Structural and transcriptional analysis of a human subtelomeric repeat

Jan-Fang Cheng, Cassandra L.Smith and Charles R.Cantor
Human Genome Center, Lawrence Berkeley Laboratory and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

Received August 7, 1990; Revised and Accepted November 19, 1990

ABSTRACT

A human subtelomeric repeat (designated as the HST repeat) has been isolated and characterized from a yeast artificial chromosome containing one human telomere. This repeat is located immediately adjacent to the telomeric T_2AG_3 repeats at the extreme termini of the human chromosomes. The DNA sequence of 3.6 kb of the HST repeat has been determined. The HST repeat spans over 3.6 kb in length, and contains one evolutionarily conserved CpG-rich region. The copy number of the HST repeat varies among telomeres. Genomic hybridization experiments suggest that the HST repeat consists of two distinct segments, and the distal portions of the HST repeat are also distributed elsewhere in the genome. In HeLa cells, the HST repeat sequence appears to be transcribed into a 6 kb polyadenylated RNA and a variety of non-polyadenylated RNA species.

INTRODUCTION

Human telomeres possess many kilobases (kb) of short, tandem repeats, which share sequence homology with other eukaryotes (1, 2). Recent studies have attempted to understand the physical structure of these simple repeats in lower eukaryotes as well as the functions of the associated proteins (3–8). DNA sequences located adjacent to the simple repeats are called telomere-associated or subtelomeric sequences. Subtelomeric sequences are usually families of moderate repetitive DNA which do not neccessary appear in all eukaryotic chromosomes (9). They are not essential to the viability of the chromosome since not all chromosomes have them, but they could be involved in structural rearrangements at chromosome ends. In lower eukaryotes, for example, there is evidence that the frequent rearrangement of yeast subtelomeric repeats has created highly polymorphic regions near the yeast telomeres (10). The chromosome-sized polymorphisms observed in *P. falciparum* may also be the result of homologous recombination occurred between subtelomeric repeats (11). Such events may also contribute to the spreading of gene families in yeast (12), and the generation of variant cell surface antigens in Trypanosomes (13, 14). Cytological studies of the Drosophila polytene chromosomes suggest that the telomere-associated repeats may be components of a common telomeric structure which mediates end-to-end associations of chromosomes (15). The first human subtelomeric clone characterized was isolated from the pseudoautosomal region of a human sex chromosome (16). It lies within 20 kb of the telomere, contains repetitive sequences, and is highly polymorphic in the population.

Human telomeric restriction fragments have been cloned in yeast by complementation of deficient yeast artificial chromosomes (YACs)(17–20). Such clones allow the determination of the DNA structure and organization of human subtelomeric regions. One human telomere YAC clone, yHT1, consists of approximately 3.6 kb of a *h*uman *sub*telomeric DNA sequence (HST repeat) and several hundred basepairs (bp) of the terminal simple repeats (17). This HST repeat is found in most, but not all, human telomeres. Part of the HST sequence does not cross-hybridize to rodent DNA. This means that most human telomeric fragments can be easily identified by hybridization experiments using DNA from human-rodent hybrid cell lines. This subtelomeric DNA sequence is sensitive to *Bal*31 exonuclease in somatic cell hybrids containing either a single human chromosome 21 (17), or a single human chromosome 4p telomere (21). Pulsed-field gel electrophoretic (PFG) analysis of genomic DNA fragments showed that most *Not* I fragments that hybridize with the simple telomeric repeats also hybridize to the subtelomeric sequence.

Here, we dissect this common HST repeat into four segments, and use them in genomic hybridizations to examine the organization of these repeats on different chromosomes. We have also detected and characterized some transcripts generated from this repeat.

MATERIALS AND METHODS

Plasmid cloning and sequence determination of the human telomeric YAC clone, yHT1

yHT1 DNA was cloned as a plasmid by using the pBR322 portion (i.e. β-lactamase and origin of replication) of the YAC vector. One half μg of gel-purified, linear yHT1 DNA was treated with 0.5 U/ml of *Bal*31 nuclease (New England Biolabs, Inc.) for 5 min at 30°C to generate blunt ends. The *Bal*31 reaction was quenched by addition of 20 mM EGTA (Ethylene Glycol-bis(β-aminoethyl Ether) N, N, N′, N′-Tetraacetic Acid), and the yHT1 DNA was ethanol precipitated after phenol and chloroform extraction. Linear yHT1 DNA was then self-ligated by T4 DNA ligase, and the resulting DNA was transformed into DH5 cells using the $CaCl_2$ method described by Maniatis *et al.* (22).

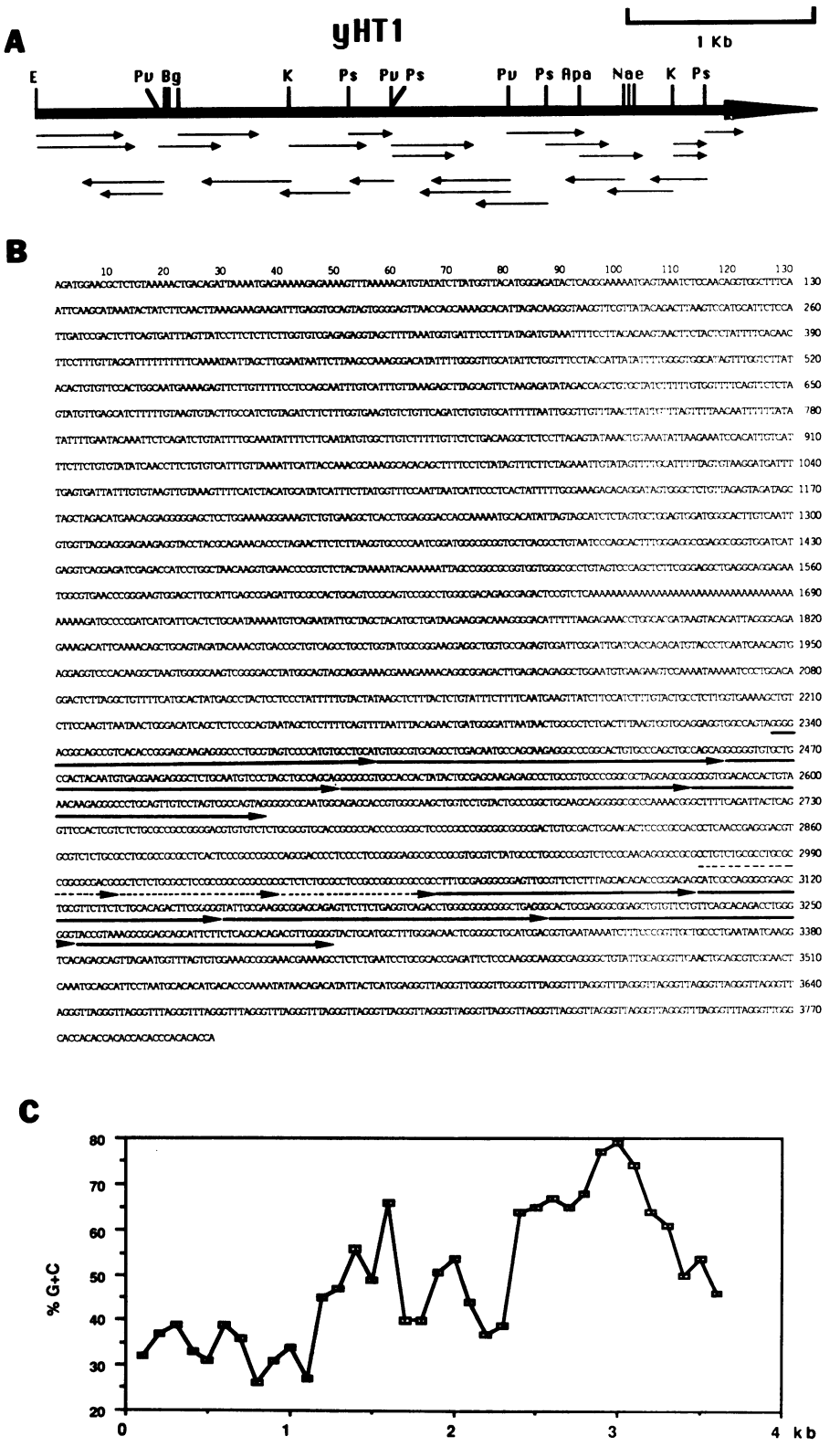

Figure 1. Sequence analysis of the HST repeat present in clone yHT1. (A) shows the strategy used to determine the DNA sequence. The thick arrow represents a restriction map of clone yHT1, and the arrowhead represents the telomeric T_2AG_3 repeats. The entire region was sequenced by the dideoxynucleotide chain termination method (23). The restriction sites used for the M13 cloning are abbreviated as E, *EcoR* I; Pv, *Pvu* II; Bg, *Bgl* II; K, *Kpn* I; Ps, *Pst* I; Apa, *ApaL* I; Nae, *Nae* I. The distances and directions of sequence determination on each fragment are indicated by the thin arrows under the map. Greater than 95% of the yHT1 DNA was sequenced either from both directions or multiple times from one direction. (B) shows the nucleotide sequence of clone yHT1. This sequence is displayed 5′ to 3′ from the centromeric end to the telomeric end. An *Alu* repeat sequence extends from nucleotide 1368 to 1695. The 61 and 46 bp tandem repeats are indicated by solid arrows. The dotted arrows indicate the repeats within the CpG island. The human telomeric simple repeat starts at nucleotide 3576, and the yeast telomeric simple repeat starts at nucleotide 3771. (C) shows the distribution of G + C content in the HST repeat region. The per cent G + C was calculated in 100-bp segments.

Clones containing the circularized yHT1 DNA were identified by hybridization with telomeric T_2AG_3 repeats. Plasmid pYHT1 was isolated from a positive clone and further subcloned into M13 bacteriophages for dideoxy sequencing (23). The sequencing reactions were performed with either Sequenase (U.S. Biochemical Corp.) at 37°C or Taq polymerase (U.S. Biochemical Corp.) at 72°C.

Genomic hybridization

Ten μg of genomic DNA was digested by restriction enzymes *Kpn* I, *Pst* I and *Pvu* II, fractionated on a 1% agarose gel by conventional gel electrophoresis, and blotted onto Nytran filters (Schleicher and Schuell) as described (22). All probes were labelled with [α-^{32}P]dCTP by Klenow fragment in random priming reactions (24). Filters were prehybridized in 0.5 M sodium phosphate, pH 7.5, 1 mM EDTA (Ethylenediaminetetraacetic Acid), 7% SDS (sodium dodecyl sulfate) for 30 min at 65°C. Hybridizations were carried out in the same solution with the addition of labelled probes at 65°C overnight. After hybridization the filters were washed twice in 2× SSC (2× = 0.3 M NaCl, 0.03 M sodium citrate, pH 7.5) and 1% SDS at 65°C.

Northern blot analysis

HeLa cells were lysed using the guanidinium method described by Ullrich *et al.* (25), and total RNA was purified in a CsCl step gradient as described in Glisin *et al.* (26). PolyA$^+$ and polyA$^-$ RNA were separated using an oligo(dT) cellulose column as described by Aviv and Leder (27). Fifteen μg of total RNA, polyA$^-$ RNA and one μg of polyA$^+$ RNA were fractionated by a 1.2% agarose/formaldehyde denaturing gel electrophoresis, and transferred to a Nytran filter (28, 29). Filter hybridization and washing were carried out using the modified Church and Gilbert procedure as described above.

RESULTS AND DISCUSSION

The HST repeat was characterized further by subcloning and sequencing the yHT1 DNA (Fig. 1). A restriction map of the yHT1 DNA and the restriction fragments used for DNA

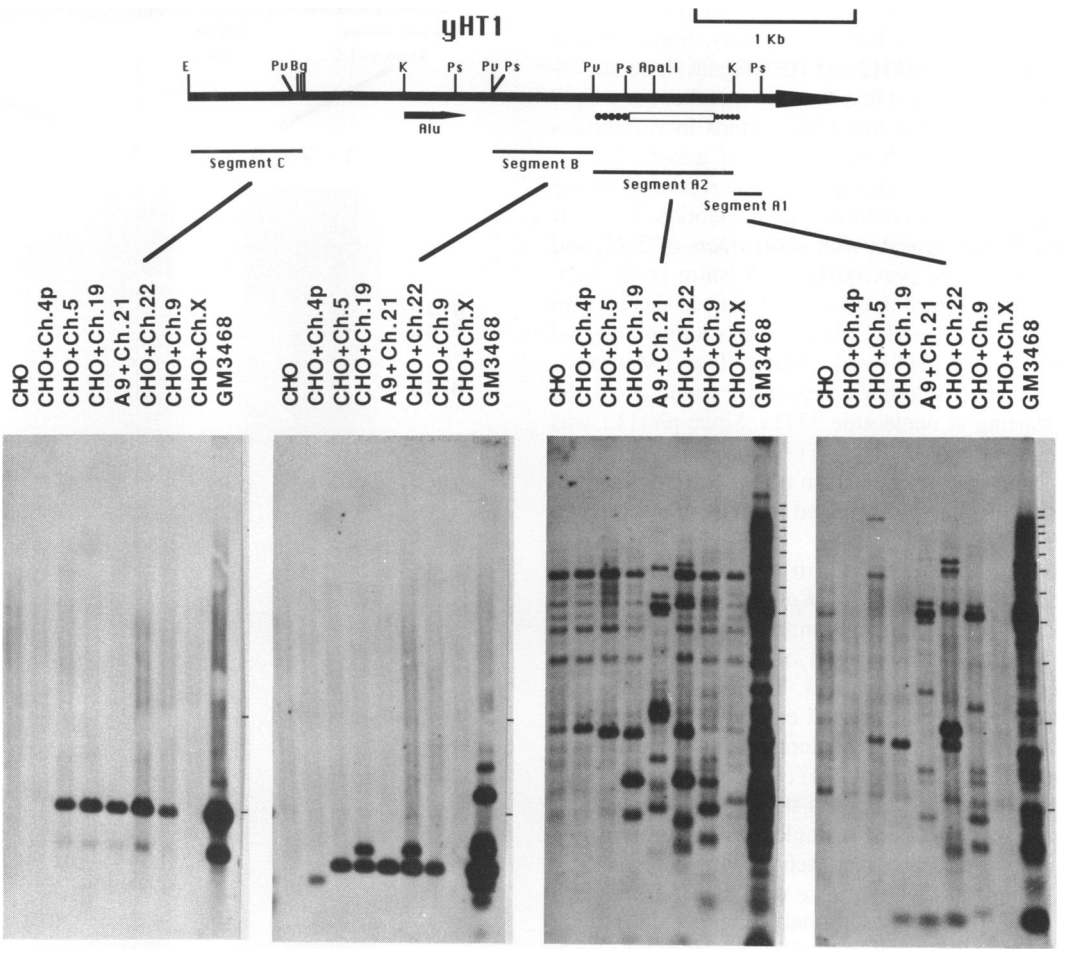

Figure 2. Genomic DNA analysis of the HST repeat. Four segments of DNA (A1, A2, B and C) isolated from the subtelomeric region of clone yHT1 were radioactively labeled, and hybridized to a panel of a *Kpn* I, *Pst* I, *Pvu* II triple digest of genomic DNAs isolated from hamster cells (CHO), human cells (GM3468) and various somatic cell hybrids with single human chromosomes either in hamster (CHO) or mouse (A9) backgrounds. The hybrid cell lines are designated by their human chromosome numbers; they are HHW693 (chromosome 4p, ref. 32), HHW599 (chromosome 5, ref. 33), PK87−9 (chromosome 9), PK87−19 (chromosome 19), WAV-17 (chromosome 21, ref. 34), EyeF3A6 (chromosome 22, ref. 35) and Y75−1B(M1) (chromosome X, ref. 36). The restriction map of yHT1 is drawn as in Fig. 1A. The corresponding positions of DNA segments A1, A2 B and C are indicated by horizontal bars below the map. The corresponding positions of the *Alu* repeat, the CpG island, and the flanking repeats are indicated as the filled arrow, the open box, and the strings of circles, respectively, below the map. Size markers are positioned every 1 kb as horizontal bars on the right side of each autoradiograph.

sequencing are shown in Fig. 1A. Approximately 3.6 kb of human subtelomeric DNA sequence has been determined (Fig. 1B). It contains an AT-rich region (nucleotides 1 to 1100 in Fig. 1B), separated by a human *Alu* repeat (nucleotides 1368 to 1695) from a CG rich region (nucleotides 2638 to 3057). The G + C content of the AT-rich region ranges from 25% to 40% (Fig. 1C). This number is significantly lower than the nearby CG rich region which has a G + C content ranging from 60% to 80%. The CG rich region is located at 519 bp 5′ to the simple repeats. It is flanked by 5 copies of a 61 bp repeat and 5 copies of a 46 bp repeat at its 5′ and 3′ ends, respectively. These repeats are imperfect, and are arranged tandemly (solid arrows in Fig. 1B). The CG rich region itself also contains at least 3 copies of a 28 bp repeat at its 3′ end (dotted arrows in Fig. 1B).

De Lange *et al.* (30) have reported the isolation of two plasmids (i.e., pTH2 and pTH14) containing both human telomeric simple repeats and flanking DNA sequences. The flanking sequences in their clones pTH2 and pTH14 share approximately 86.7% and 86.1% sequence homologies with our HST repeat sequences from nucleotide 2777 to 3080 and from nucleotide 3208 to 3570, respectively. They have also demonstrated that the CG rich sequence in clone pTH2 contains roughly equal numbers of CpG and GpC dinucleotides and they are not methylated in the germ line as would be expected in a CpG island (31). Since the CG rich regions from both clone pTH2 and HST repeat share 86.7% sequence homology, we suggest that they are members of a CpG island family. It is noteworthy that CpG islands in vertebrates are likely to be associated with the 5′ ends of genes (31).

About 200 bp of the human simple telomeric repeat sequence was determined from pYHT1 (starting at nucleotide 3576). It consists of mostly T_2AG_3 repeats with short tracts of T_2G_4 and T_3AG_3 repeats as described previously by Allshire *et al.* (32). The length of the simple repeat in plasmid pYHT1 is shorter than the original terminal repeats in yHT1 because subcloning involved a *Bal*31 degradation of the yHT1 ends (Materials and Methods). A string of $C_{1-3}A$ yeast telomeric repeats is found 3′ to the human repeats (starting at nucleotide 3771). Since pYHT1 was constructed by ligating two ends of yHT1 molecule to form a circle, the $C_{1-3}A$ repeats obtained from pYHT1 are oriented in such a way that they probably originated from the *Tetrahymena* end of the yHT1 YAC (33).

Four segments (A1, A2, B and C) of the yHT1 clone were used as hybridization probes against a *Kpn* I, *Pst* I, *Pvu* II triple digest of genomic DNAs from rodent-human somatic cell hybrids to examine the extent of conservation of the HST repeats on different human chromosomes (Fig. 2). The positions of these four segments in the HST repeat are: A1 extends from nucleotide 3249 to 3497; A2 extends from nucleotide 2448 to 3248; B extends from nucleotide 1838 to 2447; and C extends from nucleotide 1 to 693. Previous cytological studies showed that each of these hybrid cell line contained a single human chromosome (34–38). PFG analysis indicates that each cell line contains only one or two *Not* I fragments that hybridize with the HST sequence (39). This suggests that no additional, unexpected human telomeres are present in these cell lines. Chinese Hamster Ovary (CHO) cells (40) and human fibroblastoid cells (GM3468)(41) DNAs were included as controls in these hybridizations. Probe A2 cross-hybridizes strongly and distinctively to the hamster DNA. The CpG island DNA in this probe is the major component of this cross-hybridization. It has not yet been determined whether the CpG island DNA is located near hamster telomeres, or whether it codes for any conserved function. Probe A1 cross-hybridizes weakly but distinctively to hamster DNA while probes B and C do not cross-hybridize.

With total human DNA, probes A1 and A2 hybridize to a number of fragments ranging in size from less than 0.5 kb to over 12 kb. Probes B and C detect fewer fragments. For example, only 7 fragments are seen with probe B, and only 3 fragments are seen with probe C (Figure 2). A similar hybridization was carried out using somatic cell hybrids containing multiple human chromosomes. The complexity of the hybridization pattern seen with probes A1 and A2 increases as the number of human chromosomes increases, while probes B and C continue to show simple hybridization patterns (data not shown). This suggests that the A1 and A2 fragments represent a more variable portion of this subtelomeric repeat than fragments B and C.

Probes A1 and A2 detect multiple human fragments (up to 10

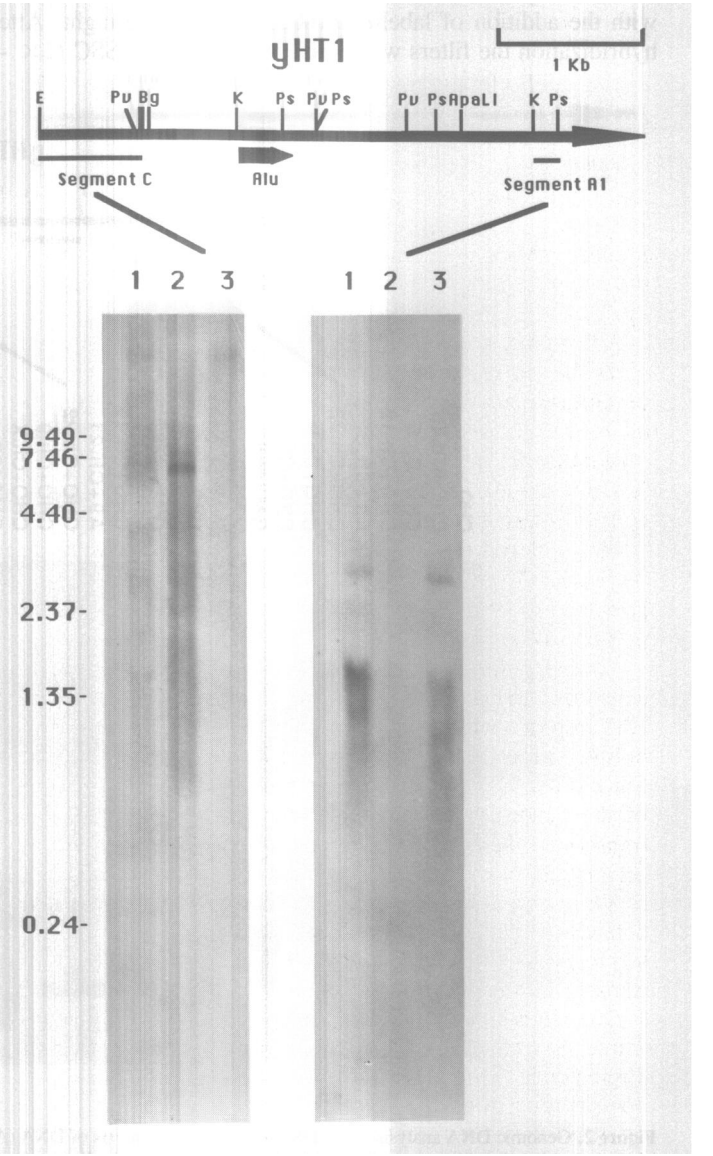

Figure 3. Northern blot analysis of the HST repeat sequence. Two segments of human subtelomeric DNA (A1 and C) were used as probes in a Northern analysis of 15 μg total RNA (lane 1), 1 μg polyA$^+$ RNA (lane 2) and 15 μg polyA$^-$ RNA (lane 3) isolated from HeLa cells. The restriction map is drawn as in Fig. 1A, and the RNA size markers are indicated in kb.

fragments) on chromosomes 9, 21 and 22 (Figure 2). Much simpler patterns are seen with chromosomes 5 and 19. No hybridization is seen with chromosomes 4p and X. Thus, it appears that this portion of the HST repeat may be present from 0 to over 5 copies per human telomere. De Lange *et al.* has used *Bal*31 digestion to determine the telomeric location of the flanking sequences of pTH14, and they found that many DNA fragments are insensitive to *Bal*31 digestion (30). However, in our previous genomic hybridizations (17), probes A1 and A2 only hybridize with the same *Not* I restriction fragments that are detected by the $(T_2AG_3)_7$ probe. Thus, the A1 and A2 repeats must occur on the same *Not* I restriction fragments as these $(T_2AG_3)_n$ repeats. To reconcile our repeats with those of De Lange et al., we suggest that some of the multiple copies of A1 DNA sequences scattered within the telomeric *Not* I fragments of some chromosomes (e.g., chromosomes 9, 21 and 22) are too proximal from the chromosome ends to show *Bal*31 sensitivity.

In contrast to the results with probes A1 and A2, at most one or two fragments per chromosome are seen with probes B and C. It is possible that each human chromosome consists only one or two copies of fragments B and C. Another possibility is that each band detected with probes B and C represents multiple copies of the same size fragments.

Probe B contains the only DNA sequence in pYHT1 that hybridizes to chromosome 4p. This indicates a surprisingly different DNA content in the telomere region of 4p compared to other autosomal telomeres. It is noteworthy that Huntington's disease gene may be located very near the tip of chromosome 4p (42).

Although at least part of the HST repeat has been detected on all of the human autosomes examined thus far, it is not present on chromosome X (Figure 2) or Y (data not shown). Cooke *et al.* (16) have identified a different type of human subtelomeric repeat that appears to be present only on the short arms of both X and Y chromosomes. The short arms of the sex chromosomes contain a pairing region (called the pseudoautosomal region) which undergoes an obligatory recombination event during meiosis. The unique sequence composition of the pseudoautosomal region may serve as a structural barrier to prevent recombination with the autosomes.

Northern blot analysis of RNA isolated from cultured HeLa cells was carried out to determine whether any sequences in the HST repeat were transcribed (Fig. 3). A predominant, 6 kb polyA$^+$ message is detected by hybridization with probe C. Since probe C is part of the HST repeat, we do not know which of the genomic sequences detected by probe C are actually transcribed. Additional faint, heterogenous transcripts are also seen on the same blot, perhaps due to the nonspecific binding of the probe DNA. In contrast, only polyA$^-$ transcripts are detected by probe A1, which contains the sequence next to the terminal T_2AG_3 repeats on yHT1. The nature of these transcripts remain to be explored.

Does the 6 kb transcript encode protein? No long open reading frame is seen in the probe C sequence. Perhaps this region corresponds to the untranslated region of a 6 kb message. Long open reading frames are seen within and near the CpG island but these may simply reflect the GC-richness of this region. So far, no other DNA sequences currently in DNA data banks are statistically similar enough to the HST repeat sequence to provide any clues about its function or evolution. If indeed the HST repeat sequence does encode protein, it will be intriguing to learn its function.

ACKNOWLEDGEMENTS

We are grateful to Dr. Dorothy Warburton for providing hybrid cell lines PK87-9 and PK87-19, and to Dr. Steve T. Warren for hybrid cell line Y75-1B(M1). This work is supported by grants from both the NCI (CA39782), and the DOE (DE-FG02-87ER-GD852).

NOTE IN PROOF

Brown and his colleagues have recently published the telomere-associated DNA sequences from three human chromosomal ends (43). We have compared their sequences (i.e., TelBam3.4, TelSau2.0 and 7a1/4) with that determined in this paper. We found that extensive homology starts at the 61 bp repeats and extends to the terminal simple repeats (including A1 and A2 repeats). No sequence similarity is seen centromeric to the 61 bp repeats. Thus, the DNA sequence in yHT1 clearly represents a different subset of human subtelomeric repeats which shares the conserved sequences, A1 and A2 repeats, with clones TelBam3.4 and TelSau2.0, but does not share regions B and C.

REFERENCES

1. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J.-R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6622–6626.
2. Allshire, R. C., Gosden, J. R., Cross, S. H., Cranston, G., Rout, D., Sugawara, N., Szostak, J. W., Fantes, P. A., and Hastie, N. D. (1988) *Nature* **332**, 656–659.
3. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., and Blackburn, E. H. (1987) *Cell* **51**, 899–908.
4. Sen, D., and Gilbert, W. (1989) *Nature* **334**, 364–366.
5. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) *Cell* **59**, 871–880.
6. Berman, J., Tachibana, C. Y. and Tye, B.-K. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3713–3717.
7. Gottschling, D. E. and Zakian, V. A. (1986) *Cell* **47**, 195–205.
8. Raghuraman, M. K. and Cech, T. R. (1989) *Cell* **59**, 719–728.
9. Blackburn, E. H., and Szostak, J. W. (1984) *Ann. Rev. Biochem.* **53**, 163–194.
10. Horowitz, H., Thorburn, P. and Haber, J. E. (1984) *Mol. Cell Biol.* **4**, 2509–2517.
11. Corcoran, L. M., Thompson, J. K., Walliker, D. and Kemp, D. J. (1988) *Cell* **53**, 807–813.
12. Carlson, M., Celenza, J. L., and Eng, F. J. (1985) *Mol. Cell Biol.* **5**, 2894–2902.
13. Borst, P., Frasch, A. C. C., Bernards, A., Van der Ploeg, L. H. T., Hoeijmakers, J. H. J., Arnberg, A. C., and Cross, G. A. M. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 935–943.
14. Borst, P., and Cross, G. A. M. (1982) *Cell* **29**, 291–303.
15. Rubin, G. M. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1041–1046.
16. Cooke, H. J., Brown, W. R. A., and Rappold, G. A. (1985) *Science* **317**, 687–692.
17. Cheng, J.-F., Smith, C. L., and Cantor, C. R. (1989) *Nucleic Acids Res.* 17, 6109–6127.
18. Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I., and Cooke, H. J. (1989) *Nature* **338**, 771–774.
19. Brown, W. R. A. (1989) *Nature* **338**, 774–776.
20. Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T., and Olson, M. V. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6240–6244.
21. Doggett, N. A., Cheng, J.-F., Smith, C. L., and Cantor, C. R. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 10011–10014.
22. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, N. Y.
23. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
24. Feinberg, A. P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
25. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J., and Goodman, H. M. (1977) *Science* **196**, 1313–1319.

26. Glisin, V., Crkvenjakov, R., and Byus, C. (1974) *Biochemistry* **13**, 2633–2637.
27. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408–1412.
28. Lehrach, H., Diamond, D., Wozney, J. M., and Boedtker, H. (1977) *Biochemistry* **16**, 4743–4751.
29. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201–5205.
30. De Lange, T., Shiue, L., Myers, R. M., Cox, D. R., Naylor, S. L., Killery, A. M. and Varmus, H. E. (1990) *Mol. Cell Biol.* **10**, 518–527.
31. Bird, A. P. (1986) *Nature* **321**, 209–213.
32. Allshire, R. C., Dempster, M., and Hastie, N. D. (1989) *Nucleic Acids Res.* 17, 4611–4627.
33. Murray, A. W., Claus, T. E., Szostak, J. W. (1988) *Mol. Cell Biol.* **8**, 4642–4650.
34. Wasmuth, J. J., Carlock, L. R., Smith, B., and Immken, L. L. (1986) *Am. J. Hum. Genet.* **39**, 397–403.
35. MacDonald, M. E., Anderson, M. A., Gilliam, T. C., Tranebjaerg, L., Carpenter, N. J., Magenis, E., Hayden, M. R., Healey, S. T., Bonner, T. I., and Gusella, J. F. (1987) *Genomics* **1**, 29–34.
36. Kozak, C. A., Lawrence, J. B., and Ruddle, F. H. (1977) *Exp. Cell Res.* **105**, 109–117.
37. Van Keuren, M. L., Hart, I. M., Kao, F.-T., Neve, R. L., Bruns, G. A. P., Kurnit, D. M., and Patterson, D. (1987) *Cytogenet. Cell Genet.* **44**, 142–147.
38. Warren S. T., Zhang, F., Licameli, G. R., and Peters, J. F. (1987) Science **237**, 420–423.
39. Cheng, J.-F. and Smith, C. L. (1990) *Genetic Analysis: Techniques and Applications* (in press).
40. Puck, T. T., and Kao, F.-T. (1982) *Ann. Rev. Genet.* **16**, 225–271.
41. Jacobs, L., and Demars, R. (1978) *Mutation Research* **53**, 29–53.
42. MacDonald, M. E., Haines, J. L., Zimmer, M., Cheng, S. V., Youngman, S., Whaley, W. L., and Bucan M. (1989) *Neuron* **3**, 183–190.
43. Brown, W. R. A., MacKinnon, P. J., Villasante, A., Spurr, N., Buckle, V. J., and Dobson, M. J. (1990) *Cell* **63**, 119–132.