FEBS Letters 489 (2001) 192–196 FEBS 24539

Structural organization, complete genomic sequences and mutational analyses of the Fukuyama-type congenital muscular dystrophy gene, *fukutin*[1]

Kazuhiro Kobayashi[a,b,2], Junko Sasaki[a,b,c,2], Eri Kondo-Iida[a], Yoji Fukuda[d], Moritoshi Kinoshita[d], Yoshihide Sunada[e], Yusuke Nakamura[f], Tatsushi Toda[a,b,*]

[a] *Laboratory of Genome Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*
[b] *Division of Clinical Genetics, Department of Medical Genetics, Biomedical Research Center, Osaka University Graduate School of Medicine, 2-2-B9 Yamadaoka, Suita, Osaka 565-0871, Japan*
[c] *Department of Pediatrics, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[d] *Gene Research Section, Gene Analysis Center, Otsuka Assay Laboratories, Otsuka Pharmaceutical Co., Ltd., 224-18 Ebisuno, Hiraishi, Kawauchi-chou, Tokushima 771-0195, Japan*
[e] *Department of Neurology, Kawasaki Medical School, 577 Matsushima, Kurashiki, Okayama 701-0192, Japan*
[f] *Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*

Received 27 September 2000; revised 27 December 2000; accepted 3 January 2001

First published online 15 January 2001

Edited by Veli-Pekka Lehto

Abstract Fukuyama-type congenital muscular dystrophy (FCMD) is an autosomal recessive severe muscular dystrophy in combination with cerebral cortical dysplasia. Previously, we identified the gene responsible for FCMD, termed *fukutin*, through positional cloning. In this study, we have sequenced 131 892 bp of genomic DNA in the region of the *fukutin* gene on chromosome 9q31 and obtained its complete genomic structure. The *fukutin* genomic sequence spans approximately 100 kb and is organized into 10 exons (41–6067 bp) and nine introns (1841–21 460 bp). Using these sequence data, we have identified three novel *fukutin* mutations in FCMD patients. We have also located a putative TATA box in the flanking 5′ region and identified numerous alternatively spliced *fukutin* mRNA transcripts. Analysis of expressed sequence tag clusters within the region revealed two novel genes upstream of the *fukutin* gene. These data provide fundamental information to support detailed genetic and functional analyses of the *fukutin* gene.

Key words: Fukuyama-type congenital muscular dystrophy; *fukutin* gene; Fukutin

1. Introduction

Fukuyama-type congenital muscular dystrophy (FCMD; MIM 253800) is an autosomal recessive disorder characterized by primary dystrophic changes in skeletal muscle associated with brain malformation, principally cerebral and cerebellar cortical dysplasia [1]. It is the second most common form of muscular dystrophy and one of the most common autosomal recessive diseases in the Japanese population. FCMD patients manifest generalized muscle weakness and hypotonia from early infancy, and most of them are incapable of standing or walking. In addition, some have seizures and all are mentally retarded. The predominant brain anomalies are micropolygyria, pachygyria and agyria. Ophthalmologic findings such as peripheral abnormalities of the retina or abnormal eye movements are often observed in FCMD patients. The course of the disease is slowly progressive and inexorable, and the majority of patients die during their second decade of life [1].

Through a positional cloning approach, we recently identified the gene responsible for FCMD, termed *fukutin* [2–7]. Most FCMD patients analyzed to date carry an ancestral mutation, which arose as a consequence of the integration of a 3 kb retrotransposon element into the 3′ untranslated region of the *fukutin* gene [7]. This ancestral FCMD mutation is the first known example of a human disease generated by an ancient retrotransposon insertion. In addition, two point mutations [6] and four other mutations [8] have been identified in rare FCMD alleles, while several other alleles show no mutations within the coding region.

The predicted protein product of the *fukutin* gene is a novel 461 amino acid protein that contains a potential N-terminal signal sequence. A fukutin fusion protein was shown to be exported from transfected cells [6]. While these observations suggest that fukutin may be an extracellular matrix component, its molecular function is not yet known.

In this study, we have determined the complete genomic sequence of *fukutin*[3] and obtained precise information pertaining to the organization and transcriptional regulation of *fu-*

*Corresponding author. Fax: (81)-6-6879 3389.
E-mail: toda@clgene.med.osaka-u.ac.jp

[1] The nucleotide sequence data reported in this paper have been submitted to the GenBank/EMBL/DDBJ nucleotide sequence databases with the accession number AB038490.
[2] These authors contributed equally to this manuscript.

Abbreviations: FCMD, Fukuyama-type congenital muscular dystrophy; SNP, single nucleotide polymorphism; EST, expressed sequence tag; ORF, open reading frame; WWS, Walker–Warburg syndrome

[3] Published as ancillary material in the web version.

PII: S0014-5793(01)02088-9

kutin that will facilitate characterization of additional mutations in FCMD patients.

2. Materials and methods

In the course of positional cloning of the *fukutin* gene, we constructed YAC and cosmid contigs encompassing the FCMD candidate region [5]. In the present study, we probed Southern blots with the *fukutin* cDNA sequence to select five overlapping cosmid clones, c2D-1, cJ5, cB12, cH1 and cE6, which constitute an approximately 130 kb contig covering the *fukutin* gene. Cosmid DNA was purified using Qiagen plasmid purification kits (Qiagen) or by equilibrium centrifugation in CsCl–ethidium bromide gradients. Cosmid DNA was sonicated for 0.2 s at 20 W in a UD-201 Sonicator (Tomy) on ice, then separated on a 0.8% agarose gel. Cosmid DNA fragments in the range of 2–5 kb were excised from the gel, purified by electro-elution, treated with T4 DNA polymerase and ligated into the *Eco*RV-digested pBluescript II vector. A shotgun-sequencing library was generated using XL1-Blue competent cells under chloramphenicol selection. More than 150 shotgun clones were sequenced using an AmpliTaq FS Dye Termination cycle sequencing kit with ABI Prism 377XL DNA Sequencer. Nucleotide sequences were assembled using the Phred/Phrap/Consed programs (http://www.phrap.org/). Gaps between the assembled segments were connected through further sequencing of linking subclones using primers designed from the end sequences of assembled segments [9].

Mutation identification was performed by direct sequence analysis using previously described primers and sequencing methods [8]. In brief, we PCR amplified and sequenced genomic DNA from each patient who showed heterozygosity for the 3 kb insertion mutation by Southern hybridization.

Total RNA was isolated from lymphoblasts using TRIzol Reagent (Gibco). mRNA of fetal brain, adult brain and skeletal muscle was obtained from commercial sources (Clontech). RNA was reverse-transcribed using random primers and Superscript II reverse transcriptase (Gibco). Primers 1F and 3R were used to amplify cDNA of the coding region of the *fukutin* gene. Primer sequences have been reported elsewhere [8].

Potential CpG islands were detected using GRAIL software (http://compbio.ornl.gov/Grail-1.3/). Analysis of upstream untranscribed sequence was carried out using the TFSEARCH, TESS, TSSG and NNPP programs (available at http://www.rwcp.or.jp/lab/pdappl/papia.html, http://www.cbil.upenn.edu/tess/index.html, http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html and http://www.fruitflynrg/seq_tools/promoter.html, respectively). Transcription start sites were also determined experimentally using the oligo-capping method [10]. The RepeatMasker program (http://ftp.genome.washingtonndu/RM/RepeatMasker.html) was used to screen DNA sequences for interspersed repeats known to exist in mammalian genomes, such as Alu and LINE 1. The BLAST program was used to determine sequence similarity with known genes and expressed sequence tag (EST) sequences using the non-redundant and dbEST compilations from GenBank, EMBL and DDBJ databases. Amino acid sequence similarity searches were carried out using FASTA and BLAST programs with non-redundant SwissProt and PIR databases.

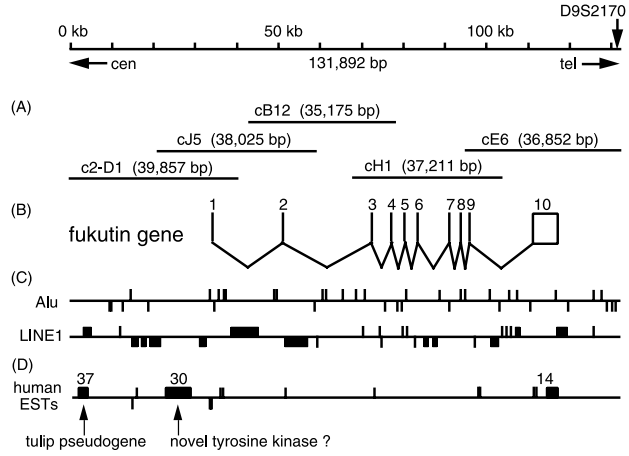

Fig. 1. Structural organization of the *fukutin* gene. The scale shown corresponds to the nucleotide sequence determined in this study. The position of the closest microsatellite, *D9S2170*, is indicated. (A) Cosmids sequenced covering a 131 kb region on chromosome 9q31 that contains the fukutin gene. (B) Organization of *fukutin*. Vertical bars and a box represent individual exons (see Table 1). (C) The locations of repetitive elements (Alu and LINE 1) in the region are shown, with those in the forward orientation above the lines and those in the reverse orientation below the lines. (D) Human ESTs identified in the region are also indicated. ESTs in the forward direction are above the line and ESTs in the reverse direction are below the line. Three clusters of ESTs were found. The number of ESTs constructing each cluster is also shown. One cluster corresponds to the 3′-region of the *fukutin* gene, another is likely to be a pseudogene ortholog of the rat *tulip 2* gene, and the third might represent a gene encoding a novel tyrosine kinase.

3. Results

The entire genomic sequence of the *fukutin* gene was determined from five cosmid clones: c2D-1, cJ5, cB12, cH1 and cE6 (Fig. 1A). The inserts of the individual cosmids were 39 857, 38 025, 35 175, 37 211 and 36 852 bp in size, respectively. Sequence data showed that the five cosmid clones covered a genomic region 131 892 bp in length[3].

The *fukutin* gene, distributed across approximately 100 kb, is comprised of 10 exons ranging in size from 41 to 6067 bp (Fig. 1B and Table 1). Exon 2 contains the upstream untranslated region, the initiation codon and the putative signal peptide. The introns in *fukutin* ranged in size from 1841 to 21 460 bp, and all had consensus AG/GT sequences at the exon–intron boundaries. Five exons displayed codon phase 0, and three codons were split at splice junctions.

Table 1
Exon–intron boundaries in the *fukutin* gene

Exon	3′ splice junction	Exon size (bp)	5′ splice junction	Intron size (bp)	Intron phase
1		41	GTTGAGTGAG/gtaaggtacg	16 777	UTR
2	ttgttcacag/AAAACAAAAT	193	ATCAACAAAG/gtaattttat	21 460	0
3	gtcttcctag/AATGGAGCTG	60	CACACAGTGG/gtatgtagaa	4 488	0
4	tctcaaacag/CGTGCAGTTA	204	GAAGAATGAG/gtaagtgact	2 866	0
5	ttggttctag/GAAGGCTGGT	278	CTTTTGACAG/gtaagttcag	3 330	2
6	ttgttttcag/GCCAGAGTTA	133	ATTCTTTCAG/gttagagaca	7 328	0
7	atatctgtag/CAGTACCTTG	130	ACTTGTCTAG/gtaaaattct	2 551	1
8	ttctttttag/GATGGTATCG	134	ATTTGGGAAG/gtcagtaaca	1 841	0
9	ttccccatag/GTAGAAGACA	128	AAAAATTCAA/gtatgaatca	14 987	2
10	tttgctgcag/ATACCTGTTT	6067			

Exon sequence is in uppercase letters and intron sequence in lowercase letters. The positions of introns between codons are indicated by phase 0, interruption after the first nucleotide by phase 1, and interruption after the second nucleotide by phase 2. UTR, untranslated region.

In addition to the FCMD families examined previously [2–8], we analyzed an additional 29 families for mutations in this study. Probands were homozygous in 19 families, and the remaining 10 were heterozygous for the 3 kb insertion mutation. Of these, three carried the previously reported C250T non-sense mutation in *fukutin* [7,8], and three others carried novel mutations: C718T non-sense, T1169A non-sense and A1223G missense mutations (Table 2). We identified a single nucleotide polymorphism (SNP) at base position 719, either A or G. The patient carrying the C718T transition was homozygous for G at position 719; thus, the C718T transition represents a non-sense mutation (CGA to TGA). To determine whether the A1223G substitution was a polymorphism or a disease-causing mutation, we tested 50 normal individuals for this change. We did not detect this change in any of the 50 controls (data not shown), indicating that the A1223G mutation is probably disease-causing. The remaining four patients who were heterozygous for the 3 kb insertion showed no alterations in the coding or the exon–intron boundary regions. For all compound heterozygous genotypes, patients' phenotypes were severe.

Examination of the flanking sequence 5′ to the *fukutin* gene, which included exon 1, revealed a relatively GC-rich sequence containing the *Bss*HII restriction site. A CpG island was predicted in this region by the GRAIL program (data not shown). Moreover, a portion of this region was previously cloned as a CpG island genomic fragment (Z57056) [11]. Therefore, the 5′ flanking sequence is likely to form a CpG island.

We analyzed a 360 bp sequence immediately upstream of the 5′ end of the *fukutin* cDNA for the presence of promoter sequences using computer programs (TFSEARCH, TESS, TSSG and NNPP). This sequence contained a number of putative transcription elements, including a TATA box and two potential transcription initiation sites (Fig. 2). Transcription start sites were also determined using the oligo-capping method with human colon mRNA [10], although the predicted start sites and the experimentally obtained sites were different. The roles of putative regulatory elements and the true sites of transcription initiation will require further evaluation.

Sunada et al. (submitted for publication) found that *fukutin* produced extensive alternative splicing products. Through RT-PCR amplification with primers specific for the *fukutin* open reading frame (ORF), we also identified a single major *fukutin* transcript along with multiple minor transcripts in this study. Transcripts showed identical expression patterns in fetal brain, adult brain, skeletal muscle and lymphoblast (Fig. 3A). Sequence analysis showed that the predominant transcript contained the entire *fukutin* coding region and that at least 18 minor transcripts were produced by alternative splicing (Fig. 3B). Most showed insertions of new exons and/or deletions of original exons, predicting amino acid changes and protein truncation. One lacked the initiating methionine codon of the *fukutin* gene and two had in-frame deletions.

```
CGTTTATCAG AAGTTCACAT TTAACGGGCT TGGAAAACGG -321
CTCAGCTGCT GCCTCCAGTT CATCTTCCGC TGTCGAAAGC -281
GTTTCAGAAC TTAAACTCTC AGATTGCCTC AGTGCGCTGC -241
TGTAGTTGTC AATTAGGCAA GAGATAAAAA AGGGCTTTGG -201
CTAAGTGGCT AGATTATAAA AGCGAGGAAT CGCCAGGGCT -161
CAGTGAACTT CTGGCGGTGA GGATGCGACA AGAGTCACAG -121
GAAAGCCGAG GCTGGGGCTG GACCGTGGGA ACGTCCCAGC -81
AGGAACCAAG CCAGCCTCCG CCGCTCAGCA CTTGACGGCG -41
GCTTTTGCGG CGGTTGCTGT GA*GTGCC*GTA* AAGCGGAGCG -1
GCTGCAGCCT GCTGTTGAGT GAGgtaaggt acgggaggct
```

Fig. 2. Nucleotide sequences in the 5′-flanking region of the *fukutin* gene. Exon 1 sequences reported previously are in boldface and intron 1 sequences in lowercase letters. Nucleotides are numbered on the right. Nucleotide position +1 is assigned to the first nucleotide of exon 1, as reported previously. Transcription initiation sites obtained by oligo-capping method are indicated by asterisks. The putative TATA box is underlined in boldface, and predicted sites for the transcription initiation are indicated in boldface italics.

We determined the copy number of repetitive sequences within the 132 kb region using the RepeatMasker program. This analysis revealed 39 copies of the Alu sequence (average of one copy per 3.4 kb) and 26 copies of the partial or full-length LINE 1 sequence (average of one copy per 5.1 kb) (Fig. 1C). A relatively high frequency of LINE 1 elements in this region is consistent with the location of the *fukutin* gene on the 9q31 G band. The interspersed repeats constitute as much as 42.89% of the DNA in this region. While such repeat sequences are thought to be associated with mutations involving gene deletion, no large deletions have been identified to date in FCMD patients. *D9S2170*, a previously identified polymorphic $(CA)_n$ repeat critical for founder-haplotype mapping of *fukutin*, is located distal to exon 10 (Fig. 1) [6].

Database searching using the genomic sequence revealed three EST clusters within the region (Fig. 1D). One cluster, consisting of 14 ESTs, corresponds to the 3′-region of the *fukutin* gene. A second cluster of 37 ESTs (GenBank: AI052024, AA115176, AA481144, etc.) shares extensive similarity with the rat *tulip 2* gene (GenBank: AF041107). However, the lack of both introns and continuous ORFs in the genomic sequence, suggests that this is likely a processed pseudogene. The third cluster, containing 30 ESTs (GenBank: AI084495, AA195108, N98696, etc.), most likely represents a gene encoding a novel tyrosine kinase. The putative translation of this sequence shares similarity with EPH-related receptor protein-tyrosine kinases (Swiss-Prot: P54753, P54760, P54762).

Table 2
Summary of *fukutin* mutations and a SNP found in this study

Nucleotide change	Coding effect	Amino acid change	Age/sex	Phenotype
Mutation:				
C718T	non-sense	R203X	3 years 7 months/female	severe
T1169A	non-sense	L353X	1 year 2 months/male	severe
A1223G	missense	T371C	0 year 2 months/male	severe
Polymorphism:				
A719G	non-synonymous	Q203R		

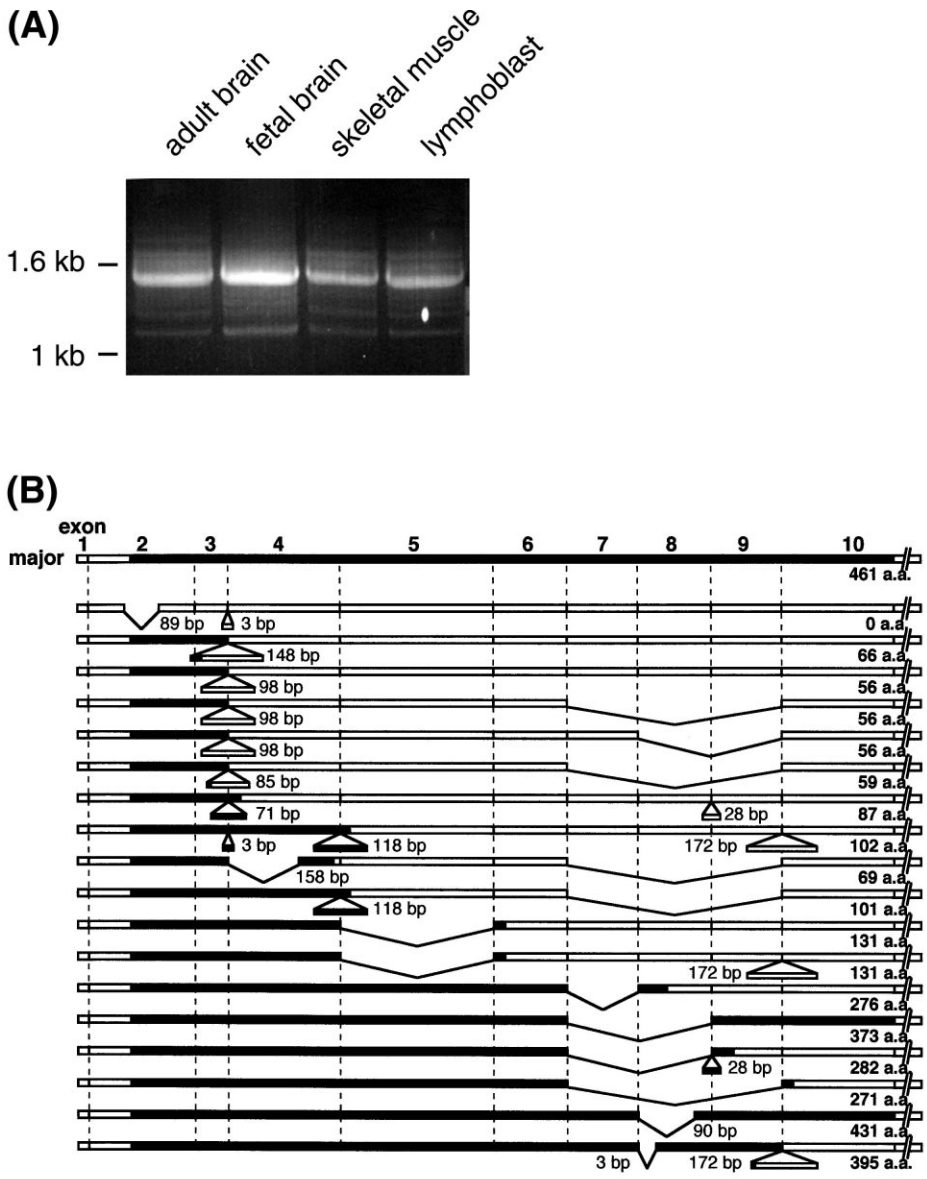

Fig. 3. Schematic representation of extensive alternative splicing of the *fukutin* gene. (A) Multiple transcripts detected by RT-PCR. (B) The major transcript is shown on the top. The black box represents the ORF of each transcript. The number of amino acids encoded is denoted on the right.

4. Discussion

The present study describes the structural organization and genomic sequencing of the *fukutin* gene[3]. These data represent a necessary step in the development of a DNA-based mutation screen for FCMD. This is particularly important for FCMD because of the wide spectrum of clinical variability, from relatively mild to severe. Walker–Warburg syndrome (WWS) and muscle–eye–brain disease are well known congenital muscular dystrophies that also involve central nervous system abnormalities and ocular malformations [12–15]. Although both syndromes are generally more severe than FCMD, it is sometimes difficult to distinguish severe FCMD from WWS clinically [16–18]. In a recent study, we used sequence data presented here to identify four novel mutations in rare alleles. Moreover, we found that compound heterozygotes for the founder mutation, including two patients with WWS-like phenotype, were clinically more severe than patients homozygous for the founder mutation [8,18].

In the present study, we have identified three additional novel mutations in compound heterozygous patients. These patients showed severe clinical phenotypes, supporting our previous observations. Thus, it may be possible to re-classify muscular dystrophies based on genetic rather than clinical criteria. Cataloguing deleterious mutations of the *fukutin* gene in patients should also highlight functionally important amino acid residues within fukutin protein, providing a basis for structure–function analyses.

The Japanese mother of the patient carrying the T1169 nonsense mutation carried the founder insertion chromosome,

while the Chinese father carried the non-sense mutation. This suggests that there may exist more patients or carriers in Asia than previously thought, although FCMD seems to be very rare in ethnic groups other than the Japanese population.

Several rare FCMD alleles show no mutation within the coding region of the *fukutin* gene. It is probable that the mutations in these alleles lie in regulatory regions such as promoter sequences or intronic sequences critical for alternative splicing. The *fukutin* gene was observed in this study to produce extensive alternative splicing transcripts: one major transcript carrying the entire coding region of *fukutin* and at least 18 minor transcripts. Although the biological significance of the minor transcripts is unknown, it is possible that they supplement the function of fukutin at the level of either mRNA or protein. Alternatively, it is possible that the expression levels of some transcripts increase, owing to the mutation in intronic sequences, and that the relative decrease of normal fukutin contributes to the disease. We have identified one FCMD-causative mutation resulting from the insertion of the LINE 1 repetitive element into intron 7. This insertion results in an increase of the amounts of minor transcripts lacking exon 7, 8 and/or 9 relative to the major transcript [8]. Using the genomic sequence of the *fukutin* gene, we are now able to search for other such mutations, analyze the regulation of *fukutin* expression and identify new genes within the region.

Acknowledgements: This work was supported by a Health Science Research Grant, 'Research on Brain Science' (H10-Brain-024), and by a Research Grant for Nervous and Mental Disorders (11B-1), both from the Ministry of Health and Welfare, Japan, and also by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] Fukuyama, Y., Osawa, M. and Suzuki, H. (1981) Brain Dev. 3, 1–29.
[2] Toda, T., Segawa, M., Nomura, Y., Nonaka, I., Masuda, K., Ishihara, T., Sakai, M., Tomita, I., Origuchi, Y., Ohno, K., Misugi, N., Sasaki, Y., Takada, K., Kawai, M., Otani, K., Murakami, T., Saito, K., Fukuyama, Y., Shimizu, T., Kanazawa, I. and Nakamura, Y. (1993) Nat. Genet. 5, 283–286.
[3] Toda, T., Ikegawa, S., Okui, K., Kondo, E., Saito, K., Fukuyama, Y., Yoshioka, M., Kumagai, T., Suzumori, K., Kanazawa, I. and Nakamura, Y. (1994) Am. J. Hum. Genet. 55, 946–950.
[4] Toda, T., Miyake, M., Kobayashi, K., Mizuno, K., Saito, K., Osawa, M., Nakamura, Y., Kanazawa, I., Nakagome, Y., Tokunaga, K. and Nakahori, Y. (1996) Am. J. Hum. Genet. 59, 1313–1320.
[5] Miyake, M., Nakahori, Y., Matsushita, I., Kobayashi, K., Mizuno, K., Hirai, M., Kanazawa, I., Nakagome, Y., Tokunaga, K. and Toda, T. (1997) Genomics 40, 284–293.
[6] Kobayashi, K., Nakahori, Y., Mizuno, K., Miyake, M., Kumagai, Y., Honma, A., Nonaka, I., Nakamura, Y., Tokunaga, K. and Toda, T. (1998) Hum. Genet. 103, 323–327.
[7] Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., Segawa, M., Yoshioka, M., Saito, K., Osawa, M., Hamano, K., Sakakihara, Y., Nonaka, I., Nakagome, Y., Kanazawa, I., Nakamura, Y., Tokunaga, K. and Toda, T. (1998) Nature 394, 388–392.
[8] Kondo-Iida, E., Kobayashi, K., Watanabe, M., Sasaki, J., Kumagai, T., Koide, H., Saito, K., Osawa, M., Nakamura, Y. and Toda, T. (1999) Hum. Mol. Genet. 8, 2303–2309.
[9] Daigo, Y., Isomura, M., Nishiwaki, T., Tamari, M., Ishikawa, S., Kai, M., Murata, Y., Takeuchi, K., Yamane, Y., Hayashi, R., Minami, M., Fujino, M.A., Hojo, Y., Uchiyama, I., Takagi, T. and Nakamura, Y. (1999) DNA Res. 26, 37–44.
[10] Maruyama, K. and Sugano, S. (1994) Gene 138, 171–174.
[11] Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P. (1994) Nat. Genet. 6, 236–244.
[12] Dobyns, W.B., Kirkpatrick, J.B., Hittner, H.M., Roberts, R.M. and Kretzer, F.L. (1985) Am. J. Med. Genet. 22, 157–195.
[13] Dobyns, W.B., Pagon, R.A., Armstrong, D., Curry, C.J., Greenberg, F., Grix, A., Holmes, L.B., Laxova, R., Michels, V.V., Robinow, M. and Zimmerman, R.L. (1989) Am. J. Med. Genet. 32, 195–210.
[14] Santavuori, P., Leisti, J. and Kruus, S. (1977) Neuropädiatrie 8, 553–558.
[15] Raitta, C., Lamminen, M., Santavuori, P. and Leisti, J. (1978) Acta Ophthalmol. 56, 465–472.
[16] Kondo-Iida, E., Saito, K., Osawa, M., Ishihara, T., Toda, T. and Fukuyama, Y. (1997) Brain Dev. 19, 181–186.
[17] Kondo-Iida, E., Saito, K., Tanaka, H., Tsuji, S., Ishihara, T., Osawa, M., Fukuyama, Y. and Toda, T. (1997) Hum. Genet. 99, 427–432.
[18] Saito, K., Osawa, M., Wang, Z.P., Ikeya, K., Fukuyama, Y., Kondo-Iida, E., Toda, T., Ohashi, H., Kurosawa, K., Wakai, S. and Kaneko, K. (2000) Am. J. Med. Genet. 92, 184–190.