Abstract  In order to identify candidate genes for Down syndrome phenotypes or monogenic disorders that map to human chromosome 21q22.3, we have used genomic sequence and expressed sequence tags mapping to an autosomal recessive deafness (DFNB10) critical region to isolate a novel 2.5-kb cDNA that maps between TFF1 and D21S49. A semi-quantitative reverse transcription/polymerase chain reaction method revealed that UBASH3A gene expression is limited to only a few tissues, with its highest expression in spleen, peripheral blood leukocytes, and bone marrow. The putative 661-amino-acid protein shows considerable homology to a hypothetical protein from Drosophila melanogaster but only domain homologies to other organisms. Both the human and D. melanogaster proteins contain protein-protein interaction domains, viz., SH3 and ubiquitin-associated (UBA) domains, in addition to a novel domain also containing a nuclear localization signal. This is the first protein described containing both UBA and SH3 domains. The gene, thus called UBASH3A, spans 40 kb and is divided into 15 exons. Mutation analysis excluded UBASH3A as being responsible for DFNB10.

Introduction

An extra copy of chromosome 21 causes Down syndrome, a major genetic cause of mental retardation, which affects approximately one liveborn child in 700. To understand the molecular pathogenesis of Down syndrome and the molecular etiology of monogenic disorders that map to human chromosome 21 (HC21), it is necessary to identify all HC21 genes. The recent publication of the sequence of HC21 cataloged 127 known genes and will aid the rapid identification of the remaining transcripts (98 predicted genes; Hattori et al. 2000). Experimental documentation of full-length transcripts is of importance, since bioinformatic analyses do not always accurately predict the full genomic structure of a gene (Claverie et al. 1997) or provide detailed information about the pattern of expression, alternative splicing, and the 5' and 3' extremities of the transcript.

The autosomal recessive nonsyndromic deafness locus, DFNB10, has previously been localized to a 12-cM region near the telomere of chromosome 21 (21q22.3) in a large consanguineous Palestinian family (BT117; Bonne-Tamir et al. 1996). By using 50 polymorphic microsatellite markers in 21q22.3, the locus for DFNB10 has been refined to an area of less than 1 Mb between markers 1016E7.CA60 and 1151C12.GT45 (Berry et al. 2000). Six cDNAs that map to this critical region, viz., ABCG1 (also known as White; Chen et al. 1996; Croop et al. 1997; Hattori et al. 2000), TFF3 (Hauser et al. 1993), TFF2 (Tomasetto et al. 1990), TFF1 (Prud’homme et al. 1985), PDE9A (Guipponi et al. 1998), and NDUFV3 (de Coo et al. 1997), have been excluded as the gene responsible for DFNB10 (Berry et al. 2000).
The DFNB10 locus lies within 21q22.3, a gene-rich region that also contains unidentified loci for several other genetic disorders, including another autosomal recessive deafness locus (DFNB8; Veske et al. 1996), the developmental disorder holoprosencephaly (Muenke et al. 1995), and a locus for susceptibility to manic depressive psychosis or bipolar affective disorder (Straub et al. 1994; Smyth et al. 1997; Gurling 1998; Aita et al. 1999; Detera-Wadleigh et al. 1999).

Here, we report the isolation and characterization of a full-length transcript, UBASH3A, that maps between markers TFF1 and D21S49, within the refined DFNB10 critical region. Homologies to UBA, SH3, and a novel domain were found in the predicted amino acid sequence, and the gene was therefore called UBASH3A (approved by the Human Gene Nomenclature Committee, http://www.ucl.ac.uk/nomenclature/). This gene contains 15 exons over 40 kb. Mutation analyses in DNAs of affected members of the Palestinian family used to define DFNB10 (Bonne-Tamir et al. 1996) have excluded UBASH3A as the gene responsible for DFNB10. UBASH3A remains a candidate for other disorders mapping to 21q22.3 and for the development of some Down syndrome phenotypes.

Materials and methods

Genomic sequencing and the genomic structure of UBASH3A

The Laboratory of Genomic Medicine, Keio University School of Medicine undertook the sequencing of the regions of human chromosome 21q22.3 (http://www.dnb.med.keio.ac.jp/seqpub/map/APECED.html). Bacterial artificial chromosomes (BACs) KB169B4 and KB994G8 from the Keio BAC library (Asakawa et al. 1997), within the DFNB10 critical region, were sequenced by the shotgun method (Kawasaki et al. 1997). The genomic structure and intron-exon junctions of UBASH3A were determined by comparison of the genomic sequence with the cDNA sequence after isolation of the full-length transcript.

Cloning of the cDNA for the human UBASH3A gene

Detailed analysis of the genomic sequence was performed by using numerous computer programs, many being interfaced with NIX (Williams et al. 2000, http://www.hgmp.mrc.ac.uk/NIX/). Primers were designed to predicted exons, and reverse transcription/polymerase chain reactions (RT-PCRs) were performed on brain, fibroblast, placenta, and small intestine cDNA by using multiple primer pairs. Oligonucleotide names correspond to the predicted gene, exon number, and orientation. For example, g1ex3exL is a primer from the 3rd exon of gene model 1. Primer pairs used were: g1ex3L (5' GAAGTCTGGAGGAGAGCAAGGC 3') and g2ex4R (3' AGCTGCTCCAGGGTCTGTG 5') for 5'RACE, and g2ex10L (5' GAGCGTGCTGGTGTTCCGC 3') for 3'RACE. Nested amplifications were performed with primers AP2 and g2ex2eR (5' GCTGCCACTGAGGAAAGGC 3') and AP2 and g2ex11eL (5' CAGGCGCAAGCCTGATTT 3').

Rapid amplification of cDNA ends (5' and 3')

RACE (5' and 3' rapid amplification of cDNA ends) was performed on human placenta poly(A)+ RNA with the Marathon cDNA Amplification Kit (Clontech K1802−1). Double-stranded cDNA synthesis and adaptor ligation to the synthesized cDNAs were carried out according to manufacturers' instructions. A first round of PCR was performed with primers AP1 and g2ex4eR (5' ACGCTGCTCCAGGGTCTGTG 3') for 5'RACE, and AP1 and g2ex10eL (5' GAGCGTGCTGGTGTTCCGC 3') for 3'RACE. Nested amplifications were performed with primers AP2 and g2ex2eR (5' GCTGCCACTGAGGAAAGGC 3') and AP2 and g2ex11eL (5' CAGGCGCAAGCCTGATTT 3'). Products were purified and directly sequenced.

Amino acid sequence analyses

The sequence of the predicted protein encoded by the UBASH3A cDNA was analyzed by using Pix (http://www.hgmp.mrc.ac.uk/Registered/Webapp/pix/index.html), which integrates a large number of analyses with a graphic interface. The analyses included cell localisation prediction by PSORT; secondary structure predictions by DSC, Predator, SimpA96, and Phd. Blast searches against sequence databases (SPTR, NRL, 3D), Blast searches against domain databases (BASE, PRODOM), motif and domain databases searches (Pfam by using hmmpfam; PRINTS and BLOCKS by using blimps; PROSITE by using prosearch and pfscan), and coiled coil predictions by Coils and Transmembrane Predictions with Tmpred, Tmap, DAS, and Phd. Domain searches were also carried out by using the Pfam A and B databases (Bateman et al. 2000).

Expression pattern studies

RT-PCR with primers g2ex10eL (5' GAGCGTGCTGGTGTTCCGC 3') and g2ex12R (5' TGGCCCTGTGACACACAGG 3') were performed on a cDNA panel containing 20 different cDNAs from human tissues to obtain a semi-quantitative estimation of the expression of the gene as described (Michaud et al. 2000). Northern analyses were also performed on commercially available blots (OriGene no. HB-1020) containing 2 µg poly(A) RNA from six different human tissues by using a probe that was generated by RT-PCR between oligos g2ex10eL and g2ex12eR and that was labelled by PCR (Scott et al. 1995).

Mutation analysis of UBASH3A in DFNB10

DNA samples from an affected individual from a large consanguineous Palestinian family (BT117) used to describe DFNB10 (Bonne-Tamir et al. 1996) were investigated in the mutation analyses of UBASH3A. Each coding exon of UBASH3A was amplified out by using the Pfam A and B databases (Bateman et al. 2000). PCR products were examined by using gap4 (Staden 1996). The sequence of the predicted protein encoded by the UBASH3A cDNA was analyzed by using Pix (http://www.hgmp.mrc.ac.uk/Registered/Webapp/pix/index.html), which integrates a large number of analyses with a graphic interface. The analyses included cell localisation prediction by PSORT; secondary structure predictions by DSC, Predator, SimpA96, and Phd. Blast searches against sequence databases (SPTR, NRL, 3D), Blast searches against domain databases (BASE, PRODOM), motif and domain databases searches (Pfam by using hmmpfam; PRINTS and BLOCKS by using blimps; PROSITE by using prosearch and pfscan), and coiled coil predictions by Coils and Transmembrane Predictions with Tmpred, Tmap, DAS, and Phd. Domain searches were also carried out by using the Pfam A and B databases (Bateman et al. 2000).

Results

Isolation of the human UBASH3A cDNA

To identify genes within the critical region for DFNB10, we examined 74 kb of genomic sequence from BACs KB169B4 and KB994G8, which map between TFF1 and D21S49 (Fig. 1). GENSCAN predicted a total of 22 exons distributed over three genes. Three of the predicted exons (g1ex3, g2ex2, and g2ex4) showed identity to expressed...