelhante ao dos tecidos embrionários, o padrão desmetilado da ilha CpG nessa linhagem não exclui o *TCOF1* como alvo desse mecanismo regulatório durante o desenvolvimento. Além disso, um trabalho recente (Frigola *et al.*, 2006) demonstrou que mesmo genes com suas ilhas CpGs desmetiladas podem ser silenciados por metilação de sua vizinhança, seja por metilação de citosinas ou de histonas de genes próximos (*LRES, long range epigenetic gene silencing*). Considerando que a transcrição de rRNA ribossômico é regulada por mecanismos epigenéticos, e responsiva a moléculas envolvidas em crescimento e proliferação celular (Grummt e Pikaard, 2003), é possível que esses mecanismos atuem também sobre a regulação do promotor do *TCOF1*.

Dado que a treacle participa da biogênese de ribossomos durante o desenvolvimento (Gozales *et al.*, 2005), e que os níveis de metilação de RNA ribossômico são diferentes nos diferentes *backgrounds* dos camundongos *Tcof1*^{+/-}, sugerimos que a variação fenotípica esteja associada a outros locos. A existência de alelos modificadores ou de genes com função semelhante à do *TCOF1*, que compensem a transcrição e o processamento ineficiente de rRNA dados pela haploinsuficiência de treacle, eventualmente podem gerar um quadro clínico mais leve. Além disso, dado que o conjunto de RNAs ribossômicos transcritos em alguns organismos e tecidos têm expressão influenciada pela origem parental (Grummt e Pikaard, 2003), é possível que a modulação fenotípica na STC tenha alguma influência parental. Julgamos importante verificar se

há expressão alélica preferencial do *TCOF1*, pois esse mecanismo pode influenciar o aparecimento do fenótipo nos indivíduos portadores da mutação patogênica.

Além disso, não podemos excluir o papel exercido por mecanismos estocásticos e ambientais. A fenocópia da STC, gerada pela administração de doses teratogênicas de ácido retinóico a camundongos (Sulik, 1987; Poswillo, 1975), demonstra o quanto o ambiente é capaz de modular o fenótipo. É possível que o ácido retinóico regule a expressão de genes envolvidos na síntese de ribossomos das células da crista neural. Segundo Uechi *et al.* (2006), os órgãos que apresentam alterações específicas em resposta ao silenciamento ou insuficiência de certas ribonucleoproteínas sugerem que a atividade dos ribossomos possa ser tecido ou órgão-específica. Assim, é possível que variações na ingestão de vitamina A (retinol) durante a gravidez possa agravar o fenótipo de um feto portador da mutação para STC.

Referências Bibliográficas

1. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002 Jan;30(1):97-101.
2. Bassam, BJ; Caetano-Anolles, G; Gresshoff, PM. (1991) Fast and sensitive silver staining of DNA in polyacrilamide gels. *Anal Biochem* 196:80-83
3. Bibikova M, Chudin E, Wu B, Zhou L, Garcia EW, Liu Y, Shin S, Plaia TW, Auerbach JM, Arking DE, Gonzalez R, Crook J, Davidson B, Schulz TC, Robins A, Khanna A, Sartipy P, Hyllner J, Vanguri P, Savant-Bhonsale S, Smith AK, Chakravarti A, Maitra A, Rao M, Barker DL, Loring JF, Fan JB. (2006) Human embryonic stem cells have a unique epigenetic signature. *Genome Res.* 16(9):1075-83. Epub 2006 Aug 9.
4. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglu S; Glocal alignment: finding rearrangements during alignment. Special Issue on the Proceedings of the ISMB 2003, Bioinformatics, 19: 54i-62i.
5. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG. (2005) Natural selection on protein-coding genes in the human genome. *Nature.* Oct 20;437(7062):1153-7
6. Cai J, Ash D, Kotch LE, Jabs EW, Attie-Bitach T, Auge J, Mattei G, Etchevers H, Vekemans M, Korshunova Y, Tidwell R, Messina DN, Winston JB, Lovett M. (2005) Gene expression in pharyngeal arch 1 during human embryonic development. *Hum Mol Genet.* 2005 Apr 1;14(7):903-12.
7. Carey, M. & Smale, S.T. (2000) Transcriptional Regulation in Eukaryotes – Concepts, Strategies and Techniques. *CSHL PRESS*
8. Carlson, BM. (2005) Human Embryology and Developmental Biology. 3rd Ed. Elsevier, Inc. Philadelphia, PA, USA.
9. Cui, Z & Dimario, PJ. (2007) RNAi Knockdown of Nopp140 Induces Minute-like Phenotypes in Drosophila. *Mol Biol Cell.* Mar 28; [Epub ahead of print]
10. den Dunnen JT, Antonarakis SE. (2001) Nomenclature for the description of human sequence variations. *Hum Genet.* Jul;109(1):121-4.
11. Dixon, J.; Gladwin, A.J.; Loftus, S.K.; Riley, J.; Perveen, R.; Wasmuth, J.J.; Anand, R.; Dixon, M.J. (1994) A yeast artificial chromosome contig encompassing the Treacher Collins syndrome critical region at 5q31.3-32. *Am J Hum Genet* 55:372-378
12. Dixon, J.; Brakebusch, C.; Fassler, R.; Dixon, M. (2000) Increased apoptosis in the prefusion neural folds underlie the craniofacial disorder, Treacher Collins. *Hum Mol Gen* 9:1473-1480
13. Dixon, J.; Hovanes, K.; Shiang, R.; Dixon, M.J. (1997a) Sequence analysis, identification of evolutionary conserved motifs and expression analysis of murine *Tcof1* provide further evidence for a potential function for the gene and its human homologue, *TCOF1*. *Hum Mol Genet* 6: 727-737
14. Dixon, J.; Edwards, S.J.; Anderson, I.; Brass, A.; Scambler, P.J.; Dixon, M.J. (1997b) Identification of the Complete Coding Sequence and Genomic Organization of the Treacher Collins Syndrome Gene. *Genome Research* 7:223-234
15. Dixon, M.J. (1995) Treacher Collins Syndrome. *J Med Genet* 32: 806-808
16. Dixon, M.J. (1998) Treacher Collins syndrome: from linkage to prenatal testing. *J Laryng and Otol* 112: 705-709
17. Dixon, J; Dixon, M (2003) *TCOF1* (Treacle) and the Treacher Collins syndrome. In: Inborn Errors of Development: The Molecular Basis of Clinical Disorders of Morphogenesis (Oxford Monographs on Medical Genetics. Epstein, Erickson and Wynshaw-Boris. Oxford University Press, USA
18. Dixon J, Jones NC, Sandell LL, Jayasinghe SM, Crane J, Rey JP, Dixon MJ, Trainor PA. (2006) Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc Natl Acad Sci U S A.* Sep 5;103(36):13403-8. Epub 2006 Aug 28.

19. Edwards, S.J.; Gladwin, A.J.; Dixon, M.J. (1997) The Mutational Spectrum in Treacher Collins Syndrome Reveals a Predominance of Mutations That Create a Premature-Termination Codon. *Am J Hum Genet* 60:515-524
20. Ellis, P.E.; Dawson, M.; Dixon, M.J. (2002) Mutation testing in Treacher Collins Syndrome. *J Orthod* 29:293-297
21. Erlich, M.S.; Cunningham, M.L.; Hudgins, L. (2000) Transmission of the Dysgnathia Complex From Mother to Daughter. *Am J Med Genet* 95:269-274
22. Emes, RD; Ponting, CP. (2001) A new sequence motif linking lissencephaly, Treacher Collins and oral-facial-digital type 1 syndromes, microtubule dynamics and cell migration. *Hum Mol Genet* 10:2813-2820.
23. Esterbauer, H. et al. (2001) A common polymorphism in the promoter of *UCP2* is associated with decreased risk of obesity in middle-aged humans. *Nat Genet* 28: 178-183
24. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 2004 Jul 1;32 (Web Server issue):W273-9
25. Frigola, J; Song, J; Stirzaker, C; Hinshelwood, R.A.; Peinado, M.A.; Clarck, S.J. (2006) Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat Genet* 38, 540-549.
26. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A.* Mar 1;89(5):1827-31.
27. Frommer et al. (2002) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *PNAS* Mar 1;89(5):1827-31.
28. Gladwin, A.J.; Dixon, J.; Loftus, S.K.; Edwards, S.; Wasmuth, J.J.; Hennekam, R.C.M.; Dixon, M.J. (1996) Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene. *Hum Mol Gen* 5:1533-1538
29. Goldenberg D, Harden S, Masayesva, BG, Ha P, Benoit N, Westra WH, Koch WM, Sidransky D, e Califano JA. (2004) Intraoperative molecular margin analysis in head and neck cancer. *Arch Otolaryngol Head Neck Surg.* 130(1); 39-44.
30. Gollop, T.R. (1981) Estudo Genético-clínico das Disostoses Mandibulofacial e Frontofacionasal. Tese de doutorado. Instituto de Biociências, Universidade de São Paulo, SP, Brasil.
31. Gonzales B, Henning D, So RB, Dixon J, Dixon MJ, Valdez BC. (2005a) The Treacher Collins syndrome (TCOF1) gene product is involved in pre-rRNA methylation. *Hum Mol Gene* Jul 15;14(14):2035-43.
32. Gonzales B, Yang H, Henning D, Valdez BC. (2005b) Cloning and functional characterization of the *Xenopus* orthologue of the Treacher Collins syndrome (TCOF1) gene product. *Gene*. Oct 10;359:73-80
33. Gordon S, Akopyan G, Garban H, Bonavida B. (2006) Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*. Feb 23;25(8):1125-42.
34. Gorlin, R.J.; Cohen, M.M.; Levin, L.S. (1990) Syndromes of the head and neck. *Oxford University Press* 3rd Ed. Oxford, UK
35. Gorlin, R.J.; Cohen, M.M.; Hennekam, R.C. (2001) Branchial Arch and oral-Acral Disorders. In: Syndromes of the head and neck, 4th edition. Oxford: Oxford University Press. p.790-849.
36. Grummt I, Pikaard CS. (2003) Epigenetic silencing of RNA polymerase I transcription. *Nat Rev Mol Cell Biol.* Aug;4(8):641-9.
37. Guion-Almeida, M.L.; Kokitsu-Nakata, N.M.; Zechi-Ceide, R.M.; Vendramini, S. (1999) Auriculo-Condilar syndrome: Further Evidence for a New Disorder. *Am J Med Genet* 86: 130-133
38. Guion-Almeida, M.L.; Zechi-Ceide, R.M.; Vendramini, S.; Kokitsu-Nakata, N.M. (2002) Auriculo-Condilar syndrome: additional patients. *Am J Med Genet* 112:209-214
39. Hansen, M.; Lucarelli, M.J.; Whiteman, D.A.H.; Mulliken, J.B. (1996) Treacher Collins Syndrome: Phenotypic variability in a family including an infant with arhinia and uveal colobomas. *Am J Hum Genet* 61:71-74
40. Haworth, K.E.; Islam, I.; Breen, M.; Putt, W.; Makrinou, E.; Binns, M.; Hopkinson, D.; Edwards, Y. (2001) Canine *TCOF1*; cloning, chromosome assignment and genetic analysis in dogs with different head types. *Mammalian Genome* 12: 622-629

41. Hayano T, Yanagida M, Yamauchi Y, Shinkawa T, Isobe T, Takahashi N. (2003) Proteomic analysis of human Nop56p-associated pre-ribosomal ribonucleoprotein complexes. Possible link between Nop56p and the nucleolar protein treacle responsible for Treacher Collins syndrome. *J Biol Chem*. Sep 5;278(36):34309-19.
42. Horiuchi K, Ariga T, Fujioka H, Kawashima K, Yamamoto Y, Igawa H, Sugihara T, Sakiyama Y. (2005) Mutational analysis of the TCOF1 gene in 11 Japanese patients with Treacher Collins Syndrome and mechanism of mutagenesis. *Am J Med Genet A*. May 1;134(4):363-7.
43. Isaac, C.; Marsh, K.L.; Panzekas, W.A.; Dixon, J.; Dixon, M.J.; Jabs, E.W.; Meier, U.T. (2000) Characterization of the nucleolar gene product, treacle, in Treacher Collins syndrome. *Mol Biol Cell* 11: 188-192
44. Jampol, M. *et al.* (1998) Prominent, Constricted Ears With Malformed Condyle of the Mandible. *Am J Med Genet* 75:499-452
45. Jones, K.L. (1997) *Smith's Recognizable Patterns of Human Malformations* 5th ed. Pennsylvania: Saunders.
46. Jones, N.C.; Farlie, P.G.; Mimichiello, J.; Newgreen, D.F. (1999) Detection of an appropriate kinase activity in branchial arches I and II that coincides with peak expression of the Treacher Collins syndrome gene product, treacle. *Hum Mol Genet* 12:2239-2245
47. Johnson, JP; Poskanzer, LS, Sherman, S (1996) Three-generation family with resemblance to Townes-Brocks syndrome and Goldenhar/oculoauriculovertebral spectrum. *Am J Med Genet* 61:134-139.
48. Joosten, P.H.L.J.; Toepoel, M.; Mariman, E.C.M.; Zoelen, E.J.J.V. (2001) Promoter haplotype of the platelet-derived growth factor alpha-receptor gene predispose to human neural tube defects. *Nat Genet* 27: 215-17
49. Kelberman D, Tyson J, Chandler DC, McInerney AM, Slee J, Albert D, Aymat A, Botma M, Calvert M, Goldblatt J, Haan EA, Laing NG, Lim J, Malcolm S, Singer SL, Winter RM, Bitner-Glindzic M. (2001) Hemifacial microsomia: progress in understanding the genetic basis of a complex malformation syndrome. *Hum Genet*. 2001 Dec;109(6):638-45. Epub 2001 Oct 26.
50. Kemper, O.; Wallach, D. (1993) Cloning and partial characterization of the promoter for human p55 tumor necrosis factor (TNF) receptor. *Gene* 134:209-216
51. Khajavi M, Inoue K, Lupski JR. (2006) Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *Eur J Hum Genet*. 2006 Oct;14(10):1074-81. Epub 2006 Jun 7. Review.
52. Kobrinski, L., Chitayat, D. Zahed, L, McGregord, D., Rochon, L., Brownstein, S. Vekemens, M, and Albert, DL. (1993). Trisomy 22 and facioauriculovertebral (Goldenhar) sequence. *Am J. Med Gent* 46: 68-71.
53. Kogai, T; Kanamoto, Y e Brent, GA. (2003) The modified firefly luciferase reporter gene (*luc+*) but not *Renilla* luciferase is induced by all-*trans* retinoic acid in MCF-7 breast cancer. *Breast Cancer Research and Treatment* 78: 119-126
54. Kuzmiak, HA and Maquat LE. (2006) Applying nonsense-mediated mRNA decay research to the clinic: progress and challenges. *TRENDS Mol Med*. 12(7): 306-316.
55. Larsen, WJ (1997) *Human embryology*. 2nd Edition. Ed. New York: Churchill Livingstone.
56. Lowry, R.B.; Morgan, K.; Holmes, T.M.; Metcalf, P.J.; Sautter, G.F (1985) Mandibulofacial dysostosis in Hutterite sibs: a possible recessive trait. *Am J Hum Genet* 22:501-512
57. Marres, H.A.M.; Cremers, C.W.R.J.; Dixon, M.J.; Huygen, P.L.M.; Joosten, B.M. (1995) The Treacher Collins syndrome: A clinical radiological and genetic linkage study on two pedigrees. *Arch Otolaryngol Head Neck Surg* 121:509-514.
58. Marsh, K.L.; Dixon, J.; Dixon, M.J. (1998) Mutations in the Treacher Collins Syndrome gene lead to mislocalization of the nucleolar protein treacle. *Hum Mol Genet* 7:1795-1800
59. Masotti C, Armelin-Correa LM, Splendore A, Lin CJ, Barbosa A, Sogayar MC, Passos-Bueno MR. (2005) A functional SNP in the promoter region of TCOF1 is associated with reduced gene expression and YY1 DNA-protein interaction. *Gene*. Oct 10;359:44-52.
60. McDonald, J. H., and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351: 652-654.
61. Meier, UT; Blobel, G. (1992) Nopp140 shuttles on tracks between nucleolus and cytoplasm. *Cell* 70:127-138.

62. Miller, S.A.; Dykes, D.D.; Polesky, H.F. (1988) A simple testing out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
63. Nielsen, R. (2005) Molecular Signatures of Natural Selection. *Annual Review of Genetics*, 39: 197-218.
64. Nikaido I, Saito C, Mizuno Y, Meguro M, Bono H, Kadomura M, Kono T, Morris GA, Lyons PA, Oshimura M, Hayashizaki Y, Okazaki Y; RIKEN GER Group; GSL Members.(2003) Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome Res.*Jun;13(6B):1402-9.
65. Nobrega MA, Pennacchio LA. (2004) Comparative genomic analysis as a tool for biological discovery. *J Physiol.* Jan 1;554(Pt 1):31-9. Review.
66. Oefner, P.J. (2000). Allelic discrimination by denaturing high performance liquid chromatography. *J. Chromatogr. B. Biomed. Sci. Appl.* Mar 10; 739(2): 345-55.
67. Online Mendelian Inheritance in Man (OMIM) Johns Hopkins University, Baltimore, M.D. *1545500 [3/5/1999] <http://www.ncbi.nlm.nih.gov/omim>
68. Orita, M.; Suzuki, Y.; Sekiya, T.; Hayashi, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5:874-879
69. Poswillo, D. (1975) The pathogenesis of Treacher Collins syndrome (mandibulofacial dysostosis). *Br J Oral Surg* 13:1-26
70. Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, Pennacchio LA. (2006) Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res.* 16:855-63
71. Priolo, M. *et al.* (2000) Question mark eras, temporo-mandibula joint malformation and hipotonia: auriculo-condylar syndrome or a distinct entity? *Clin Dysmorph* 9: 277-280
72. Read, M., 1998. Electrophoretic mobility shift assay (EMSA). In: Docherty, K. (Ed.), *Essential Techniques*. Wiley & Sons, NewYork, US.
73. Reik, W; Dean, W; Walter, J. (2001) Epigenetic reprogramming in mammalian development. *Science*, 293: 1089-1093.
74. Richards, B.; Skoletsky, J.; Shuber, A.P.; Balfour, R.; Stern, R.C.; Dorkin, H.L.; Parad, R.B.; Witt, D.; Klinger, K.W. (1993) Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum Mol Genet* 2:159-163
75. Richieri-Costa, A.; Bortolozo, M.A.; Lauris, J.R.P.; Lauris, R.C.M.C.; Gion-Almeida, M.L.; Marques, D.; Moretti, D. (1993) Mandibulofacial dysostosis: report on two brazilian families suggesting autosomal recessive inheritance. *Am J Méd Genet* 46: 659-664
76. Rogers, BO. (1964) Berry-Treacher-Collins syndrome: a review of 200 cases. *Br J Plast Surg*, 17:109.
77. Rogers, BO. (1976) The surgical treatment of mandibulofacial dysostosis. *Clin Plast Surg*, 3:653.
78. Roux PP, Shahbazian D, Vu H, Holz MK, Cohen MS, Taunton J, Sonenberg N, Blenis J. (2007) RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. *J Biol Chem.* Mar 14; [Epub ahead of print]
79. Rovin, S; Dachi, SF, Borestein, DB, Cotter, WR. (1964) Mandibulofacial dysostosis, a familial study of five generations. *J Pediatr*, 64:215.
80. Rozas, J, Sánchez-DelBarrio, J.C, Messeguer, X, e Rozas, R. (2003) *Bioinformatics* 19: 2496-2497
81. Seto, E; Shi, Y; Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. *Nature* Nov 21; 354 (6350): 241-5.
82. Shows KH, Ward C, Summers L, Li L, Ziegler GR, Hendrickx AG, Shiang R. (2006) Reduced TCOF1 mRNA level in a rhesus macaque with Treacher Collins-like syndrome: further evidence for haploinsufficiency of treacle as the cause of disease. *Mamma Genome* Feb;17(2):168-77. Epub 2006 Feb 7.
83. Smale, S.T. (1997) Transcription incitaion from TATA-less promoters within eukaryotic protein-coding genes. *Bioch Bioph ACTA* 1351: 73-88
84. So RB, Gonzales B, Henning D, Dixon J, Dixon MJ, Valdez BC: Another face of the Treacher Collins syndrome (*TCOF1*) gene: identification of additional exons. *Gene* 2004 **328**: 49-57.
85. Splendore, A.; Silva, E.O.; Alonso, L.G.; Richieri-Costa, A.; Alonso, N.; Rosa, A.; Carakushansky G.; Cavalcanti, D.P.; Brunoni, D.; Passos-Bueno, M.R. (2000) High mutation detection rate in *TCOF1* among

Treacher Collins syndrome patients reveals clustering of mutations in 16 novel pathogenic changes. *Hum Mutat* 16:315-322

86. Splendore, A.; Passos-Bueno, M.R. (2001) Síndrome de Treacher Collins: Aspectos clínicos, genéticos e moleculares. *Revista de Medicina* (editada pelo Departamento Científico do Centro Acadêmico Oswaldo Cruz da Faculdade de Medicina da Universidade de São Paulo) 80:52-56 Jan./Fev./Mar.
87. Splendore, A.; Jabs, E. W.; Passos-Bueno, M.R. (2002a) Screening of *TCOF1* in patients from different populations: confirmation of mutational hot spots and identification of a novel missense mutation that suggests an important functional domain in the protein treacle. *J Med Genet* 39:493-495
88. Splendore, A.; Jabs, E. W.; Maldergem, L.V.; Wulfsberg, E.; Passos-Bueno, M.R. (2002b) *TCOF1* mutations excluded from a role in other first and second branchial arch related disorders. *Am J Med Genet* 111:324-327
89. Splendore, A.; Jabs, E. W.; Félix, T.M.; Passos-Bueno, M.R. (2003) Parental origin of mutations in sporadic cases of Treacher Collins syndrome. *Eur J Hum Genet* Sep 11(9): 718-22
90. Stephens, M., and Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am.J. Hum. Genet.* 73: 1162-1169.
91. Storm AL, Johnson JM, Lammer E, Green GE, Cunniff C. (2005) Auriculo-condylar syndrome is associated with highly variable ear and mandibular defects in multiple kindreds. *Am J Med Genet* Oct 1;138(2):141-5
92. Sulik, KK.; Johnston, MC.; Smiley, MC; Speight, HS; Jarvis, BE. (1987) Mandibulofacial dysostosis (Treacher Collins syndrome): a new proposal for its pathogenesis. *Am J Med Genet* 27:359-372.
93. Szlazak, J. (1953) Treacher-Collins syndrome. *Canad Med Assoc J*, 69:274.
94. Tan DP, Nonaka K, Nuckolls GH, Liu YH, Maxson RE, Slavkin HC, Shum L. YY1 activates *Msx2* gene independent of bone morphogenetic protein signaling. *Nucleic Acids Res.* 2002 Mar 1;30(5):1213-23.
95. Teber OA, Gillesen-Kaesbach G, Fischer S, Bohringer S, Albrecht B, Albert A, Arslan-Kirchner M, Haan E, Hagedorn-Greife M, Hammans C, Henn W, Hinkel GK, König R, Kunstmann E, Kunze J, Neumann LM, Prott EC, Rauch A, Rott HD, Seidel H, Spranger S, Sprengel M, Zoll B, Lohmann DR, Wiczorek D. (2004) Genotyping in 46 patients with tentative diagnosis of Treacher Collins syndrome revealed unexpected phenotypic variation. *Eur J Hum Genet.* 12(11):879-90.
96. Treacher Collins, E. (1900) Cases with symmetrical congenital notches in the outer part of each lid and defective development of the malar bones. *Trans Ophthalmol Soc UK* 20: 190-192.
97. The Treacher Collins Collaborative Group (1996) Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat Gen* 12:130-136
98. Thomsom, A. (1846) Notice of several cases of malformations on the external ear together with experiments of hearing in such persons. *Monthly J Med Sci*, 3: 419.
99. Uechi, T; Nakajima, Y; Nakao, A; Torihara, H; Chakraborty, A; Inoue, K; Kenmochi, K. (2006) Ribosomal Protein Gene Knockdown Causes Developmental Defects in Zebrafish. *Plos ONE*, Dez (1), e37.
100. Valdez BC, Henning D, So RB, Dixon J, Dixon MJ: The Treacher Collins syndrome (*TCOF1*) gene is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc Natl Acad Sci* 2004; 101 (29): 10709-10714.
101. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002 Jun 18;3(7):RESEARCH0034. Epub 2002 Jun 18.
102. Vendramini, S; Richieri-Costa, A; Guion-Almeida, ML. (2007) Oculoauriculovertebral spectrum with radial defects: a new syndrome or an extension of the oculoauriculovertebral spectrum? Report of fourteen Brazilian cases and review of the literature. *Eur J Hum Genet* 15, 411–421
103. Vidal M, Legrain P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* 1999 Feb 15;27(4):919-29. Review.
104. Winokur, S.T. and Shiang, R. (1998) The Treacher Collins syndrome (*TCOF1*) gene product, treacle, is targeted to the nucleolus by signals in the C-terminus. *Hum Mol Genet* 7: 1947-1952
105. Wise, C.A.; Chiang, L.C.; Paznekas, W.A.; Sharma, M.; Musy, M.M.; Ashley, J.A.; Lovett, M. Jabs, E.W. (1997) *TCOF1* gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treaches Collins Syndrome throughout its coding region. *Proc Natl Acad Sci USA* 94:3110-3115

106. Yan, H.; Dobbie, Z.; Gruber, S.B.; Markowitz, S.; Romans, K.; Giardello, F.M.; Kinzier, K.W.; Vogelstein, B. (2002) Small changes in expression affect predisposition to tumorigenesis. *Nature Genet* 30: 25-26

Anexo

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(maiores de 18 anos)

ESTUDO: “*Estudo de mecanismos regulatório e mapeamento de genes associados malformações craniofaciais*”

Você está sendo convidado(a) a participar do projeto de pesquisa acima citado. O documento abaixo contém todas as informações necessárias sobre a pesquisa que estamos fazendo, então leia atentamente e caso tenha dúvidas, vou esclarece-las (se não souber ler, fique tranquilo(a) que leio para você). Se concordar, o documento será assinado e só então daremos início a pesquisa. Sua colaboração neste estudo será de muita importância para nós, mas se desistir a qualquer momento, isso não causará nenhum prejuízo a você.

Eu _____, RG _____,
nascido(a) em ____ / ____ / _____, abaixo assinado(a), concordo de livre e espontânea vontade em participar como voluntário do estudo “***Mapeamento e caracterização de mecanismos regulatórios de genes associados malformações de primeiro e segundo arcos faríngeos***”, esclareço que obtive todas as informações necessárias e fui esclarecido(a) de todas as dúvidas apresentadas.

Estou ciente que:

- I) O estudo se faz necessário para que possam descobrir as possíveis causas da doença denominada “*Estudo de mecanismos regulatório e mapeamento de genes associados malformações craniofaciais*”.
- II) Será feita uma coleta de 5 a 10 mL de sangue venoso com o uso de uma seringa (), ou uma coleta de células da mucosa bucal com o uso de espátulas (), ou coleta de amostra de tecido descartado durante cirurgia plástica ().
- III) Essas coletas serão feitas apenas para este estudo e em nada influenciará no meu tratamento; não vai me curar; não vai me causar nenhum problema, exceto a dor da picadinha da agulha no local da coleta;
- IV) A participação neste projeto não tem objetivo de submeter a um tratamento terapêutico e será sem custo algum para mim;
- V) Tenho a liberdade de desistir ou interromper a colaboração neste estudo no momento em que desejar, sem necessidade de qualquer explicação;
- VI) A desistência não causará nenhum prejuízo a mim, e sem que venha interferir no atendimento ou tratamento médico;
- VII) Os resultados obtidos durante este ensaio serão mantidos em sigilo, mas concordo que sejam divulgados em publicações científicas, desde que meu nome não seja mencionado;
- VIII) Quanto ao registro fotográfico de meu (minha) filho (a), eu
 - () autorizo que seja feito, podendo ser arquivado no prontuário e ser eventualmente utilizado em publicações científicas, sem que o nome de meu (minha) filho (a) seja relacionado.
 - () autorizo que seja feito, mas apenas deve permanecer arquivado no prontuário, não podendo ser exposto ou publicado de alguma forma.
 - () não o autorizo.
- IX) Caso eu desejar, poderei tomar conhecimento dos resultados ao final desta pesquisa
 - () Desejo conhecer os resultados desta pesquisa.
 - () Não desejo conhecer os resultados desta pesquisa.

São Paulo, de de 2007.

() Paciente / () Responsável
.....

Testemunha 1 : _____
Nome / RG / Telefone

Testemunha 2 : _____
Nome / RG / Telefone

Responsável pelo Projeto: _____
Dra. Maria Rita Passos-Bueno/ Cibele Masotti