Characterization and organization of DNA sequences adjacent to the human telomere associated repeat $(TTAGGG)_n$

Bernhard Weber*, Colin Collins, Carolyn Robbins, R.Ellen Magenis[3], Allen D.Delaney[1], Joe W.Gray[2] and Michael R.Hayden
Department of Medical Genetics and [1]Biomedical Research Centre, University of British Columbia, Vancouver V6T 2B5, Canada, [2]Lawrence Livermore National Laboratories, University of California and [3]Department of Medical Genetics, Oregon Health Sciences University, Portland, OR, USA

Received December 18, 1989, Revised and Accepted April 25, 1990

ABSTRACT

We present a strategy for the cloning of DNA sequences adjacent to the tandemly repeated DNA sequence $(TTAGGG)_n$. Sequence analysis of 14 independently isolated clones revealed the presence of non-repetitive sequences immediately adjacent to or flanked by blocks of the simple repeat $(TTAGGG)_n$. In addition, we provide sequence information on two previously undescribed tandemly repeated sequences, including a 9 bp repeat and a modification of the $(TTAGGG)_n$ repeat. Using different mapping approaches six sub-clones, free of the TTAGGG repeat, were assigned to a single human chromosome. Moreover, in situ hybridization mapped one of these subclones, G2 – 1H, definitively to the telomeric band on chromosome 4q. However, Bal 31 insensitivity suggests a location in a more subterminal region. All the $(TTAGGG)_n$-adjacent unique sequences tested are highly conserved among primates but are not present in other mammalian species. Identification and mapping of TTAGGG-adjacent sequences will provide a refined insight into the genomic organization of the $(TTAGGG)_n$ repeat. The isolation of chromosome specific TTAGGG-adjacent sequences from subtelomeric regions of all human chromosomes will serve as important end points for the genetic maps and will be useful for the molecular characterization of chromosomal rearrangements involving telomeres.

INTRODUCTION

The ends of all eucaryotic chromosomes are capped by telomeric structures which are composed of tandem reiterations of simple repeats (1). These simple repetitive sequences have a G- and C-rich strand with a 5′ to 3′ orientation of the G-rich strand toward the telomere (2). The common structural features likely confer upon the telomeres the capacity for replication (3–5) meiotic pairing (6) and stability (7,8).

Recently a tandemly repeated DNA sequence $(TTAGGG)_n$ has been cloned and shown to reside in major clusters at the ends of all human chromosomes (9). This hexanucleotide sequence is identical to the tandem repeat sequence previously identified at the telomeres of trypanosomes (10) and has recently been shown to be highly conserved among 91 different species of vertebrates including representative orders of bony fish, reptiles, amphibians, birds and mammals (11). In addition a human telomerase activity that catalyzes the addition of $(TTAGGG)_n$ repeats onto the 3′ ends of telomeric oligonucleotide primers has been characterized (12) and provides further evidence that tandem arrays of the $(TTAGGG)_n$ repeat constitute the extreme ends of all human chromosomes. Southern hybridization experiments identified two additional G-rich repeats $(TTGGGG)_n$ and $(TGAGGG)_n$ and suggested that these sequences reside at most or all human telomeres with a non random distribution (13). More recently, the successful cloning of human telomeres in modified yeast artifical chromosome (YAC) vectors has been reported (14,15,16). However, more detailed analysis including sequence data from TTAGGG-adjacent sequences has not yet been reported.

In this manuscript we describe a new strategy for the cloning of sequences adjacent to the human telomeric repeat $(TTAGGG)_n$. This method employs one synthetic oligonucleotide primer derived from the human telomeric repeat $(TTAGGG)_n$ in a polymerase mediated primer extension reaction. Priming occurs within the block of telomeric repeats and can extend into adjacent non-$(TTAGGG)_n$ sequences. We provide detailed sequence analysis confirming the presence of non-$(TTAGGG)_n$ sequences adjacent to telomeric repeats. A combination of different mapping techniques has localized six of the unique sequences to different human chromosomes. Fine

* To whom correspondence should be addressed at Department of Medical Genetics, University Hospital-UBC site, Room A68-2211 Wesbrook Mall, Vancouver, BC V6T 2B5, Canada

mapping by in situ hybridization assigned one clone to the telomeric region of 4q. Bal 31 experiments suggest a location of at least 30 kb from the end of the chromosome.

MATERIALS AND METHODS

Strategy for cloning sequences adjacent to the human telomeric repeat $(TTAGGG)_n$

In an effort to isolate sequences adjacent to human telomeric repeats on chromosome 4 a variation of the polymerase chain reaction (PCR) was applied to flow purified chromosomes. The strategy outlined in figure 1 involves rounds of Taq-polymerase mediated linear amplifications with either primer $(TTAGGG)_4$ or $(CCCTAA)_4$. These primer sequences were derived from cloned human repeats which have been shown to reside in major clusters at all human telomeres (9). The synthetic oligonucleotides were designed to exploit the inverse polarity of the G- and C-rich strands known to exist in lower eucaryotic telomeres and hypothesized to be conserved in human. Theoretically, 2 types of sequences can be expected. Priming should occur in the telomeric sequences and the primer can either extend within contiguous blocks of the repeat or into adjacent subtelomeric sequences. The single-stranded extension products were made double-stranded by Klenow polymerase-mediated randomly primed synthesis (17). T4 polymerase blunt-ending and ligation of Eco RI adaptors resulted in fragments which were cloned into the bluescript vector pBT-KS(+)(Stratagene).

DNA sources

Dual-beam high-speed sorting (18) has been used to sort approximately 500 human chromosome 4 per vial from a human × hamster cell line (UV 20 HL 21−27) containing human chromosomes 4,8 and 21. Initial experiments were done with human genomic DNA which was sheared by passing through a 25 gauge needle.

Primer extension reaction

Human genomic DNA or flow-sorted chromosome 4 (approximately 500 chromosomes per reaction) served as templates throughout the rounds of amplification. With one primer only $(TTAGGG)_4$ or $(CCCTAA)_4$, 40 pmole/100 μl reaction), 100 μM of each dNTP and 2 units of Taq polymerase in 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, 0.01% gelatin, the telomeric repeat was extended in 30 (genomic DNA) or 60 cycles (flow sorted chromosomes) of 94°C denaturation (1 min) 51°C annealing (1 min) and 72°C extension (1 min).

Second strand synthesis

The reaction product was boiled for 2 min then chilled on ice. The randomly primed second strand synthesis was done as previously described but with all four nucleotides at 100μM and 10 μCi ^{32}P-dCTP for monitoring the DNA yield and size.

Blunt-ending reaction

The reaction mix was heated to 70°C for 10 min and subsequently chilled on ice. After addition of 10 units of T4 DNA polymerase, the mix was incubated at 37°C for 10 min. The enzyme was inactivated by adjusting the mix to 20mM EDTA.

Adaptor ligation and transformation

DNA was extracted twice from phenol/chloroform/isoamylalcohol and precipitated with 0.1 volume of 3M NaOAc, 2.5 volume of 96% ethanol. Adaptors NTB p18/22 (19) were ligated in a 3:1 molar excess to the resuspended DNA in 0.05M Tris-HCl pH 7.6, 10mM $MgCl_2$, 1mM ATP, 1mM DTT, 5% polyethylene glycol (PEG) 8000 with 10 units of T4 DNA ligase at 15°C for 14 hours. Removing of the unreacted adaptors was achieved by running the ligation mix through a Sepharose CL-4B column. The fractions were monitored for fragment sizes on an agarose gel and DNA fragments larger than 100 bp were collected and ligated into the Eco RI site of the bluescript KS(+) vector (Stratagene). Transformation in MAX Efficiency DH5α Competent Cells (BRL) was achieved following the supplier's protocol. The colonies were screened with radio-labeled human DNA to identify recombinant clones.

The polymerase chain reaction (PCR)

Oligonucleotides were synthesized on a PCR-Mate 391 DNA-synthesizer (Applied Biosystems, Inc.), or the Gene Assembler Plus (Pharmacia LKB) and purified according to the manufacturers recommendations.

PCR reaction conditions were done as described previously (20) with minor modifications. Briefly, automated cycles (30−35) of PCR were performed with denaturation (94°C/1 min), annealing (see table 1) and extension (72°C/1 min) in a total volume of 100 μl per reaction using genomic DNA as template (100 ng), each primer at 0.50μM, each dNTP at 100μM, 1× PCR-buffer (Perkin-Elmer-Cetus) with varying concentrations of $MgCl_2$ (see table 1) and 2 units of Taq DNA polymerase. A 10 μl aliquot of each reaction mixture was subjected to electrophoresis on a 1.5% agarose gel at 80 V for 3h and was subsequently stained with ethidium bromide.

Plasmid DNA preparation

The alkaline lysis protocol for rapid plasmid DNA preparation was used as described (21).

Sequencing of primer-extended clones

The plasmid DNA was PEG-purified and subsequently alkaline denatured according to Hattori and Sakaki (22). Double-strand dideoxy sequencing using the KS (CGAGGTCGACGGTATCG) and SK primers (TCTAGAACTAGTGGATC) which flank the EcoRI insertion site of the bluescript KS(+) vector was done as described (23,24) with the heat stable Sequenase Version 2.0 DNA Polymerase (UBS, Cleveland, Ohio).

PCR mapping

Genomic DNA from a panel of 27 human-hamster hybrid cell lines (BIOS Corp., New Haven) plus human and hamster controls were used as templates for chromosome localization of the cloned fragments. The conditions for the PCR are given in table 1. A panel of 15 human-rodent somatic cell hybrids and a panel containing translocations and derivatives of chromosome 10 were kindly provided by Drs G Bruns, Boston and P Goodfellow, Vancouver respectively.

Chromosomal panel blots

Commercially prepared chromosome panels containing human, human × hamster somatic cell hybrids, together representing the entire human karyotype and hamster DNAs were used in this study (BIOS-Corp, New Haven). Prehybridization, hybridization and washing was performed following the manufacturers instructions.

In situ hybridization

In situ hybridization to metaphase chromosomes on microscope slides was carried out according to the technique of Harper and Saunders (25). The chromosomes were treated for 1 hour with RNase heat denatured and dehydrated in an alcohol series. Hybridization was done in 50% formamide /2×SSPE, 10% dextran sulphate, 0.8 mg/ml sonicated denatured human placental DNA and 0.5 mg/ml denatured probe. DNA probe G2-1H was radiolabeled by random priming using 3H-dTTP and 3H-dCTP (17). After washing the slides were covered with Kodak autoradiographic NTB-2 liquid emulsion and exposed for 4 to 6 days. Chromosomes were identified after R-banding (26). For illustrative purposes, preparations were destained in methanol and stained with Wright's (27).

RESULTS

We have devised a variation of the PCR technique to identify sequences adjacent to the human telomere associated repeat $(TTAGGG)_n$ (fig 1). 4032 recombinant clones were isolated from 2 independently constructed libraries using either primer $(TTAGGG)_4$ or $(CCCTAA)_4$ and flow sorted human chromosome 4 in a polymerase mediated amplification reaction. The clones were archived in 96 well microtiter plates and screened with radiolabelled total human DNA. A total of 139 clones gave positive hybridization signals and 62 of these had inserts greater than 100 bp.

Structural characterization of primer-extended clones

A total of 14 independently isolated clones were analyzed in more detail. Sequence analysis of these clones is summarized schematically in figure 2. All clones characterized consist of varying stretches of telomeric repeats $(TTAGGG)_n$ and except for clone Y9, of adjacent non-telomeric sequences. Clone Y9 consists exclusively of 181 bp of the telomeric repetitive sequence $(TTAGGG)_n$. The remaining 13 clones can be classified into 3 main groups with (i) unique sequences immediately adjacent to a block of $(TTAGGG)_n$ repeats, (ii) unique sequences flanked by telomeric repeats with an inverse 5′ to 3′ orientation of the $(TTAGGG)_n$ sequence, (iii) simple or highly complex repetitive sequences adjacent to the telomeric repeat.

Clone G2 is 1096 bp in length and represents a chimeric clone resulting from ligation of noncontiguous sequences during the cloning procedure. Sequence analysis and subsequent southern analysis has unambiguously defined the junction of these two clones and allowed their independent analysis (referred to as G2-1 and G2-2).

X157, G2-1, G7, Y2 and Y11 represent clones of the first group. These clones are all characterized by the presence of unique sequence DNA adjacent to variable length tandem repeats of the form $(TTAGGG)_n$. Figure 4b shows the nucleotide sequence of clone Y2. This sequence is remarkable in that the adjacent unique DNA in this clone is composed of 78% G+T residues.

Sequence analysis revealed significant sequence similarity

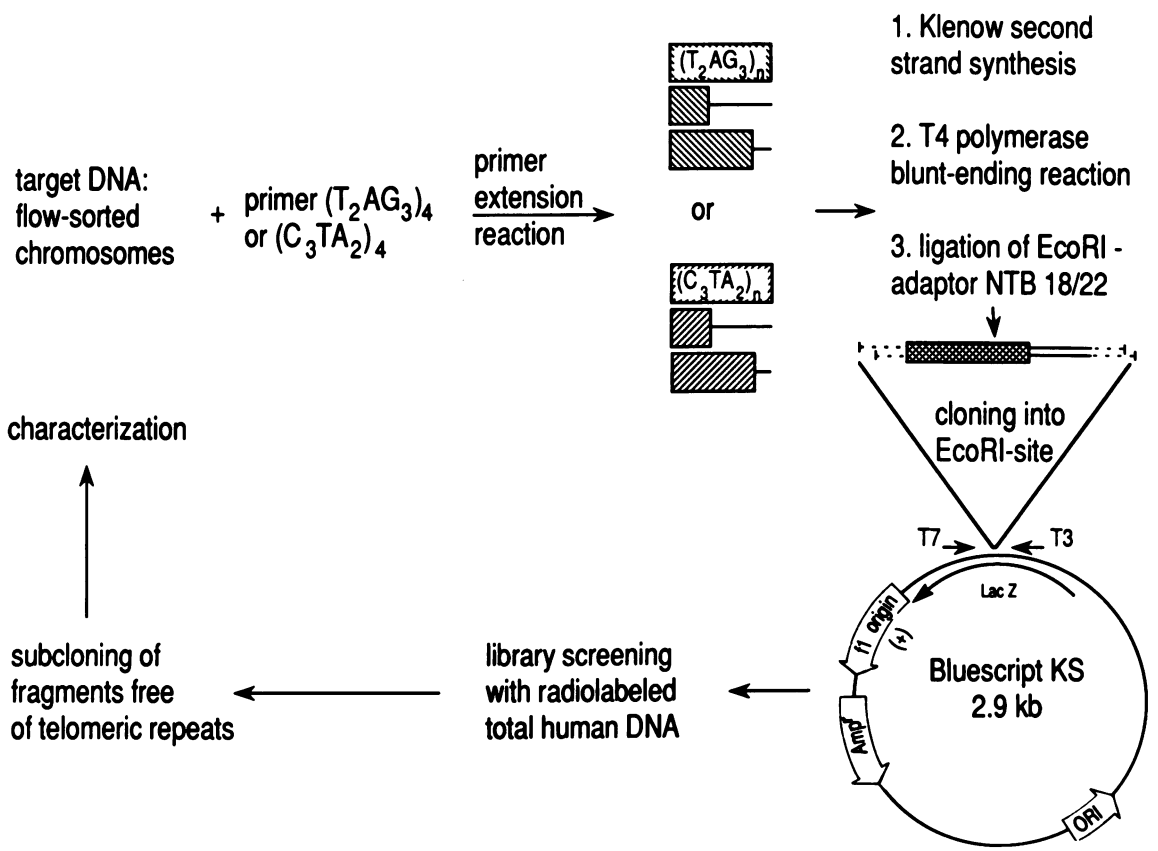

Figure 1. Schematic outline of the strategy for cloning sequences adjacent to the telomeric repeat $(TTAGGG)_n$. The Klenow second strand synthesis was accomplished by random priming of the denatured primer extension products.

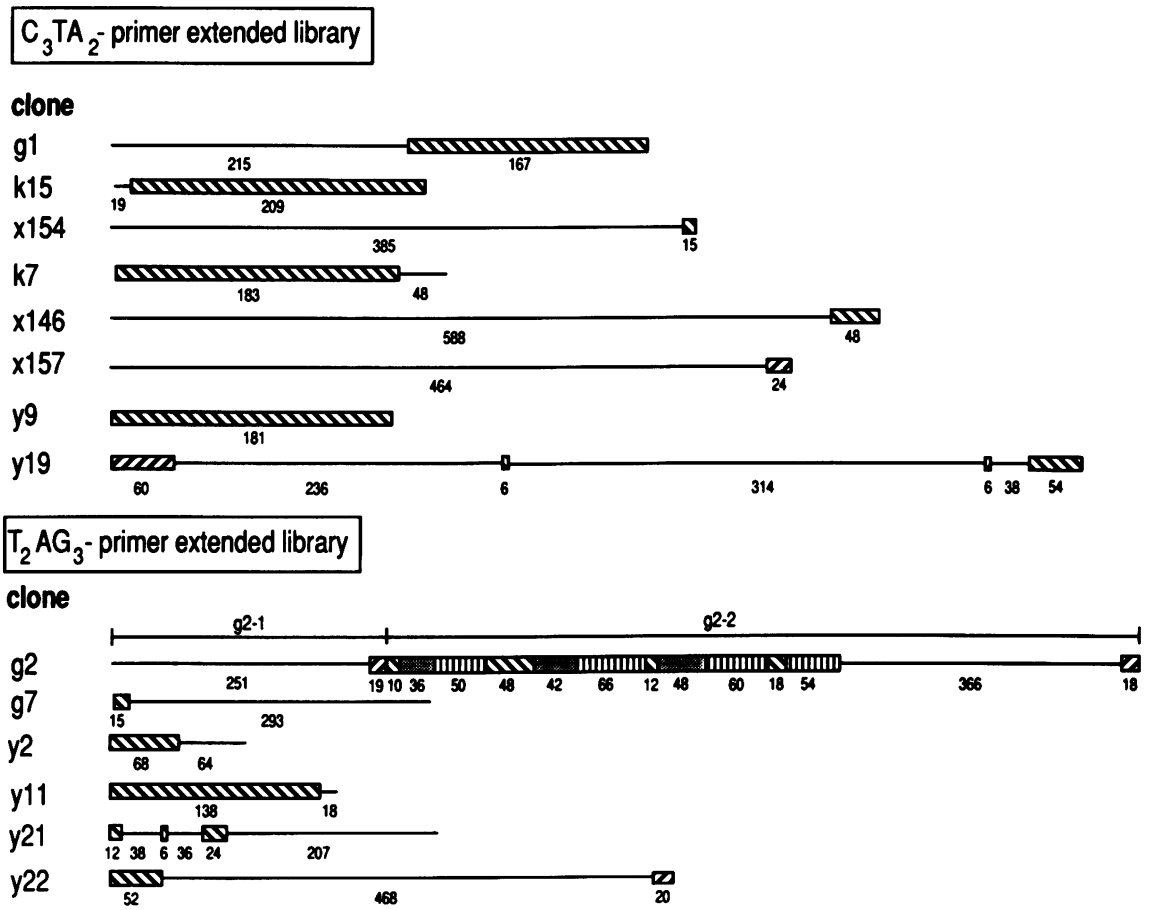

Figure 2. Schematic representation of the structure of the primer extended clones using the $(CCCTAA)_4$ or the $(TTAGGG)_4$ primer, respectively. Numbers indicate sequence length in bp. Rectangle boxes with diagonal lines to the left = 5'-$(TTAGGG)_n$- 3', with diagonal lines to the right = 5'-$(CCCTAA)_n$-3', shaded in = 5'-$(TTAGAG)_n$-3', with verticial lines = sequences with significant similarities. Note the inverted polarity of the flanking telomeric repeats in clone Y19, G2−2 and Y22.

between the first 42 bases at the 5' end of clone G2−1 and the 3' end of the right monomer of the Alu consensus sequence (28).

Clones Y19, Y21, and Y22 despite being isolated from different libraries generated with either primer $(CCCTAA)_4$ or $(TTAGGG)_4$ all share a common organization and are in group 2 together with G1, K15, X154 and X146. Both clones Y19 and Y22 contain large regions of unique sequence flanked by blocks of telomeric repeat which are in inverse polarity to each other. Clone Y21 shares a similar organization with clones Y19 and Y22 in that it is predominately nontelomeric sequence punctuated by blocks of telomeric repeat (fig 4a). In contrast to Y19 and Y22 the blocks of telomeric repeat in clone Y21 all have the same polarity (fig 2).

Significant sequence homologies were found in clones G1, K15, X154 and X146 (fig 3a). All but 37 residues of G1 match exactly with X154. Those 37 residues are the beginning of a 44 base pair exact hairpin structure at the 5' terminus of G1. Clone K15 contains a long segment of telomeric repeat $(TTAGGG)_n$ and only 19bp of unique sequence DNA which is identical to sequences in G1 and X154. In contrast, X154 is composed of 15 bp of $(TTAGGG)_n$ repeat and 385 bp of adjacent unique sequence DNA. A 40 bp match between the 5' end of X154 and the 3' end of X146 is of particular interest. Alignment of these two clones results in a structural organization similar to those seen in clones Y19 and Y22. Blocks of telomeric repeats with inverted orientation of the TTAGGG sequence are flanking a stretch of 933 bp of unique sequence DNA. PCR amplification with primers G1A and X146A and genomic DNA as template yields the predicted 905 bp fragment (fig 3b). In addition, southern hybridization with an internal oligonucleotide, X154A results in a positive signal for the 905 bp band (fig 3b).

Clones K7 and G2-2 form the third group. Clone K7 is 231 bp in length with only 48 bp of non-$(TTAGGG)_n$ repeats. These 48 bp themselves comprise a repetitive domain composed of 9 bp monomers of the form (a) and (b) (fig 4c). These two monomers differ only in one position with (a) having a T and (b) having a C at nucleotide 8 and appear to be a low copy repeat based on southern analysis utilizing several enzymes (data not shown). Clone G2-2 has a more complex organization. This clone is composed of 364 bp of unique sequence flanked at the 5' end by a highly complex domain of repetitive sequences and 18 bp of telomeric repeat at the 3' end (fig 2). In figure 4d the nucleotide sequence is shown for a single repeat unit which itself is composed of three structural motifs. These three motifs are telomeric repeats of the form $(TTAGGG)_n$, a modified telomeric repeat of the form $(TTAGAG)_n$ and a nontelomeric sequence.

This nontelomeric sequence is itself repeated four times in the clone G2-2, however with varying sequence similarities within the four repetitions.

Conservation

Unique sequences from clones G1, X154, X146, X157 and G7 showed hybridization to DNAs from primates including great apes and old world monkeys under stringent washing conditions as detected by southern blot analysis. No signals were detected using DNAs from other mammalian species including sheep, whale, cat and hamster (data not shown).

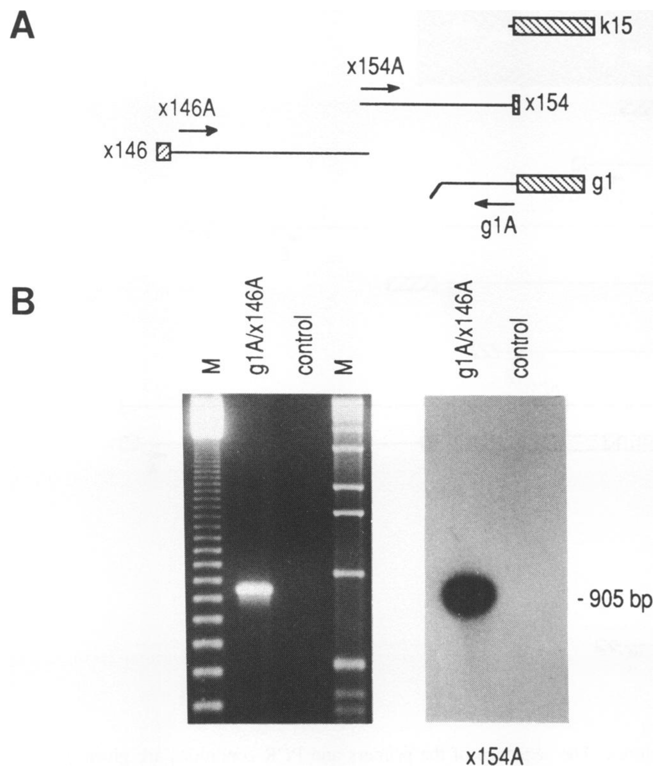

Figure 3. The independent cloned sequences G1, K15, X154 and X146 represent a single locus in the human genome. **(A)** Diagramatic alignment indicating sequence homology between the four clones. 37 residues of G1 do not match with X154 and are part of a 44 base pair exact hairpin at the 5′ terminus of G1. **(B)** PCR amplification product using primers G1A and X146A and genomic DNA, confirms the proposed organization. **(C)** Gel shown in (B) was transferred to a nylon membrane and hybridized with the internal oligonucleotide X154A at 42°C. Final washing was carried out in 1×SSC, 0.1% SDS at 42°C Localization and orientation of primers G1A, X146A and X154A are shown in (A).

Characterization of the genomic organization of the cloned sequences

In order to verify the structural colinearity of the cloned sequences with their counterparts in the human genome oligonucleotides were synthesized flanking the unique sequences in each clone (fig 5a and c) and the junction between the telomeric repeats $(TTAGGG)_n$ and the unique sequence DNA (table 1 and fig 5b and c).

PCR analysis of sequences adjacent to the $(TTAGGG)_n$ *repeats*

For each of the clones shown in figure 5c PCR amplification with total human DNA and oligonucleotide primers flanking the unique sequence DNA was performed. Analysis of the PCR products revealed (i) a single amplified product for each clone and (ii) a product whose size, for each clone, corresponds exactly to the distance between flanking oligonucleotide primers in the cloned sequences (fig 5a).

PCR analysis of the junction between unique sequences and the telomere repeat $(TTAGGG)_n$

Primer sets G1A/$(CCCTAA)_4$, X154A/$(CCCTAA)_4$, and $(TTAGGG)_4$/G7B were tested in a PCR reaction in order to confirm the cloned structure of telomeric repeats immediately adjacent to unique sequences in the human genome. These primer sets flank the junction between the unique sequences and the telomeric repeats (fig 5c) and result in PCR products which are slightly larger than the respective products amplified with primer sets G1A/G1B, X154A/G1B and G7A/G7B, respectively (fig 5b and c). The annealing of the $(CCCTAA)_4$ or $(TTAGGG)_4$ primer should occur randomly within a block of the telomeric repeat and therefore, in combination with an anchored 'unique' primer (such as G1A, X154A or G7B) leads to amplified products with varying fragment sizes but with a strong preference for the shortest PCR product possible (fig 5b).

Clone Y22 is flanked on both its 5′ and 3′ends by tandem repeats of $(TTAGGG)_n$ in inverse polarity. Both primer sets Y22A/$(TTAGGG)_4$ and $(TTAGGG)_4$/Y22B result in the

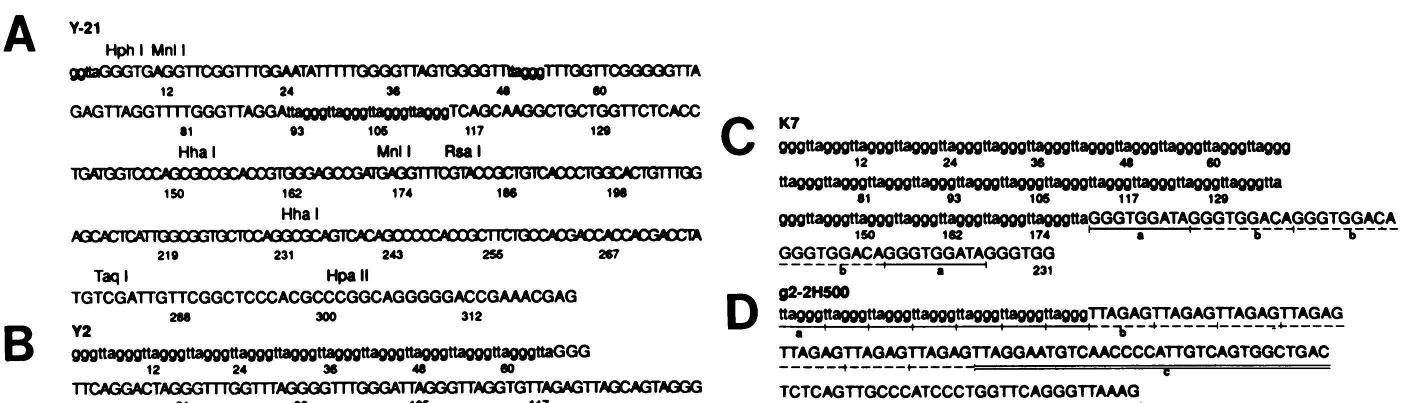

Figure 4. Nucleotide sequence and restriction enzyme sites of inserts Y21 **(A)**. Y2 **(B)**, K7 **(C)** and G2-2 H500 **(D)**. The TTAGGG repeat is printed in small letters, respectively. In **(C)** the 9 bp repeat is shown with (-a-) having a T and (--b--) having a C at nucleotide position 8. **(D)** represents one repeat unit found in clone G2-2 which itself is composed of three structural motifs (-a-) the TTAGGG-repeat, (--b--) a modified repeat of the form $(TTAGAG)_n$ and a 66 bp sequence (=c=) which occurs four times in clone G2-2 (fig.2) with varying sequence homology.

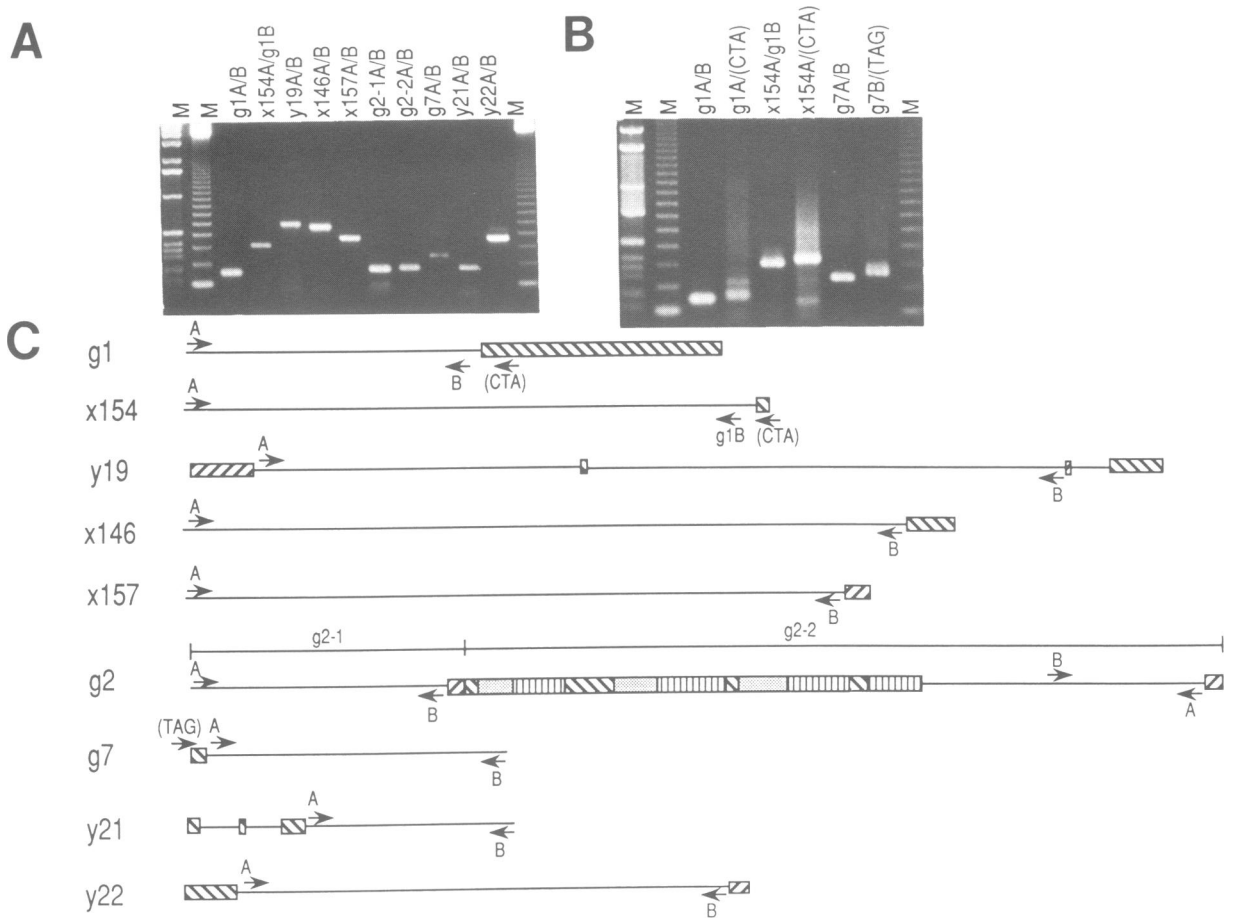

Figure 5. PCR analysis of the genomic organization of the telomere primer extended clones. The sequence of the primers and PCR conditions are given in table 1. Total human DNA served as template. **(A)** PCR products reveal the expected fragment sizes in genomic DNA. The priming site of each clone−specific primer is given in **(C)**. (CTA) = 5′-$(CCCTAA)_4$-3′ and (TAG) = 5′-$(TTAGGG)_4$-3′. For rectangle boxes see legend to fig.2. **(B)** Amplified PCR products spanning the junction of unique sequences and the telomeric repeats are shown for three primer sets derived from clones G1, X154 and G7. Note the slight increase in size of the major amplification product and 1 or 2 minor products of larger fragment size when one priming event occurs within the telomeric repeats.

expected PCR product. The amplification product with primer set Y22A/$(TTAGGG)_4$ is about 490bp and with primers $(TTAGGG)_4$/Y22B about 450bp in size. These products are 80bp and 40bp respectively larger than the amplification product of primer set Y22A/Y22B (data not shown). Primer sets Y22A/$(CCCTAA)_4$ and $(CCCTAA)_4$/Y22B fail to prime amplification. This reflects and confirms the inverse polarity of the flanking telomere repeats in clone Y22 (data not shown).

Chromosomal localization of the $(TTAGGG)_n$-adjacent sequences

Three different panels of human/hamster and human/mouse hybrid cell lines representing all human chromosomes were used as template DNAs in PCR reactions for chromosomal assignment of the cloned $(TTAGGG)_n$-adjacent sequences. All three panels were tested and their chromosome content confirmed in control PCR amplifications with several primer sets of known chromosomal localizations (data not shown). In addition some cloned fragments were analyzed in southern blot hybridizations to human/hamster hybrid cell lines. Under the washing conditions used (0.5×SSC, 0.1% SDS, at 65°C) no cross-hybridization to hamster control DNA was observed. Six cloned sequences could be assigned unambiguously to a single human chromosome. Results are summarized in table 2. However, with clones G1, X154 and X146 both Southern blot and PCR mapping gave negative results indicating that these sequences are not represented in the human-hamster hybrid panel.

In situ-hyridization of subclone G2-1H

A subclone of G2-1 (G2-1H), free of telomeric repeats and the 42 bases of the Alu sequence, was hybridized in situ to human chromosome spreads from a normal male (46,XY). A detailed analysis of the grain distribution in 56 metaphase spreads revealed a significant grain accumulation in the most distal band of the long arm of chromosome 4. (12 of 83) of the total number of grains were scored in the telomeric band 4q35 (fig 6).

Bal 31 analysis and terminal heterogeneity

Restriction fragments that terminate at telomeres have been shown to decrease steadily in size following increased exposure times to the exonuclease Bal 31. In addition, restricted telomere fragments from a particular chromosome within a cloned cell population vary from cell to cell whereas the extent of this heterogeneity seems to be the same within different tissues of a single individual (29). However, telomere fragments derived from germline DNA of the same individual are consistently about

5kb larger in size (29). We used these two approaches, Bal 31 sensitivity (with increasing enzyme concentrations and additional restriction enzymes yielding larger fragments) and germline/somatic DNA size differences, to estimate the location of our subclone G2-1H relative to the telomere. Simultaneous southern hybridization with probe pBS674 localized on 4p16.3(30) and the telomeric repeat probe $(TTAGGG)_{10}$ reveals the integrity of internal restriction fragments over the time course of the Bal 31 digestion and the sensitivity of the telomeric repeats, which disappear completely after 120 min of enzymatic digestion with Bal 31. Subsequent hybridization of subclone G2-1H detects a consistent band over the time course of Bal 31 digestion. Moreover, hybridization of clone G2-1H to sperm and blood DNA of the same individual yields a DNA fragment of identical size (data not shown).

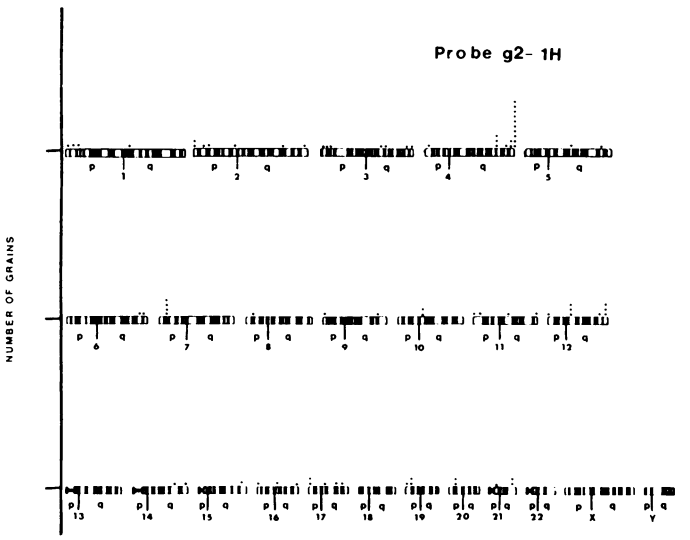

Figure 6. In situ hybridization to human metaphase chromosomes from a normal male (46,XY) with ^{3}H-labeled probe G2-1H. Statistical evaluation showing the grain distribution in 56 metaphase spreads. Significant grain accumulation was found in 4q35 (14.5% of the total number of grains).

Subclones of g1, X157, g2-2 and Y21 were hybridized to the same set of filters and all of these sequences gave similar results with no sensitivity to Bal 31 digested high molecular weight DNA and no differences in fragment size between sperm and blood DNA from a single individual (data not shown).

DISCUSSION

In this manuscript we describe a strategy for identifying DNA sequences adjacent to the human telomeric repeat $(TTAGGG)_n$. The method employs multiple rounds of a polymerase mediated primer extension reaction with a single primer derived from the telomeric repeat sequence $(TTAGGG)_n$ (fig 1). 14 independently isolated primer-extended products were sequenced (fig 2 and 4) and subjected to further analysis including mapping studies. A total of 6 of these clones were unambiguously assigned to a specific human chromosome (table 2).

Initially we used flow-purified chromosomes 4 as substrate in the primer extension reaction in our initial goal to isolate unique sequences from the telomeric region of the short arm of chromosome 4. These sequences would definitely flank the mutant gene for Huntington Disease (HD). At this point we cannot exclude the possibility of contamination with genomic DNA. This contamination may have different sources, including the carry over of telomere extended products from pilot experiments with human genomic DNA or impurities from the

Table 1. Oligonucleotide sequences and PCR conditions for primer sets used in this study. Sequences are derived from telomere primer extended clones and the binding sites within the cloned fragments are given in fig.5c.

PRIMER SET	OLIGONUCLEOTIDE SEQUENCE	ANNEALING TEMPERATURE [°C]	$MgCl_2$ CONCENTRATION [mM]
G1A/G1B	TCTCTGAGGCAACTGGAACTT/ TAATCATAGACTGGTTTCTTG	52	1.0
G1A/(CTA)	TCTCTGAGGCAACTGGAACTT/ $(CCCTAA)_4$	53	1.5
X154A/G1B	CAGATCCAGGGATCAAGCAGG/ TAATCATAGACTGGTTTCTTG	50	1.5
X154A/(CTA)	CAGATCCAGGGATCAAGCAGG/ $(CCCTAA)_4$	56	1.0
Y19A/Y19B	CTCCACAGGGATAGAACCCCA/ AAGGCTGTAAGGATGCAGAAG	53	1.0
X146A/X146B	AATAGCTCAATGGAGTTGGTC/ GCTTGATTCCTGGATCTGTAA	55	1.25
X157A/X157B	GCGGAGTGATTAATACTGCTG/ CTGGGCCAGTGAGTCTGGAGT	55	1.05
G2-1A/G2-1B	AATTCTTTCCCTCTCTAGACC/ CAACGTCAAGATTCTCTAGAA	50	0.75
G2-2A/G2-2B	ATCAAGGGGTCTACACCTTTC/ CCCTCCTTCATCCAGGAGGAC	55	1.25
G7A/G7B	GTGAGGTGGGGAAAGTCATGT/ AGAGTGAGCCTCCGAACTTTC	55	1.25
G7B/(TAG)	AGAGTGAGCCTCCGAACTTTC/ $(TTAGGG)_4$	55	0.75
Y21A/Y21B	CTCGTTTCGGTCCCCCTGCCG/ GGTCAGCAAGGCTGCTGGTTC	60	0.75
Y22A/Y22B	TTTACAGGAATGTTTCTGTGA/ TCATCTCTTGCTCTGATTAAA	54	1.5
G1A/X146A	TCTCTGAGGCAACTGGAACTT/ AATAGCTCAATGGAGTTGGTC	58	1.0

Table 2. Chromosomal localization of telomere primer extended clones. The DNA of the somatic cell hybrid panel for PCR mapping and southern blot hybridization was provided from BIOS Inc., New Haven. Fine localization of G2-2 was achieved by using somatic cell hybrid panels containing translocations and derivatives of chromosome 10. Dashes indicate not done.

	CHROMOSOMAL LOCALIZATION		
CLONE	PCR MAPPING	SOUTHERN BLOT HYBRIDIZATION	IN SITU HYBRIDIZATION
X157	chromosome 1	–	–
Y19	chromosome 8	chromosome 8	–
G2-2	10q11.2-qter	–	–
G7	chromosome 6	–	–
Y22	chromosome 3	–	–
G2-1	–	–	4q35

debris continuum which is estimated to be 4% to 5% in the high-speed flow sorting process (31). Currently, we are testing the hybrid cell line which served as source for the flow-purified chromosomes 4 for presence or absence of our cloned sequences.

Four of fourteen clones showed significant sequence homologies (fig 3a and b) and therefore are assumed to be generated by priming and extension events within a small area of one chromosome. This finding would suggest that random contamination with genomic telomeric sequences cannot explain the preferential occurrence of one particular DNA segment. Instead it would favour selective inclusion of a particular chromosomal segment in the flow sorted material or selective priming events.

We have provided direct evidence for a distal location for one of the cloned $(TTAGGG)_n$ adjacent sequences (subclone G2-1H) by using in situ hybridization to human metaphase spreads. DNA sequences originating from the physical ends of a chromosome are sensitive to Bal 31 digestion (9,10). One such sequence has been reported which terminates within 20–30 kb of the telomeres of the sex chromosomes (29,32,33). Clone G2-1 localized to the distal region of the telomeric band 4q35 (fig 6) does not show either Bal 31 sensitivity or tissue specific heterogeneity. These findings suggest a location of G2-1 at least 30 kb proximal from the terminus of 4q.

Recently, indirect evidence has been provided that the proterminal region of most human chromosomes is composed of a mosaic of G- and T-rich repeats (13). The identification of a previously undescribed 9 bp repeat and a modified telomeric repeat of the form $(TTAGAG)_n$ is in direct support of these findings (fig 4).

It has been estimated that between 95% (14) and 99% (9) of sequences detected by the $(TTAGGG)_n$ repeat are sensitive to Bal 31 digestion of intact human DNA. We have now demonstrated that small arrays of $(TTAGGG)_n$ repeats do occur in internal positions of the human chromosomes. The structural integrity of the unique sequences and the junction fragments between the $(TTAGGG)_n$ repeats and the unique sequences in the human genome was clearly demonstrated by using PCR analysis of human genomic DNA (fig 5a and b).

The cloning strategy described in this manuscript is not dependant on a given restriction enzyme site. The small fragment sizes of the inserts generated by the primer extension reaction might contribute to their stability in the plasmid vector and facilitate their rapid sequencing. YAC cloning is a powerful new technique for cloning human telomeres. However, potential problems may arise in subcloning as some repetitive sequences may lead to cloning instabilites (34). Moreover complex rearrangements during cloning in YACs might complicate the analysis (15). The method presented here has led to cloning of unique sequences adjacent to the $(TTAGGG)_n$ repeat regardless of their location within the chromosome. Refined mapping of these sequences will be of great importance in regard to understanding the genetic organization of the $(TTAGGG)_n$ repeats.

All 14 primer-extended clones contain telomeric repeats and with the exception of one also adjacent non-$(TTAGGG)_n$ sequences (fig 2). The use of a single telomeric primer appears to select for generating DNA sequences outside the block of tandemly repeated $(TTAGGG)_n$ sequences. In contrast, sequence analysis of 8 clones generated as outlined in figure 1 but using two complementary telomeric primers revealed exclusively the telomeric repeat $(TTAGGG)_n$ as expected (unpublished results). Therefore, the primer extension technique provides a powerful tool (though for reasons not yet understood) for preferentially cloning sequences adjacent to blocks of the telomeric repeat.

The identification of chromosome specific sequences from subtelomeric regions of all human chromosomes will be essential for completion of the chromosomal maps. For genetic diseases localized closely to a human telomere such as Huntington disease or polycystic kidney disease, these sequences will constitute definitive flanking markers. These subtelomeric sequences will also facilitate molecular analysis of specific chromosomal rearrangements involving telomeres.

ACKNOWLEDGEMENTS

The authors wish to thank Drs G.Bruns, Boston and P.Goodfellow, Vancouver for kindly providing somatic cell hybrid panels and Dr P.Goodfellow for critical comments on the manuscript. We thank S.Howe and K.Jack for secretarial assistance and C.Munro for preparation of the figures. B.W. is supported by a fellowship from the Deutsche Forschungsgemeinschaft (We 1259/1-1). C.C. is a predoctoral student of the Huntington Society of Canada. This work has been supported by MRC (Canada No.5-99133).

REFERENCES

1. Blackburn,E.H. (1984) Cell, **37**,7–8
2. Blackburn,E.H., Budarf,M.L., Challoner,P.B., Cherry,J.M., Howard,E.A., Katzen,A., Pan,W-C, Ryan,T. (1983) Cold Spring Harbor Symp. Quant. Biol., **47**,1195–1207
3. Blackburn,E.H., Szostak,J.W. (1984) Ann. Rev. Biochem., **53**,163–194
4. Greider,C.W., Blackburn,E.H. (1989) Nature, **337**,331–337
5. Pluta,A.F., Zakian,V.A. (1989) Nature, **337**,429–433
6. Sen,D., Gilbert,W. (1988) Nature, **334**,364–366
7. Muller,H.J. (1938) The Collecting Net-Woods Hole, **13**,181–198
8. McClintock,B. (1941) Genetics, **26**,234–282
9. Moyzis,R.K., Buckingham,J.M., Cram,L.S., Dani,M., Deaven,L.L., Jones,M.D., Meyne,J., Ratcliff,R.L., Wu,J-R. (1988) Proc. Natl. Acad. Sci. USA, **85**, 6622–6626
10. Blackburn,E.H., Challoner,P.B. (1984) Cell, **36**,447–457
11. Meyne,J., Ratcliff,R.L., Moyzis,R.K. (1989) Proc. Natl. Acad. Sci., USA, **86**,7049–7053
12. Morin,G.B. (1989) Cell, **59**,521–529
13. Allshire,R.C., Dempster,M., Hastie,N.D. (1989) Nucl. Acids Res., **17**,4611–4627
14. Cross,S.H., Allshire, R.C., McKay,S.J., McGill, N.I. Cooke, H.J. (1989) Nature, **338**,771–774
15. Brown,W.R.A. (1989) Nature, **338**,774–776
16. Cheng,S.F., Smith,C.L., Cantor,C.R., (1989) Nucl. Acids Res., **17**,6109–6127
17. Feinberg,A.P., Vogelstein,B. (1984) Analyt Biochem, **137**,266–267

18. Peters,D, Branscomb,E, Dean,P.,Merrill,T., Pinkel,D., VanDilla,M., Gray,J.W. (1985) Cytometry, **6**,290–301
19. Bevins,C.L., Roberts,J.T., McKeon,C. (1988) DNA and Protein Engineering Techniques, **1**,12–15
20. Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B., Erlich,H.A. (1988) Science, **239**,487–491
21. Maniatis,T., Fritsch,E.F., Sambrook,J. (1982) Cold Spring Harbor, New York
22. Hattori,M., Sakaki,Y. (1986) Anal Biochem, **152**,232–238
23. Henikoff,S. (1984) Gene, **28**,351–359
24. Sanger,F., Nicklen,S., Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, **74**,5463–5467
25. Harper,M.E., Saunders,G.F. (1981) Chromosoma, **83**,431–439
26. Schweizer,D. (1980) Cytogenet. Cell Genet, **27**,190–193
27. Yunis,J.J., Chandler,M.E. (1977) Prog. Clin. Pathol., **7**,267–288
28. Kariya,Y., Kato,K., Hayashizaki,Y., Himeno,S., Tarui,S., Matsubara,K. (1987) Gene **53**,1–10
29. Cooke,H.J., Smith,B.A. (1986) Cold Spring Harbor, Symp. Quant. Biol., **51**,213–219
30. Smith,B. Skarecky,D., Bengtsson,U., Magenis,R.E., Carpenter,N., Wasmuth,J.J. (1988) Am. J. Hum. Genet., **42**,335–344
31. Gray,J.W., Dean,P.N., Fuscoe,J.C., Peters,D.C., Trask,B.J., Van Den Engh,G.J., Van Dilla,M.A. (1987) Science, **238**,323–329
32. Cooke,H.J., Brown,W.R.A., Rappold,G.A. (1985) Nature, **317**,687–692
33. Petit,C., Levilliers,J., Weissenbach,J. (1988) EMBO J., **7**,2369–2376
34. Wyman,A.R., Wolfe,L.B., Botstein,D. (1985) Proc. Natl. Acad. Sci. USA, **82**,2880–2884