Identification of previously unreported *TCOF1* orthologous exons in the mouse and monkey and expression analysis of an alternatively spliced exon in the mouse

Greg Ziegler

Abstract

Treacher-Collins syndrome is an autosomal dominant craniofacial disorder which affects approximately 1 in 50,000 live births. Treacher-Collins is caused by mutations in the protein treacle, which is coded for by the gene *TCOF1*. *TCOF1* has known orthologs in the canine, mouse, and monkey species. Reported in this paper is evidence of the expression of exon 16c in the mouse and exon 19 in the monkey. Whole mount in-situ hybridization was performed to determine the expression levels of mouse exon 16c during embryogenesis.

Introduction

Alternative splicing of RNA transcripts is a method by which many different protein products are coded for by the same gene. It is predicted that approximately 50% of genes are alternatively spliced in humans (Brett et al, 2001). Alternative splicing often is a result of an omission or addition of an exon. Alternative splice products can be identified by expressed sequence tag (EST) analysis. One gene which undergoes alternative splicing is *TCOF1*. *TCOF1* is found on the long arm of chromosome 5 and codes for a protein called treacle, mutations in treacle (usually resulting in premature stop codons) are the cause of the craniofacial disorder Treacher-Collins syndrome (Gladwin et al, 1996). Treacher-Collins syndrome is an autosomal dominant disorder which affects approximately 1 in 50,000 live births. The protein treacle has been shown to be a nucleolar phosphoprotein (Wise et al., 1997) and has close structural similarities to the nucleolar phosphoprotein Nopp140 (Isaac C. et al. 2000). Nucleolar phosphoproteins thought to aid in the transport of proteins or ribosomal subunits between the cytoplasm and the nucleolus, thus they likely have a large role in ribosome construction. A recent study indicates that treacle is involved in ribosomal DNA gene transcription by interacting with upstream binding factor (Valdez et al. 2004).

The *TCOF1* gene has known orthologs in canine, bovine, mouse, and rhesus monkey genomes (monkey sequence currently unpublished). Currently little is known about the various structural similarities and differences between the human TCOF1 gene and its orthologs. By creating a comprehensive cDNA analysis for each species using the EST data banks and experimental methods, a complete alignment of the orthologous sequences can be developed. This will provide information on how well conserved the splice variants of *TCOF1* are across species and allow for further analysis of the alternatively spliced products expression and function in other organisms. Expression of an alternatively spliced exon, 16c, of the mouse is also analyzed using whole-mount *in situ* hybridization of embryos to determine whether its expression varies from that of the entire *TCOF1* gene during embryogenesis (Dixon et al, 1997).

This paper presents new sequence for alternatively spliced exons in the mouse and monkey. Additionally, it reports the expression of one of the alternatively spliced mouse exons during embryogenesis.

Materials and Methods

Sequence Analysis

Published cDNA sequences of the *TCOF1* gene in the human, mouse, and canine were available from GenBank and by searching the expressed sequence tag (EST) data bank on the NCBI internet server (http://www.ncbi.nih.gov/) as well as the UCSC genome browser (http://genome.ucsc.edu/). The rhesus monkey *TCOF1* cDNA sequence was generated in the laboratory (unpublished data).

Once a complete cDNA sequence for each species was found, the protein translations of each were aligned using DNAstar sequence analysis software (http://www.dnastar.com/). The program used the Clustal W slow algorithms for alignment. Further manual alignment was also performed to fix alignment where exons were missing.

The BLAST program was used to search the non-redundant database to find whether missing orthologous sequence could be found in the genomic DNA for a specific organism (Altschul et al, 1990). If the sequence was found in the genomic DNA, but not in the EST database, the expression of the missing exon was tested.

Reverse Transcription-Polymerase Chain Reaction

Primers were developed to span the gaps for the missing exons (Figure 1). RNA was isolated from brain and liver tissue in the mouse, monkey spleen tissue, and canine white blood cells (isolated using Lymphoprep tube) using the Ambion RNAqueous kit and protocol. Reverse Transcription (RT) was performed using 100ng random primers, and approximately 1µg RNA, diluted to 12µL and incubated at 70°C for 10 min. 4µL of 5x M-MLV RT buffer (BRL), 2µL of 0.1 M DTT, 2µL of 10mM dNTP's and 0.25µL of 40U/µL RNAsin (Promega) were added and equilibrated at 37°C for 2 min. Then 1µL of M-MLV RT (BRL) was added and incubated for 1 hour at 37°C and heated to 95°C for 5 min. This method was used for the mouse and monkey samples. The Qiagen One-Step RT-PCR kit and protocol was used for the canine RNA.

Canine Exon 6b:
(F) CAGTGATGAGACCGATGTGGA
(R) CCTCCTTTGGCTTCTCTGTCTT
Mouse Exon 16b and 16c:
(F) AGTCTGACAGTGACACAG
(R) GGCAGTGGTGCCTGAGGCCTT
Mouse Exon 19:
(F) AGCAAGAGTGAGCAGTCTAG
(R) AGTGTCACTGCTGTCCT
Monkey Exon 16b:
(F) AGCAGCAGCGAGGAGTCAAA
(R) GCTGGGCCAGCTGGACTACG
Monkey Exon 19:
(F) AGGGACCAGCCACTCAGG
(R) GCCCTCAGGAAGCTTAG
Mouse 16c (for hybridization probe):
(F) GATGCTATCTCCCAGCCT
(R) CTGGCCGGAGGGCAGAT

Figure 1: PCR primers used to amplify gaps for missing exons. All are 5'-3'.

Polymerase Chain Reaction (PCR) of the RT products for mouse and monkey were carried out in a 25µL solution containing 2.5µL rt product, 2.5µL of 10X reaction buffer, 4µL of 1.25mM dNTPs, 1.5µL of 25mM $MgCl_2$, 2µL of 6.25 picoM primers, and 0.1µL Taq Polymerase. PCR conditions were 95°C for 3 min, 94°C for 30 s, 70°C for 30 s (-1°C each cycle), 72°C for 30 s, cycle step 2-4 10 times, 94°C for 30 s, 60°C for 30 s, 72°C for 30 s, cycle 24 times, 72°C for 7 min. The canine PCR was carried out using a

12.5µL solution containing 3.0µL rt product, 1.25µL of 10X reaction buffer, 2µL of 1.25mM dNTPs, 1.75µL of 25mM $MgCl_2$, 1µL of 6.25 picoM primers, and 0.05µL Taq Polymerase. PCR conditions were 94°C for 1:30 min, 94°C for 30 s, 55°C for 30 s, 72°C for 30 s, cycle steps 2 to 4 29 times, 72°C for 7 min. PCR results were run on 6% polyacrylamide gels to determine product sizes.

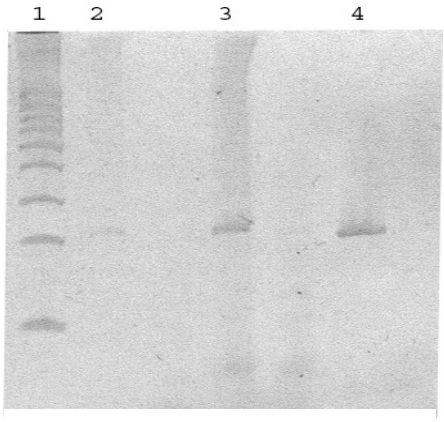

Figure 2 – RT-PCR Results Showing Splice Variants for Mouse Exon 19. 1: DNA Marker Hyperladder IV (BioLine), 2:Mouse Brain, 3:Mouse Liver 4: Control cDNA strand (pmtcof1 plasmid)

Individual RT-PCR products were inserted into pCR 2.1-topo vector (Invitrogen). The cloning reaction was transformed into DH5-α *E. coli* cells and plated on LB. Ampicillin plates were streaked with Blue-Gal and IPTG for Blue-White cloning and incubated at 37˚C overnight.

Eight white colonies were chosen from each plate, grown overnight in 5 mLs of LB with ampicillin and the plasmid DNA was extracted by using the QuantumPrep Plasmid Miniprep kit (BioRad). Insert size was determined by digesting the plasmid with 10 units of EcoR1 in 1X EcoR1 buffer (New England Biolabs) and running on a 0.9% agarose gel.

Clones were then sequenced on an ABI 377 Sequencer using M13F and M13R primers.

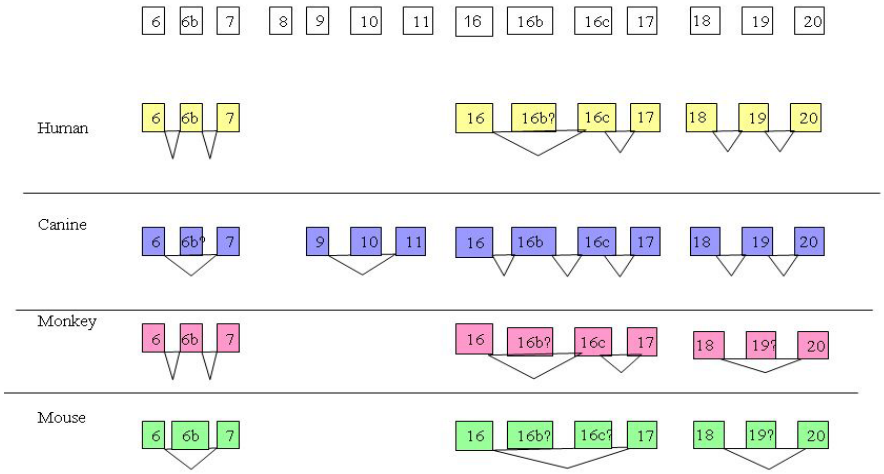

Figure 3 - Known and Unknown Sequences in Each Species After GenBank cDNA analysis

Whole Mount in-situ Hybridization

Mouse exon 16c and the TCOF1 cDNA 9AB5 was used for whole mount *in-situ* hybridization. Primers were developed which span the start and the end of exon 16c (Figure 1). The fragment was cloned into pBluescript II SK(+) and the plasmid isolated using Eppendorf FastPlasmid Mini Kit. The cDNAs were digested with HindIII and Not I and then run on a 1% agarose gel to determine effectiveness. A clone with the correct product size was chosen and a large yield of the plasmid was made using the Qiagen Qiafilter Plasmid Midi kit. Eight and a half and 9 $^1/_2$ dpc embryos were dissected and prepared using the protocol described by Wilkinson, D.G. in the article "Whole mount in situ hybridization of vertebrate embryos (1992)". The probe was prepared per the protocol guidelines using T3/T7 polymerase (Amersham).

Results

Sequence Analysis

The human *TCOF1* gene has 24 major form exons, as well as two identified alternatively spliced exons, exon 6b and exon 16c. Exon 19 also undergoes alternative splicing. After analysis of the GenBank cDNA libraries and the unpublished monkey sequence the following gaps were found: there was no evidence of exon 6b in canine or mouse, no evidence of exon 16b in human, mouse, or monkey, no evidence of exon 16c in mouse, no evidence of exon 19 in the mouse or the monkey. There was also no exon 10 in the canine. Mouse exon 6b was missing from both the cDNA library as well as the mouse genomic DNA (Figure 3).

Missing Exon Expression

All sequences missing from the cDNA library, but present in the genomic database were tested for experimentally. RT-PCR revealed multiple splice variants across most of the regions (Figure 3). Analysis of sequencing shows no evidence of exon 16b in the mouse or monkey. There is also no evidence of expression of exon 19 in the mouse (Figure 2). The monkey exon 19 gap showed multiple splice variants after the RT-PCR (Figure 4). Sequence analysis of this RT-PCR confirms that the larger band contains exon 19 (Lane 3, Figure 4; Figure 5). Analysis of the splice variants shown for the mouse tissues for exon 16b and 16c revealed the presence of exon 16c (Figure 6).

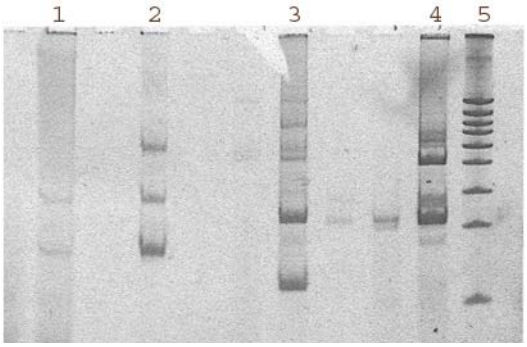

Figure 4 - RT-PCR Results Showing Splice Variants – Each band indicates a possible variant. 1: Mouse Liver (Region between exon 16 and 17 amplified), 2: Mouse Brain (Region between exon 16 and 17 amplified), 3: Monkey Spleen (Region between exon 18 and 20 amplified), 4: Monkey Spleen (Region between exon 16 and 17 amplified), 5:DNA Marker Hyperladder IV (BioLine)

Discussion

The results indicate that there are possibly many more alternatively spliced products contained in the *TCOF1* gene than can be found in the databases. Because of the high degree of mutation and alternative splicing in *TCOF1*it is difficult to obtain a comprehensive cDNA sequence for orthologous species.

ACAGTTTCTTCTTGGGTCGAAGGGAGGGT
GACTGGGTCCGACGGGACTTCCAGGAGCG
GGTCTTTCGGTCACTCCGAGTCGGAGGAC
AACGGTCCTGGGTCGCAAGTGCACCC
Figure 5 - Monkey Exon 19, 113bp.

GATGCTATCTCCCAGCCTGCCAGAGGCAA
AAGCGTCAGGTCCAGCGTCCCCAGAGAAG
AGCATAGAAGGGTCCTCANAGAGCAGTGA
TGAGGATCTGCCCTCCGGCCAG
Figure 6 - Mouse Exon 16c, 109bp.

Exon 16c has now been found to be expressed in all of the species analyzed. The existence of the exon in all of the species indicates that it may have a significant role in the function of treacle. Analysis with Prosite (http://ca.expasy.org/) reveals that Exon 16c has a potential casein kinase 2 (CK2) phosphorylation site which means that it could be used as a phosphorylation regulator. The treacle protein is highly phosphorylated and it has been hypothesized that CK2 may be a factor in the high degree of phosphorylation present (Isaac, C. et Al, 2000).

It is interesting that the canine, human, and monkey all express exon 19, while the mouse shows no evidence of expression of this exon. Exon 19 contains a phosphorylation site and it has been proposed that expression of exon 19 could be used to regulate phosphorylation of TCOF1 (Haworth et. Al, 2001). However, the major alternative splice form is without exon 19 (Haworth et Al, 2001), and the exons apparent absence from the mouse indicate that the exon may have little to no functional significance.

Acknowledgements

Thank you to the Dr. Rita Shiang and the Shiang lab for this opportunity and all of the help and support. Also, thank you to Dr. Jeff Elhai for setting up the BBSI-VCU program and NIH for funding BBSI.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. "Basic local alignment search tool." J. Mol. Biol. 215:403-410

Brett, D., Pospisil, H., Valcárcel, J., Reich, J., and Bork, P. "Alternative splicing and genome complexity." Nature. January 2002. 30:29-30.

Gladwin, A.J., Dixon, J, Loftus, S.K., Edwards, S, Wasmuth, J.J., Hennekam, R.C., and Dixon, M.J. "Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene." Human Molecular Genetics. 1996. 5:1533-1538.

Haworth, K.E., Islam, I., Breen, M., Putt, W., Makrinou, E., Binns, M., Hopkinson, D., Edwards, Y. "Canine TCOF1; cloning, chromosome assignment and genetic analysis in dogs with different head types." Mammalian Genome. 2001. 12:622-629.

Isaac, C., Marsh, K.L., Paznekas, W.A., Dixon, J., Dixon, J.M., Ethylin, W.J., Wang Jabs, E., Meier, U.T. "Characterization of the Nucleolar Gene Product, Treacle, in Treacher Collins Syndrome." Molecular Biology of the Cell. 2000. 11:3061-3071.

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 Sci USA. 101:10709-14.

Wilkinson, D.G. 1992. "Whole Mount in situ Hybridization of Vertebrate embryos," in *In Situ Hybridization: A Practical Approach*. IRL Press, Oxford 75-83

Wise C. Chiang L. Paznekas W. Sharma M. Musy M et al. 1997. "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treacher Collins Syndrome throughout its coding region." Proc Natl Sci USA 94:3110-3115.