BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **235,** 185–190 (1997)
ARTICLE NO. RC976758

Genomic Organization and Complete Nucleotide Sequence of the TMEM1 Gene on Human Chromosome 21q22.3[1]

Kentaro Nagamine,*,† Jun Kudoh,* Kazuhiko Kawasaki,* Shinsei Minoshima,* Shuichi Asakawa,* Fumiaki Ito,† and Nobuyoshi Shimizu*,[2]

**Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan; and †Department of Biochemistry, Faculty of Pharmaceutical Sciences, Setsunan University, Hirakata, Osaka 573-01, Japan*

Received May 6, 1997

TMEM1 (EHOC-1) gene encodes a putative transmembrane protein and is located on human chromosome band 21q22.3. Analysis of a 122,638-bp genomic sequence revealed that TMEM1 gene consists of 23 exons spanning approximately 94 kb and is transcribed in the direction of centromere to telomere. The 5′ region of the TMEM1 gene was associated with a CpG island and the 3′ end of the TMEM1 gene was mapped just proximal to the 5′ end of the neighboring gene PWP2. We determined that the TMEM1 gene encodes a protein of 1,259 amino acids, which is 69-amino acids longer than the previously reported sequence. Since TMEM1 gene is considered to be a candidate for genetic disorders mapped in the 21q22.3 region, the information including complete nucleotide sequence and genomic organization of the TMEM1 gene should be invaluable for the mutation analysis of the corresponding genetic disorders.

The TMEM1 (transmembrane protein 1; originally called EHOC-1) gene that encodes a putative transmembrane protein was isolated from human chromosome 21q22.3 (1). Within this chromosomal region, five hereditary disorders have been localized by genetic linkage studies. These include progressive myoclonus epilepsy of the Unverricht-Lundborg type (EPM1) (2), autoimmune polyglandular disease type I [also known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED)] (3), one form of bipolar affective disorder (4), one form of nonsyndromic hereditary deafness for which two loci DFNB8 (5) and DFNB10 (6) are independently mapped, and Knobloch syndrome that is characterized by a high degree of myopia, vitreoretinal degeneration with retinal detachment, macular abnormalities and occipital encephalocele (7). Furthermore, this region was included within the critical region (D21S113-21qter) for holoprosencephaly (HPE) (8). Although the cystatin B was recently determined as a responsible gene for EPM1 (9), the pathogenic genes for other disorders have not yet been identified. Thus, the TMEM1 gene can still be considered as an important candidate for the genes of these disorders.

We have constructed a cosmid/BAC (bacterial artificial chromosome) contig of 450 kb covering four markers D21S1460 (LJ104)-D21S25-PFKL-D21S154 and mapped the TMEM1 gene just proximal to a *Not* I site of linking clone LJ104 within the contig (10). Furthermore, we recently started sequencing of the 450-kb cosmid/BAC contig as one of the initial targets of our large-scale genomic DNA sequencing project of human chromosome 21. In this paper, we report the complete nucleotide sequence and genomic organization of the TMEM1 gene.

MATERIALS AND METHODS

DNA sequencing and analysis of nucleotide and amino acid sequence. A BAC clone KB86A5 containing the TMEM1 gene was isolated from a total human BAC library (11) and was assigned as a member of the 450-kb cosmid/BAC contig (Fig. 1) which covers a region containing markers D21S1460-D21S25-PFKL-D21S154

[1] Sequence data from this article is available through the Advanced Lifescience Information Systems (ALIS) project Web site (http://www-alis.jst-c.go.jp) of Japan Science and Technology Corporation (JST) and it has been deposited with the DDBJ/EMBL/GenBank DNA Libraries under Accession No. AB001523 (BAC KB86A5).

[2] To whom correspondence should be addressed. Fax: 81-3-3351-2370. E-mail: shimizu@dmb.med.keio.ac.jp.

Abbreviations: TMEM1, transmembrane protein 1; APECED, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (autoimmune polyglandular disease type I); BAC, bacterial artificial chromosome; EST, expressed sequence tag; UTR, untranslated region; PCR, polymerase chain reaction.

0006-291X/97 $25.00

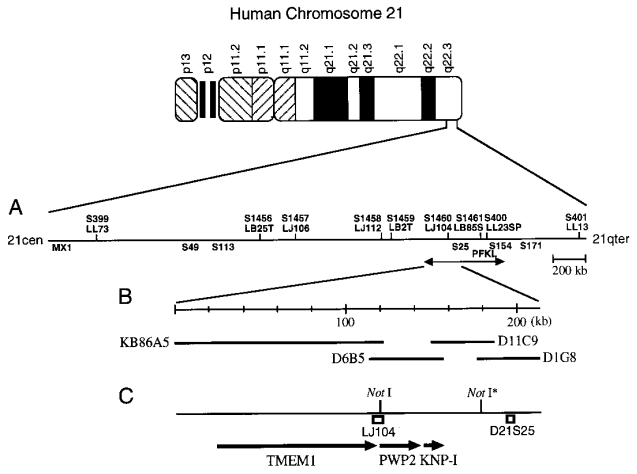

FIG. 1. Physical map of the TMEM1 gene locus. (A) A long-range *Not* I restriction map of the D21S49-D21S171 region of human chromosome 21q22.3 (18). Vertical bars show *Not* I sites with D number and name of linking clone. Locations of markers are also shown. A line with bidirectional arrows indicates the 450-kb cosmid/BAC contig we have constructed (10). (B) A cosmid/BAC contig containing the TMEM1 gene locus. KB86A5 is a BAC clone isolated from a total human BAC library (11) and others are cosmid clones isolated from a chromosome 21-specific KU21D cosmid library (10). (C) Transcript map of the TMEM1 gene locus. *Not* I* indicates a *Not* I site that is uncleavable in WAV17 DNA but cleavable in cloned cosmid DNA. Locations of the *Not* I linking clone LJ104 and the DNA marker D21S25 are shown as open boxes below line. The TMEM1 (1), PWP2 (14), and KNP-I (13) genes are shown by solid arrows.

(10). Genomic sequencing of BAC DNA was performed with the accuracy of over 99.98% using the shotgun method as described previously (12). The nucleotide sequence and amino acid sequence homologies were analyzed through the GenomeNet WWW server using BLAST search of the non-redundant database as described (13,14). Prediction of transmembrane segments from amino acid sequence was performed using SOSUI program (http://www.tuat.ac.jp/~mitaku/sosui).

PCR amplification of cDNA fragments from cDNA libraries. PCR amplification of cDNA fragments from various cDNA libraries was performed as described (10) using a pair of primers synthesized according to sequences of exon 22 of the TMEM1 gene (5′-ATCATTCTGCACACTCCTCCCAACT-3′; nt 116,027-116,051 of BAC KB86A5, DDBJ Accession No. AB001523) and an expressed sequence tag (EST) (N94245) (5′-CCTTCTCTCCACAACTTGCA-3′; complementary to nt 117,962-117,981 of BAC KB86A5).

RESULTS AND DISCUSSION

We have previously constructed a 450-kb cosmid/BAC contig covering markers D21S1460 (LJ104)-D21S25-PFKL-D21S154 and mapped the TMEM1 gene just proximal to a *Not* I site of linking clone LJ104 within the contig as shown in Fig. 1 (10). Both 5′ and 3′ end of TMEM1 cDNA were mapped within the BAC KB86A5 (10). Nucleotide sequence of a cosmid clone D6B5 (DDBJ Accession No. AB001517) and genomic organization of adjacent genes, PWP2 and KNP-I, were reported previously (13,14). We performed genomic sequencing of the BAC KB86A5 using the shotgun sequencing method with the accuracy of over 99.98% (12). A total insert DNA was determined to be 122,638 bp (DDBJ Accession No. AB001523). We found the entire sequence of the TMEM1 cDNA (1) and the 5′ end of the PWP2 cDNA (14-17) in the genomic sequence of the KB86A5 (Fig. 2). There is a CpG island that includes exon 1 of the TMEM1 gene and recognition sites of the rare-cutting enzymes such as *Bss*HII, *Eag* I, *Sac*II, *Nae* I, *Nar* I, and *Sma* I (Fig. 2). Just distal to the 3′ end of the TMEM1 gene, another CpG island is found to be associated with exon 1 of the PWP2 gene (14). The 5′-flanking sequence of TMEM1 gene has features of the housekeeping gene promoter, including lack of consensus TATA or CCAAT sequences and a high G+C content, and these characteristics are consistent with ubiquitous expression of the TMEM1 gene (1).

Comparison of the TMEM1 cDNA sequence (GenBank Accession No. U19252) with the nucleotide sequence of KB86A5 revealed that the TMEM1 gene consists of 23 exons and is transcribed in the direction of centromere to telomere (Fig. 2). Exon-intron boundary sequences are shown in Table 1. All introns are bordered by the consensus splice site sequence, gt . . . ag. These introns ranged from 310 bp to 19,530 bp in size. Furthermore, three expressed sequence tags (ESTs) (GenBank Accession Nos. N94245, L25499 and N65981) were found to locate between the 3′ end of TMEM1 cDNA and the 5′ end of PWP2 cDNA. These ESTs which together span 1,709-bp region (nt 117,929-119,637) on the genomic sequence, were found to locate only 102 bp downstream of the 3′ end (nt 117,827) of the TMEM1 cDNA. Based on their location and comparison of size of the TMEM1 cDNA (5.1 kb) and those of transcripts (8, 7.5, and 5.3 kb), these 3 ESTs were considered to be derived from the 3′-untranslated region (UTR) of the TMEM1 gene (1). To confirm whether these ESTs in fact belong to the 3′ UTR of TMEM1, we performed PCR amplification using various cDNA libraries as templates and a pair of primers which correspond to exon 22 of the TMEM1 gene and one of these ESTs. The expected size (1,635 bp) of the cDNA fragment is shorter than the size (1,955 bp) of PCR products from genomic DNA, because there is an intron 22 sequence (320 bp) between these two primers. The 1.64-kb cDNA fragment was amplified from various cDNA libraries including fetal brain, fetal liver, fetal heart, and fetal kidney (data not shown). This result indicated that these ESTs in fact belong to the 3′ UTR of TMEM1 gene and therefore exon 23 is extended further 1,810 bp, resulting in a total length of transcript 6,936 nucleotides (Table 1).

The nucleotide sequence of coding region and predicted amino acid sequence of TMEM1 protein are shown in Fig. 3. A comparison of our genomic se-

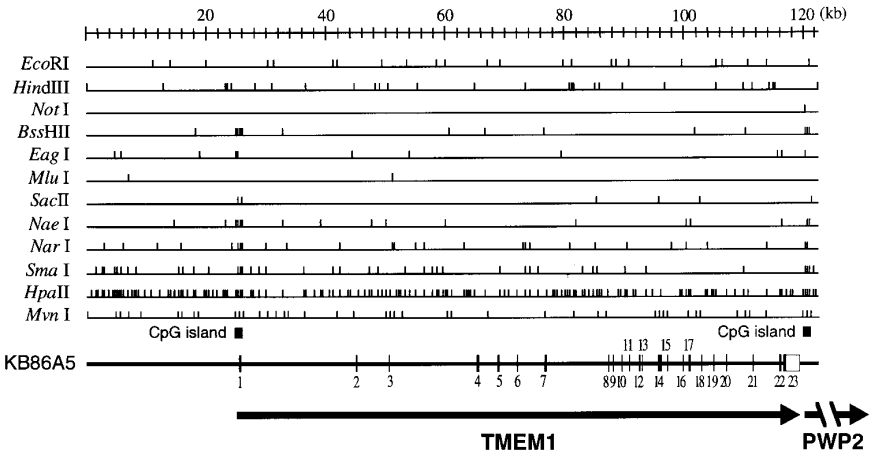

FIG. 2. Genomic structure of the TMEM1 gene. Restriction maps generated by the nucleotide sequence of BAC clone KB86A5 are shown. CpG islands are shown by closed boxes under restriction maps. Structure of the TMEM1 gene is schematically shown. Boxes represent exons in which coding regions are blackened. The TMEM1 gene is shown by a solid arrow.

quence with the TMEM1 cDNA sequence (U19252) revealed differences in 4 positions. Two differences were found in the 3′ UTR, insertions of 17 bp between nt 117,111 and 117,127 in KB86A5 and "A" at nt 117,342. The other two differences were in the coding region, causing amino acid changes. These are: a "G to T" nucleotide substitution at codon 813 in exon 16 causing amino acid change of Asp to Tyr and an insertion of "A" between nt 3,696 and 3,697 in codon 1,187 causing a frameshift mutation which replaces carboxy-terminal 4 amino acids of the reported TMEM1 protein (1) with novel 73 amino acids that are rich in Ser (15/73 residues) (Fig. 3). Genomic sequence of an overlapping cosmid clone D6B5 (AB001517) that was derived from different human DNA source is consistent with the sequence of BAC

TABLE 1

Exon-Intron Junctions of the Human TMEM1 Gene

3′-Acceptor site	Exon (bp)	5′-Donor site	Intron (bp)
(5′ end of U19252[a]) GCGGCGCAAC	1 (204)	ATCGTCACCTgtgagtgccc	1 (19530)
cccttcatagGTGCTGGAGA	2 (82)	AATGGAGAAGgtatgagttg	2 (5619)
ttgtttacagGTCCTATGGC	3 (136)	AGAGTGCTGTgtgagtacca	3 (14352)
ttgtttttagGATACCGAAG	4 (197)	AGAGTGACAGgtaagtgtat	4 (3221)
gtctttgcagGTGTGTTGTG	5 (196)	CATGGTTCAGgtacttgact	5 (3209)
gggcgaatagGAGGAGCTTG	6 (112)	GGGGCCGGGGgtgagtagtg	6 (4323)
gtttgtttagATGGTGCCAA	7 (248)	GCTCTTAGAAgtgagtcggc	7 (10538)
acccccacagGTCTCTGTCC	8 (147)	CACAGAAAAGgtgcctacct	8 (568)
atcttaacagTTAAAGTCCT	9 (115)	CCAGAAACAGgtacttttttt	9 (1398)
tctgtttaagCCAACACAGC	10 (77)	ACACTACTTAgtaagtatta	10 (1064)
ttccccacagGATTTGTCCC	11 (92)	AGTTTTACATgtaattgatt	11 (1779)
actttgcaagGAGGAAAAAG	12 (141)	AAATTGAAAAgtatccttta	12 (310)
tccgctccagCTACCTGCAG	13 (113)	GACAGCCCAGgtaagaccag	13 (2660)
ctatttacagGTCATAAGAT	14 (515)	CAGGACTCAGgtatgcgttc	14 (819)
ttgactctagGCCAAGGAAC	15 (142)	CCGCTGGCTGgtgagtgggg	15 (2538)
tttcctttagATAGCCTTCT	16 (159)	AACACGAGAGgtgaggtgcc	16 (738)
tccttcgtagAACAGTCTTC	17 (230)	AGACCACAAAgtgagtaggg	17 (1904)
tcctcctcagGTGTCGATTG	18 (101)	CAGGAACACGgtaacggagg	18 (1989)
tccactttagGAAATATGTT	19 (127)	GTCCCAGCAGgtaaacattg	19 (2013)
ctctttccagCCCATCTACA	20 (171)	AAATTTTTTTgtaagtgatg	20 (4123)
tcctgcccagACATTATACA	21 (178)	ATGTATGAAGgtaggtggtc	21 (4246)
ctcgtttcagTTGTCGACAA	22 (195)	CTGGACGCTGgtaaggactt	22 (320)
tgttttgcagACAGCTGGAT	23 (1448)	AAAGCGATGT (3′ end of U19252[a])	
	23 (3258)	TTTTAAGAAG (3′ end of N65981[a])	
	Total (6936)		Total (87261)

[a] U19252 and N65981 are the GenBank Accession Numbers of the TMEM1 cDNA (1) and EST, respectively.

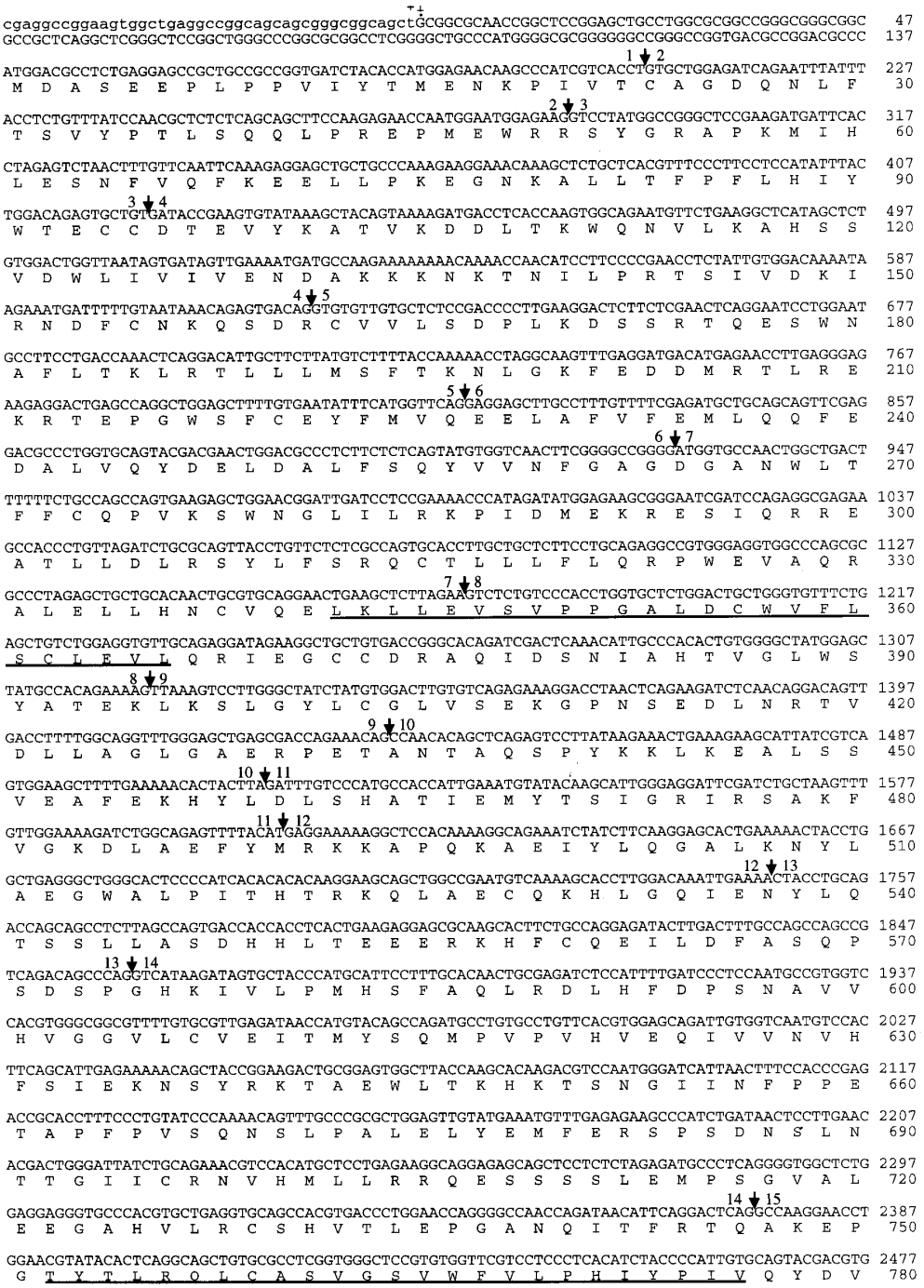

FIG. 3. Nucleotide and predicted amino acid sequences of the TMEM1 gene. Nucleotide position +1 corresponds to the 5′ end of the reported cDNA (GenBank Accession No. U19252). Nucleotides in the 5′ UTR and the coding region of the TMEM1 gene are indicated by uppercase letters. Lowercase letters indicate the 5′-flanking sequence. The positions of 22 introns are indicated by vertical arrows. Nucleotides different from the reported cDNA sequence (U19252) are indicated by bold letters (a "G to T" substitution at nt 2574 and an insertion of "A" between nt 3696 and nt 3697). Amino acids different from the reported 1190 amino acids (1) are indicated by bold letters under which revised amino acids are also shown. Amino acids that constitute the putative transmembrane regions are underlined.

KB86A5 in this region. These results indicate that the TMEM1 protein consists of 1,259 amino acid residues instead of 1,190 amino acid residues (1). Two regions of the TMEM1 protein (residues 342-366 and 752-776 in Fig. 3) were predicted as transmembrane domains by SOSUI program. Although the TMEM1 protein was proposed to have multiple putative transmembrane domains and have partial similari-

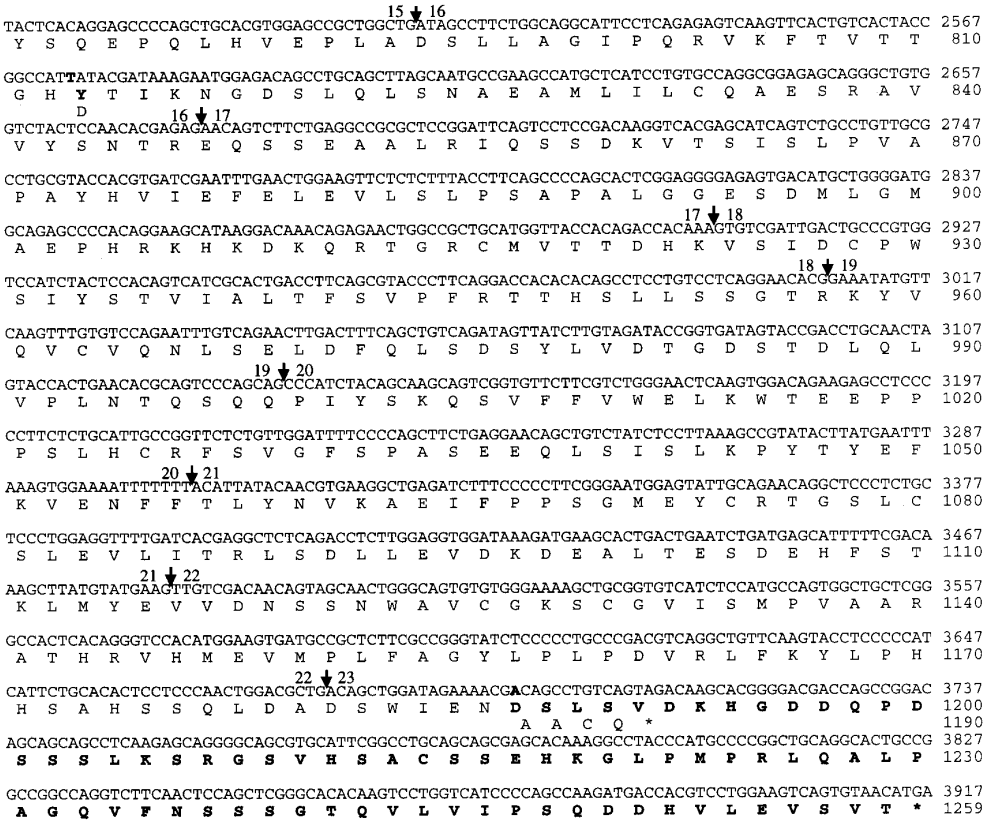

FIG. 3—*Continued*

ties with ion channel proteins (1), the exact function of TMEM1 protein remains to be elucidated.

In this study, we determined the genomic structure of TMEM1 gene and revised amino acid sequence of the TMEM1 protein. The information including nucleotide sequence and genomic organization of the TMEM1 gene should be invaluable for the mutation analysis of the corresponding genetic disorders.

ACKNOWLEDGMENTS

The authors thank Drs. E. Nakato and K. Shibuya, and all the members of genomic sequencing team in the Laboratory of Genomic Medicine for their contribution to this work. This work was supported in part by Fund for Human Genome Sequencing Project from the Japan Science and Technology Corporation (JST), and Grants in Aid for Scientific Research on Priority Areas and Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan.

REFERENCES

1. Yamakawa, K., Mitchell, S., Hubert, R., Chen, X.-N., Colbern, S., Huo, Y.-K., Gadomski, C., Kim, U.-J., and Korenberg, J. R. (1995) *Hum. Mol. Genet.* **4,** 709–716.
2. Lehesjoki, A.-E., Koskiniemi, M., Norio, R., Tirrito, S., Sistonen, P., Lander, E., and de la Chapelle, A. (1993) *Hum. Mol. Genet.* **2,** 1229–1234.
3. Aaltonen, J., Björses, P., Sandkuijl, L., Perheentupa, J., and Peltonen, L. (1994) *Nature Genet.* **8,** 83–87.
4. Straub, R. E., Lehner, T., Luo, Y., Loth, J. E., Shao, W., Sharpe, L., Alexander, J. R., Das, K., Simon, R., Fieve, R. R., Lerer, B., Endicott, J., Ott, J., Gilliam, T. C., and Baron, M. (1994) *Nature Genet.* **8,** 291–296.
5. Veske, A., Oehlmann, R., Younus, F., Mohyuddin, A., Müller-Myhsok, B., Mehdi, S. Q., and Gal, A. (1996) *Hum. Mol. Genet.* **5,** 165–168.
6. Bonné-Tamir, B., DeStefano, A. L., Briggs, C. E., Adair, R., Franklyn, B., Weiss, S., Korostishevsky, M., Frydman, M., Baldwin, C. T., and Farrer, L. A. (1996) *Am. J. Hum. Genet.* **58,** 1254–1259.
7. Sertié, A. L., Quimby, M., Moreira, E. S., Murray, J., Zatz, M., Antonarakis, S. E., and Passos-Bueno, M. R. (1996) *Hum. Mol. Genet.* **5,** 843–847.
8. Muenke, M., Bone, L. J., Mitchell, H. F., Hart, I., Walton, K., Hall-Johnson, K., Ippel, E. F., Dietz-Band, J., Kvalfy, K., Fan, C.-M., Tessier-Lavigne, M., and Patterson, D. (1995) *Am. J. Hum. Genet.* **57,** 1074–1079.
9. Pennacchio, L. A., Lehesjoki, A.-E., Stone, N. E., Willour, V. L., Virtaneva, K., Miao, J., D'Amato, E., Ramirez, L., Faham, M., Koskiniemi, M., Warrington, J. A., Norio, R., de la Chapelle, A., Cox, D. R., and Myers, R. M. (1996) *Science* **271,** 1731–1734.
10. Kudoh, J., Nagamine, K., Asakawa, S., Abe, I., Kawasaki, K., Maeda, H., Tsujimoto, S., Minoshima, S., Ito, F., and Shimizu, N. (1997) *DNA Res.* **4,** 45–52.
11. Asakawa, S., Abe, I., Kudoh, Y., Kishi, N., Wang, Y., Kubota, R., Kudoh, J., Kawasaki, K., Minoshima, S., and Shimizu, N. (1997) *Gene,* in press.

12. Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J. L., Wang, J., and Shimizu, N. (1997) *Genome Res.* **7,** 250–261.

13. Nagamine, K., Kudoh, J., Minoshima, S., Kawasaki, K., Asakawa, S., Ito, F., and Shimizu, N. (1996) *Biochem. Biophys. Res. Commun.* **225,** 608–616.

14. Nagamine, K., Kudoh, J., Minoshima, S., Kawasaki, K., Asakawa, S., Ito, F., and Shimizu, N. (1997) *Genomics,* in press.

15. Lalioti, M. D., Chen, H., Rossier, C., Shafaatian, R., Reid, J. D., and Antonarakis, S. E. (1996) *Genomics* **35,** 321–327.

16. Yamakawa, K., Gao, D.-Q., and Korenberg, J. R. (1996) *Cytogenet. Cell Genet.* **74,** 140–145.

17. Lafrenière, R. G., Rochefort, D. L., Chrétien, N., Neville, C. E., Korneluk, R. G., Zuo, L., Wei, Y., Lichter, J., and Rouleau, G. A. (1996) *Genome Res.* **6,** 1216–1226.

18. Ichikawa, H., Hosoda, F., Arai, Y., Shimizu, K., Ohira, M., and Ohki, M. (1993) *Nature Genet.* **4,** 361–366.