Greg Ziegler
BBSI Academic Year Proposal
8/5/03

Identifying Splice Variants in the Treacher-Collins gene: TCOF1

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 called Treacher-Collins syndrome (Gladwin et al, 1996). Treacher-Collins 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 and has close structural similarities to the nucleolar phosphoprotein Nopp140 (Isaac et al, 2000 et al, 2000; Winokur and Shiang, 1998). Nucleolar phosphoproteins 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.

The *TCOF1* gene has known orthologs in canine, bovine, mouse, and rhesus monkey genomes (unpublished). Currently little is known about the various structural similarities and differences between the human *TCOF1* gene and its orthologs. This study will seek to analyze and compare the various splice variants that have been identified in the different organisms. This will help provide us with an idea of what the similarities and differences are in the function of the treacle protein between the species as well as information on how well the gene has been evolutionarily conserved between the species. After identifying the different splice variants and the differences between the orthologs, we can analyze the differential expression of the genes in the lab to find the role of treacle throughout the development process and hopefully help shed some light on the role that treacle plays on craniofacial development.

Summer Progress Report

The first part of the summer was spent working with the nucleotide database at NCBI and the UCSC genome browser to find a complete TCOF1 cDNA sequence for the human, mouse, and canine. The rhesus monkey sequence is unpublished but on file in the lab. However, the complete bovine sequence hasn't yet been published in the database, so its sequence couldn't be used in the comparison. After identifying the TCOF1 sequence for each species the DNAstar sequence analysis software (http://www.dnastar.com/) was used to align the translated protein sequence of each species to one another. This way it would be easy to see if the sequences were missing in any exon. This proved to be somewhat difficult because the TCOF1 gene contains many very similar repeats in the middle of the gene. Often it was very difficult to determine which exon was missing in a gene. An example of one such problem is in the canine

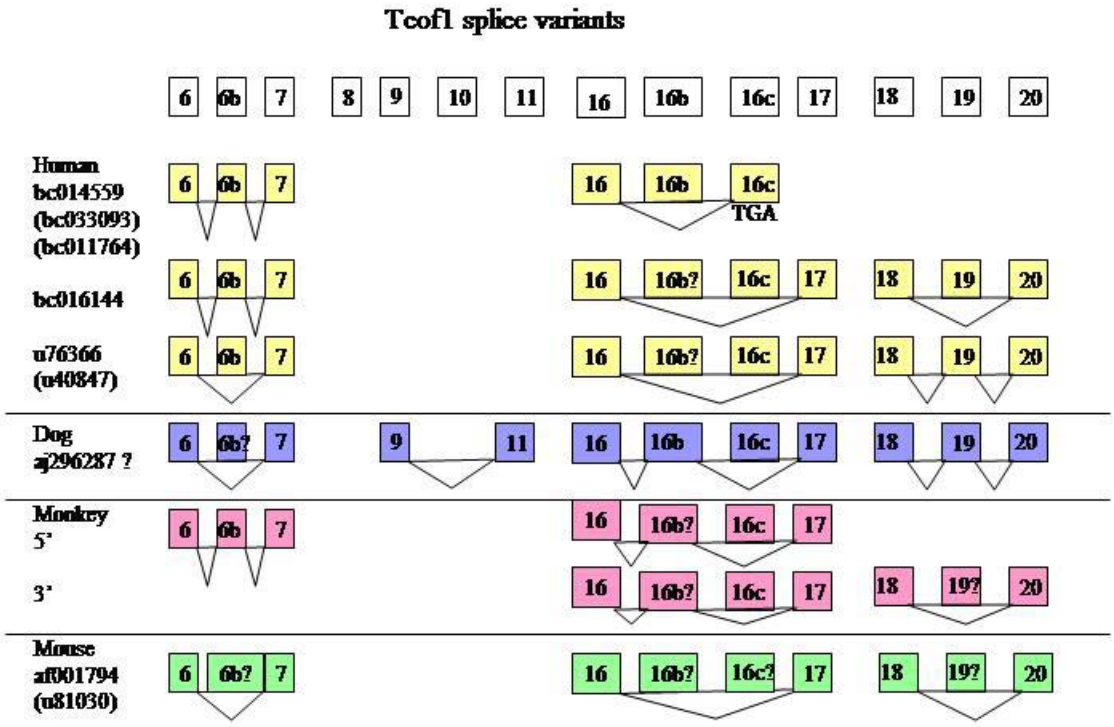

Figure 1 – Splice variants found from analysis of NCBI, UCSC Databases and unpublished data. Question marks indicate variants which were or will be searched for experimentally.

sequence. The canine is known not to have exon 10 (Haworth et al, 2001). However, the program wasn't able to account for this, especially because exon 11 is very similar. The program aligned canine exon 11 with human, mouse, and monkey exon 10 thus causing the rest of the sequence to be shifted. Due to these problems it was necessary to further analyze the aligned sequence by hand to be sure that gaps were placed where they belong. Once aligned, the missing gaps of each animal's sequence could be filled as best possible by using BLAST to search the EST database for the missing sequence.

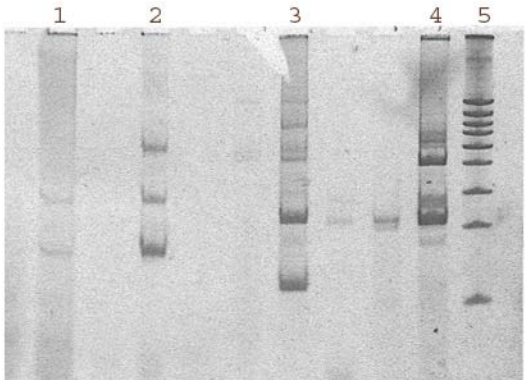

Figure 2 - RT-PCR Results Showing Splice Variants – Each band indicates a possible variant. 1: Mouse Liver (Region between exon 16 and 17 amplified), 2: Mouse Brain (Region between exon 16 and 17 amplified), 3: Monkey Spleen (Region between exon 18 and 20 amplified), 4: Monkey Spleen (Region between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

After developing the most complete cDNA sequence of the TCOF1 gene possible for each organism from the databases (figure 1) the second step was to begin looking for the missing sections of the sequence as well as splice variants

experimentally. Only tissue for the monkey and mouse was currently available. Exon 19 was not identified in any mouse or monkey transcripts in the databases. Exon 19 is known to be commonly spliced out of the gene (Haworth et al, 2001). Primers were developed and ordered which flank where exon 19 should be for both the mouse and monkey. Exon 16 also has alternative splice forms (16b and 16c). Exon 16b has only this far been found in the canine, so to determine if it is present in either the mouse or the monkey, primers were developed that go forward in exon 16 and reverse in exon 17. Exon 16c is found in the monkey but not in the mouse, so this would also help show whether exon 16c exists in the mouse. PCR was used with each of the primer sets on appropriate control cDNAs to find the correct conditions. After conditions were found for each set of primers, RT-PCR was performed using mouse RNA from both mouse brain and mouse liver tissues and RNA from monkey spleen. The RT-PCR revealed multiple products (splice variants) for each of the primers (figure 2) except exon 19 in the mouse (Figure 3). So it appears that the mouse does not have an exon 19 or possibly exon 19 is not expressed in the tissues tested. To isolate the individual fragments for cloning the RT-PCR products were inserted into pCR 2.1-topo vector which has its cloning site between Plac and lacZα in the plasmid. The cloning reaction was transformed into DH5-α *E. coli* cells. This made it possible to do blue-white selection so that the colonies which contained an insert would appear white on the plate and those that didn't would appear blue. Eight white colonies for each plasmid set were chosen from the plates at random and plasmid DNA was extracted by using the QuantumPrep Plasmid Miniprep kit. After the plasmid DNA was isolated the EcoR1 restriction enzyme was used in a digestion reaction to determine the size of the insert. Inserts which matched RT-PCR products were then sequenced to determine if they correspond to a specific exon. Results from the sequencing are currently being analyzed. From preliminary analysis it appears that exon 19 has been found in the rhesus monkey gene.

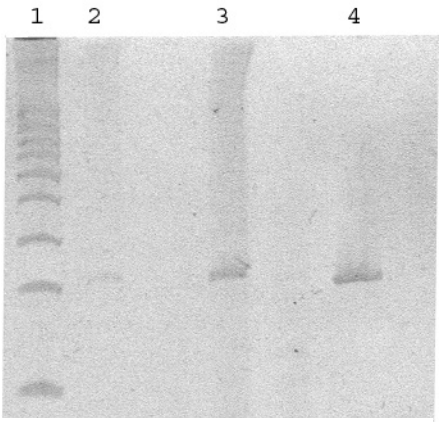

Figure 3 – RT-PCR Results Showing Splice Variants for Mouse Exon 19. Note only one band per lane, thus probably no splice variants. 1: DNA Marker Hyperladder IV (BioLine), 2:Mouse Brain (Region between exon 18 and 20 amplified), 2:Mouse Liver (Region between exon 18 and 20 amplified) 3: Control DNA strand

Goals for the Academic Year

During the academic year I hope to further analyze the results generated during the summer as well as identify additional splice variants. The canine and human gene needs to be analyzed much the way that the mouse and monkey genes were analyzed during the summer.

Plan for the Academic Year

1. **Experiments**

During the academic year I will continue the same work I have done during the summer with canine muscle tissue and human placenta tissue. Canine tissue was recently acquired by the Shiang Lab. First I will need to make RNA from the canine muscle tissue by using an RNA isolation kit. Next I will perform RT-PCR using various primers to look for splice variants in the canine muscle tissue. Once splice variants are found I will clone the RT-PCR product and isolate and sequence the inserts.

2. **Computation**
   During the academic year I will also use the NCBI and UCSC databases to continually search for updates and new sequences of different species being added to the database. New information from the databases may provide new insights into future projects. I will also use the BLAST program at NCBI to analyze the results from experimentation.

Budget

- RNA Isolation Kit: $207
- RT-kit: $205
- Taq polymerase (2 tubes) (for PCR): $220
- dNTPs: $160
- Topo-Cloning Kit: $341
- Lab Supplies (pipettes, tubes, etc.): $200

Works Cited

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.

Winokur, S.T. and Rita Shiang. "The Treacher Collins syndrome (TCOF1) gene product, treacle, is targeted to the nucleolus by signals in its C-terminus." Human Molecular Genetics. 1998. 7:1947-1952.