Identification of a Novel Human Gene Containing the Tetratricopeptide Repeat Domain from the Down Syndrome Region of Chromosome 21

Miki OHIRA,1,* Akira OOTSUYAMA,2,† Emiko SUZUKI,1 Hitoshi ICHIKAWA,2 Naohiko SEKI,1 Takahiro NAGASE,1 Nobuo NOMURA,1 and Misao OHKI2

Laboratory of Gene Structure 1, Kazusa DNA Research Institute, 1532-3 Yanauchino, Kisarazu, Chiba 292, Japan 1 and Radiobiology Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104, Japan 2

(Received 22 January 1996)

Abstract

The Down syndrome (DS) region on chromosome 21, which is responsible for the DS main features, has been defined by analysis of DS patients with partial trisomy 21. Within the DS region, we constructed a 1.6-Mb P1 contig map previously. To isolate gene fragments from the 1.6-Mb region, we performed direct cDNA library screening and exon trapping using the P1 clones and a human fetal brain cDNA library, and obtained 67 cDNA fragments and 52 possible exons. Among them, 23 cDNA fragments and 4 exons were interpreted to be derived from a single gene by localization on P1 clones and by Northern analysis. To obtain the full-length cDNA sequence, longer cDNA clones were further screened from another human cDNA library which was enriched with longer cDNA species. These clones were sequenced and assembled to a sequence of 9045 bp. This transcribed sequence encodes a novel 2025 amino-acid protein containing tetratricopeptide repeat (TPR) motifs and therefore the gene was designated as TPRD (a gene containing the TPR motifs on the Down syndrome region). The TPR domain has been found in a certain protein phosphatase and in other proteins involved in the regulation of RNA synthesis or mitosis. The TPRD gene, the novel gene which was proved to be in the 1.6-Mb region and to have the interesting features described above, is a candidate for genes responsible for the DS phenotypes.

Key words: Down syndrome region; exon trapping; direct cDNA library screening; tetratricopeptide repeat

1. Introduction

Down syndrome (DS) is the most frequent birth defect (1 in 1000 newborns) caused by trisomy 21. Patients exhibit certain clinical features, such as mental retardation, congenital heart defect, and distinct facial and physical appearances. Although it is still unknown why the acquisition of an extra copy of chromosome 21 results in these complex phenotypes, studies of DS patients with partial trisomy 21 suggested the existence of an essential region for the pathogenesis of DS. This region is called the Down syndrome region. Now the region from D21S17 to ERG (approximately 2.5 Mb) is thought to contain the genes responsible for DS main features, although the association of other regions cannot be excluded.9 The cloning and characterization of genes involved in the region is a necessary step to understand the pathogenesis. As the first step, the several types of physical maps such as a radiation hybrid map,6,7 a YAC-STS map,8 a Not I restriction map,9,10 a ‘cosmid pocket’ map,11,12 and a high-resolution restriction map13 have already been well developed. Now the efforts toward the construction of transcription maps are advancing by collecting expressed-sequence tags (ESTs),14 cDNA selection,15–16 and exon trapping19,20 (for review, see ref. 21).

Within the 2.5-Mb region, we have focused on the more limited region of 1.6 Mb between LA68 and ERG (see Fig. 1a) from the analysis of a Japanese DS family with partial trisomy 21.22 This 1.6-Mb region was triplicated in most of the patients with partial trisomy 21 described above and therefore, may be associated strongly with the phenotypes commonly exhibited by
1.6 Mb region

**Figure 1.** a. cDNA clones and exon fragments mapped on the P1 BamHI fragments. Not I restriction map of human chromosome 21q22.2 region and DNA markers are shown on the top. Vertical bars show the Not I sites. The 1.6-Mb region we focused on is represented by a bracket. Below the Not I map, a BamHI restriction map and a P1 contig of PT1212, PT1601, and PS310 are shown. The location of cDNAs and exons are represented by open boxes and shaded boxes, respectively. The TPRD gene is shown by a thick arrow. b. Sequencing templates of TPRD cDNA. A closed box shows the open reading frame of TPRD gene. Relative locations of FB17-4, FB18-5, and kg-63 cDNAs are represented by shaded boxes. Exons contained in the 9-kb cDNA sequence are also shown.

these patients. We constructed a P1 contig map and a high-resolution BamHI map of the 1.6-Mb region, and subsequently started the gene isolation by direct cDNA library screening and exon trapping. Several genes including *hSIM*, *SLC5A3*, *Isk*, GABPA (E4TF1-60), ATP50, DSCR1, and *KATP-2* were reported to be contained in the 2.5-Mb region or to be candidate genes for DS. However, by using our map, we found that all of these genes are outside of the 1.6-Mb region except the *KATP-2* gene. We report here the cloning of the second novel gene which resides in the 1.6-Mb region and codes for a large protein containing the tetratricopeptide repeat (TPR) motifs.

2. Materials and Methods

2.1. Exon trapping

Exon trapping was carried out as suggested by a manufacturer (Gibco/BRL). Briefly, each of P1 clone DNA was Bgl II- and BamHI-digested, subcloned into vector pSPL3, and introduced into COS7 cells by electroporation. Total RNA was isolated from COS7 cells 24 h after transfection. Splicing products were amplified by RT-PCR with vector-specific primers, and subcloned into se-
quences, under the conditions described previously. To reduce the background hybridization by *Escherichia coli* chromosome DNA contamination, 5×10⁵ plaques of the library were screened by each of P1 clone [³²P]DNA with human placenta DNA for suppression of repetitive sequences, under the conditions described previously.

2.2. Direct cDNA library screening

A human fetal brain cDNA library was purchased from Stratagene. P1 clone DNA used as a probe was prepared by a standard alkaline lysis method and then purified by ethidium bromide-CsCl centrifugation⁹ to prepare the background hybridization by *Escherichia coli* chromosome DNA contamination. 5×10⁵ plaques of the library were screened by each of P1 clone [³²P]DNA with human placenta DNA for suppression of repetitive sequences, under the conditions described previously.⁹

2.3. Cloning and sequencing of TPRD cDNA

Construction of a size-fractionated cDNA library from a human immature myeloid cell line KG-1 was described previously.⁶ Several sublibraries enriched with longer sequences and, by hybridization to the BamHI map of the 1.6-Mb region between LA68 and ERG in the DS region, we identified two large transcripts of approximately 8 kb and 9 kb (see Fig. 3). It suggested the existence of a large gene flanking these three P1 clones.

2.4. Northern and Southern blot analyses

Insert DNA of the cDNA clone was labeled with [³²P]dCTP by random priming and used for Northern and Southern hybridization. Human multiple tissue Northern blots were purchased from Clontech.

2.5. Computer analysis

Homology search and other analyses were carried out with the GCG software package.⁴⁴

3. Results and Discussion

3.1. Isolation of gene fragments from the 1.6-Mb region of the Down syndrome region

We have constructed P1 contigs and a high-resolution BamHI map of the 1.6-Mb region between LA68 and ERG in the DS region.²² This map contains 46 P1 clones, among which 31 P1 clones are required for minimum tiling path. Using these 31 P1 clones, we carried out isolation of gene fragments from the region. Two approaches were applied; one is direct cDNA library screening and the other is exon trapping. For cDNA screening, we used a human fetal brain cDNA library. Exon trapping was carried out using the method of Buckler et al.³¹ To date, we obtained 100 cDNA clones and 80 unique trapped sequences and, by hybridization to the BamHI-digested P1 fragments, 67 cDNAs and 52 exons were mapped back to the P1 clones (unpublished). Among them, we found that many of the isolated cDNAs (23 clones) and 4 exons were mapped in the limited region of approximately the 100-kb range covered with three overlapping P1 clones PT1212, PT1601, and PS310 (Fig. 1a). Northern analysis using some of these clones commonly identified two large transcripts of approximately 8 kb and 9 kb (see Fig. 3). It suggested the existence of a large gene flanking these three P1 clones.

3.2. Determination of the complete cDNA sequence and expression analysis

To determine the full-sized sequence of 23 cDNA clones, single-pass sequencing from both ends of inserts and assembly of these end sequences were performed. However, the obtained sequences were smaller than the predicted transcript sizes. We then screened another cDNA library from a human immature myeloid cell line KG-1 to obtain a longer cDNA clone. This library was prepared from oligo(dT)-primed and size-fractionated cDNAs, and contained relatively longer cDNA species. Three positive clones were identified by hybridization with the insert of a fetal brain cDNA clone FB18-5 as a probe (Fig. 1a, b) and among them, a clone kg-63 had the longest insert of 6.5 kb. Single-pass sequencing and mapping data indicated that kg-63 contained most of the cDNA clones already identified (Fig. 1a) and extended to poly(A) tail. From these mapping data of all cDNA clones, we selected the three cDNAs FB17-2, FB18-5, and kg-63 (Fig. 1a) for compilation of a full-length cDNA sequence. The complete nucleotide sequences of FB17-2, FB18-5, and kg-63 produced a combined sequence of 9009 bp followed by a 36-bp poly(A) tail (Fig. 1b, 2a). It contains a single open reading frame of 6075 nucleotides, encoding a 2025-amino acid protein. The nucleotide sequence surrounding the predicted initiation codon at nucleotide
Figure 2. a. Nucleotide sequence and predicted amino acid sequence of the TPRD gene. This sequence was composed of FB17-4 (1-2311), FB18-5 (1267-3927), and kg-63 (3052-9045). The stop codon is denoted by an asterisk. The possible polyadenylation signals (7676, 8987) are underlined. Exon sequences E16-4, E16-8, and E18-1 are italicized. Three TPR motifs [(1), (2), and (3)] are also underlined. Three basic domains are shown in bold. b. Comparison of TPR motifs of TPRD with those of rat PPT, human IEF SSP 3521, and S. cerevisiae STI1. The 34 amino acid TPR unit was aligned with similar regions of proteins identified in a database search. IEF SSP 3521 and STI1 have two TPR domains. The position of the first amino acid of each TPR is given in parentheses. Residues identical to the TPRD are reverse-printed. Similarities of each protein against TPRD are: PPT (37%), IEF SSP 3521-(1) (31%), IEF SSP 3521-(2) (34%), STI1-(1) (31%), STI1-(2) (25%).
A Novel Human Gene in the Down Syndrome Region

1397 agreed well with the Kozak consensus sequence and contained an in-frame stop codon located 48 bp upstream. Thus, the predicted 5' and 3' untranslated regions (UTRs) are 1396 bp and 1577 bp long, respectively. Two consensus polyadenylation signals AATAAA were detected within the 3' UTR at nucleotides 7676 and 8987, the latter signal being located 17 bp upstream of the poly(A) sequence.

Expression patterns of this gene were examined in 16 human adult tissues and 4 fetal tissues by Northern blot analysis. The cDNA clone kg-63 identified two transcripts of about 8 kb and 9 kb (Fig. 3). Although these transcripts were seen in all tissues examined, the expression levels were slightly different in each tissue. For example, the 8-kb transcript was expressed higher in brain, testis, and colon. This suggested that the expression of each transcript is regulated, to some extent, in a tissue-specific manner.

Mapping of the cDNA clones on the P1 map indicated that this gene extends over 100 kb, more likely ranging from 109 kb to 128 kb, on the genome (Fig. 1a). In this region, four possible exons were mapped (Fig. 1a). Comparing their sequences with the 9-kb cDNA, three of them (E16-4, E16-8, and E18-1) were contained in the cDNA sequence (Fig. 2b). Although the other exon E18-10 (75 bp) was not presented in the 9-kb cDNA sequence, it also detected both of the 8-kb and 9-kb transcripts in all tissues by Northern analysis (data not shown), suggesting the existence of multiple alternatively spliced transcripts.

3.3. Predicted TPRD protein

The predicted protein (Fig. 2a) is quite hydrophilic as estimated from a Kyte-Doolittle hydropathy plot. The Chou-Fasman analysis indicates the presence of many α-helical regions. Database searching with BLASTX revealed a significant similarity to many unrelated proteins of diverse species, including a rat protein serine/threonine phosphatase PPT (P(N)=7.0e-15), a human heat shock and transformation-sensitive protein IEF SSP (P(N)=3.6e-12), and a stress-inducible protein from yeast STI1 (P(N)=1.6e-8). These proteins possess commonly the tetratricopeptide repeat (TPR) sequences in the regions of similarity (Fig. 2b). Three such imperfect 34 amino acid repeats were found in the NH2 terminal region of the predicted protein. Therefore, we designated this gene as TPRD (a gene containing the TPR motifs in the Down syndrome region). We found no further significant similarities in other regions except some human ESTs. Other members of this TPR family provide some hints as to the function of this novel gene. Saccharomyces cerevisiae STI1 and human IEF SSP 3521 are as well as a S. cerevisiae SSN6 and a SKI3, which are essential for normal growth, mating, and sporulation, may regulate transcription. Other proteins are highly associated with mitosis; S. pombe nuc2 is assumed to be a component of nuclear scaffold, S. cerevisiae CDC16 and CDC25 are involved in chromosome segregation and are essential for passage through the G2-M phase of the cell cycle, and Drosophila crooked neck (ern) gene is required for the proliferation of brain neuroblasts. A human protein phosphatase PP5 (a ho-
molog of a rat PPT) localizes to the nucleus, and may also be involved in RNA synthesis or mitosis. In addition to the TPR motifs, TPRD has three highly basic regions (residue 1018-1029, 1172-1185, and 1563-1579) which resemble the consensus of bipartite nuclear localization signal.\(^{47}\) Taken together, TPRD might have a nuclear function similar to that described above.

DS patients exhibit many complicated features and therefore, these features may be caused by multiple genes. Investigation of all genes in the DS region is important to elucidate the mechanism of DS. Several genes located on the DS region were reported so far. However, only one gene, ATP-sensitive K\(^+\) channel KATP-2,\(^{29}\) was included in the 1.6-Mb region. Therefore, TPRD, the second gene which was proved to be in the 1.6-Mb region and accompanied with interesting features described above, could be one of DS candidate genes. Further analysis is necessary to discuss the possible association between TPRD and Down syndrome.

Acknowledgments: We thank T. Kato, K. Sato, and S. Takahashi for sequencing analysis. We are also grateful to K. Ozawa, T. Tajino, and K. Yamada for technical assistance. This work was supported in part by Grants-in-Aid for Creative Basic Research (Human Genome Program) and Scientific Research on Priority Areas from the Ministry of Education, Science, and Culture; by a grant from the Special Coordination Funds for the Promotion of Science; and by grants from the Kazusa DNA Research Institute.

References

chromosome 21, *Genomics*, in press.


