© 2000 Oxford University Press Human Molecular Genetics, 2000, Vol. 9, No. 8 1201–1208

Co-localisation of CCG repeats and chromosome deletion breakpoints in Jacobsen syndrome: evidence for a common mechanism of chromosome breakage

Christopher Jones+, Roman Müllenbach, Paul Grossfeld[1], Rebecca Auer, Remi Favier[2], Ken Chien[1], Michael James[3], Alan Tunnacliffe[4] and Finbarr Cotter

Royal London and St Bartholomew's School of Medicine and Dentistry, Department of Experimental Haematology, Turner Street, London E1 2AD, UK, [1]Department of Medicine, Cardiac Molecular Biology Program, University of California San Diego, La Jolla, CA 92093-0613, USA, [2]Service d'Haematologie Biologique, Hopital d'Enfants Armand Trousseau, 26 Avenue du Docteur Arnold Netter, 75571 Paris Cedex 12, France, [3]Wellcome Centre of Human Genetics, Nuffield Orthopaedic Centre, Headington, Oxford OX3 7BN, UK and [4]Institute of Biotechnology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QT, UK

Received January 11 2000; Revised and Accepted March 1 2000

Folate-sensitive fragile sites are associated with the expansion and hypermethylation of CCG-repeats. The fragile site in 11q23.3, *FRA11B*, has been shown to cause chromosome deletions *in vivo*, its expression being associated with Jacobsen (11q–) syndrome. However, the majority of Jacobsen deletions are distal to *FRA11B* and are not related to its expression. To test the hypothesis that other unidentified fragile sites might be located in 11q23.3–24 and may cause these deletions, we have identified and characterised CCG-trinucleotide repeats within a 40 Mb YAC contig spanning distal chromosome 11q. Only eight CCG-repeats were identified within the entire YAC contig (not including *FRA11B*), six of which map to the region of 11q23.3–24 that includes Jacobsen deletions. We have previously collated the deletion mapping data of 24 Jacobsen patients with the physical map of chromosome 11q, and accurately localised six breakpoints to short intervals corresponding to individual YAC clones. We now show that in each of these cases, YAC clones found to contain a deletion breakpoint also contain a CCG-repeat. The improved analysis of one of these deletions, together with those of several new Jacobsen cases, further strengthens this association by localising five breakpoints to individual PAC clones containing CCG-repeats. These data provide strong evidence for the non-random clustering of chromosome deletion breakpoints with CCG-repeats, and suggests that they may play an important role in a common mechanism of chromosome breakage.

INTRODUCTION

There are many developmental syndromes caused by partial chromosome deletion, involving most chromosome arms. However, little is known about the general mechanisms giving rise to chromosome breakage. We have chosen to study the chromosome deletions of Jacobsen syndrome patients as a model to investigate the mechanisms of chromosome breakage. Jacobsen (11q–) syndrome is a rare example of such a disease (~1 in 100 000 live births) characterised by deletions of the long arm of chromosome 11, typically from band 11q23 to the telomere (1). The subsequent loss of a large number of genes results in a complex and varied phenotype, typically including trigonocephaly, cardiac anomalies and Paris Trousseau thrombocytopenia (1,2).

Folate-sensitive fragile sites are inherited chromosome abnormalities that are characterised by a constriction of the metaphase chromosome and which can cause chromosome breakage under certain experimental conditions (3). They are caused by the extensive expansion of CCG-repeats, typically from several copies of the CCG-trinucleotide in the normal population to several hundred copies on chromosomes expressing the fragile site. Intermediate repeat expansions which predispose to instability of the repeat in future generations, but which do not cause fragile site expression, are known as premutations. In addition, the larger CCG-repeat expansions that cause fragile site expression are also associated with the hypermethylation of adjacent CpG islands (4–11).

Until recently, the observation that fragile sites cause chromosome breakage *in vitro* was thought to be an artefact of the conditions used to visualise their expression, and was not expected to occur *in vivo*. However, the fragile site at chromosome band 11q23.3 (*FRA11B*) has been shown to have caused chromosome deletions in at least two Jacobsen syndrome patients. In both of these cases, the deletion derived from a maternal chromosome carrying a *FRA11B* CCG-repeat expansion, and the deletion breakpoint was found to map close to, and include, the CCG-repeat (11). Considering the rarity of these two chromosome abnormalities, it is extremely unlikely that their relationship in these two cases is due to chance.

+To whom correspondence should be addressed. Tel: +44 207 377 7000 ext. 3152; Fax: +44 207 377 7629; Email: cjones2@hgmp.mrc.ac.uk

Whilst many chromosome deletions in Jacobsen syndrome have been shown to map close to *FRA11B*, most appear not to be caused by expression of this site but are located ~1–2 Mb distal to it (12,13). Many other folate-sensitive fragile sites on other chromosomes have been reported and in addition to *FRA11B*, four have previously been characterised at the molecular level (*FRAXA*, *FRAXE*, *FRAXF* and *FRA16A*) and shown to be caused by extensive CCG-repeat expansions (4–10). However, unlike *FRA11B*, none of these sites has been shown to cause chromosome deletions *in vivo*. In addition, several *FRA11B*-expressing families have been reported with no history of Jacobsen syndrome (11). While such discrepancies in the effects of fragile site expression remain to be explained, a more general role for fragile site expression in causing chromosome deletions needs to be examined.

A mechanism by which fragile sites may mediate chromosome breakage has been suggested following the analysis of their replication timing. Folate-sensitive fragile site loci are late replicating regions of the genome, and their replication is delayed further with fragile site expression (14,15). *In vitro* experiments have demonstrated that unusual DNA structures, including hairpins (16–18) and intrastrand tetraplexes (19), are formed by expanded CCG-repeats, which may form a barrier to DNA synthesis and hence cause the delay of replication at a fragile site (20,21). Replication delay has been implicated in causing deletions at the aphidicolin-inducible fragile site, *FRA3B*; under conditions that bring about expression of the fragile site, some cells enter G_2-phase without having completed replication of the region (22). In addition, replication timing of *FRA3B* has been shown to be allele specific; two distinct alleles are replicated at different stages of the cell cycle, and chromosome rearrangements preferentially occur on the late replicating chromosome (23). Whilst the mechanisms of expression of aphidicolin-inducible and folate-sensitive fragile sites are apparently unrelated, the common feature of their delayed replication would appear to be pertinent to their potential role in causing chromosome rearrangements.

We have recently reported the assembly of a 40 Mb YAC contig covering the distal long arm of chromosome 11 from 11q22 to the telomere, and described its multi-functional use, including the mapping of Jacobsen syndrome chromosome deletion breakpoints (13). The contig was used to map two chromosome deletions by fluorescence *in situ* hybridisation (FISH) analysis using YACs from the contig as probes, and also allowed another 22 deletions, previously analysed by haplotype analysis and described in the literature, to be placed on the physical map. Whilst it was clear from this analysis that the majority of deletions have only been relatively crudely mapped, the breakpoints of several deletions could be accurately localised to individual YAC clones and suggested clustering of breakpoints in some regions (13). The accurate localisation of deletion breakpoints in this way therefore provides the opportunity to investigate the mechanisms of their genesis.

In order to test the hypothesis that previously unidentified folate-sensitive fragile sites in chromosome bands 11q23–24 might be responsible for some of the chromosome deletions distal to *FRA11B*, we have analysed the YAC contig for the presence of CCG-trinucleotide repeats. A total of eight CCG-repeats were identified, six of which were within the region known to contain Jacobsen syndrome deletions. These repeats were found to co-localise with the best-mapped deletion breakpoints described in the literature, and also with those of several new Jacobsen cases described in this study. These data provide strong evidence for the non-random clustering of chromosome deletion breakpoints with CCG-repeats and we discuss the possible mechanisms by which CCG-repeats may be involved in causing chromosome deletions.

RESULTS

Previous reports have described protocols for CCG-repeat detection in cDNA using hybridisation of CCG-oligonucleotide probes of 30 bases or more, representing 10 or more copies of the CCG-trinucleotide (24,25). However, we have found that these probes often give false positive signals at regions of high GC-content when used on genomic DNA, especially where the trinucleotides CCG, CGC or GCC are in abundance (data not shown). We have examined the efficacy of several CCG-oligonucleotide probes to determine the optimum size that gives specific signals (even at short stretches of contiguous CCG-repeat) and conclude that a probe of only 15 bases (five copies of the CCG-trinucleotide) gives the most specific signals. Larger probes appear not to dissociate effectively from the CCG-rich CpG island templates, even at high temperatures, presumably due to the high melting temperature of large GC-tracts. In contrast, the CCG probe used here gives highly specific signals in normal hybridisation conditions, even with short repeat targets. Hybridisation to a PAC clone containing a CCG-repeat and other clones containing CpG islands demonstrates the specificity of this probe (see Fig. 1a); a clone containing a $(CCG)_{11}$ repeat from the *CBL2* gene (Fig. 1a, lane 1) provides a strong signal with the CCG probe, whereas clones containing the CpG islands of the *p16/MTS-1* and *MLL* genes (Fig. 1a, lanes 3 and 4, respectively) which are both rich in CCG-trinucleotides but do not contain a contiguous CCG-repeat, do not provide a hybridisation signal.

Identification of CCG-repeats within the YAC contig

Southern blots of 120 YAC clones from a 40 Mb contig of distal chromosome 11q, spanning the region 11q22–ter, were hybridised with the CCG probe. Three faint bands of hybridisation are observed in all YAC DNA samples, representing CCG-repeat sequences present in yeast; additional bands of hybridisation are hence specific to the human DNA sequences contained within the YAC insert (see Fig. 1b). From the entire 40 Mb contig analysed, only eight chromosome 11q-specific CCG-repeats were identified (not including *FRA11B*), six of which are within the region known to contain Jacobsen deletion breakpoints (see Figs 1b and 3). We have designated these repeats TNR/11q#1–8 (TriNucleotide Repeat on 11q).

YAC clones containing the CCG-repeats were used as hybridisation probes for direct screening of the RPCI-1 PAC library (26) and specific clones containing six of the CCG-repeats were isolated. Subsequent FISH analysis with these clones confirmed their chromosome localisation (CCG-repeats from other regions of the genome were also identified due to the high incidence of chimerism in the CEPH YAC library, but discounted at this stage; data not shown) and mapping of other PACs identified in these screens allowed the precise localisation of five of the CCG-repeats with respect to local markers. CCG-repeats were then cloned and sequenced in order to

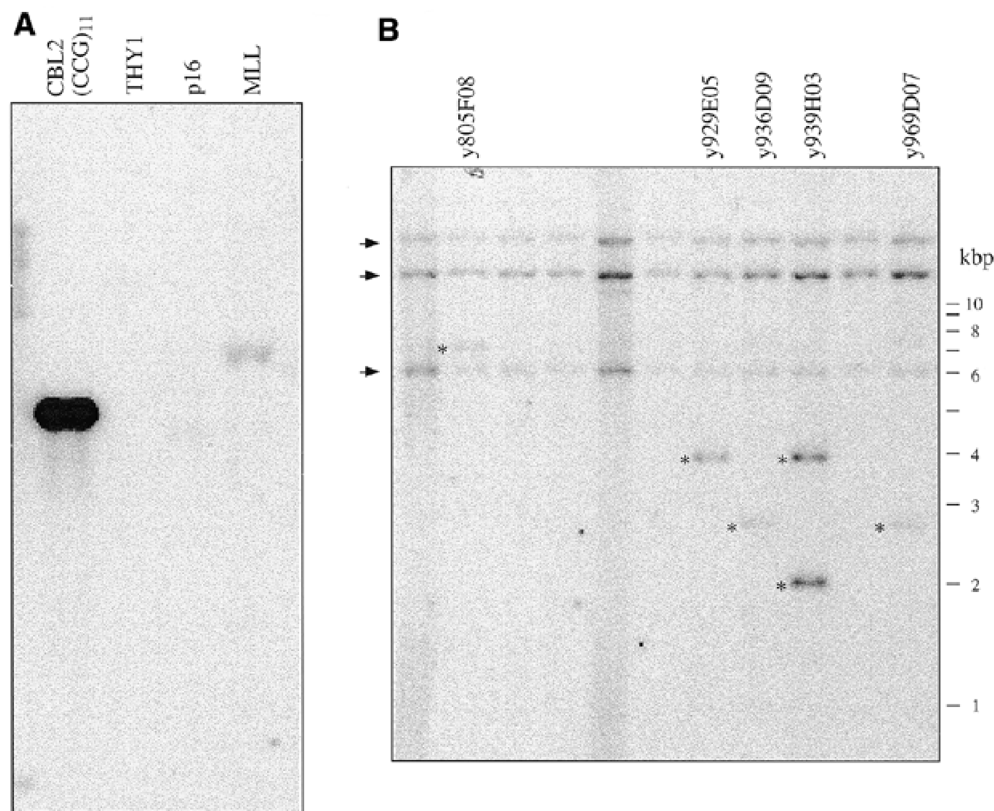

Figure 1. Autoradiograph of YAC and PAC restriction digests, hybridised with the CCG probe. YAC or PAC DNA was digested with *Eco*RI and transferred to nylon filters by Southern blotting, and the CCG probe was end-labelled with [γ^{32}-P]ATP and hybridised to the filters under stringent conditions, as described in Materials and Methods. (**a**) Hybridisation to various PAC clones, demonstrating the specificity of the CCG probe. Lane 1, dJ44-n4 (*CBL2* [CCG]$_{11}$); lane 2, dJ189-a16 (THY1); lane 3, dJ76-i7 (p16/MTS-1); lane 4, dJ217-a21 (5′ MLL). (**b**) Hybridisation to various YAC clones from chromosome 11q, demonstrating the detection of CCG-repeats. The locations of the yeast-specific CCG-repeats are denoted by the arrows at the left of the figure, and the positions of size markers on the gel are shown at the right of the figure. Human (YAC) specific CCG-repeats are identified by an asterisk to the left of the hybridising band, and the identity of those clones containing CCG-repeats are denoted above the figure. The reader will note the presence of two CCG-repeats detected in the YAC clone y939H03, where only one is referred to in the text: we have previously reported that this YAC is chimeric, containing material from chromosome 3p in addition to 11q23. The lower of these two bands of hybridisation was found to derive from chromosome 3p and hence is not discussed.

confirm their identity. Figures 2 and 3 show the mapping data assembled, including the locations of each of these CCG-repeats with respect to nearby markers.

It is of interest to note that of the CCG-repeats present within the YAC contig, all except TNR/11q#2 are located in regions containing one or more unstable YAC clones, accounting for most regions of YAC instability in the contig (including a chimeric YAC that contains a CCG-repeat from another chromosome). Analysis of YACs spanning *FRA11B* demonstrated that the two clones that should contain the *CBL2* CCG-repeat (y856B09 and y720A01) were highly unstable and rarely contained that repeat (11). It is clear from this analysis that there is a fundamental problem concerning the stability of CCG-repeats cloned in YACs, and that this instability might account for many YAC rearrangements. For this reason we suspect that there may be additional repeats within this region of chromosome 11 that we have failed to identify.

CCG-repeats co-localise with previously mapped chromosome deletion breakpoints

The locations of 23 Jacobsen syndrome deletion breakpoints have been described in the literature (1,11,12) and placed on the physical map (13). Whilst the accuracy of these analyses varies widely, several breakpoints have been accurately mapped to short regions corresponding to single YAC clones or less (reviewed in 13; see Fig. 3). Comparison of the locations of these breakpoints and of the CCG-repeats that map distal to *FRA11B* is revealing: all but one of the minimal intervals that contain these accurately mapped deletion breakpoints contain one or more of the CCG-repeats identified. Moreover, all of the CCG-repeats distal to *FRA11B* coincide with a YAC containing a deletion breakpoint (see Fig. 3). This coincidence is further strengthened by the improved analysis of a previously reported deletion, together with the analysis of deletions from five new patients.

These repeats and their coincidence with chromosome deletion breakpoints are described below, starting with the most centromeric of the repeats identified. It should be noted that the sequences are described in order to reflect the longest contiguous repeat, and that regardless of orientation the trinucleotides listed as GGC, CCG, GCC and CGC are effectively equivalent.

TNR/11q#1

TNR/11q#1 was identified within a panel of overlapping YAC clones, accurately localising the repeat to between the markers D11S1897 and D11S2105 in chromosome band 11q22–23.1.

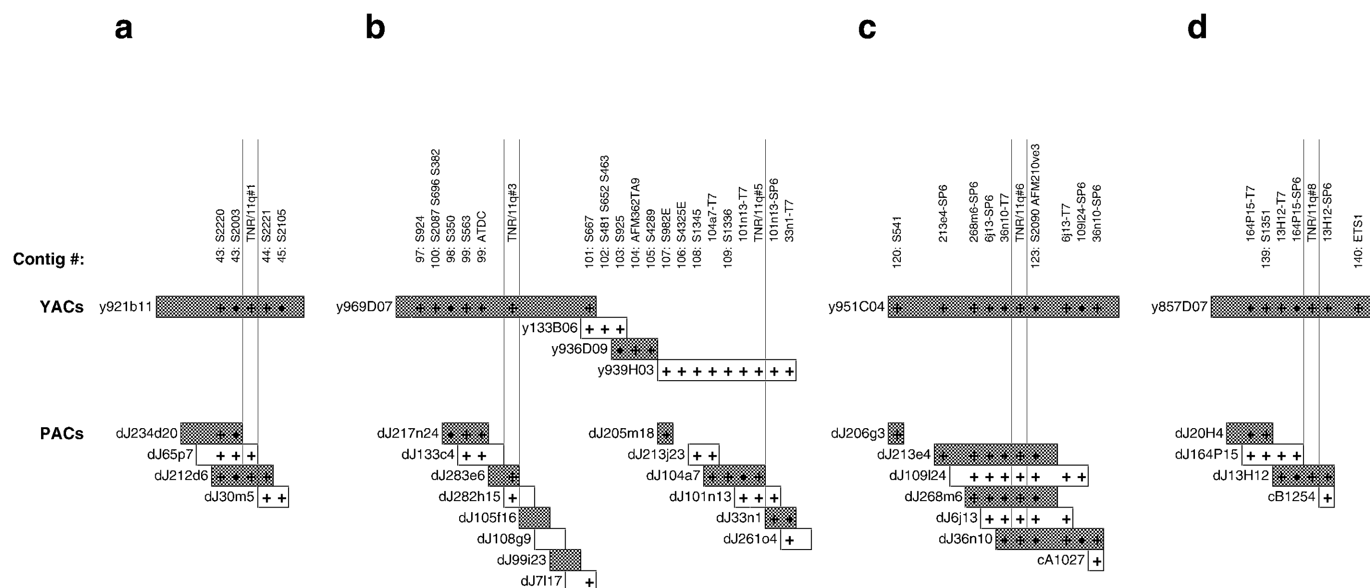

Figure 2. Localisation of each of the CCG-repeats that were mapped in detail (i.e. beyond the resolution of a single YAC clone) with respect to local markers. Not all markers present within each YAC are necessarily shown, but only those markers that are pertinent to the mapping of the CCG-repeat or are described in the text. The orientations of two contig positions within panel (b) have been changed from their original reported positions in Tunnacliffe *et al.* (13) in both this figure and Figure 3. Markers at position 100 are not contained within the PAC contig that extends between positions 98 and 101, and therefore must actually be located between positions 97 and 98. The marker at position 107 (D11S982E) has previously been shown to lie between 105 and 106 (13). Clones prefixed with the letter y are from the CEPH mega-YAC library (34) and have been described previously (13); clones prefixed with the letter c are from the RZPD chromosome 11-specific cosmid library no. c107 (35) (http://www.rzpd.de/) and clones prefixed with the letters dJ are from the RPCI-1 PAC library (26) (http://bacpac.med.buffalo.edu/). Details of markers are available online from http://shows.med.buffalo.edu/

Extension of a PAC contig between these two markers further localised the repeat to between markers D11S2003 and D11S2221 (see Fig. 2a). Sequence analysis revealed the presence of a **(GCC)**$_9$TTC**(GCC)**$_2$ repeat in one of these YACs, y975H06, and PCR analysis demonstrated that it is polymorphic (R. Auer, C. Jones, R. Müllenbach, D. Milligan, C. Fegan and F. Cotter, manuscript in preparation) varying in length between **(GCC)**$_3$TTC**(GCC)**$_2$ and **(GCC)**$_{12}$TTC**(GCC)**$_2$. This demonstrates that even short stretches of interrupted trinucleotides should be considered as a repeat, and also provides clues as to their origins; we have not found any repeat length between **(GCC)**$_3$TTC**(GCC)**$_2$ and **(GCC)**$_9$TTC**(GCC)**$_2$ in the normal population, suggesting that the longer repeat may have evolved from a duplication of the entire **(GCC)**$_3$TTC**(GCC)**$_2$ motif, followed by the subsequent mutation of the first TTC to GCC.

TNR/11q#2

TNR/11q#2 is a **(CGC)**$_5$(CAC)$_2$**CGC** repeat located within YAC y911F02, close to the *APOA* gene in chromosome band 11q23.2. Whilst both of these repeats are located centromeric to *FRA11B* and all other published Jacobsen deletion breakpoints, there is evidence of a deletion breakpoint cluster within the same YAC containing TNR/11q#2. A recent report described the mapping of a region of loss of heterozygosity within 11q23, observed in mantle cell lymphoma (27). The majority of the deletions characterised in this study were interstitial, with the most telomeric breakpoint localised between YACs y785E12 and y911F02 in 12 out of 20 deletions analysed (27), an ~750 kb region that contains TNR/11q#2 (see Fig. 3).

TNR/11q#3

This interrupted repeat **(GGC)**$_2$GAC**GGC**TCC**GGC**AGC**GGC**TCCCGC**(GGC)**$_4$ was identified within YAC y969D07, which is located ~1 Mb distal to *FRA11B* in chromosome band 11q23.3. A PAC contig across part of this region further localised TNR/11q#3 between markers D11S563/ATDC and D11S667 (see Figs 2b and 3).

The YAC y969D07 has been shown to contain the deletion breakpoint of at least one Jacobsen patient [VH (12); see Fig. 3]. FISH analysis demonstrated that y969D07 was lost from the deletion chromosome of this patient. However, the polymorphic marker D11S924 (which is contained within y969D07) was shown by haplotype analysis to be heterozygous in the patient, demonstrating that the YAC must be partially retained on the deletion chromosome (12). Together, these data suggest that whilst the breakpoint is distal to D11S924, it is close to the proximal end of y969D07 which contains TNR/11q#3.

TNR/11q#4

A CCG-repeat was identified within YAC y936D09 in chromosome band 11q23.3 (see Figs 2b and 3) but has not been characterised further. However, the deletions of two previously undescribed patients with del(11) (q23.3–ter) karyotypes were characterised and found to be located within short intervals containing this YAC (patients BS and RS; see Fig. 3). FISH analysis was used to characterise the deletion of patient RS; YAC y133B06 was retained on the del(11) derivative, whereas a PAC containing marker D11S982E (dJ205-m18) was lost. This places the deletion breakpoint between D11S667 (the most

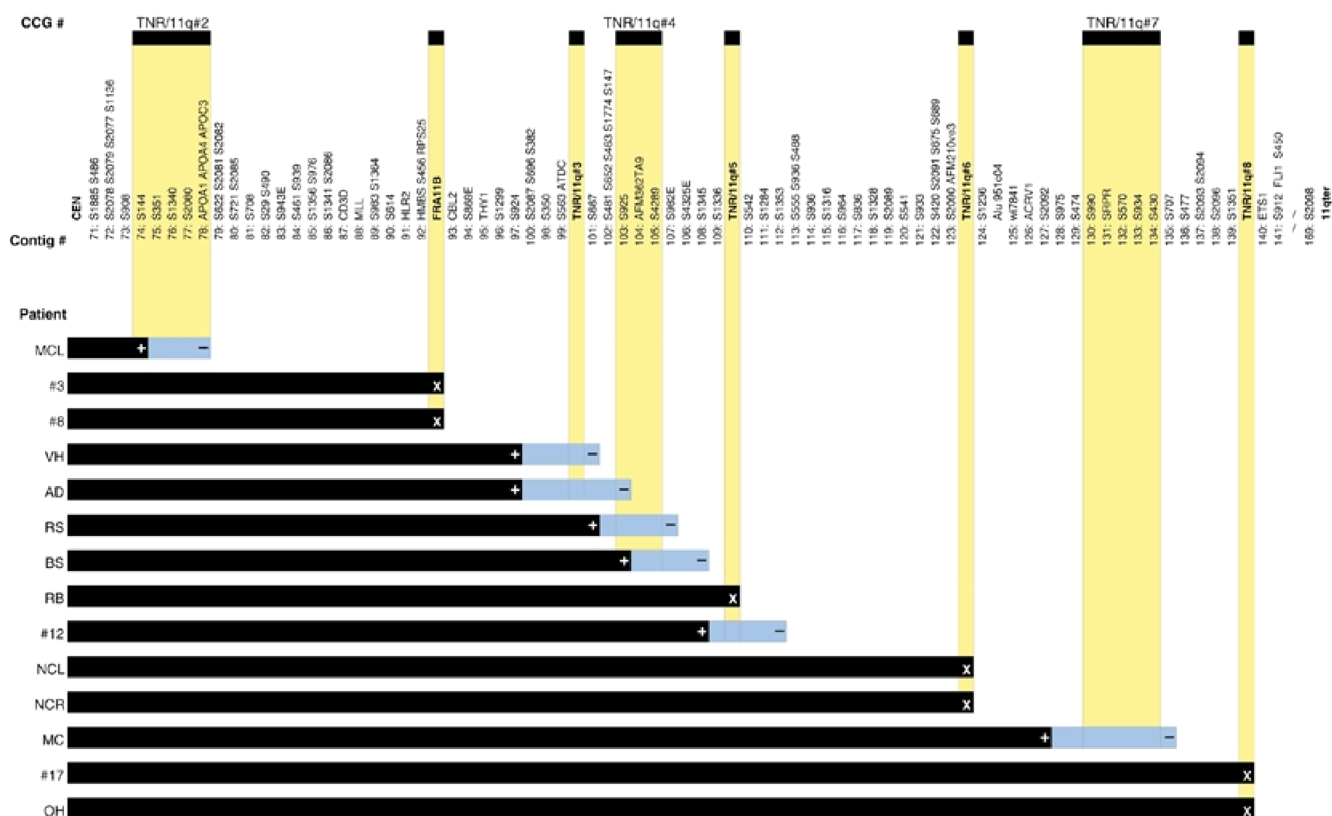

Figure 3. Co-localisation of chromosome deletion breakpoints and CCG-repeats from distal chromosome 11q. Part of the distal chromosome 11q YAC contig is represented by the contig positions and respective markers, including CCG-repeats characterised in this study (denoted by the TNR/11q locus names as described in the text). The intact region of the deleted chromosome is represented by a solid (black) bar. Breakpoints mapped to the resolution of a single position/PAC clone are represented by a white X. Where breakpoints have been mapped to a region spanning multiple markers, the minimal interval containing the breakpoint is represented by a shaded (grey) bar; the retained marker is denoted as (+) and the deleted marker denoted as (–). Note: the results of FISH analyses with YACs are always the most pessimistic interpretation; the most proximal marker of a retained YAC is counted as the retained signal, and the most distal marker of a deleted YAC is counted as the deleted signal. Patient IDs: MCL, mantle cell lymphoma and refers to the common region of chromosome breakage in that disease (27); patients #3 and #8 were originally described by (1) and were shown to have breakpoints at *FRA11B* by (11); patients #12 and #17 are from (1); patients VH and AD are from (12); patients RB and MC are from (13); patients RS, BS, NCL, NCR and OH are described in this study.

proximal marker of y133B06) and D11S982E (see Fig. 3). PCR with microsatellite markers from 11q23 was used to identify common haplotypes between patient BS and his mother, and demonstrated that the deletion breakpoint was between markers D11S925 (retained) and D11S1345 (deleted) (see Fig. 3). The shared interval containing these two breakpoints therefore includes the 0.8–1.4 Mb region spanned by YAC y936D09.

TNR/11q#5

A CCG-repeat was identified within the overlapping YAC clones y939H03 and y929E05, in chromosome band 11q23.3. We have previously described the construction of a PAC contig across this region and the subsequent mapping of a deletion breakpoint to within a single PAC clone (patient RB) (13). Hybridisation analysis of this PAC contig with the CCG probe identified a $(GCC)_6$ repeat within two overlapping PAC clones dJ104-a7 and dJ101-n13, the latter of which was also shown to contain the deletion breakpoint of patient RB and is estimated to be 75 kb long.

In addition, the breakpoint of patient #12 (1) lies between markers D11S1345 and D11S1353 (see Fig. 3): an interval of ~850 kb that overlaps with that of patient RB and hence also contains TNR/11q#5.

TNR/11q#6

This repeat was identified within YAC y951C04, in chromosome band 11q23.3. Five overlapping PAC clones containing TNR/11q#6 were isolated and found to also contain markers D11S2090 and AFM210ve3 (see Fig. 2c). Sequencing of the region identified two repeats, $(CCG)_7$ and **CCG**$(CAG)_3$ **(CCG)**$_3$CG**(CCG)**$_5$, separated by 295 bp and inverted with respect to each other. This presented extreme difficulties in sequencing the region, which is discussed in detail in Materials and Methods.

We had previously localised the chromosome deletions of two Jacobsen syndrome patients, a mother and her son, to y951C04 by FISH analysis [patients NCL and NCR; both with del(11) (q23.3–ter) karyotypes and previously undescribed]. In both cases the PAC clone dJ206-g3 (marker D11S541) which is contained within y951C04, was retained on the del(11) derivative, whereas y951C04 itself appeared to be lost. This suggested that the deletion breakpoint was likely to be within y951C04 and hence TNR/11q#6 was a strong candidate for the location of the breakpoint in this family. FISH analysis was used to demonstrate that the breakpoint was indeed within the same PAC clones that contain the CCG-repeat; four of the PAC clones containing TNR/11q#6 were lost from the deleted

chromosome, whereas one of the PAC clones (dJ213-e4) gave a retained signal in ~50% of metaphases analysed. This maps the breakpoint to the same 120 kb region spanned by these PACs that also contains TNR/11q#6.

TNR/11q#7

A CCG-repeat was identified within a 780 kb YAC y805F08 in chromosome band 11q23.3–24 (see Fig. 3) but has not been characterised further. However, this YAC clone also contains the breakpoint of patient MC (13); FISH analysis demonstrated that an overlapping YAC clone (y975F11) was retained, and the immediately distal marker (D11S707) was lost from the del(11) derivative of this patient (13).

TNR/11q#8

This short $(GGC)_4$ repeat is contained within YAC y857B09, close to the *ETS1* and *FLI1* genes in chromosome band 11q24. The breakpoint of patient #17 (1) lies between the markers D11S1351 and D11S912, which are also contained within this YAC (13). A PAC contig containing TNR/11q#8 was extended to include marker D11S1351, and clones from this contig were used in FISH analysis of metaphase chromosomes from patient #17 and another patient (OH) whose chromosome deletion was shown by karotype analysis to be in chromosome band 11q24. The breakpoint of both of these patients was found to map to the same 75 kb PAC clone (dJ13-h12) that contains the CCG-repeat; the centromeric-overlapping PAC clone dJ164-p15 (D11S1351) was retained on the del(11) derivative, and the telomeric-overlapping cosmid clone cB1254 was deleted from the del(11) derivative of both patients.

DISCUSSION

The locations of eight CCG-trinucleotide repeats on the long arm of chromosome 11 have been determined. Of these, six are in the region of chromosome 11q that is known to contain Jacobsen syndrome chromosome deletions. In addition, we have shown that all of the most accurately mapped deletion breakpoints each co-localise with these CCG-repeats; to three regions representing single PAC clones of between 75 and 120 kb in length, and to four regions representing single YAC clones of between 750 kb and 1.0 Mb in length. Whilst further analysis of these breakpoints will be required to precisely correlate these observations, these data provide strong evidence for the non-random clustering of chromosome deletion breakpoints with CCG-repeats.

The observation that CCG-repeats may be common sites of chromosome breakage raises the possibility that these loci may represent previously unidentified folate-sensitive fragile sites. If this is the case, it is surprising that none of these sites have been recognised cytogenetically, although the possibility that their expression has been confused for *FRA11B* may explain this discrepancy. Indeed our laboratory has identified one such potential case, of an individual who expresses a folate-sensitive fragile site in 11q23.3, but who lacks CCG-repeat expansion at the *FRA11B* locus. The failure to detect these fragile sites cytogenetically could also be explained by differences in the stability of folate-sensitive fragile sites. The majority of *FRA11B*-expressing individuals have no family history of Jacobsen syndrome (11), indicating that whilst *FRA11B* can cause chromosome deletions, this is actually a rare event. In contrast, other fragile sites may be highly unstable and may only ever be manifested as a chromosome deletion.

One of the patients analysed in this study who suffered a chromosome deletion at TNR/11q#5 (RB) has also been found to have a *FRA11B* premutation of approximately $(CCG)_{80}$. In addition, we have also detected an anonymous methylated CCG-repeat in one of the families described previously (Family 2, C. Jones, unpublished data). These data present the possibility that some individuals may exhibit a general CCG-repeat instability and express more than one fragile site, suggesting an underlying DNA-repair defect. Interestingly the role of the DNA repair gene *MSH2* in the destabilisation of trinucleotide repeat length in Huntington disease has recently been demonstrated (28), further suggesting that *MSH2* or similar genes may also be involved in the regulation of CCG-repeat stability.

PCR and Southern analysis of CCG-repeat lengths of the families of Jacobsen syndrome patients will demonstrate whether or not expansions at these loci does indeed occur, and we are collecting family materials with a view to answering this question. Alternatively, it is possible that expansion of these repeats does not occur, and they are not the locations of folate-sensitive fragile sites as described by cytogenetic criteria. The mechanism of the *in vitro* cytogenetic manifestation of folate-sensitive fragile sites is not well understood, and it is still not clear whether their *in vitro* expression fully reflects their *in vivo* behaviour. We should not discount the possibility that the expression of folate-sensitive fragile sites *in vitro* is merely an artefact of CCG-repeat expansion, and that other characteristics of these sites are reproduced at other CCG-repeats, regardless of their length.

The role of replication delay in causing chromosome deletions has been described for the aphidicolin-inducible fragile site *FRA3B* (22,23). The folate-sensitive fragile site loci are also late replicating regions, and their expression has been shown to further delay their replication (14,15). However, it is not clear whether repeat expansion as opposed to hypermethylation of the CCG-repeat (or necessarily both) is the cause of replication delay at folate-sensitive fragile sites. Whilst it has been suggested that CCG-repeat expansion alone would be sufficient to delay replication (20,21) the role of methylation should not be overlooked. Imprinted regions of the genome have been shown to be replicated asynchronously, demonstrating delayed replication of the imprinted allele (29,30). It is of interest to note that there is a strong parental bias towards the 11q deletions in Jacobsen syndrome (deletion breakpoints below marker D11S925 are all paternal in origin) (1) and genetic imprinting has been suggested as a potential cause of this bias. Considering the role of methylation in the imprinting process, it is possible that imprinting-mediated methylation of CCG-repeats on chromosome 11 leads to further delay of replication at these sites, and therefore to chromosome deletion. Analysis of the timing of replication across chromosome 11q would help to determine whether late replication is a common feature of all CCG-repeats, and hence whether this could be the underlying cause of deletion at these loci.

In summary, this study has demonstrated a distinct correlation between the sites of CCG-trinucleotide repeats and of chromosome deletion in Jacobsen syndrome. The data presented implicate CCG-repeat sequences in a common mechanism of chromosome breakage.

MATERIALS AND METHODS

Identification of CCG-containing YAC and PAC clones

One microgram of YAC or PAC miniprep DNA was digested with *Eco*RI and electrophoresed on a 1% agarose gel, which was transferred to Zeta-probe nylon membrane (Bio-Rad, Hercules, CA) by alkaline capillary transfer (31). Pre-hybridisation with 0.2 mg/ml sonicated salmon sperm DNA was performed in a standard SSPE/dextran-sulphate/Denhardt's hybridisation buffer (31) at 65°C for 2 h. The pre-hybridisation solution was removed and the filter rinsed in hybridisation buffer at 65°C for 15 min to remove excess salmon sperm DNA. The $(CCG)_5$ oligonucleotide probe was end-labelled with 20 µCi [γ-^{32}P]ATP (ICN) using polynucleotide kinase (New England Biolabs, Beverley, MA) and then hybridised to the filter at 50°C for 16 h. Filters were washed for 15 min in 2× SSC/0.1% SDS, followed by washing for 15 min in 0.2× SSC/0.1% SDS at 55°C, before exposure to X-OMAT AR X-Ray film (Kodak, Rochester, NY).

Identification of PAC clones by direct YAC hybridisation

High density filters of the RPCI-1 PAC library (obtained from the Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, UK) (26) were hybridised with 150 ng of YAC miniprep DNA, labelled with [α-^{32}P]dCTP (ICN) by oligo-labelling (32). YAC-specific PAC clones were grown in 50 ml 'terrific broth' containing kanamycin (25 µg/ml) and DNA was prepared by a protocol involving multiple precipitations. Briefly, the alkaline lysate is precipitated with isopropanol, and the pellet resuspended in TE to which a one-third volume of 10 M ammonium acetate is added. Following incubation on ice, the precipitate is removed by centrifugation and the supernatant is added to 2 vol of ethanol. DNA and RNA are precipitated at room temperature and the pellet resuspended in TE and treated with RNase, after which the DNA is precipitated with 13% PEG_{8000}/10 mM $MgCl_2$.

Secondary screening was performed by PCR of 1 ng PAC DNA using appropriate STS markers and standard protocols, and also hybridised with the CCG probe to determine which PACs contain a CCG-repeat. Ordering of PACs was refined either by hybridisation of PAC end-probes (generated by vectorette-PCR) (33) or by PAC to PAC hybridisation, using standard techniques.

Fragments containing CCG-repeats were cloned from PACs into either pUC or M13 cloning vectors, and sequenced using BigDye sequencing chemistry on an ABI377 automated sequencer. In many cases, the region immediately surrounding the CCG-repeat was very difficult to sequence from double-stranded templates, and sequencing from a single-stranded M13 template was necessary. In the case of TNR/11q#6, sequencing from both M13 and pUC templates proved impossible; sequencing dramatically halted in a reproducible manner regardless of template preparation, primer or sequencing buffer. Despite these difficulties we found that it is possible to PCR across this region, using a deaza–dGTP nucleotide mix and 5% DMSO in addition to the standard PCR buffer. The resulting deaza–dGTP-template was then able to be sequenced (albeit sub-optimally). The reason for this apparent contradiction is that *Taq*-mediated cycle sequencing is of course a linear reaction; the starting template, which when using pUC or M13 templates is a dGTP template, has to be re-used at each cycle. However, PCR is an exponential reaction; a single successful round of PCR using the deaza–dGTP analogue will provide a template that will then be re-used in successive cycles. The cause of the difficulty in sequencing this region is discussed in the text.

Fluorescence *in situ* hybridisation (FISH)

Metaphase chromosomes of patients were obtained following short-term culturing of fresh blood. Cells were grown in RPMI/10% fetal calf serum (Gibco BRL, Gaithersburg, MD) and metaphase chromosomes and spreads were prepared using standard techniques. Probes were labelled with either biotin-16-dUTP or digoxygenin-11-dUTP using a nick-translation protocol (Gibco BRL). Labelled PAC DNA (500 ng) was pre-annealed with 50 µg Cot1 DNA (Gibco BRL) at 37°C for 90 min in hybridisation solution (2× SSC, 50% formamide, 10% dextran sulphate). Metaphase spreads were denatured in 70% formamide/2× SSC at 70°C for 3 min and pre-annealed probe was hybridised in a moist chamber at 37°C overnight. A Texas-Red-conjugated chromosome 11 α-satellite probe (Intergenco, New York, NY) was co-hybridised to specifically detect both chromosomes 11.

Detection of biotin-labelled probes was achieved using 'sandwiches' of FITC-conjugated avidin and biotinylated anti-avidin (Vector, Burlingame, CA); and detection of digoxygenin-labelled probes was achieved using Texas Red-conjugated anti-digoxygenin (Boehringer Mannheim, Mannheim, Germany). Slides were mounted in antifadant solution (AF1; Citifluor, London, UK) containing DAPI (4,6-diamino-2-phenylindole; 10 µg/ml). Staining and hybridisation signals were analysed under a computer-controlled Nikon epifluorescence microscope equipped with a cooled charge-coupled device camera. Computer images were merged and Q-banding (generated by DAPI staining) converted to G-banding, using IP Lab Spectrum software (Scanalytics, Fairfax, VA).

For all probes, hybridisation signals were only counted where the identity of both chromosomes 11 was clear; if the deleted chromosome 11 was not easily distinguished from the normal chromosome 11, then no signal was recorded. At least 15 informative metaphases were counted for each probe.

ACKNOWLEDGEMENTS

We would like to thank all of the patients and their families for the generous gift of blood samples that made this work possible. The very special support of Mrs Annet van Betuw and all of the families of The European Chromosome 11q Network (http://home.wxs.nl/~avbetuw/) is once again gratefully acknowledged, as is the support of the 11q Research and Resource Group (http://ucsu.colorado.edu/~biasca/11q.html). We would also like to thank Nigel and Val Barrett (Unique; http://members.aol.com/rarechromo/index.htm) for their enthusiasm. We thank Dr Laura Penny for providing new patient samples and for communicating unpublished results and The Phyllis and Sidney Goldberg Medical Research Trust for support to F.C. This research was supported by a Leukaemia Research Fund grant to F.C., C.J., R.A. and R.M. This work is dedicated to Billy Hendrix Jones.

REFERENCES

1. Penny, L.A., del Aquila, M., Jones, M.C., Bergoffen, J.A., Cunniff, C., Fryns, J.P., Grace, E., Graham, J.M., Kousseff, B., Mattina, T. *et al.* (1995) Clinical and molecular characterization of patients with distal 11q deletions. *Am. J. Hum. Genet.*, **56**, 676–683.
2. Breton-Gorius, J., Favier, R., Guichard, J., Cherif, D., Berger, R., Debili, N., Vainchenker, W. and Douay, L. (1995) A new congenital dysmegakaryopoietic thrombocytopenia (Paris-Trousseau) associated with giant platelet alpha-granules and chromosome 11 deletion at 11q23. *Blood*, **85**, 1805–1814.
3. Sutherland, G.R. and Hecht, F. (1985) *Fragile Sites on Human Chromosomes.* Oxford University Press, Oxford.
4. Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J.C., Warren, S.T., Schlessinger, D. *et al.* (1991) Fragile X genotype characterised by an unstable region of DNA. *Science*, **252**, 1179–1182.
5. Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F. *et al.* (1991) Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–914.
6. Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R. and Richards, R.I. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p (CCG)n. *Science*, **252**, 1711–1714.
7. Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, R.G., Warren, S.T. *et al*. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.
8. Knight, S.J., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton Price, H.R., Barnicoat, A., Pembrey, M.E. *et al.* (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in *FRAXE* mental retardation. *Cell*, **74**, 127–134.
9. Parrish, J.E., Oostra, B.A., Verkerk, A.J.M.H., Richards, C.S., Reynolds, J., Spikes, A.S., Shaffer, L.G. and Nelson, D.L. (1994) Isolation of a GCC repeat showing expansion in *FRAXF*, a fragile site distal to *FRAXA* and *FRAXE. Nature Genet.*, **8**, 229–235.
10. Nancarrow, J.K., Kremer, E., Holman, K., Eyre, H., Doggett, N.A., Le Paslier, D., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1994) Implications of *FRA16A* structure for the mechanism of chromosomal fragile site genesis. *Science*, **264**, 1938–1941.
11. Jones, C., Penny, L., Mattina, T., Yu, S., Baker, E., Voullaire, L., Langdon, W.Y., Sutherland, G.R., Richards, R.I. and Tunnacliffe, A. (1995) Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene *CBL2*. *Nature*, **376**, 145–149.
12. Michaelis, R., Velagaleti, G.V.N., Jones, C., Pivnick, E.K., Phelan, M.C., Boyd, E., Tarleton, J., Wilroy, R.S., Tunnacliffe, A. and Tharapel, A.T. (1998) Most Jacobsen syndrome deletion breakpoints occur distal to *FRA11B*. *Am. J. Med. Genet.*, **76**, 222–228.
13. Tunnacliffe, A., Jones, C., Le Paslier, D., Todd, R., Cherif, D., Birdsall, M., Devenish, L., Yousry, C., Cotter, F.E. and James, M.R. (1999) Localisation of Jacobsen syndrome breakpoints on a 40-Mb physical map of distal chromosome 11q. *Genome Res.*, **9**, 44–52.
14. Subramanian, P.S., Nelson, D.L. and Chinault, A.C. (1996) Large domains of apparent delayed replication timing associated with triplet repeat expansion at *FRAXA* and *FRAXE*. *Am. J. Hum Genet.*, **59**, 407–416.
15. Hansen, R.S., Canfield, T.K., Fjeld, A.D., Mumm, S., Laird, C.D. and Gartler, S.M. (1997) A variable domain of delayed replication in *FRAXA* fragile X chromosomes: X inactivation-like spread of late replication. *Proc. Natl Acad. Sci. USA*, **94**, 4587–4592.
16. Mitas, M., Yu, A., Dill, J. and Haworth, I.S. (1995) The trinucleotide repeat sequence d (CGG)15 forms a heat-stable hairpin containing Gsyn.Ganti base pairs. *Biochemistry*, **34**, 12803–12811.
17. Pearson, C.E. and Sinden, R.R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**, 5041–5053.
18. Yu, A., Barron, M.D., Romero, R.M., Christy, M., Gold, B., Dai, J., Gray, D.M., Haworth, I.S. and Mitas, M. (1997) At physiological pH, d (CCG)15 forms a hairpin containing protonated cytosines and a distorted helix. *Biochemistry*, **36**, 3687–3699.
19. Usdin, K. (1998) NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucleic Acids Res.*, **26**, 4078–4085.
20. Usdin, K. and Woodford, K.J. (1995) CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.*, **23**, 4202–4209.
21. Samadashwily, G.M., Raca, G. and Mirkin, S.M. (1997) Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.*, **17**, 298–304.
22. Le Beau, M.M., Rassool, F.V., Neilly, M.E., Espinosa, R., Glover, T.W., Smith, D.I. and McKeithan, T.W. (1998) Replication of a common fragile site, *FRA3B*, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum. Mol. Genet.*, **7**, 755–761.
23. Wang, L., Darling, J., Zhang, J.S., Huang, H., Liu, W. and Smith, D.I. (1999) Allele-specific late replication and fragility of the most active common fragile site, *FRA3B*. *Hum. Mol. Genet.*, **8**, 431–437.
24. Li, S.H., McInnis, M.G., Margolis, R.L., Antonarakis, S.E. and Ross, C.A. (1993) Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. *Genomics*, **16**, 572–579.
25. Kleiderlein, J.J., Nisson, P.E., Jessee, J., Li, W.B., Becker, K.G., Derby, M.L., Ross, C.A. and Margolis, R.L. (1998) CCG repeats in cDNAs from human brain. *Hum. Genet.*, **103**, 666–673.
26. Ioannou, P.A. and de Jong, P.J. (1996) Construction of bacterial artificial chromosome libraries using the modified P1 (PAC) system. In Dracopoli, N.C., Haines, J.L., Korf, B.R., Moir, D.T., Morton, C.C., Seidman, C.E., Seidman, J.G. and Smith, D.R. (eds), *Current Protocols in Human Genetics*. John Wiley and Sons, New York, NY, Unit 5.15.
27. Monni, O., Zhu, Y., Franssila, K., Oinonen, R., Hoglund, P., Elonen, E., Joensuu, H. and Knuutila, S. (1999) Molecular characterization of deletion at 11q22.1–23.3 in mantle cell lymphoma. *Br. J. Haematol.*, **104**, 665–671.
28. Manley, K., Shirley, T.L., Flaherty, L. and Messer, A. (1999) *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genet.*, **23**, 471–473.
29. Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D. and Cedar, H. (1993) Allele-specific replication timing of imprinted gene regions. *Nature*, **364**, 459–463.
30. Lin, M.S., Zhang, A. and Fujimoto, A. (1995) Asynchronous DNA replication between 15q11.2q12 homologs: cytogenetic evidence for maternal imprinting and delayed replication. *Hum. Genet.*, **96**, 572–576.
31. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
32. Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**, 6–13.
33. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C. and Markham, A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.*, **18**, 2887–2890.
34. Cohen, D., Chumakov, I. and Weissenbach, J. (1993) A first generation physical map of the human genome. *Nature*, **366**, 398–400.
35. Nizetic, D., Monard, S., Young, B., Cotter, F., Zehetner, G. and Lehrach, H. (1994) Contruction of cosmid libraries from flow-sorted human chromosomes 1, 6, 7, 11, 13 and 18 for reference library resources. *Mamm. Genome*, **5**, 801–802.