



Characterization of Short Tandem Repeats from Thirty-One Human Telomeres

Marjorie Rosenberg, Lester Hui, Junli Ma, et al.

Genome Res. 1997 7: 917-923

Access the most recent version at doi:[10.1101/gr.7.9.917](https://doi.org/10.1101/gr.7.9.917)

References This article cites 34 articles, 5 of which can be accessed free at:
<http://genome.cshlp.org/content/7/9/917.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press



LETTER

Characterization of Short Tandem Repeats from Thirty-One Human Telomeres

Marjorie Rosenberg,¹ Lester Hui,² Junli Ma,² Harris C. Nusbaum,² Kevin Clark,³ Louise Robinson,³ Laura Dziadzio,¹ Pamela M. Swain,⁴ Tim Keith,⁴ Thomas J. Hudson,² Leslie G. Biesecker,¹ and Jonathan Flint^{3,5}

¹National Human Genome Research Institute, Laboratory for Genetic Disease Research, National Institutes of Health, Bethesda, Maryland 20892; ²Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142; ³Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, OX3 9DS, UK; ⁴Genome Therapeutics, Waltham, Massachusetts 02154

Completion of genetic and physical maps requires markers from the ends (telomeres) of every human chromosome. We have searched for short tandem repeats (microsatellites) in cosmid and P1 clones and generated 661 sequence-tagged sites (STS) from the terminal 300 kb of 31 human chromosome ends. PCR assays were successfully designed for 58 microsatellites and mapped both genetically and on radiation hybrids (RHs) to confirm their telomeric location. Sequence analysis revealed marked variation in sequence composition, consistent with the hypothesis that even very highly GC-rich chromosome bands (the T bands) are not homogenous. The STSs that we have generated will be a necessary resource for the construction of physical maps of these complex regions of the genome.

[Information about the microsatellites is available electronically at <http://www.cshl.org/gr> and sequence has been deposited in the Genome Database (GDB).]

Polymorphic markers from the ends of human chromosomes are valuable for a number of reasons. Without them, genetic maps are not complete, leading to the possibility of a disease gene being missed in the course of a mapping experiment. So far, polymorphic markers have been reported for 15 chromosome ends (Blouin et al. 1995; Browne et al. 1995; De Lange et al. 1990; Freije et al. 1992; Hing et al. 1993; Vocero-Akbani et al. 1996; Wintle et al. 1997), but for some clinical applications multiple markers will be needed at every chromosome end. For instance, subtelomeric rearrangements have been found in ~7% of cases of idiopathic mental retardation (Flint et al. 1995), but single markers do not have sufficiently high heterozygosity to detect all cases of monosomy and trisomy (Wilkie 1993). Terminal markers will also be required for high resolution analysis and physical characterization of regions that are involved in a number of human genetic diseases (Ledbetter 1992). Molecular probes from the terminal regions are already proving to be

useful diagnostic tools for a number of disorders involving the telomeric regions (NIH/IMM Collaboration 1996).

Unfortunately, characterization of the ends of human chromosomes is complicated by the presence of subtelomeric repetitive sequences. These sequences lie immediately adjacent to the telomere proper, which consists of a tandemly repeated 6-mer (TTAGGG) that protects chromosomes from degradation and fusion (for review, see Blackburn and Greider 1995; Kipling 1995). Subtelomeric sequences contain a wide variety of repetitive DNA, ranging from interspersed repeats, where one or a few copies are found on a restricted number of chromosome ends, to highly repetitive sequences present in many subtelomeric regions (Brown et al. 1990; Cross et al. 1990; De Lange et al. 1990; Wilkie et al. 1991; Royle et al. 1992; Youngman et al. 1992). The repetitive nature of subtelomeric regions has made it difficult to clone these regions using conventional approaches. However, the discovery that human telomeric sequences can function in yeast made it possible to clone human telomeres as yeast artificial chromosomes (half-YACs; Brown 1989; Ri-

⁵Corresponding author.
E-MAIL jf@worf.molbiol.ox.ac.uk; FAX 44 1865 222500.

ROSENBERG ET AL.

ethman et al. 1989). Previous work has demonstrated that the DNA cloned in half-YACs can faithfully represent human subtelomeric structure (Brown et al. 1990; Riethman et al. 1993; Macina et al. 1994; Negorev et al. 1994; Martin-Gallardo et al. 1995).

In this present study a variety of strategies were used to screen half-YAC DNA for polymorphic di-, tri-, and tetranucleotide repeats. In addition, a random collection of sequence was obtained to investigate sequence composition and sequence similarities between telomeres. Markers were genetically mapped on Centre d'Etude du Polymorphisme Humain (CEPH) panels and physically mapped using a set of radiation hybrids (RHs).

RESULTS AND DISCUSSION

We have developed PCR assays for 58 newly characterized markers from 28 half-YACs, 7 cosmids, and 5 P1 clones, representing 31 human telomeres. Thirty-eight of these markers are polymorphic. Table 1 shows the markers developed in this study, as well as those identified by others that we have found to be present in the clones we characterized. For completeness, telomeric markers reported by other workers are given. Table 1 shows that there are now unique sequence-tagged sites (STSs) for 28 human telomeres.

Short tandem repeat (STR) (microsatellite) markers were derived from three sources. First, we had complete sequence from the 16p (286 kb) (Flint et al. 1997a) and 4p (118 kb) (Youngman et al. 1992; Flint et al. 1997b) telomeres and knew which parts of the sequence were not repetitive. PCR assays could therefore be designed for microsatellites lying in unique DNA within 200 kb of the chromosome terminus. Second, hybridization screening with oligonucleotides and *Alu*-PCR identified sequence from half-YACs that contained microsatellites (see Methods). Third, we subcloned P1s and cosmids that had been isolated using half-YAC vector-insert junction sequence (for 2q, 13q, 6q) or were known from other data to contain subtelomeric sequence (19p, 19q, 22q, 20p) (NIH/IMM Collaboration 1996; Ning et al. 1996).

By combining these resources, we detected 167 dinucleotide microsatellites with more than 6 repeats and 20 tri- or tetramicrosatellites with more than 4 repeats. Eleven microsatellites had been or were reported during the course of this study. To design PCR assays for the remainder, we first excluded repetitive sequences. In particular, no primers were designed for microsatellites lying within

sequence that showed a similarity of >95% either to other sequence from the half-YACs or to known subtelomeric repeats. However, of the remaining 89 microsatellites, only 69 amplified a single chromosome from a monochromosomal hybrid panel.

The majority of the polymorphic markers were mapped on at least eight CEPH families. With the exception of the 2q markers, the best ordering of the new markers placed them in the most distal bin of each chromosome arm. The 2q markers mapped proximally to the most distal framework marker *D2S125*. However, *D2S125* did not amplify the 2q half-YAC or P1 clones, suggesting that there is insufficient genotyping data to order these markers correctly. Thirty-nine markers were placed with lod scores of 15 or greater on an RH panel, extending the map at 11 telomeres. At least one marker from every telomere was given a placement on the RH map. RH mapping put microsatellites from 6q, 8q, and 11q five centiRays (cR) or less centromeric of the existing distal markers, whereas the 18p marker mapped 10.5 cR centromeric and the 18q marker mapped 6.9 cR centromeric. However, existing distal markers did not amplify the corresponding telomere containing half-YACs. Because RecA-assisted restriction endonuclease (RARE) mapping has confirmed the fidelity of 6q, 8q, and 18q half-YACs, the RH map order may be incorrect. The RH mapping data are available electronically at <http://www-genome.wi.mit.edu> and a summary of RH and genetic mapping is available at <http://www.cshl.org/gr>.

The availability of complete sequence data for two telomeres (16p and 4p) allows us to assess the results of the mapping data. In line with other reports (Blouin et al. 1995; Donis-Keller et al. 1987) we find that genetic distance is inflated at the telomeres: Although the 16p markers are separated by 86,283 bp the genetic distance is >3 cM. However, neither genetic nor RH mapping data returned the correct order for the markers at 4p or 16p.

One marker showed an anomalous pattern of inheritance. Although detecting a single chromosome on the somatic cell panel and mapping by RH panel analysis to 2q, the marker *2QTEL44* was hemizygous in some CEPH families. This is consistent with the observation that there is a 50-kb chromosome length polymorphism at 2q (Macina et al. 1994) and predicts that *2QTEL44* lies within the polymorphic region. Because of this polymorphism, *2QTEL44* is not useful in all families for genetic analyses.

We investigated the sequence composition of subtelomeric regions. There is evidence that these

CHARACTERIZATION OF STRs FROM 31 HUMAN TELOMERES

Table 1. STSs for 31 Human Telomeres

| Name ^a | Het. | Size | Primer F | Primer R | Reference ^b |
|-------------------|------|---------|---------------------------|----------------------------|------------------------|
| 1QTEL10 | 0.00 | 119 | GATCCATTCCTGTATGATGG | ACTCATTCTACGAAGTCAGC | |
| 1QTEL19 | 0.75 | 156-189 | GGAGTTAAGGTTGAAGAGCC | TTCACGTACAACAGTATCTC | |
| 1QTEL36 | 0.00 | 146 | CAAATACACTGTTGGGCATG | GTGCCTCCAGAGATATGTC | |
| 2PTEL12 | 0.79 | 134-148 | CTGTGCTTCTGCAGGTTAGA | TACCTAGGTGAGAGTTATCC | |
| 2PTEL23 | 0.56 | 164-200 | GCCCTTCCCTACTTATTCTT | CCTATAGCAGTGGGATCCTA | |
| 2PTEL27 | 0.53 | 230-242 | TATATACATTAGGAATAATGAGAGT | GTCCCTCCATGTATAGA | |
| 2PTEL31 | 0.27 | 167-176 | TCCAGAACATGTGCTCCTGT | AGCTGAGATCGCACAAATTC | |
| 2P* | 0.82 | 119-135 | CTCCACTTAAGCTTGGTTTACA | CCCTAAAAGCCTCACTACATG | 1 |
| 2QTEL37 | 0.23 | 179-183 | ATCCCGGCCATCTGCCTAC | CACCAGTCCCTGATGGTAAC | |
| 2QTEL44 | 0.56 | 175-185 | AGCCTGGGGTGACAAAATCT | AGACGTCTCCAAAGCCAGTT | |
| 2QTEL47 | 0.70 | 146-162 | GTGAGGTGAGAACGCTAAG | CTGTATCGCTATGAGCCAG | |
| 2QTEL86 | 0.59 | 226-258 | AAACGTGGCAGGAGCTATGG | CATTGTTAGTTTTGCATCGC | |
| 2QTEL98 | 0.00 | 169 | CCITTTCTCCACATCCCTCAA | TCCCAGACGTTTTGTTAGGG | |
| 3PTEL01 | 0.52 | 186-230 | AGAGTTCCTAGAGGGACAG | TTGCCTGCAGTCTTCTGCC | |
| 3PTEL02 | 0.00 | 179 | TGAACTGTGGATAACTGAG | GGAAGCTGAAAAGCCAAACA | |
| 3PTEL25 | 0.63 | 182-190 | GCTAGCATGAGAATGTATAG | TACGCCTAATGAGATAACAG | |
| 3QTEL05 | 0.79 | 150-164 | TCACAGTGGCCAGATATCA | TCCATGTTGCTGCAAATGAC | |
| 3QTEL06 | 0.78 | 262-276 | TCACAAGGGAAAATAACTGTTTTAC | TTCTGTAAACCTCCAAAAT | 2 |
| 3QTEL07 | 0.75 | 134-152 | GGAAGTTTCAGCCAACG | TTAGTCCCCTGATGTTACATTT | 2 |
| 4PTEL01 | 0.00 | 179 | TATTGCTCTCTATGTCTTGC | GCTCATAAGAGACAAAAGAAG | |
| 4PTEL02 | 0.74 | 153-169 | ACAGCTAAAGCAGGTAATGG | GAGCGAGACTCAGTCTCAA | |
| 4PTEL04 | 0.67 | 166-182 | CTAGCTTTGATTCTATTGACC | GGTCTAAATCAATGACCTAAGC | |
| 4PTEL05 | 0.00 | 140 | CCATGTCTCATGCTAAGGCT | AACCGCCCTACTACTCAG | |
| 5QTEL70 | 0.00 | 191 | CTATTTTTATTTCAGTTGGCTGTTT | AAGGAAACGTTCTCTAAGTTATTA | |
| 6QTEL53 | 0.74 | 103 | GGCTAATAAATGCTTAGAGCC | GGGCTTAGTTGTTTTCCATTAG | 1 |
| 6QTEL54 | 0.50 | 206-210 | CAGAACAGATTAAGACTCAG | GCATTTATCAACTTGTGTCC | |
| 6QTEL57 | 0.00 | 169 | TGTGTTGCTCATCCACATGG | CATGCATTAACCTGCAG | |
| 7Q* | 0.84 | 217-235 | GCCCAGCCTGTGTATGGTTA | CTAAAGGAAGTTCTACAGG | 3 |
| 8PTEL91 | 0.00 | 144 | CCCAGCATCTGACATGAGAA | TTATCGCTCAAGAAGTAGAC | |
| 8QTEL11 | 0.72 | 94-112 | TGAGATCTGTCTCTATACAC | CTACAGGTGTGTGCTACCAT | |
| 8QTEL25 | 0.61 | 150-164 | GGGCTTATTAACCTATGAGCAC | GAGTTTTACCTATCTCATTGC | |
| 8QTEL54 | 0.79 | 159-177 | GTTCTCTGAACCCTAATGCA | GTTTCCAGCATAGCACATCT | |
| 9QTEL33 | 0.17 | 122-130 | ATCTGTGTTGGATTCTTGGC | ACTGAACACACCTGTACAGG | |
| 10PTEL35 | 0.76 | 134-150 | CAGAGACTGGCATTCCCAA | GCTTTGAAAGTCACCAAGTCC | |
| 10PTEL45 | 0.58 | 186-198 | TGGGTAACAGAGTGAGACTGTCTC | AGTACACTGATGATTTTCATTTCCCT | |
| 10PTEL62 | 0.69 | 116-130 | GAAAGAAATGTCAATTCTGCA | TAGGGATCCCCTCTCCTGTT | |
| 10P* | 0.81 | 151-185 | GGCTTAAAAGTAGGTAATTAC | CGCTTGGAACTTTTCAAGAT | 1 |
| 10QTEL23 | 0.37 | 138-164 | AGCAGAAGTCCAGTGTCAAGT | TCCATAAGTGCATGAGCCAA | 1 |
| 10QTEL24 | 0.00 | 163 | CTCCAAGGGTAACATTGAAC | CGTAAACGTTGTGAGATGGT | |
| 11PTEL03 | 0.85 | 168-202 | AGGGCAATGAGGACATGAAC | ATGTGGCTGGTCCACCTG | 4 |
| 11QTEL38 | 0.00 | 158 | AGGGTCTAAGCTACCCAAAG | TATGCAGGTTACAGTCAAG | |
| 12QTEL04 | 0.81 | 177-203 | AGGTGCGAGTGAGCCAAAGATT | TTCTTGGGTCTCAAGAGATC | |
| 12QTEL07 | 0.00 | 197 | GGAAACCAGCGATGAACAGA | GGACATCCTCAGATGTAGAC | |
| 12QTEL36 | 0.47 | 230-234 | AGCTAAAATCATGGGATCTCTACTA | CCAGATTATTTCTCATCTGTTTTAT | |
| 12QTEL82 | 0.75 | 208-258 | GTTCCAAGGGGAGAGTTTCAT | TAAAATGATAGTTGCACAATAATGG | |
| 12QTEL87 | 0.49 | 170-174 | TGAGACTGCAGTGAACCATG | TCTTTCTTGAAGAGGAAGCC | |
| 13QTEL10 | 0.36 | 100-150 | TACAAAGCACTGTGAGAAATGG | ATGCTGGAGCAAGCTAGGAC | 1 |
| 13QTEL56 | 0.30 | 162-166 | TTGCAGTGAGCTGATATCGC | TAAACAGGATCTCTGAAGCG | |
| 14QTEL01 | 0.64 | 174-182 | GTGCCTGTAGGTATCTATGC | GCTCCCTATTTGCAAGATAC | 5 |
| 14QTEL04 | 0.74 | 145-161 | TGCTGTTGGACTCAGGTAGCTA | TCTCTAAAGCTACTATAACCCAG | 1 |
| 14QTEL23 | 0.14 | 194-196 | GATCACGCCAAATAGTATGT | TGAGATCTGTCTTGGAAACC | |
| 16PTEL02 | 0.75 | 156-172 | GAGCGAGACTCCGTCTAAA | CAGCAGCCTCAGGGTT | 2 |
| 16PTEL03 | 0.81 | 174-186 | ACCTAGATCCCTCCAGGTTT | GGGCCATTATTCAGCCAATC | |
| 16PTEL05 | 0.76 | 170-182 | AGCTGAGATCACGCCATTGC | TGTGAGGAGACAGGAAAGAG | |
| 16PTEL06 | 0.84 | 149-185 | GAACAGGACATGGCTGTCAT | TCTACATGTGCTTGAAGAGC | |
| 16PTEL07 | 0.00 | 158 | GGTGGACAAATGTACGAAGC | CTGAGATCGTGTCTTGCAC | |
| 17PTEL37 | 0.00 | 196 | CTGCCTTAGTGTCTCTGAGT | CACAGGGACGCACATGCTTT | |
| 17PTEL49 | 0.00 | 261 | AGTAGGTTTCAGTTGCCTTTTC | AGAGACACACAAATGACAATTAG | |
| 17PTEL80 | 0.00 | 190 | ACGCACATGCTTTCACGAAC | CCTGCCTTAGTGTCTCTGAG | |
| 17QTEL13 | 0.50 | 233-235 | CGTGCCACTCAAATATAAAC | CAAAAATAAAACTGCAAGCAATATA | |

(Table 1 continued on following page.)

Table 1. (Continued)

| Name ^a | Het. | Size | Primer F | Primer R | Reference ^b |
|-------------------|------|---------|------------------------------|---------------------------|------------------------|
| 18PTEL02 | 0.00 | 220 | TCATTTATATGAAATTTCAAGGGAT | AAAAATAAAGCAATTGCTCATAGAA | |
| 18P* | 0.80 | 225-237 | GCCTCATTTTCATGGAGGTGG | CTCATTTTTCAGTCCATCTGCAAG | 1 |
| 18QTEL11 | 0.76 | 150-172 | CCTATTTAAGTTTCTGTAAAGG | ATGGTGTAGACCCTGTGGAA | |
| 18QTEL69 | 0.84 | 118-144 | ATTGCCATTCAAGGCTGAAC | GTTTTGGGAATGTCAAGAAGTACC | 2 |
| 20PTEL04 | 0.84 | 152-194 | GAATTCAGACTTGAGCTGG | GCTCAGTAATCTATCACTTC | |
| 20PEND | 0.70 | 149-173 | CCCAATCTCCATAGTTCTGG | AGCTCTTCAATTAATTTCTAGG | |
| 20PTEL18 | 0.50 | 139-141 | GATAGTCACTTCAACAGATGG | GTTTGCCAGGCTCACATTTA | |
| 20PTHY33 | 0.81 | 245-261 | GGTTCAATGCTACTCAATGGC | AACCACACTGACATCGTAGTGG | |
| 21QTEL07 | 0.78 | 311-343 | GAACCCATCTCACATGCAG | GAAGTGCTCTAAGAACTTGC | 6 |
| 21QTEL08 | 0.00 | 190 | TCAAATCTTTGTCAGAGGCG | GGAATGGGCAAATATTTCTGTG | |
| 21QTEL14 | 0.39 | 182-194 | CTAAGGACACATGCCCAATG | ACAGAGAAGGTGGGAGATTG | |
| 22QTEL31 | 0.00 | 183 | TGCATTAGGTAGATGCTGGG | TACACTAGACCAGGTGAAG | |
| 22QTEL32 | 0.00 | 189 | GACTGAGCCAAGCATGATGG | GCGTTATCCGCTCACCCATG | |
| 22Q* | 0.76 | 194-210 | TTGCAGACAGCAGACTACAGG | TTCAGTCTGTGGCTGTCCAG | 1 |
| XQYQ* | 0.70 | 220-244 | GGCCTGAATTCATTTATTATTCTAATAG | GAACAGGCAAGATGCCCACTCTC | 7 |

^aListed are markers that have been mapped to human telomeres. These markers are identified by an asterisk (*).

^bMarkers identified during this project have been reported by other workers. (1) Vocero-Akbani et al. (1996); (2) Gyapay et al. (1994); Dib et al. (1996); (3) Hing et al. (1993); (4) Browne et al. (1995); (5) Blouin et al. (1995); (6) Wintle et al. (1997); (7) Freije et al. (1992).

regions are enriched for *Alu* repetitive sequences, GC content, and genes (Bickmore and Sumner 1989; Holmquist 1992), but the data are derived from low resolution methods. Our sequence data provide a more detailed picture, indicating considerable variation in these parameters among telomeres.

This study generated 267,549-bp sequence data. On average, 6% of each half-YAC was sequenced (range 2%–15%); approximately half derived from subclones that hybridized to oligonucleotide repeat probes, and the other half was a random sample. A summary of the sequence composition is given in Figure 1 and on-line at <http://www.cshl.org/gr>. We found that ~30% of the sequence was present on more than one chromosome end (this figure excludes *Alu*, L1, and other repetitive sequences dispersed throughout the genome), as judged by sequence comparison and PCR analysis of a monochromosomal somatic cell hybrid panel. Thus, ~30% of our sequence contains subterminal repeats. With the exception of *2QTEL44*, we believe that all of the microsatellites that we mapped lie centromeric to these subterminal repeats.

Analysis of sequence composition showed that the GC content for all sequence was 0.45; however, there was marked variation among telomeres: The GC content of 3p was 37% compared to a GC content of 52% for 17p. These differences were maintained when microsatellite sequences and other repeats were excluded from the analysis. The interchromosomal differences were also independent of

the total amount of sequence and the number of individual sequences obtained.

All sequence was screened for *Alu* and non-*Alu* repetitive DNA. Again, there was variation in the

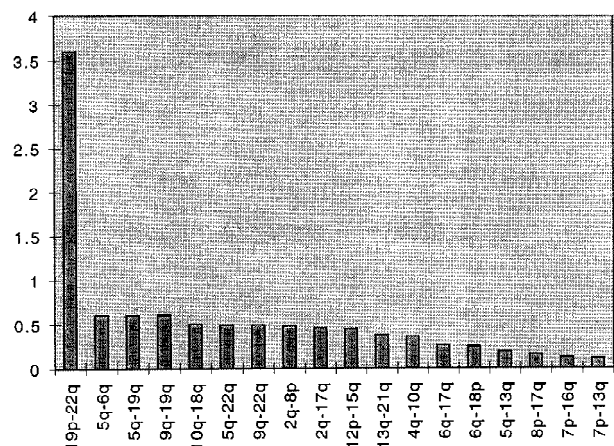


Figure 1 Intertelomeric sequence comparisons were performed with MSPcrunch. Intersequence comparisons were regarded as significant for BLASTN scores >150. The similarity score (y-axis) was then calculated by dividing the number of matches between two arms by the product of the number of clones sequenced for each arm and then multiplied by 100. The high 19q-22q similarity score is attributed to the fact that there were four matches among sequences from these arms and relatively few clones sequenced (7 for 19q and 17 for 22q).

CHARACTERIZATION OF STRs FROM 31 HUMAN TELOMERES

Alu frequency at each telomere, but there was no correlation between the two types of repeat. Regression analysis of GC content on non-*Alu* repeats was marginally significant ($P < 0.018$, $F = 6.275$ with 32 df, correlation coefficient -0.4), indicating that GC content was partly determined (inversely related) by the presence of non-*Alu* repeats.

Our analyses confirmed the expected sequence similarity among subtelomeric repeats on different chromosomes. For example, a large proportion of sequence from 2q was found to be >95% identical with sequence from 8p (Macina et al. 1994). We also detected extensive similarities between 4p and 4q; 4q and 10q; 16p and 10p, 18p, 9q (Wilkie et al. 1991); 5q and 6q; and 10q and 18q.

In conclusion, we have characterized 58 microsatellites from 31 human telomeres; 38 of these are polymorphic. The telomeric localization of these markers is attested to by data from three sources. First, molecular characterization of some half-YACs has demonstrated that they contain faithful representations of human telomeres. RARE site-specific cleavage has confirmed that half-YACs contain human telomeric and subtelomeric DNA for chromosomes 1q, 2p, 2q, 6q, 8q, 12q, 13q, 14q, 18p, 18q, and 21q (NIH/IMM Collaboration 1996). Cloned contigs of the 16p, 4p, and 22q telomeres have been compared with genomic data and, in the case of the 16p and 4p telomeres, have been completely sequenced (Flint et al. 1997a,b). Second, sequence analysis presented here and fluorescent in situ hybridization data from cosmid subclones derived from the half-YACs are consistent with the telomeric origin of a further 15 half-YACs (3p, 3q, 4q, 5q, 7p, 8p, 9q, 10p, 10q, 11p, 11q, 16q, 17p, 17q, 20p) (NIH/IMM Collaboration 1996). Finally, we present genetic and RH mapping data demonstrating a telomeric location for the markers (available electronically at <http://www.cshl.org/gr>).

Our data also allow us to draw some conclusions about the sequence organization and composition of the terminal regions of human chromosomes. The variation in sequence composition at human telomeres that we have observed is consistent with the hypothesis that even very highly GC-rich chromosome bands (the T bands) are not homogenous (Holmquist 1992). We conclude that these bands contain regions with markedly different base composition and repeat content. It remains to be shown how this relates to larger scale chromosome structural features. Nevertheless, the sequence analyses described here and the markers that have been developed provide a rich resource for physical and genetic mapping of the human genome.

METHODS

Acquisition of Sequence from Half-YACs

The complete sequence of two telomeres (285 kb from 16p and 118 kb from 4p) was obtained (Flint et al. 1997a,b), permitting the recognition of di- and trinucleotide repeats by computer analysis. To obtain sequence from half-YACs, cosmid libraries were constructed from *Mbo*I partial digests of yeast DNA in agarose plugs (NIH/IMM Collaboration 1996). Each cosmid was fingerprinted by digestion to completion with *Bam*HI endonuclease, and 6–12 cosmids with minimal overlap were chosen for sequence analysis. Cosmid DNA from each half-YAC was pooled, sonicated, and a 1- to 2-kb fraction cloned into M13. M13 libraries were plated out and plaques transferred to Hybond N (Amersham International) nylon membranes. Filters were screened by hybridization to oligonucleotides for all combinations of di- and trinucleotide repeats and to GGAA(3) and GAAA(3) oligonucleotides. In addition, 96 randomly selected M13 clones were sequenced from each half-YAC library. M13 clones were sequenced using an ABI model 373 sequencer.

Seven markers were developed from sequences obtained by two alternate methods. *Alu*-CA PCR (Mutirangura et al. 1993) was used to isolate CA repeats and the unique sequence flanking one side of markers *2PTEL12*, *2QTEL44*, *8QTEL11*, *8QTEL25*, and *12QTEL04*. Using a primer selected from the *Alu*-CA PCR sequence and a primer from the vectorette sequence, vectorette PCR (Riley et al. 1990) was used to isolate the entire microsatellite sequence of each marker. 20PA36, 20PEND, and 20PTHY33 were isolated from 20p telomeric cosmids by standard subcloning methods. Fragments hybridizing to di- or tetranucleotide repeat oligonucleotides were subcloned into plasmid vectors and sequenced.

Identification of Specific and Polymorphic Repeats

Amplifications were performed in 96-well polycarbonate using a PTC-100 thermocycler (MJ Research, Watertown, MA). DNA was amplified in a 15- μ l reaction containing the following: 10 mM Tris-HCl (pH 8.3), 50 mM KCl; 1.5 mM MgCl₂; 0.01% gelatin; 0.2 mM dATP; 0.2 mM dGTP, 0.2 mM dTTP; 0.0024 mM dCTP; 1 μ Ci of [α -³²P]dCTP (3000 Ci/mole, Amersham Corp. Arlington Heights, IL); 0.45 units of *Taq* polymerase (Perkin Elmer, Foster City, CA); 20 ng of DNA; and 0.67 μ M of each primer. After initial denaturation at 95°C for 3 min, amplification was performed for 30 cycles at 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min. A final extension step of 10 min at 72°C was also performed. Products were separated by electrophoresis at 80 W through gels containing 6% polyacrylamide and 8 M urea (Sequagel-6, National Diagnostics, Atlanta, GA). The gels were exposed directly to autoradiography film at -80°C without drying.

Markers were tested for heterozygosity and specificity using a single gel format containing the products of amplification from 20 unrelated CEPH parents and either National Institute of General Medical Sciences (NIGMS) human/rodent mapping panel 2 (Coriell Cell Repositories, Camden, NJ) or a monochromosomal somatic cell hybrid panel acquired from the UK Human Genome Mapping Project (HGMP) Resource Center. Those markers that were specific for the expected chromosome and had multiple alleles were used to amplify DNA from 8–18 CEPH families to perform linkage analysis. Heterozygosity scores and allele frequencies were determined

ROSENBERG ET AL.

using genotypes from unrelated individuals from either the initial screening gels (40 chromosomes) or from the mapping gels (70–128 chromosomes).

Genetic and RH Mapping

CRI-MAP (version 2.4; P. Green, unpubl.) was used to construct small linkage maps for several telomeres. Genotypes for distal markers were obtained from the Genotype CEPH database (<http://www.cephb.fr/cephdb/>). Markers were chosen primarily from the Cooperative Human Linkage Center (CHLC) maps (Murray et al. 1994), although other known telomeric markers were sometimes included in the analysis. Orders were generated by starting with the reference markers from the CHLC framework maps. These were entered as fixed markers in the CRI-MAP program. Additional markers were inserted one to two at a time and accepted in the order if they gave 1000:1 odds of being correct. They were then fixed in order, and the next markers were added. After all markers were ordered, the CRI-MAP fixed program was used to determine map position. For most telomeres, the order of those markers present within the half-YAC could not be determined with 1000:1 odds. Several of the most distal reference markers could not be placed with 1000:1 odds either. In those cases, the best order was used for determining map position. RH mapping was carried out using established protocols (Hudson, et al. 1995).

Data Analysis and Primer Design

After excluding *Alu* and other repetitive DNA, sequences were screened against dbEST and EMBL with BLASTN and against a nonredundant compilation of SWISS-PROT, PIR, and WORM-REP (*Caenorhabditis elegans* genes) with BLASTX. BLAST output was filtered using MSPcrunch (Sonnhammer and Durbin 1994), requiring a minimum of 90% identity for dbEST and EMBL matches. Sequence was viewed from within an ACEDB database. BLASTN/MSPcrunch was also used to identify sequence matches between telomeres. Sequences were then processed automatically for analysis by the PRIMER program (vs. 3) using scripts developed in the course of this project (available on request from jf@worf.molbiol.ox.ac.uk).

ACKNOWLEDGMENTS

We thank Maria R. Marinetti, Susie Chang, and Deborah Stone for technical assistance and Harold Riethman, William Brown, and Yi Ning for providing clones. This work was supported by the Wellcome Trust (J.F.) and National Institute of Health award HG00098 to E. Lander. T.J.H. is a recipient of a Clinical Scientist Award from the Medical Research Council of Canada.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Bickmore, W.A. and A.T. Sumner. 1989. Mammalian

chromosome banding: An expression of genome organization. *Trends Genet.* 5: 144–148.

Blackburn, E.H. and C.W. Greider, eds. 1995. *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Blouin, J.L., D.H. Christie, A. Gos, A. Lynn, M.A. Morris, D.H. Ledbetter, A. Chakravarti, and S.E. Antonarakis. 1995. A new dinucleotide repeat polymorphism at the telomere of chromosome 21q reveals a significant difference between male and female rates of recombination. *Am. J. Hum. Genet.* 57: 388–394.

Brown, W.R.A. 1989. Molecular cloning of human telomeres in yeast. *Nature* 38: 774–776.

Brown, W.R.A., P.J. MacKinnon, A. Villasanté, N. Spurr, V.J. Buckle, and M.J. Dodson. 1990. Structure and polymorphism of human telomere-associated DNA. *Cell* 63: 119–132.

Browne, D.L., E.A. Smith, J. Dietz-Band, H.C. Riethman, T. Phromchotikul, and M. Litt. 1995. Dinucleotide repeat polymorphism at the human 11p telomere (D11S2071). *Genomics* 25: 600–601.

Cross, S., J. Lindsey, J. Fantes, S. McKay, N. McGill, and H. Cooke. 1990. The structure of a subterminal repeated sequence present on many human chromosomes. *Nucleic Acids Res.* 18: 6649–6657.

De Lange, T., L. Shiue, R.M. Myers, D.R. Cox, S.L. Naylor, A.M. Killery, and H.E. Varmus. 1990. Structure and variability of human chromosome ends. *Mol. Cell. Biol.* 10: 518–527.

Dib, C., S. Faure, C. Fizames, D. Samsom, N. Drouot, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun, M. Lathrop, G. Gyapay, J. Morissette, and J. Weissenbach. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380: 152–154.

Donis-Keller, H., P. Green, C. Helms, S. Cartinhour, B. Weiffenbach, K. Stephens, T.P. Keith, D.W. Bowden, D.R. Smith, E.S. Lander et al. 1987. A genetic linkage map of the human genome. *Cell* 51: 319–337.

Flint, J., A.O. Wilkie, V.J. Buckle, R.M. Winter, A.J. Holland, and H.E. McDermid. 1995. The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nature Genet.* 9: 132–140.

Flint, J., K. Thomas, G. Micklem, H. Raynham, K. Clark, N. Doggett, A. King, and D.R. Higgs. 1997a. The relationship between chromosome structure and function at a human telomeric region. *Nature Genet.* 15: 252–257.

Flint, J., G.P. Bates, K. Clark, A. Dorman, D. Willingham, B.A. Roe, G. Micklem, D.R. Higgs, and E.J. Louis. 1997b. Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* 6: 1305–1314.

CHARACTERIZATION OF STRs FROM 31 HUMAN TELOMERES

- Freije, D., C. Helms, M.S. Watson, and H. Donis-Keller. 1992. Identification of a second pseudoautosomal regions near the Xq and Yq telomeres. *Science* 258: 1784–1787.
- Gyapay, G., J. Morissette, A. Vignal, C. Dib, C. Fizames, P. Millasseau, S. Marc, G. Bernardi, M. Lathrop, and J. Weissenbach. 1994. The 1993-94 Genethon human genetic linkage map. *Nature Genet.* 7: 246–249.
- Hing, A.V., C. Helms, and H. Donis-Keller. 1993. VNTR and microsatellite polymorphisms within the subtelomeric region of 7q. *Am. J. Hum. Genet.* 53: 509–517.
- Holmquist, G.P. 1992. Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.* 51: 17–37.
- Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.-H. Xu et al. 1995. An STS-based map of the human genome. *Science* 270: 1945–1954.
- Kipling, D. 1995. *The telomere*. Oxford University Press, Oxford, UK.
- Ledbetter, D.H. 1992. Minireview: Cryptic translocations and telomere integrity. *Am. J. Hum. Genet.* 52: 451–456.
- Macina, R.A., D.G. Negorev, C. Spais, L.A. Ruthig, X.-L. Hu, and H.C. Riethman. 1994. Sequence organization of the human chromosome 2q telomere. *Hum. Mol. Genet.* 3: 1847–1853.
- Martin-Gallardo, A., J. Lamerdin, P. Sopapon, C. Friedman, A.L. Fertitta, E. Garcia, A. Carrano, D. Negorev, R.A. Macina, B.J. Trask, and H.C. Riethman. 1995. Molecular analysis of a novel subtelomeric repeat with polymorphic chromosomal distribution. *Cytogenet. Cell Genet.* 71: 289–295.
- Murray, J.C., K.H. Buetow, J.L. Weber, S. Sudwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V.C. Sheffield, S. Sunden, G.M. Duyk et al. 1994. A comprehensive human linkage map with centimorgan density. *Science* 265: 2049–2054.
- Mutirangura, A., F. Greenberg, L.G. Butler, S. Malcolm, R.D. Nicholls, A. Chakravarti, and D.H. Ledbetter. 1993. Multiplex PCR of three dinucleotide repeats in the Prader-Willi/Angelman critical region (15q11-q13): Molecular diagnosis and mechanism of uniparental disomy. *Hum. Mol. Gene.* 2: 143–141.
- Negorev, D.G., R.A. Macina, C. Spais, L.A. Ruthig, X.-L. Hu, and H.C. Riethman. 1994. Physical analysis of the terminal 270 kb of DNA from human chromosome 1q. *Genomics* 22: 569–578.
- NIH/IMM (National Institutes of Health/Institute of Molecular Medicine) Collaboration, Y. Ning, A. Roschke, A.C. Smith, M. Macha, K. Precht, H. Riethman, D. Ledbetter, J. Flint, S. Horsley, R. Regan, K. Kearney et al. 1996. A complete set of human telomeric probes and their clinical application. *Nature Genet.* 14: 86–89.
- Ning, Y., M. Rosenberg, L.G. Biesecker, and D.H. Ledbetter. 1996. Isolation of the human chromosome 22q telomere and its application to detection of cryptic chromosomal abnormalities. *Hum. Genet.* 97: 765–769.
- Riethman, H.C., R.K. Moyzis, J. Meyne, D.T. Burke, and M.V. Olson. 1989. Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast artificial chromosome vector. *Proc. Natl. Acad. Sci.* 86: 6240–6244.
- Riethman, H.C., C. Spais, J. Buckingham, D. Grady, and R. Moyzis. 1993. Physical analysis of the terminal 240 kb of DNA from chromosome 7q. *Genomics* 17: 25–32.
- Riley, J., R. Butler, D. Ogilvie, D. Jenner, S. Powell, R. Anand, J.C. Smith, and A.F. Markham. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosomes (YAC) clones. *Nucleic Acids Res.* 18: 2887–2890.
- Royle, N.J., M.C. Hill, and A.J. Jeffreys. 1992. Isolation of telomere junction fragments by anchored polymerase chain reaction. *Proc. R. Soc. Lond. B Biol. Sci.* 247: 57–61.
- Sonnhammer, E. and R. Durbin. 1994. An expert system for processing sequence homology data. *Proc. Int. Soc. Mol. Biol.* 94: 363–368.
- Vocero-Akbani, A., C. Helms, J.-C. Wang, F.J. Sanjurjo, J. Korte-Sarfaty, R.A. Veile, L. Liu, A. Jauch, A.K. Burgess, A.V. Hing et al. 1996. Mapping human telomere regions with YAC and P1 clones: Chromosome-specific markers for 27 telomeres including 149 STSs and 24 polymorphisms for 14 proterminal regions. *Genomics* 36: 492–506.
- Wilkie, A.O.M. 1993. Detection of cryptic chromosomal abnormalities in unexplained mental retardation: A general strategy using hypervariable subtelomeric DNA polymorphisms. *Am. J. Hum. Genet.* 53: 688–701.
- Wilkie, A.O.M., D.R. Higgs, K.A. Rack, V.J. Buckle, N.K. Spurr, N. Fischel-Ghodsian, I. Ceccherini, W.R.A. Brown, and P.C. Harris. 1991. Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* 64: 595–606.
- Wintle, R.F., T.G. Nygaard, J.-A. Herbrick, K. Kvaloy, and D.W. Cox. 1997. Genetic polymorphism and recombination in the subtelomeric region of chromosome 14q. *Genomics* 40: 409–414.
- Youngman, S., G. Bates, S. Williams, A.I. McClatchey, S. Baxendale, Z. Sedlacek, M. Altherr, M.E. MacDonald, J.F. Gusella, D. Sheer, and H. Lehrach. 1992. The telomeric 60 kb of chromosome arm 4p is homologous to telomeric regions on 13p, 15p, 21p and 22p. *Genomics* 14: 350–356.

Received May 22, 1997; accepted in revised form July 24, 1997.