BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **240,** 239–243 (1997)
ARTICLE NO. RC977643

Cloning of Novel Trinucleotide-Repeat (CAG) Containing Genes in Mouse Brain[1]

Sun Jung Kim,*[,2] Bo Hwa Shon,* Joo Hyun Kang,† Kyung-Soo Hahm,†
Ook Joon Yoo,‡ Young Sik Park,§ and Kyung-Kwang Lee*

**Developmental Biology and †Peptide Engineering Research Unit, Korea Research Institute of Bioscience & Biotechnology, Taejon 305-333, Korea; ‡Department of Biological Sciences, BioMedical Research Center, Korea Advanced Institute of Science & Technology, Taejon 305-701, Korea; and §Department of Animal Science, Kyungpook National University, Taegu 702-701, Korea*

Received October 3, 1997

CAG trinucleotide repeat (CTR) sequence often appears in mammalian genome including transcription-regulatory protein and homeobox genes. Its expansion is associated with six genetic disorders in human. To identify novel CTR-containing genes expressed in mouse brain, a brain cDNA library was screened using an oligonucleotide, $(CTG)_{10}$. Eight clones were novel mouse genes and they were sequenced on both strands. The size of the cloned DNA ranged from 0.5 to 2.1 kb. The number of the CAG repeats in the clones ranged from 6 to 25. The inserts of the clones were analyzed for open reading frames and the peptide sequences were used for a GenBank homology search. Of the clones, one (CAG-6) shared 13 consecutive identical amino acid residues with the OB-cadherin gene, a member of cadherin family. CAG-14 showed high homology (657 nucleotides identity in 1022 nucleotides; 64%) with the 3′-untranslated region of rat leukocyte common antigen-related (LAR) tyrosine phosphatase receptor. All the 8 clones were originated from mouse DNA as judged by Southern blot analysis of mouse genomic DNA. The expression of the clones in mouse brain was addressed by RT-PCR and 4 clones showed specific expression.

Genes with trinucleotide repeats are frequently found in mammalian genome (1,2). Some of them are involved in neuropsychiatric disorders in human. Expansions of CTG triplet repeats in genes are discovered in patients with myotonic dystropy (3). Expansions of CAG triplet are discovered in six genetic disorders; Huntington's disease (4), spinocerebellar ataxia type 1 (5), dentatoribral-pallidoluysian atrophy Haw River syndrome (6), Machado-Josep disease (7), spinal and bulber muscular atrophy (8), and spinocerebellar ataxia type 2 (9). In mouse and *Drosophila*, genes with CAG or CAA repeats encoding the amino acid glutamine are called *opa* repeats and are frequently found in homeobox genes or other DNA binding protein-encoding genes (10,11). Trinucleotide repeat sequences are highly non-random in abundance and location in genes. CAG triplet repeat (CTR) sequence shows a dramatic enrichment in coding sequence (12). In addition, surveys of trinucleotide repeat found in GenBank indicate that the CTR sequences are not found in the intronic sequences. The bias of the CTR sequence to the exon region raises the possibility to isolate novel CTR-containing genes when a cDNA library is screened. A few methods to isolate genes with triplet repeat sequences (including expanded triplet repeat sequences) have been reported; positional cloning (13-15), a method using an antibody raised against polyglutamine (16), DIRECT (direct identification of repeat expansion and cloning technique) method (9), RRACE (random rapid amplification of cDNA ends) method (17), and screening of a cDNA or genomic DNA library (18). Human genes with CTR sequence have been screened extensively in brain (19), pancreatic islets (18), and Jurkat cells (17). However, no systematic studies to isolate mouse genes containing CTR sequences have been tried and relatively a small number of mouse genes with CTR sequence have been reported in GenBank compared to human genes. In this study, we adopted screening of a mouse brain cDNA library and cloned 8 novel genes containing CTR sequence.

[1] The DNA sequences described in this paper have been deposited with the GenBank Database under accession numbers U80888, U80889, U80890, U80891, U80892, U80893, U80894, and U80895.

[2] Correspondence to Sun Jung Kim: Developmental Biology Research Unit, Korea Research Institute of Bioscience & Biotechnology, Taejon 305-333, Korea. Fax: 82-42-860-4608. E-mail: sunjungk@kribb4680.kribb.re.kr.

Abbreviations: CTR, CAG trinucleotide repeat; RT-PCR, reverse-transcriptase polymerase chain reaction.

0006-291X/97 $25.00

MATERIALS AND METHODS

cDNA cloning. To identify CTR sequence-containing genes expressed in mouse brain, a brain cDNA library in λgt10 (Clontech) was screened with an oligonucleotide $(CAG)_{10}$. The oligonucleotides were labeled with DIG-11-ddUTP (Boehringer Mannheim) and terminal transferase. The library was plated at a density of 35,000 plaques per 150 mm plat. The filters lifted from plaques were hybridized with a hybridization buffer containing 5× SSC (1× SSC= 0.15M NaCl, 0.015M sodium citrate, pH7.0), 0.1% N-lauryl sarcosine, 0.02% SDS, and 1% blocking reagent (Boehringer Mannheim) at 60°C for 6 hrs. The filters were washed twice at 60°C in 1× SSC/0.1% SDS for 20 min. Colorimetric detection of positive plaques were performed by following the supplier's protocol (Boehringer Mannheim). Positive plaques were purified by two additional rounds of plating and probing.

Sequence analysis. The inserts in the λ DNA were amplified by polymerase chain reaction using Pwo DNA polymerase (Boehringer Mannheim) with λgt10 forward and reverse primers (Promega). To confirm that the positive clones have trinucleotide repeat sequences, the PCR products were subject to electrophoresis on a 1.0% agarose gel and Southern hybridization with the same oligonucleotide probe used for the cDNA cloning. The inserts of the clones showing positive signal were subcloned into the SmaI site of the pUC19 vector (New England Biolabs). Nucleotide sequences of both strands were determined in their entirety by dideoxy chain termination method (20) using synthesized oligonucleotide primers and Sequenase 2.0 kit (Amersham) and analyzed for open reading frames. The nucleotide and protein sequences were then used for homology search against GenBank data base.

Southern hybridization. Mouse genomic DNA was isolated from the liver as described previously (21). Six μg of DNA was digested with appropriate restriction enzymes and the fragments were separated on a 0.7% agarose gel. It was then transferred onto nylon membrane (Boehringer Mannheim) and hybridized to ^{32}P-labelled cDNA probe. The probe DNAs were prepared by PCR amplification of each insert DNA in the positive λ clones and labeled by nick translation. The oligonucleotide primers for the PCR were designed to bias the trinucleotide repeat sequence in the insert so that the probe could not hybridize to the non-specific CTR sequence in the mouse genome. In case the initial hybridization did not produce any specific signal, probe DNA from other region of the insert was prepared and used. Hybridizations were performed in the Quick Hybridization solution (Clontech) at 65°C for 4 hrs. The filters were then washed twice at 65°C in 0.1× SSC/0.1% SDS for 20 min and exposed to X-ray films for 12-16 hrs at −70°C.

RT-PCR. Mouse brain mRNA was purchased from Clontech Co. One μg of mRNA was reverse transcribed using oligo-dT primer and AMV reverse transcriptase (Promega). PCR was performed from 1 μl aliquot of the reaction using specific set of primers. The primer sets used were for CAG-6, 5′-CTGCCTAGAGCTCCTCATTTAGCCC and 5′-GTCTTTCACACCACCAC; CAG-9, 5′-TGATTAACAGATAGGTAAAGGTGAT and 5′-CTTGAGGTATTTAGCAC; CAG-14, 5′-TCTGCTCCTCCAGCTCC and 5′-GGAGGAACAGGAAGACC; CAG-18, 5′-CGTATCAAAGAATCAGC and 5′-TCATCAACCAGTTCTCC. PCR reaction went through 30 cycles at 94°C for 30 sec, 50°C for 30 sec, and 72°C for 30 sec.

RESULTS AND DISCUSSION

Isolation of CTR Sequence-Containing Genes in Mouse Brain

To identify novel mouse genes with CAG (alternatively CTG, the 5′ sequence on the other strand) trinucleotide repeats (CTR) expressed in the mouse brain, a λgt10 cDNA library of mouse brain was screened with a 30 mer oligonucleotide containing ten repeats of CTG triplet. The oligonucleotide was labeled with nonradioactive DIG-11-ddUTP using terminal transferase. About 3×10^5 plaques of the mouse brain cDNA library were screened and 34 (0.01%) were revealed positive at the final screening. The inserts in the DNA were amplified by PCR using Pwo DNA polymerase and they were subjected to Southern hybridization with the same oligonucleotide probe (data not shown). Of the positives, 11 clones showing strong signal were selected and subcloned into the pUC19 vector for sequence analysis. The size of the inserts ranged from 0.5 to 2.1 kb. The nucleotide sequences of the clones were deter-

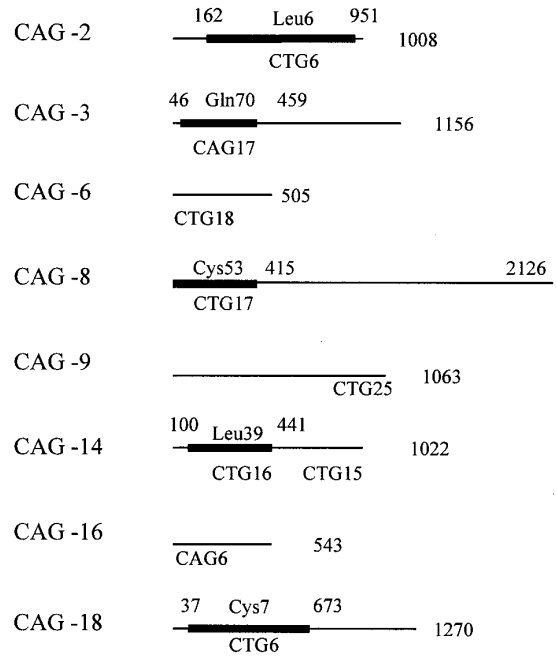

FIG. 1. Maps of the inserts of the CTR sequence-containing genes. To identify CAG triplet containing genes expressed in the mouse brain, a brain cDNA library in λgt10 (Clontech) was screened with oligonucleotide $(CTG)_{10}$ using standard technique (21). The oligonucleotides were labeled with DIG-11-ddUTP (Boehringer Mannheim) using terminal transferase. The filters lifted from plaques were hybridized with a hybridization buffer containing 5× SSC (1× SSC= 0.15M NaCl, 0.015M sodium citrate, pH 7.0), 0.1% N-lauryl sarcosine, 0.02% SDS, and 1% blocking reagent (Boehringer Mannheim) at 60°C. The filters were washed twice at 60°C in 1× SSC/0.1% SDS for 20 min. Colorimetric detection of positive plaques were performed by following the supplier's protocol (Boehringer Mannheim). The inserts of the positive clones were amplified by PCR and subcloned into the pUC19 vector (New England Biolabs). Nucleotide sequences were determined by sequencing both strands of the inserts. Tentative open reading frames and untranslated regions are indicated by thick and thin lines, respectively. The length of the clones as well as the first and last nucleotides of the open reading frames are numbered. The locations of triplet repeat are denoted by CAG or CTG with the frequency of unbroken repeat sequence and encoded amino acids.

TABLE 1

Characterization of the Novel CAG Repeat Containing Clones

Clones	Size (bp)	Type of repeat*	Repeat location	GenBank Accession No.
CAG-2	1008	$(CTG)_6$	Coding	U80888
CAG-3	1156	$(CAG)_{17}CAA(CAG)_3CAA(CAG)_8CAA(CAG)_9$ $TGC(CAG)_2TGC(CAG)_2TGCCAGCAACTG$ $(CAG)_3GAG(CAG)_{16}$	Coding	U80889
CAG-6	505	$(CTG)_{18}$		U80890
CAG-8	2126	$(CTG)_{16}CTC(CTG)_3CAGTTGCTGGCA$ $(CTG)_2GCA(CTG)_2GCA(CTG)_9TTG(CTG)_3$ $TTG(CTG)_{17}$	Coding	U80891
CAG-9	1063	$(CTG)_{25}$		U80892
CAG-14+	1022	$(CTG)_{16}CTC(CTG)_3CAGTTGCTGGCA$ $(CTG)_2GCA(CTG)_9TTG(CTG)_3$, $(CTG)_{15}TTC(CTG)_9CTCCTGCAGTTC(CTG)_6$	Coding 3′-UTR	U80893
CAG-16	543	$(CAG)_2CAA(CAG)_4CAA(CAG)_2CAA(CAG)_6$ $(CAA)/_2(CAG)_5CAACAGCAA(CAG)4CAA$ CAGCAACAG		U80894
CAG-18	1270	$(CTG)_6$	Coding	U80895

* CAG or CTG repeat sequences with more than two repeats are included.
† CAG-14 clone has two regions of CTG repeat separated by 295 nucleotides.

mined by sequencing both strands. It was revealed that three clones were mouse homologues of known human genes including two clones of AF-9 gene involved in common chromosome abnormality, t(9:11)(p22:q23) (22) and one clone of myopathy gene (23). Locations of the CTR sequence in the two homologous mouse and human genes were well conserved where it appeared in the coding region. The repeat number of the trinucleotides was same as 42 in the case of human and mouse AF-9 genes, whereas for the myopathy gene, it was 25 in mouse and 11 in human.

Sequence Analysis of the Novel CTR Sequence-Containing Clones

Eight clones were novel mouse genes, confirmed by BLAST searching of GenBank and finding no identical or close homology with any other entries. The inserts of the eight clones are presented schematically in Fig. 1. For five clones which showed enough length of open reading frames, their plausible long open reading frames and untranslated regions are indicated as thick and thin lines, respectively. The remaining three clones did not show appropriate open reading frame and indicated as thin lines. The DNA sequences have been submitted to GenBank and the accession numbers are indicated in Table 1. Trinucleotide repeat sequences can be classified as either simple (one tandem repeating triplet unit representing at least two-third of the total length) or cryptic (24). The number of the longest simple CAG repeats in our clones ranged from 6 to 25 (Table 1). The clone CAG-3 had 64 CAG repeats with a few base interruptions and the clone CAG-14 had 16 and 15 CTG repeats at two distinct regions. In 5 of 8 clones, the triplet repeats were in the long open reading frames although the definitive assignment of coding regions must await full-length clones. Amino acids encoded by these triplet repeats would include glutamine (1 clone), cysteine (2 clones), and leucine (2 clones). One clone had CTR sequence in the 3′-untranslated region. It is noteworthy that the amino acid sequences of CAG-6 deduced from the nucleotide sequence shared 13 consecutive identical residues located at just upstream region of the CTR sequence with the OB-cadherin gene, a member of cadherin family which is highly expressed in osteoblasts and weakly in brain (25). The nucleotide sequence of this region is 5′-GGCATAGAACTGTTT-GAAATC (T in case of OB-cadherin) ACAACAGAC-TATGAAACA-3′ and the amino acid sequence is Gly-Ile-Glu-Leu-Phe-Glu-Ile-Thr-Thr-Asp-Tyr-Glu-Thr. No other genes or sequences from the GenBank showed such a significant matches with the 13 amino acids. CAG-14 showed high homology (657 nucleotides identity in 1022 nucleotides; 64%) with the 3′-untranslated region of rat leukocyte common antigen-related (LAR) tyrosine phosphatase receptor (26). Both genes have triplet repeat sequences at two distinctive regions in common while the number of triplet repeat as well as the distance between the repeat regions are different. In rat LAR tyrosine phosphatase receptor, the CTR sequence is found in only one of five alternatively spliced transcripts. The presence of CTR in the clone CTG-14 may imply the expression of the specific transcript in mouse brain as well. However, whether they are homologues of rat and mouse genes should be deter-

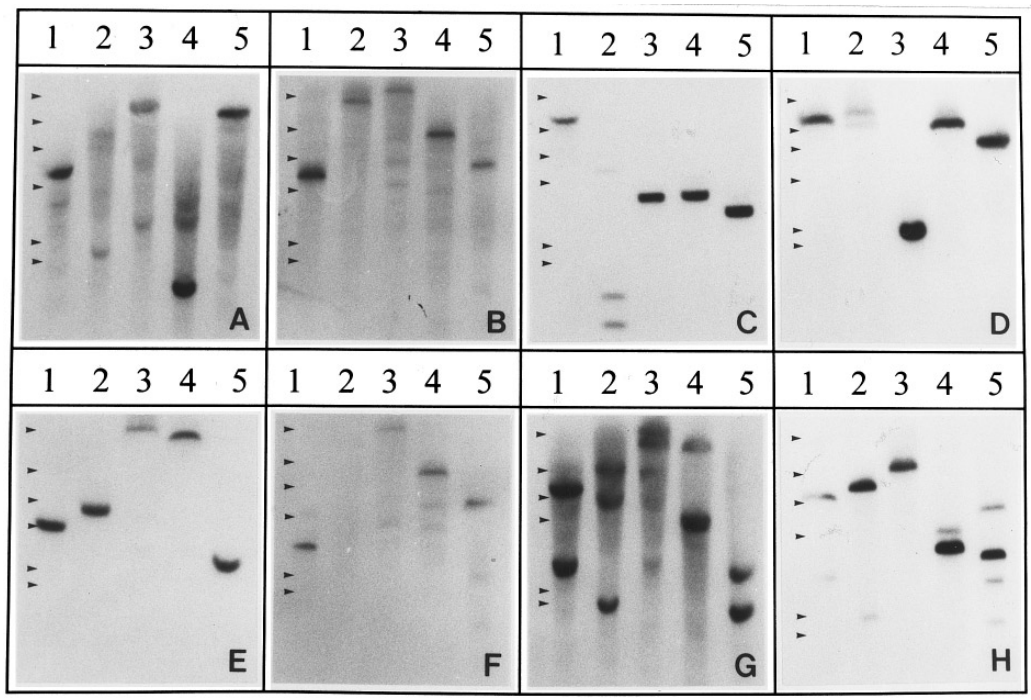

FIG. 2. Southern hybridization of mouse genomic DNA for the CTR sequence-containing genes. Mouse genomic DNA was isolated from the liver as described previously (21). Six μg of DNA was digested with restriction enzymes and the fragments were separated on a 0.7% agarose gel. It was transferred onto nylon membrane (Boehringer Mannheim) and hybridized to ^{32}P-labelled cDNA probe. Lanes 1-5 are DNAs digested with EcoRI, HindIII, BamHI, PstI, and BglII, respectively. Panels A-H are for CAG-2, 3, 6, 8, 9, 14, 16, and 18, respectively. The arrow heads at the left margin of each panel indicate λ DNA size marker digested with HindIII. Hybridizations were performed in the Quick Hybridization solution (Clontech) at 65°C for 3 hrs. The filters were then washed twice at 65°C in 0.1× SSC/0.1% SDS for 20 min and exposed to X-ray films for 12-16 hrs at −70°C.

mined after the full-length clone of CAG-14 including the coding region is obtained.

Genomic Southern Hybridization and RT-PCR Analysis

To determine that the cDNA clones are originated from mouse genes and to get information about their genomic structure, Southern blot analyses were performed for the mouse genomic DNA. Probes were prepared from PCR-amplified DNA fragments which do not contain the CTR sequence. Under high stringency, all the 8 novel clones detected single or two major bands from restriction enzyme-digested genomic DNA and in some cases, especially in case of CAG-18 clone, a few extra weak bands appeared, suggesting that the clones are likely to exist as a single copy genes and there may be also some related genes (Fig. 2). Interestingly, in CAG-3 and CAG-14 clones (Fig. 2B and F), it was found that four of the five restriction enzymes showed same size of positive bands implying that they are originated from the same gene. However, we could not find any extensive homology between the two genes except for the CTR region. So, they are thought to be from different fragments of the same gene.

RT-PCR was carried out to determine whether the cDNA clones were expressed in the mouse brain. Sets of oligonucleotide primers were designed complementary to the sequences of the cDNA clones. Of the eight clones, four clones showed correct-sized PCR products (Fig. 3). The same primer sets did not produced any specific product from the RNA itself even though in CAG-6 clone, there appeared a faint band presumably due to the contamination of genomic DNA (lane 2 of Fig. 3). This fact implies that many of the trinucleotide repeat containing clones newly identified in this study are expressed in the mouse brain tissue. For the four clones of which expression was not detected by the RT-PCR analysis, it could be considered that the amount of message was not enough to be detected, or the PCR condition was not adjusted.

In conclusion, with the nonradioactively labeled oligomer we could efficiently isolate trinucleotide repeat-containing cDNAs of mouse brain. We expect that more novel CTR sequence-containing genes could be obtained with this approach. The analyzed genes could be used to clone the full-length cDNA sequence. They could also be used to find their human homologues and to examine any polymorphism in the length of the trinucleotide repeat.

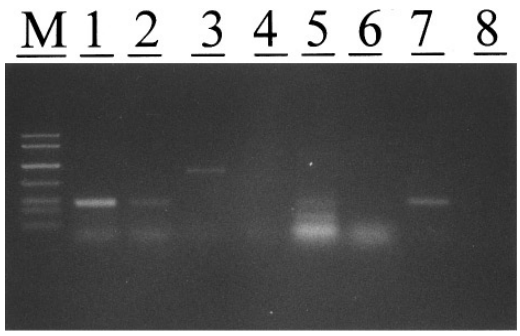

FIG. 3. RT-PCR analysis of mouse brain mRNA for the CTR sequence-containing genes. One μg of mouse brain mRNA (Clontech) was reverse transcribed using oligo-dT primer and AMV reverse transcriptase (Promega). PCR was performed from 1 μl aliquot of the reaction using specific set of primers. Lanes 1, 3, 5, and 7 are for CAG-6, 9, 14, and 18, respectively. Lanes 2, 4, 6, and 8 are PCR products from RNA *per se* for CAG-6, 9, 14, and 18, respectively. PCR reaction went through 30 cycles at 94°C for 30 sec, 50°C for 30 sec, and 72°C for 30 sec. Lane M is a DNA size marker; 1510, 1250, 870, 510, 330, 240, and 140 bp from the top band.

ACKNOWLEDGMENTS

This work was partly supported by the Korea Ministry of Science and Technology grant HS1100. We are grateful to Drs P.-L. Han and S. Kang for critically reviewing the manuscript.

REFERENCES

1. Tautz, D. (1989) *Nucleic Acids Res.* **17,** 6463–6471.
2. Ricke, D. O., Liu, Q., Gostout, B., and Sommer, S. S. (1995) *Genomics* **26,** 510–520.
3. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hey, K., Leblond, S., Earle-MacDonald, J., de Jong, P. J., Wieringa, B., and Korneluk, R. G. (1992) *Science* **255,** 1253–1255.
4. Huntington's Disease Collaborative Research Group (1993) *Cell* **72,** 971–983.
5. Orr, H. T., Chung, M., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. W., and Zoghbi, H. Y. (1993) *Nature Genet.* **4,** 221–226.
6. Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F., and Tsuji, S. (1994) *Nature Genet.* **6,** 9–13.
7. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S., and Kakizuka, A. (1994) *Nature Genet.* **8,** 221–228.
8. Bingham, P. M., Scott, M. O., Wang, S., McPhaul, M. J., Wilson, E. M., Garbern, J. Y., Merry, D. E., and Fischbeck, K. H. (1995) *Nature Genet.* **9,** 191–196.
9. Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., Wakisaka, A., Tashiro, K., Ishida, Y., Ikeuchi, T., Koide, R., Saito, M., Sato, A., Tanaka, T., Hanyu, S., Takiyama, Y., Nishizawa, M., Shimizu, N., Nomura, Y., Segawa, M., Iwabuchi, K., Eguchi, I., Tanaka, H., Takahashi, H., and Tsuji, S. (1996) *Nature Genet.* **14,** 277–284.
10. Wharton, K. A., Yedvobnick, B., Finnerty, V. G., and Artavanis-Tsakonas, S. (1985) *Cell* **40,** 55–62.
11. Duboule, D., Haenlin, M., Galliot, B., and Mohier, E. (1987) *Mol. Cell. Biol.* **7,** 2003–2006.
12. Stallings, R. L. (1994) *Genomics* **21,** 116–121.
13. Pulst, A.-M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.-N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., DeJong, P., Rouleau, G., Auburger, G., Korenberg, J., Figueroa, C., and Sahba, S. (1996) *Nature Genet.* **14,** 269–276.
14. Allotey, R., Twells, R., Cemal, C., Norte, B. S., Weissenbach, J., Pook, M., Williamson, R., and Chamberlain, S. (1995) *Am. J. Hum. Genet.* **57,** 185–189.
15. Gispert, S., Lunkes, A., Santos, N., Orozco, G., Ha-Hao, D., Ratzlaff, T., Aguiar, J., Torrens, I., Heredero, L., Brice, A., Cancel, G., Stevanin, G., Vernant, J.-C., Durr, A., Lepage-Lezin, A., Beial, S., Ben-Hamida, M., Pulst, S., Rouleau, G., Weissenbach, J., LePasleir, D., Kucherlapati, R., Montgomery, K., Fukui, K., and Auburger, G. (1995) *Am. J. Hum. Genet.* **57,** 972–975.
16. Imbert, G., Saudou, S., Yvert, G., Devys, D., Trottier, Y., Garnier, J.-M., Weber, C., Mandel, J.-L., Cancel, G., Abbas, N., Durr, A., Didierjean, O., Stevanin, G., Agid, Y., and Brice, A. (1996) *Nature Genet.* **14,** 285–291.
17. Carney, J. P., McKnight, C., VanEpps, S., and Kelley, M. R. (1995) *Gene* **155,** 289–292.
18. Aoki, M., Koranyi, L., Riggs, A. C., Wasson, J., Chiu, K. C., Vaxillaire, M., Froguel, P., Gough, S., Liu, L., and Donis-Keller, H., and Permutt, M. A. (1996) *Diabetes* **45,** 157–164.
19. Li, S.-H., McInnis, M. G., Margolis, R. L., Antonarakis, S. E., and Ross, C. A. (1993) *Genomics* **16,** 572–579.
20. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
21. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular cloning: A laboratory manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., Kamada, N., Gale, R. P., Lange, B., Crist, W. M., Nowell, P. C., Croce, C. M., and Canaani, E. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4631–4635.
23. Laporte, J., Hu, L. J., Kretz, C., Mandel, J. L., Kioschis, P., Coy, J. F., Klauck, S. M., Poustka, A., and Dahl, N. (1996) *Nature Genet.* **13,** 175–182.
24. Jacobson, D. P., Schmeling, P., and Sommer, S. S. (1993) *Am. J. Hum. Genet.* **53,** 443–450.
25. Okazaki, M., Takeshita, S., Kawai, S., Kikuno, R., Tsujimura, A., Kudo, A., and Amann, E. (1994) *J. Biol. Chem.* **269,** 12029–12098.
26. Zhang, J. S., and Longo, F. M. (1995) *J. Cell Biol.* **128,** 415–431.