© 1993 Oxford University Press
Nucleic Acids Research, 1993, Vol. 21, No. 25 5865–5874

The organisation of repetitive sequences in the pericentromeric region of human chromosome 10

Michael S.Jackson*, Predrag Slijepcevic and Bruce A.J.Ponder
Cancer Research Campaign, Human Cancer Genetics Research Group, Department of Pathology, University of Cambridge, Tennis Court Road,Cambridge CB2 1QP, UK

Received September 20, 1993; Revised and Accepted November 1, 1993 EMBL accession no. X74413

ABSTRACT

Three satellite DNA families are present in the pericentromeric region of chromosome 10; the alpha satellite and two 5 bp satellite families defined here as satellites 2 and 3. Pulsed field gel electrophoresis (PFGE) demonstrates that these sequences are organised into five discrete arrays which are linked within a region of approximately 5.3 Megabases (Mb) of DNA. The alpha satellite is largely confined to a 2.2 Mb array which is flanked on its p arm side by two 100–150 kb satellite 3 arrays and on its q arm side by a 900 kb satellite 2 array and a further 320 kb satellite 3 array. This linear order is corroborated by fluorescent *in situ* hybridisation analyses. In total, these arrays account for 3.6 Mb of DNA in the pericentromeric region of chromosome 10. These data provide both physical information on sequences which may be involved in centromere function and a map across the centromere which has the potential to link yeast artificial chromosome (YAC) contigs currently being developed on both arms of this chromosome.

INTRODUCTION

Understanding the organisation of repetitive sequences present at human centromeres is an important step in efforts to characterise centromere function. In budding and fission yeasts, centromere DNA sequences have been successfully identified by physical and mutational analyses (reviewed in ref. 1). However, in mammalian chromosomes the centromere, seen cytogenetically as the primary constriction of the chromosomes during metaphase, is an ill-defined region which may span several megabases of DNA, most or all of which is highly repetitive in nature.

The best candidate for a centromere sequence in humans is the alpha satellite, or alphoid DNA, which is present at the centromeres of all chromosomes (2). This satellite family is based on a 170 bp monomeric unit organised into higher order repeat structures which can be specific to a single human chromosome (reviewed in ref. 3). Focussing on the chromosome-specific aspect of alphoid repeats probably underestimates the sequence heterogeneity which this satellite shows both within and between chromosomes (4,5). The involvement of alpha satellite in centromere function is suggested both by its distribution and by the fact that it contains a 17 bp binding site specifically recognised by one protein component of the centromere, CENP-B (6). In addition, when human alpha satellite DNA from chromosome 17 was transfected into African green monkey cells, the integrated DNA bound CREST antiserum and caused segregational abnormalities (7). These data imply a major role for alphoid DNA in centromere function. However, it is not clear that this is the only sequence involved. For instance, a functional marker chromosome has been observed in the absence of alphoid DNA and CENP-B (8); Y chromosome alphoid DNA does not bind CENP-B (9); and CENP-B is found at the inactive centromeres of dicentric chromosomes (10).

Other candidates for sequences with centromere function include the classical satellite families, satellites I, II and III, originally identified by isopycnic centrifugation. The basic genomic location of these satellites has been known for some time with major sites found at or near the centromeres of the acrocentric chromosomes, chromosomes 1, 9, 16 and on the long arm of the Y chromosome (11). Sequences from these families have also been identified on other chromosomes using hybridisation techniques (12–14), suggesting that they may be represented on most, if not all, human chromosomes. Satellite I is based on a 42 bp AT-rich repeat unit, while satellites II and III consist of a heterogeneous family of sequences based on the 5 bp motif GGAAT (15). A family with a 68 bp periodicity, the *Sau*3A or b satellite (16,17), has been identified in the heterochromatin of the acrocentric chromosomes and the pericentromeric region of chromosomes 1 and 9. A further family, with a 48 bp periodicity, has been localised to the pericentromeric region of chromosome 22 by *in situ* hybridisation (18). There is good evidence that higher order repeat structures exist within some or all of these satellites (12,14,17,19), suggesting that they share the basic organisational features of alphoid DNA. More recently, several studies involving microdissection of a marker chromosome (20) and analysis of large genomic clones (4,21,22) have led to the identification of new repetitive families at the centromeres of human chromosomes and previously unidentified arrays of exisiting satellites. Thus, while our knowledge of the sequences present at human centromeres is extensive, it is far from complete.

The advent of pulsed field gel electrophoresis (PFGE, 23) has allowed the organisation of these sequences to be investigated. Alphoid DNA is primarily arranged as long, tandem arrays

* To whom correspondence should be addressed at: University of Newcastle upon Tyne, Division of Human Genetics, Newcastle upon Tyne NE2 4AA, UK

uninterrupted by other sequence types (e.g. 24,25). The length of these arrays is known to vary considerably between homologous copies of each chromosome (e.g. 25,26). *In situ* hybridisation has also provided some information on the relative position of satellite families within individual chromosomes, placing them in discrete, non-overlapping domains (27,28). In addition, PFGE and sequence analyses have shown that different satellite families are tightly linked on some human chromosomes (13,29,30).

An integrated picture of sequence organisation at a human centromere has only been obtained for the Y chromosome where a patchwork of the 5 bp, 48 bp and 68 bp satellites, together with some novel interspersed repeats, have been linked to the main centromeric alphoid array (22). With no comparable maps from other chromosomes the significance, or generality, of this organisation is difficult to assess.

The pericentromeric region of chromosome 10 has been the subject of intense study since the gene(s) responsible for the multiple endocrine neoplasia type 2 syndromes was mapped to 10p11.2−q11.2 (31), and repetitive sequences in this region of the chromosome have been partially characterised. Members of a chromosome 10-specific alphoid family have been cloned (32) and sequenced (33), and a satellite III clone (34) has been used to identify a satellite array in 10q11.2 which is separated from alphoid sequences by a 485 bp rearranged L1 sequence (13).

Here we describe the identification of further arrays of satellite sequences in this region, and determine the position of these sequences relative to each other in a somatic cell hybrid which contains chromosome 10 as its only human material.

RESULTS

Identification of repetitive sequences on chromosome 10

The somatic cell hybrid line R342A4 contains chromosome 10 as its only human component (35). The 68 bp and 48 bp satellites and the satellite I probe oligo-sat 1 (14) do not identify sequences in this hybrid by Southern hybridisation (13, data not shown). However, a satellite III clone derived from the Y chromosome, pHS5 (34), does cross-hybridise with sequences in R342A4. This probe produces a smear in total human DNA (Figure 1a, lane 1), hybridises weakly to a 1.2 kb fragment in the hamster cell line CHO-K1 (lane 2) and identifies a series of fragments of between 2.5 kb and 7 kb in the hybrid R342A4 (lane 3). For comparative purposes, Figure 1(b and c) shows identical filters hybridised with a chromosome 10 satellite III clone p375M2.4 (13) and a chromosome 10 alphoid clone pa10RP8 (32), respectively. It is clear from a comparison of all 3 figures that each probe recognises fragments of different sizes in the hybrid R342A4, despite the fact that pHS5 and p375M2.4 are both satellite III-related clones.

The hybridisation signal obtained when R342A4 DNA is probed with the clone pHS5 is greatly reduced under washing conditions of high stringency (data not shown). It was therefore desirable to obtain a chromosome 10-derived clone for further work. A cosmid clone cMEN219 (representing locus *D10S130*) maps to 10cen−q11.2 (36,37) and shows homology to pHS5 (36). Subclones containing pHS5 related DNA were obtained from this cosmid and one, mC219.2, was sequenced (see Materials and Methods).

Comparison and classification of satellite III-related sequences on chromosome 10

Figure 2(a) shows the frequency of the 5 bp repeat GGAAT (characteristic of satellites II and III) in 275bp of satellite III related sequence obtained from mC219.2 and the first 275 bp of satellite III related sequence from p375M2.4, The mC219.2 sequence contains 22 perfect matches for this sequence with only 11 matches in the p375M2.4 sequence. When the stringency required to register a match is reduced to 4 out of 5 bp, the number of observed matches increases in both sequences (to 42 in mC219.2 and 30 in p375M2.4) showing that this 5 bp motif is degenerate in both clones. When the sequence from mC219.2 is compared to itself (Figure 2b), the 5 bp repeating unit on which it is based can be clearly defined, with no higher order structure being observed. However, the self comparison of sequences from p375M2.4 (Figure 2c) demonstrates the existence of a repeat structure which is longer than 5 bp in this clone. A more detailed analysis shows that this repeating unit is 26 bp, with individual repeat units varying from 22 bp to 27 bp in length (data not shown). When the sequences are directly compared (Figure 2d) there is little evidence of homology between the two. This is consistent with the fact that these clones do not cross-hybridise in Southern analyses.

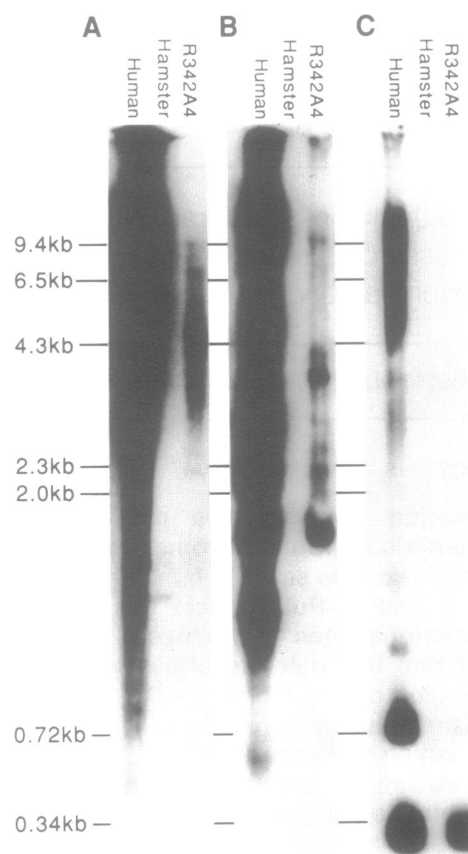

Figure 1. Southern analyses of repetitive sequences in human, hamster and the derivative cell line R342A4. All lanes contain 5 μg of *Eco*RI digested DNA. All filters were washed in 2×SSC at 50°C. **A;** pHS5 probe. **B;** p375M2.4 probe **C;** pα10RP8 probe.

Satellites II and III have been referred to collectively as the 5 bp family (22) to highlight the GGAAT repeat on which they are based. Both mC219.2 and p375M2.4 were identified by satellite III clones (34), and share their highest sequence identities with satellite III sequences present in the EMBL database (13,30). However, consensus sequences have been presented (15) which

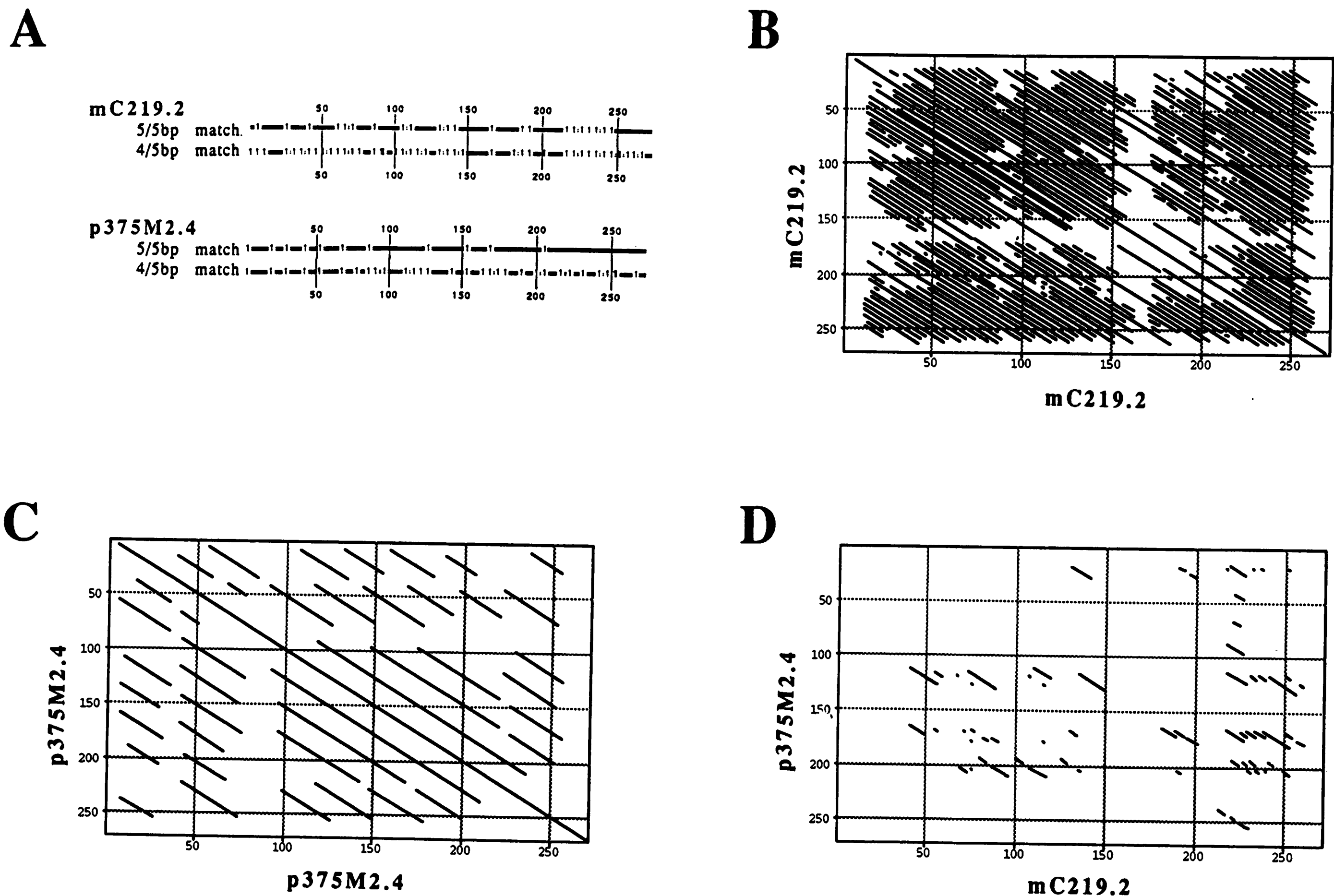

Figure 2. A; Frequency of GGAAT sequence in chromosome 10 clones. The presence of a 1 indicates a match of the test sequence (GGAAT) with the target sequence. Perfect matches and matches of 4 out of 5 bp are shown. **B**; dot matrix showing self comparison of sequence data from mC219.2. **C**; dot matrix showing self comparison of sequence data from p375M2.4. **D**; dot matrix showing comparison between mC219.2 and p375M2.4. All comparisons have a window size of 30 bp with a stringency of 66.6%.

identify simple sequence components (termed satellites 2 and 3) enriched in satellites II and III. Comparison of mC219.2 and p375M2.4 with these consensus sequences results in the classification of the 5 bp/26 bp repeat in p375M2.4 as a satellite 2 sequence and the 5 bp repeat in mC219.2 as satellite 3 sequence (data not shown). We therefore adopt this terminology as it discriminates between these two 5 bp satellites.

Thus, three distinct satellite sequences have been identified on chromosome 10: alphoid sequences (identified by pa10RP8), satellite 2 sequences (identified by p375M2.4) and satellite 3 sequences (identified by mC219.2 and mC219.28).

Array structure of satellite sequences on chromosome 10

The satellite 2 sequences on chromosome 10 are known to be arranged as a single array which is 900 kb long and identified by the restriction enzymes *Hin*dIII, *Bgl*II and *Bst*EII in the hybrid R342A4 (13). We have used the chromosome 10 alphoid and satellite 3 clones to further investigate the organisation of satellite sequences in R342A4.

The alphoid probe does not resolve any hybridising fragments under standard PFGE conditions (50 kb−1.5 Mb) using the restriction enzymes *Bam*HI, *Bgl*I, *Nsi*I, *Bst*EII and *Sfi*I (data not shown). Under electrophoretic conditions which separate larger DNA fragments, this probe detects a single fragment of approximately 2.2 Mb when DNA from R342A4 is digested with these enzymes both singly and in combination (Figure 3a and b). This demonstrates that sequences homologous to the alphoid probe are present as a single 2.2 Mb array in this hybrid, with clusters of restriction enzyme sites at both ends. Differences in the size of the hybridising fragment obtained with some enzymes (Figure 3b) suggest that the enzyme sites which identify this array are not tightly clustered.

Satellite 3 sequences in R342A4 cannot be resolved as a single fragment, even when infrequently-cutting enzymes are used (data not shown). However, these sequences are cleaved into a consistent pattern in R342A4 with the enzymes *Bgl*II, *Bst*EII and *Hin*dIII (Figure 3c, lanes 1−3). A fragment of 320−370 kb, two fragments of between 100 and 150kb and several smaller fragments less than 50 kb long are present in each digest. In the *Bgl*II/*Bst*EII double digest (lane 4) and the *Bst*EII/*Hin*dIII double digests (lane 6), this pattern is not significantly altered, with minor changes in the size of the fragments in the 100−150 kb range being seen. In the *Bgl*II*Hin*dIII double digest (lane 5), one of the fragments between 100 and 150kb is lost, with no smaller hybridising fragments being seen which are not in the single digests (lanes 1 and 3). The most likely explanation for this result is that the single band of hybridisation of 125 kb in the *Bgl*II/*Hin*dIII digest (lane 5) consists of two co-migrating

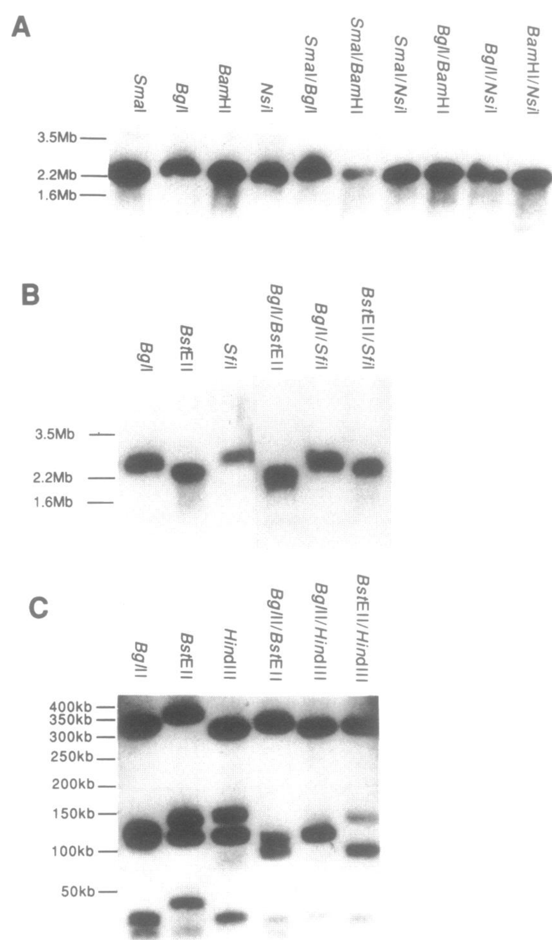

Figure 3. PFGE analyses of satellite arrays in R342A4. **A**; alphoid probe (pα10RP8), electrophoresis performed using CHEF DRII. **B**; alphoid probe (pα10RP8), electrophoresis performed using CHEF Mapper. **C**; satellite 3 probe (mC219.28), electrophoresis performed using CHEF DRII. All filters were washed in 0.5×SSC/0.1% SDS at 65°C. For details of electrophoretic conditions see Materials and Methods.

fragments. Despite this ambiguity, it is clear that these sequences are not present as a single homogeneous array.

The fact that the three sequence types are present on different *BstE*II restriction fragments (alphoid: 2.2 Mb (Figure 3a), satellite 2: 900 kb (13) and satellite 3: 370 kb, 140 kb, 120 kb, and <40 kb (Figure 3c)) demonstrates that these sequences are organised into a series of discrete, non-overlapping arrays in the hybrid R342A4.

Mapping the satellite arrays using somatic cell hybrids

Somatic cell hybrids exist with breakpoints at, or near, the centromere of chromosome 10 (36), but the existence of other human chromosomes in these cell lines can complicate the interpretation of hybridisation data, especially when repetitive probes are used. Recently, three independently derived subclones of the chromosome 10 only hybrid R342A4 have been isolated (termed R432A4-B14, R342A4-B15 and R342A4-C3) which have lost 10q as assayed by *in situ* hybridisation and PCR analyses (38). The mapping of the alphoid, satellite 2 and satellite 3 sequences relative to the breakpoints in these subclones is presented in Figure 4.

*Xba*I cuts the alphoid array in R342A4 several times to produce a series of fragments ranging in size from approximately 20 kb to 900 kb (Figure 4a, lane 1). The subclone R342A4-B14 (lane 3) has lost the 250 kb fragment present in R3423A4 and has gained a 340 kb fragment not seen in the parent hybrid. R342A4-B15 (lane 4) has lost both the 250 kb fragment and a weakly hybridising 170 kb fragment from R342A4, while R342A4-C3 (lane 5) has lost only the 250 kb fragment. The fact that the same 250 kb fragment is rearranged, or lost, from the parent hybrid in all three sublines suggests that the breakpoints in these lines lie close to the q arm boundary of the 2.2 Mb alphoid array identified by this probe.

The satellite 2 probe identifies a single fragment of approximately 900 kb in *Hin*dIII-digested R342A4 DNA (>450 kb in Figure 4b, lane 1). The three sublines contain no sequences homologous to this probe (lanes 3–5) showing that this array is not present in any of the sublines. This places the 900 kb satellite 2 array on the q arm side of the alphoid array. This is consistent with a previous map location of 10q11.2 for this array (13).

The satellite 3 probe identifies 4 fragments in R342A4 of 320 kb, 150 kb, 120 kb and 40 kb (Figure 4c, lane 1). All 3 sublines have lost the 320 kb hybridising fragment present in R342A4 but retain the smaller fragments (lanes 3–5). Thus, satellite 3 sequences lie on both sides of the breakpoints present in these cell lines. No hybridising fragments are seen in the hamster cell line (lane 2) confirming that the signals seen in R342A4 and the sublines are of human origin. To determine if the 320 kb *Hin*dIII fragment lost from the sublines (Figure 4c) defines a discrete array of sequences physically separated from the smaller *Hin*dIII fragments which are retained in these sublines, R342A4-B15 was analysed further (Figure 4d). Fragments ranging in size from 180 kb to <50 kb are present in all single and double digests. By comparing these fragment sizes with those seen in identical digests using the parent hybrid R342A4 (Figure 3c), it is clear that the 320–370 kb fragment which is present in all 6 digests of R342A4 is absent from R342A4-B15 (Figure 4d). Therefore, this fragment must define a single array of satellite 3 sequence, of approximately 320 kb, which is on the q arm side of the breakpoint in R342A4-B15. Since the breakpoints in the R342A4 sublines disrupt the alphoid array (Figure 4a), the satellite 3 sequences retained in these sublines (Figure 4d) must lie on the p arm side of the alphoid array. In addition, the satellite 3 probe identifies a minimum of 2 hybridising fragments in digests of R342A4-B15 DNA (Figure 4e) suggesting that the satellite 3 sequences retained in this subline are present in two discrete arrays. A rough estimate of the size of both of these arrays, derived from the fragment sizes in Figures 3(c) and 4(c and d), is 100–150 kb each.

Long range mapping of the satellite arrays on chromosome 10

We have performed a series of experiments using infrequently-cutting restriction enzymes to investigate linkage between the satellite arrays. As with other workers (22), we find that most methylation-sensitive infrequently cutting enzymes cleave very poorly around the satellite arrays, particularly the alphoid array, producing smeared and inconsistent fragment patterns. Several enzymes, such as *Sma*I and *Nru*I, give consistent results with the satellite 2 and satellite 3 probes, but the fragment sizes are not significantly larger than the arrays identified by these probes

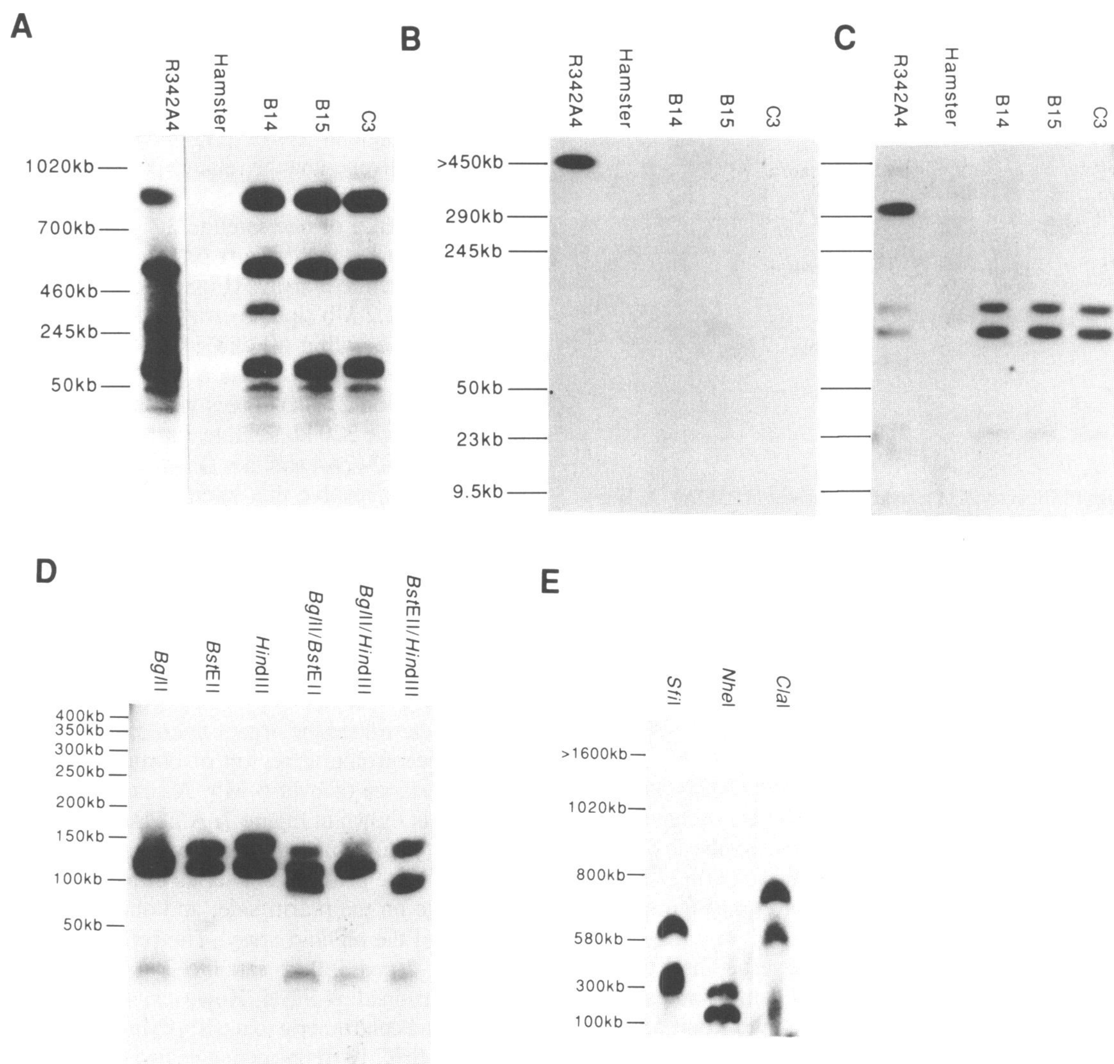

Figure 4. Southern analyses of repetitive sequences in R342A4 and derivative cell lines B14, B15 and C3. **A**; alphoid probe (pα10RP8), *Xba*I digested DNA. **B**; satellite 2 probe (p375M2.4), *Hin*dIII digested DNA. **C**; satellite 3 probe (mC219.28), *Hin*dIII digested DNA. **D** and **E**; R342A4-B15 DNA, satellite 3 probe (mC219.28). Electrophoresis was performed using the CHEF DRII (see Materials and Methods). All filters were washed in 0.5×SSC/0.1% SDS at 65°C.

and so provide no evidence of linkage (data not shown). However, the enzyme *Bss*HII gives unambiguous results with all the satellite probes and these are presented in Figure 5.

The alphoid probe gives a strong hybridising band of approximately 3.0 Mb and a weaker hybridising band of approximately 4.8 Mb in *Bss*HII-digested R342A4 DNA (Figure 5a, lane 1). When the filter used for this hybridisation is stripped and rehybridised with the satellite 2 probe (Figure 5a, lane 2) a strong hybridising band of approximately 2 Mb and a less intense band of approximately 4.8 Mb are observed. The satellite 3 probe gives a more complex pattern (lane 3), hybridising to four fragments. Three of these are the same size as those recognised by the alphoid and satellite 2 probes. The fourth band of hybridisation lies in a zone of compression under the electrophoretic conditions used here (less than 1.1 Mb in size). To determine the number and size of fragments in this zone of compression, and their mapping position relative to the breakpoints in the R342A4 sublines, hybrid DNAs were digested with *Bss*HII and electrophoresed under conditions which separate fragments of less than 1.5 Mb (Figure 5b). In the size range of interest (< 1.1 Mb) only a single fragment of 300 kb is present in R342A4 (lane 1). This fragment is lost from the subclone R342A4-B15 (lane 2).

The simplest interpretation of these results is that the approximately 4.8 Mb band, to which all three probes hybridise, is a partial digestion product which can be cleaved to give the 2 Mb fragment, to which both the satellite 2 and satellite 3 probes hybridise, and the 3 Mb fragment to which both the alphoid and satellite 3 probes hybridise. This is supported by the fact that the enzyme *Bss*HII does not cut within the 2.2Mb alphoid array, or the 900 kb satellite 2 array (data not shown). The 300 kb *Bss*HII fragment homologous to the satellite 3 probe which is lost from R342A4-B15 (Figure 5b) must be derived from the 320 kb array which is the only satellite 3 array lost from the R342A4 sublines (Figures 3c and 4d). Therefore, if the 4.8 Mb fragment is a composite of the 3.0 Mb and 2.0 Mb fragments then it follows that the 300 kb *Bss*HII fragment must lie on the q arm side of the alphoid and satellite 2 arrays present on these two fragments (it cannot lie on the p arm as it is lost from the R342A4 sublines). The fact that the satellite 3 probe hybridises to both the 300 kb

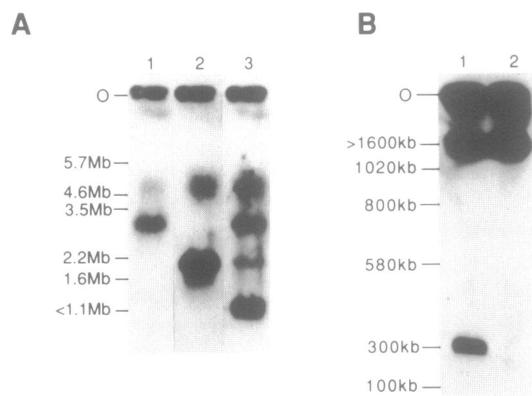

Figure 5. Long range PFGE analysis of repetitive sequences in R342A4. **A; lane 1**—alphoid probe (pα10RP8), **lane 2**—satellite 2 probe (p375M2.4), **lane 3**—satellite 3 probe (mC219.28). All lanes contain R342A4 DNA digested with *Bss*HII. Electrophoresis was performed using the CHEF mapper. **B**; satellite 3 probe (mC219.28), **lane 1**—R342A4 DNA, **lane 2**—R342A4-B15 DNA. Electrophoresis was performed using the CHEF DRII (see Materials and Methods). The wells are shown (○) to facilitate comparison between lanes. All filters were washed in 0.5×SSC/0.1% SDS.

and the 2.0 Mb *Bss*HII fragments suggests that the 320 kb satellite 3 array contains a *Bss*HII restriction site. However, we have not ruled out the possibility that the hybridisation to the 2.0 Mb *Bss*HII fragment is due to a second, much smaller, q arm satellite 3 array not detected in our initial PFGE analyses (Figures 3c and 4d).

Thus the PFGE and hybrid mapping data suggest a linear order for the satellite arrays we have identified and demonstrates that they span a region of at least 5.3 Mb in size. Two small arrays (approximately 100–150 kb) of satellite 3 are present on the p arm side of a 2.2 Mb alphoid array. A 900 kb satellite 2 array is present on the q arm side of the alphoid array and a further 320 kb satellite 3 array lies telomeric to this array.

Confirmation of sequence arrangement using fluorescent *in situ* hybridisation

To confirm the mapping information obtained from the PFGE data we performed a series of fluorescent *in situ* hybridisation (FISH) experiments on metaphase spreads of R342A4 using all three probes. Initially, each probe was used individually to confirm the specificity of these probes to the human chromosome 10 in R342A4 (data not shown). Double labelling experiments were then performed to investigate the relative position of these sequences on chromosome 10. These are presented in Figure 6.

The combination of the alphoid (FITC green) and satellite 2 (Texas red) probes produces one major signal from each (Figure 6d.1). The two signals overlap, but the satellite 2 signal is clearly on the q arm side of the alphoid signal. In Figure 6(d.2), where the alphoid (Texas red) and satellite 3 (FITC green) probes are used in combination, it is apparent that the alphoid signal lies between two satellite 3 signals (the q arm satellite 2 signal appears as a symmetrical doublet). In Figure 6(d.3) where the combination of the probes is satellite 2 (Texas red) and satellite 3 (FITC green), one satellite 3 signal is clearly on the p arm side of the satellite 2 signal and appears as a symmetrical doublet. However, the second satellite 3 signal and the satellite 2 signal overlap are too close together to be distinguished by metaphase cytogenetics.

The observation of a single signal with the alphoid and satellite 2 probes (Figure 6d.1) is consistent with these sequences being arranged as single arrays of 2.2 Mb (Figure 3a and b) and 900 kb (13) respectively, and the relative position of the signals agrees with the mapping data (Figure 4) and the linkage data (Figure 5a). The presence of one satellite 3 signal on the p arm side of the main alphoid signal (Figure 6d.2) is expected since the small (< 150 kb) satellite 3 arrays (Figure 4c) must, by virtue of their linkage to the 2.2 Mb alphoid array on a 3.0 Mb *Bss*HII fragment (Figure 5a), lie within 600 kb of each other on the p arm. A single satellite 3 signal on the q arm side of the main alphoid signal (Figure 6d.2) also agrees with the PFGE analyses which identifies a single 320 kb satellite 3 array distal to the breakpoints present in the R342A4 sublines (Figures 3c, 4c and d). However, the inability to resolve this satellite 3 signal from the satellite 2 signal (Figure 6d.3), while being consistent with the linkage of these two sequences on a 2 Mb *Bss*HII fragment (Figure 5a) fails to confirm the linear order of the 900 kb satellite 2 and 320 kb satellite 3 arrays.

DISCUSSION

We have determined the organisation of three satellite sequences in the pericentromeric region of chromosome 10, spanning a physical distance of over 5 Mb. A schematic representation of this region is shown in Figure 7. A 2.2 Mb alphoid array is linked to a 900 kb satellite 2 array which maps to 10q11.2 (13). Satellite 3 sequences are present as a series of at least three arrays, two of which lie on the p arm side, and one of which lies on the q arm side, of the alphoid array. The relative position of the 900 kb satellite 2 and 320 kb satellite 3 arrays on the q arm could not be determined by FISH. However, satellite 2 sequences have been mapped centromeric to a t(X;10) breakpoint in 10q11.2 (13) whereas cMEN219 (representing locus *D10S130*), which contains satellite 3 sequences, maps telomeric to this breakpoint (37). This supports the inference from the PFGE data that the 320 kb satellite 3 array lies telomeric to the 900 kb satellite 2 array in 10q11.2. The size of satellite arrays and the restriction enzymes which identify them can vary considerably between different copies of one chromosome (25–27), although the 2.2 Mb alphoid array falls within previous estimates for this array on chromosome 10 (25). Because of this, the detailed information on this map is specific to the chromosome 10 present in R342A4, with only the approximate size and linear order of the arrays likely to be shared with other chromosomes.

The five satellite arrays present in R342A4 account for 3.7 Mb of DNA (2.2 Mb + 0.9 Mb + 0.3 Mb + 0.15 Mb + 0.15 Mb) in a region of approximately 5.3 Mb in size (3.0 Mb + 2.0 Mb + 0.3 Mb), leaving 1.6 Mb of DNA to be accounted for in this region of chromosome 10. Some of this DNA is likely to be sequences which have diverged from the consensus sequence of the satellites discussed here. For instance, the cosmid clone cMEN269 (representing locus *D10S134*), contains diverged alphoid DNA sequences, maps to 10cen–q11.2 (36) and identifies several *Bss*HII fragments of between 100 and 400 kb in the hybrid R342A4 (M.S.J. unpublished). The cross-hybridisation of this clone to the 2.2 Mb alphoid array complicates the interpretation of PFGE results, but this result, together with the presence of alphoid DNA in mC219.2, demonstrates that further alphoid sequences remain to be characterised in this

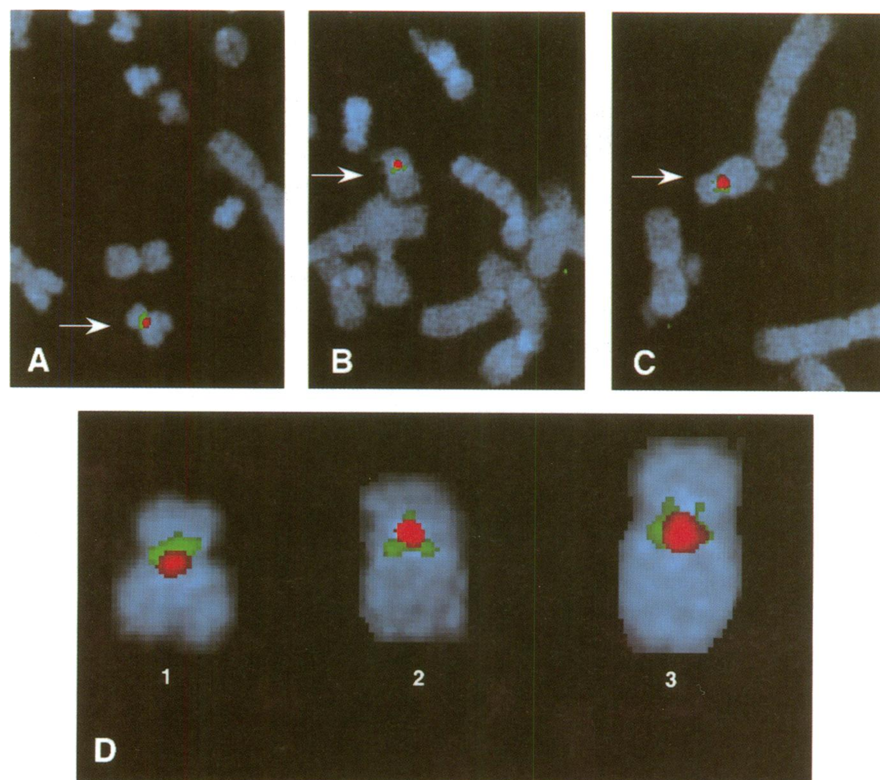

Figure 6. FISH analyses of repetitive sequences in R342A4. The digital images (see Materials and Methods) have been pseudocoloured to distinguish the different fluorochromes. In **A**, **B** and **C** a partial metaphases from each double labelling experiment is presented. Magnified images of the human chromosome 10 from each of these metaphases is presented in **D**. Detection methods and probes used in each case are as follows: **D.1** (magnified from metaphase A): Texas red; satellite 2 probe (p375M2.4), FITC green; alphoid probe (pα10RP8). **D.2** (magnified from metaphase B): Texas red; alphoid probe (pα10RP8), FITC green; satellite 3 probe (mC219.28) **D.3** (magnified from metaphase C): Texas red; satellite 2 probe (p375M2.4), FITC green; satellite 3 probe (mC219.28). For details of metaphase preperation and hybridisation conditions see Materials and Methods.

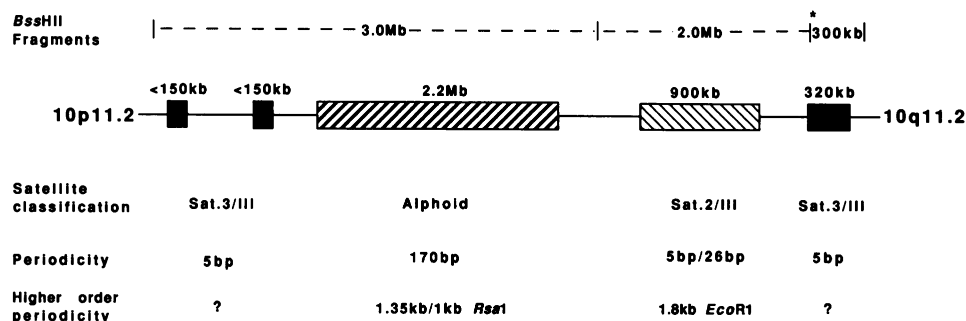

Figure 7. Schematic representation of sequence organisation in the pericentromeric region of chromosome 10 in R342A4. The precise distance beween arrays is unknown and their spacing is arbitrary. The satellite arrays are classified both in terms of the sequence definitions of Prosser *et al.* (satellite 2 or 3, ref. 15) and in terms of the origin of the clones originally used to identify them (satellite II or III, ref. 34). The presence of a single *BssHII* site (*) between the 2.0 Mb and 300 kb fragments has not been confirmed. The periodicities and higher order periodicities are based both on data presented here and on other work (32,33).

region. The analysis of YAC and cosmid clones can circumvent the problems of probe cross-hybridisation and has resulted in the identification of relatively low copy number, novel, repetitive elements on other chromosomes (4,22). Recently a PCR tag derived from *D10S130* (37) has been used to identify a series of YAC clones which contain both satellite 2 and satellite 3 sequences (M.S.J. unpublished). This confirms that arrays of these satellites are linked on chromosome 10 and provides an opportunity to investigate the sequences which lie between them.

The organisation of sequences found here is consistent with current knowledge of satellite organisation on other chromosomes. On the Y chromosome the centromeric alphoid array is closely linked to a large q arm array of the 5 bp satellite and a patchwork of the 48 bp, 68 bp and 5 bp satellite also on the q arm (22). Two alphoid arrays on chromosome 7 have been linked in a region of approximately 3.5 Mb (25) and a novel AT-rich repeat has been identified in the pericentromeric region of this chromosome (4). On the short arms of the acrocentric

chromosomes a series of 5 bp satellite families have been identified (19,30,39) together with the 68 bp satellite (16,17) and a series of alphoid arrays (40). Thus, each human centromeric region appears to possess at least one alphoid array, together with a unique combination of other satellites arranged into discrete, linked, arrays.

The tandemly arranged repeats identified in the pericentromeric region of chromosome 10 are of two basic types: alphoid and the 5 bp satellite family (satellites II and III or 2 and 3). The sequence relationships within the alphoid family have been extensively studied (e.g. 3,5) and the involvement of these sequences in centromere function is well established (7). The sequence organisation of the 5 bp satellite family is less well established, although there is extensive evidence of higher order structures within this family. The 26 bp periodicity observed in the satellite 2 clone p375M2.4 has been reported in clones from chromosomes 1 and 16 (34,41). The abundant 1.8 kb fragment which p375M2.4 identifies in *Eco*R1-digested R342A4 DNA (Figure1) has been observed in pure preparations of classical satellite III (34), on chromosome 15 (39), and with a variety of other restriction enzymes (12,42). In addition, other higher order structures have been identified with 5 bp satellite clones from the acrocentric chromosomes (19,30). However, the functional role, if any, of these sequences is unclear. It has been suggested that satellite DNAs stabilise the chromatin fibre in the cell nucleus (43) and the enrichment of the 5 bp satellite in highly stable DNA–polypeptide complexes (44) supports this. Recently, members of this family have been shown to bind proteins from HeLa cell nuclear extracts with high affinity, and to exhibit unusual hydrogen bonding properties (45). Although the biological significance of these data is unclear, they have been cited, together with the fact that the 5 bp satellites share sequence similarity to the *S. cerevisiae* CDE III element, as evidence that these sequences are an integral part of human centromeres (45). It is clear from our data that discrete arrays of this sequence type are closely linked to alphoid DNA and account for a large proportion of the DNA in the pericentromeric region of chromosome 10.

The data presented here are also relevent to the broader mapping efforts on chromosome 10. Two groups have used markers in the pericentromeric region of this chromosome to identify YACs and to construct contigs (38,46,47). The mapping information presented here will act as an anchor for the physical map of chromosome 10, with the potential to link p and q arm contigs across sequences known to be unstable in YAC vectors (48). Much of the mapping effort has been directed towards the identification of the gene(s) responsible for the MEN type 2 syndromes (49,50). However, the region as a whole is of interest due to the discovery of extensive gene duplication and a pericentric inversion event which have occurred during primate evolution (38). The observation of related arrays of satellite 3 sequences on both sides of the alphoid array may be of particular interest in this respect as inversion events have been invoked to account for the physical separation of related satellite sequences on the Y chromosome (22). The discovery of inversions within alphoid cosmid clones derived from chromosome 17 (4) implies that these may be common events in satellite arrays. Thus, a full understanding of the sequence organisation in this region of chromosome 10 is necessary if the genetic events which have shaped this complex region are to be elucidated.

MATERIALS AND METHODS

Isolation of subclones from cosmid cMEN219

The library from which cMEN219 was isolated, and the mapping of this cosmid is described elsewhere (36,37).This cosmid contains 9 kb of pHS5-related sequence separated from 8 kb of uncharacterised sequence by less than 1 kb of alphoid DNA (data not shown). To rapidly generate subclones, cMEN219 was digested to completion with *Eco*R1 and the digest was re-ligated to itself and transformed into DH5a bacteria. After an initial cracking miniprep screen (51), clones containing inserts were analysed to identify those consisting of single *Eco*R1 fragments re-inserted into the cosmid vector. These were then hybridised with pHS5 and two clones were identified which contained sequences related to this satellite. A 4 kb subclone, mC219.28, consisted entirely of pHS5-related sequences as assayed by Southern analysis (data not shown), and was used for all subsequent hybridisation studies. The second, mC219.2, was mapped to the boundary between pHS5-related sequences and alphoid sequences present in cMEN219 by virtue of an internal *Stu*I restriction site. This clone was sequenced and found to contain 275 bp of satellite III-related sequence abutting directly onto 263 bp of alphoid DNA.

Preparation of PFGE blocks

Cells were harvested, washed twice in PBS and resuspended at a concentration of 5×10^5 cells per 50 μl. An equal volume of 1% LMP agarose in PBS was then added and the cell suspension was immediately aliquoted into block moulds. After chilling, the blocks were incubated for 48 hours in 0.4 M EDTA pH 8.0, 1% lauryl sarcosine and 2 mg/ml proteinase K (Boerhinger Mannheim). Blocks were then washed 3 times in TE and incubated at 50°C in TE containing 0.04 mg/ml PMSF. The blocks were then stored in 0.5 M EDTA at 4°C and washed 3 times in TE before use.

Electrophoresis, Southern transfer and hybridisations

Electrophoresis, restriction enzyme digestion and Southern blotting techniques were carried out using manufacturers recommendations and standard methods (52).

Pulsed Field Gel Electrophoresis (PFGE) was performed using the CHEF DRII system (BioRad). The conditions, and the ranges of separation were as follows.

50 kb–1.6 Mb: 1% agarose, 0.5×TBE buffer with a pulse time of 60 sec. for 15 hrs. at 200 V followed by a pulse time of 90 sec. for 9 hrs. at 200 V.

5–450 kb: 1% agarose, 0.5×TBE buffer with a pulse time of 0.9–29 sec. linearly ramped over 20 hrs. at 200 V.

1.5–6 Mb: 0.6% Rapid agarose (Bethesda Research. Laboratories), 1×TAE, with a pulse time of 30 min. for 72 hrs. at 50 V.

The CHEF Mapper system (Bio-Rad) was also used to resolve DNAs in the size range 2–6 Mb using the conditions recommended by the manufacturers.

Size standards appropriate for each range of seperation were used, these being chromosomes from the *S. cerevisiae* strain YNN295 (Bio-Rad), chromosomes from *S. pombe* strain 972h (Bio-Rad), phage lambda concatamers (New England Biolabs) and phage lambda DNA digested with *Hind*III. Gels were depurinated for 30 minutes in 0.25 M HCl prior to denaturation

and transfer. DNA fragments were transferred to Hybond-N membranes (Amersham International).

DNA probes were labelled using the random oligonucleotide-priming method (53). Filters were washed in 2×SSC at 42°C for low stringency and 0.5×SSC, 0.1% SDS at 65°C for high stringency. Where multiple probeings were required filters were stripped by washing in 0.1 M NaOH/1% SDS at room temperature. The removal of probe was confirmed using autoradiography.

DNA sequencing

Double stranded sequencing templates were prepared according to Kraft *et al.* (54) and di-deoxy sequencing reactions were performed with [^{35}S]dATP as the labelled deoxynucleotide using Sequenase (United States Biochemicals) according to the manufacturers instructions. Both strands of mC219.2 were sequenced using the following primers which flank the *Eco*R1 sites in pWEX15:

pBR4341 5′ AGGCCCTTTCGTCTTCA 3′
WEXT7 5′ CGATGATAAGCGGTCAAACATGAG 3′

Gel readings were assembled into contigs using the IBI AssemblyLign program and analysed using the Macvector software package (International Biotechnologies Inc.).

The EMBL accession number for the sequence of mC219.2 is X74413.

Fluorescent *in situ* hybridisation

Metaphase spreads were prepared from hybrid cells in the standard way (2 hrs. colcemid treatment, 10 min hypotonic treatment and 3×fixation). Slides were stored dessicated at room temperature for between seven and 14 days before use.

The probes pα10RP8, p375M2.4, and mC219.28 were labelled with biotin-16-dUTP (Boehringer Manheim) by nick translation, and hybridised with metaphase spreads in 2×SSC/50% formamide, 10% dextran sulphate, 0.5 mM Tris−HCl, pH 7.6, 0.1 mM EDTA, 0.1 mg/ml sonicated salmon sperm DNA, at 42°C for 16 h. Posthybridisation washing involved 3 washes at 42°C in 2×SSC/50% formamide followed by 3 washes in 0.1×SSC. The biotinylated probe was detected by the two layer avidin−fluorescein isothiocyanate (FITC) detection system (avidin−FITC, biotin−antiavidin, and avidin−FITC) (55), which produced a fluorescent signal at the site of probe hybridisation. Slides were mounted in antifadant solution (AF1; Citiflour) containing DAPI (4′,6-diamidino-2-phenylindole; 0.8 mg/ml) and propidium iodide (0.4 mg/ml) and anlysed under a confocal laser scanning microscope (MRC-600; Bio-Rad Microscience) in order to confirm the presence of the signal on human chromosome 10.

For double-labelling experiments the probes were labelled either with digoxigenin-11-dUTP (Boehringer Manheim) or biotin-16-dUTP (Boehringer Manheim). Detection of digoxigenin was facilitated by using primary mouse monoclonal anti-digoxin antibodies (Sigma), and then secondary antibody FITC conjugated anti-mouse (Sigma). Biotin was detectected by avidin−Texas red (Vector Laboratories), and signal amplified once more with biotinylated antiavidin and another layer of avidin−Texas red. Chromosomes were counterstained with DAPI (4′,6-diamidino-2-phenylindole; 10 mg/ml), and mounted in antifadant solution (AF1; Citiflour). Triple colour images were collected using a computer-controlled Nikon epifluorescence microscope equipped with a cooled charge-coupled device (CCD) camera. FITC, Texas red, and DAPI fluorescences were recorded separately, pseudocoloured to distinguish the different fluorochromes, and merged using a software program developed by Digital Scientific UK. In each experiment a minimum of 10 metaphase spreads was examined

ACKNOWLEDGEMENTS

We would like to thank Dr. C. Tyler-Smith for access to satellite clones, Drs. A. Tunnacliffe and C. Eng for critical reading of the manuscript and two anonymous referees for valuable comments. This work was funded by Medical Research Council H.G.M.P. Grant number G8916654 to B.A.J.P. and a Birthright grant to B.A.J.P. B.A.J.P. is a Gibb Fellow of the Cancer Research Campaign.

REFERENCES

1. Clarke, L. (1990) *Trends Genet,.* **6**, 150−154.
2. Mitchell, A.R., Gosden, J.R. and Miller, D.A. (1985) *Chromosoma*, **92**, 369−377.
3. Choo, K.H., Vissel, B., Nagy, A., Earle, E. and Kalitsis, P. (1991) *Nucleic Acids Res.*, **19**, 1179−1182.
4. Weverick, R., Willard, V.P. and Willard, H. F. (1992) *Genomics*, **14**, 912−923.
5. Alexandrov, I.A., Medvedev, T.D., Mashkova, T.D., Kisselev, L.L, Romanova, L.Y. and Yurov, Y.B. (1993) *Nucleic Acids Res.*, **21**, 2209−2215.
6. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) *J. Cell Biol.*, **109**, 1963−1973
7. Haaf, T., Warburton, P.E. and Willard, H.F. (1992) *Cell*, **70**, 681−696.
8. Voullaire, L.E., Slater, H.R., Petrovic, V. and Choo, K.H.A. (1993) *Am. J. Hum. Genet.*, **52**, 1153−1163.
9. Pluta, A.F., Cooke, C.A. and Earnshaw, W.C. (1990) *Trends Biochem. Sci.*, **15**, 181−185.
10. Earnshaw, W.C., Ratrie, H., and Stetten, G. (1989) *Chromosoma*, **98**, 1−12.
11. Gosden, R.J., Mitchell, A.R., Buckland, R.A., Clayton, R.P. and Evans, H.J. (1975) *Expl. Cell. Res.*, **92**, 148−158.
12. Beauchamp, R.S., Mitchell, A.R., Buckland, R.A. and Bostock, C.J. (1979) *Chromosoma*, **71**, 153−166.
13. Jackson, M.S., Mole, S.E. and Ponder, B.A.J. (1992) *Nucleic Acids Res.*, **20**, 4781−4787.
14. Kalitsis, P., Earle, E., Vissel, B., Shaffer, L.G. and Choo, K.H.A. (1993) *Genomics*, **16**, 104−112.
15. Prosser, J., Frommer, M., Paul, C. and Vincent, P.C. (1986) *J. Mol. Biol.*, **187**, 145−155.
16. Agresti, A., Rainaldi, G., Lobbiani, A., Magnani, I., DiLernia, R., Menervi, R., Siccardi, A.G. and Ginelli, E. (1987) *Hum. Genet.*, **75**, 326−332.
17. Waye, J.S. and Willard, H. F. (1989) *Chromosoma*, **97**, 475−480.
18. Metzdorf, R., Gottert, E. and Blin, N. (1988) *Chromosoma*, **97**, 154−158.
19. Choo, K.H., Earle, E., Vissel, B. and Kalitsis, P. (1992) *Am. J. Hum. Genet.*, **50**, 706−716.
20. Johnson, D.H., Kroisel, P.M,. Klapper, H.J. and Rosenkranz, W. (1992) *Hum. Molec. Genet.* **1**, 741−747.
21. Lin, C.C., Sasi, R., Lee, C. and Court, D. (1993) *Chromosoma*, **102**, 333−339.
22. Cooper, K.F., Fisher, R.B. and Tyler-Smith, C. (1992) *J. Mol. Biol.*, **228**, 421−432.
23. Schwartz, D.C. and Cantor, C.R. (1984) *Cell*, **37**, 67−75.
24. Tyler-Smith, C. and Brown, W.R.A. (1987) *J. Mol. Biol.*, **195**, 457−470.
25. Wevrick, R. and Willard, H.F. (1991) *Nucleic Acids Res.*, **19**, 2295−2301.
26. Oakey, R. and Tyler-Smith, C. (1990) *Genomics*, **7**, 325−330.
27. Rocchi, M., Archidiacono, N., Ward, D.C. and Baldini, A. (1991) *Genomics*, **9**, 517−523.
28. Mitchell, A.R., Jeppesen, P., Hanratty, D. and Gosden, J. (1992) *Chromosoma*, **101**, 333−341.
29. Tyler-Smith, C. (1987) *Development*, **101**, 93−100.

30. Vissel, B., Nagy, A. and Choo, K.H.A. (1992) *Cytogenet. Cell Genet.*, **61**, 81–86.
31. Nakamura, Y., Mathew, C.G.P., Sobol, H., Easton, D.F., Telenius, H., Bragg, T., Chin, K., Clark, J., Jones, C., Lenoir, G.M., White, R. and Ponder, B.A.J. (1989) *Genomics*, **5**, 199–203.
32. Devilee, P., Kievits, T., Waye, J.S., Pearson, P.L. and Willard, H.F. (1988) *Genomics*, **3**, 1–7.
33. Looijenga, L.H.J., Oosterhuis, J.W., Smit, V.T.H.B.M., Wessels, J.W., Mollevanger, P. and Devilee, P. (1992) *Genomics*, **13**, 1125–1132.
34. Cooke, H.J. and Hindley, J. (1979) *Nucleic Acids Res.*, **6**, 3177–3197.
35. Trask, B.J., Ger van den Engh, M.C., Massa, H.F., Gray, J.W. and Van Dilla, M. (1991) *Som. Cell Mol. Genet.*, **17**, 117–136.
36. Mole, S.E., Jackson, M.S., Tokino, T., Nakamura, Y. and Ponder, B.A.J. (1993) *Genomics*, **15**, 457–458.
37. Tunnacliffe, A., Jackson, M.S., Gardner, E., Moore, J.K., Mole, S.E., Mulligan, L.M., Graham, A., Finocchiaro, G., Ørstavik, S and Ponder, B.A.J. (1993) *Hum. Genet.* (in press).
38. Tunnacliffe, A., Liu, L., Moore, J.K., Leversha, M.A., Jackson, M.S., Papi, L., Ferguson-Smith, M.A., Thiesen, H.-J. and Ponder, B.A.J. (1993) *Nucleic Acids Res.* **21**, 1409–1417.
39. Higgins, M.J., Wang, H., Shtromas, I., Haliotis, T., Roder, J.C., Holden, J.J.A. and White, B.N. (1985) *Chromosoma*, **93**, 77–86.
40. Vissel, B. and Choo, K.H.A. (1992) *J. Mol. Evol.*, **35**, 137–146.
41. Moyzis, R.K., Albright, M.F., Bartholdi, M.F., Cram, L.S., Deaven, L.L, Hildebrand, C.E., Joste, N.E., Longmire, J.L., Meyne, J. and Schwarzacher-Robinson, T. (1987) *Chromosoma*, **95**, 375–386.
42. Jeanpierre, M., Turleau, C., Aurius, A. Prieur, M., Ledeist, F., Fischer, A. and Viegas-Pequignot, E. (1993) *Hum. Mol. Genet.*, **2**, 731–735.
43. Vogt, P. (1990) *Hum. Genet.*, **84**, 301–336.
44. Pfutz, M., Gileadi, O. and Werner, D. (1992) *Chromosoma*, **101**, 609–617.
45. Grady, D.L., Ratliff, R.L., Robinson, D.L., McCanlies, E.C., Meyne, J. and Moyzis, R.K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1695–1699.
46. Mole, S.E., Mulligan, L.M., Healey, C.S., Ponder, B.A.J. and Tunnacliffe, A. (1993) *Hum. Mol. Genet.* **2**, 247–252.
47. Lairmore, T.C., Dou, S., Howe, J.R., Chi, D., Carlson, K., Veile, R., Mishra, S.K., Wells, S.A. and Donis-Keller, H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 492–496.
48. Neil, D.L., Villasante, A., Fisher, R.B., Vetrie, D., Cox, B. and Tyler-Smith, C. (1990) *Nucleic Acids Res.*, **18**, 1421–1428.
49. Mulligan, L.M., Kwok, J.B.J., Healey, C.S., Elsdon, M.J., Eng, C., Gardner, E., Love, D.R., Mole, S.E., Moore, J.K., Papi, L., Ponder, M.A., Telenius, H., Tunnacliffe, A. and Ponder, B.A.J. (1993) *Nature (London)*, **363**, 458–460.
50. Donnis-Keller, H., Dou, S., Chi, D., Carlson, K.M., Toshima, K., Lairmore, T.C., Howe, J.R., Moley, J.F., Goodfellow, P. and Wells, S.A. (1993) *Hum. Mol. Genet.* **2**, 851–856.
51. Barnes, W. (1977) *Science*, **195**, 393.
52. Sambrook, J., Fitch, E.F. and Maniatis, E.F. (1989) Molecular cloning: a laboratory manual. 2nd Ed. Cold Spring Harbor Laboratory Press, New York.
53. Feinberg, A. and Vogelstein, B. (1984) *B. Anal. Biochem.*, **137**, 266–267.
54. Kraft, R., Tardiff, J., Krauter, K.S. and Leinwand, L.A. (1988) *Biotechniques*, **6**, 544–546, 549.
55. Pinkel, D., Straume, T. and Gray, J.W. (1986) *Proc. Natl. Acad. Sci. USA.* **83**, 2934–2938.