© 1996 Oxford University Press
Human Molecular Genetics, 1996, Vol. 5, No. 2 **187–195**

FRA3B extends over a broad region and contains a spontaneous HPV16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites

Charles M. Wilke*, Bryan K. Hall, Ann Hoge, William Paradee[1], David I. Smith[1] and Thomas W. Glover

Department of Pediatrics and Department of Human Genetics, University of Michigan, 2818 MS II 0618, Ann Arbor, MI 48109, USA and [1]Departments of Molecular Biology, Genetics and Internal Medicine, Wayne State University School of Medicine, 540 East Canfield, Detroit, MI 48201, USA

Received August 22, 1995; Revised and Accepted November 8, 1995

The common fragile site at 3p14.2 (FRA3B) is the most sensitive site on normal human chromosomes for the formation of gaps and breaks when DNA replication is perturbed by aphidicolin or folate stress. Although rare fragile sites are known to arise through the expansion of CCG repeats, the mechanism responsible for common fragile sites is unknown. Beyond being a basic component of chromosome structure, no biological effects of common fragile sites have been convincingly shown, although suggestions have been made that breakage and recombination at these sites may sometimes be mechanistically involved in deletions observed in many tumors and in constitutional deletions. In an observation related to the high rate of recombination at fragile sites, a number of studies have shown a statistical association between the integration of transforming DNA viruses and chromosomal fragile sites. Using FISH analysis we recently identified a 1.3 Mb YAC spanning both FRA3B and the t(3;8) translocation associated with hereditary RCC. Here we report the further localization of FRA3B within this YAC. Using lambda subclones of the YAC as FISH probes, gaps and breaks were found to occur over a broad region of at least 50 kb. Neither CCG nor CAG repeats were found in this region suggesting a different mechanism for fragility than seen with rare fragile sites. We further show that an area of frequent gaps and breaks within FRA3B, defined by a lambda contig, coincides with a previously characterized site of HPV16 integration in a primary cervical carcinoma. The HPV16 integration event gave rise to a short chromosomal deletion limited to the local FRA3B region within 3p14.2. Interestingly, 3p14.2 lies within the smallest commonly deleted region of 3p in cervical cancers, which are often HPV16 associated. To our knowledge this is the first molecular characterization of an *in vivo* viral integration event within a confirmed fragile site region, supporting previous cytogenetic observations linking viral integration sites and fragile sites.

INTRODUCTION

Chromosomal fragile sites are loci which show gaps or breaks on metaphase spreads of cells that have been grown in the presence of inhibitors of DNA replication (1). Following induction, these sites show increased sister chromatid exchange (2,3), a high rate of translocation and deletion in somatic cell hybrid systems (4–6) and an increased rate of recombination with pSV2neo DNA artificially introduced into cells (7). It has also been suggested that fragile sites may play a mechanistic role in chromosome breakage and rearrangements involved in cancer (8,9).

The sequences of five ‘rare’ fragile sites have been cloned and characterized and in all cases the mutation leading to the expression of fragility is the expansion of a CCG trinucleotide repeat (10–14). The rare fragile site mutations therefore represent a subclass of the recently discovered repeat expansion mutations involved in a number of human diseases. In the case of rare fragile sites, the molecular instability observed at the nucleotide level is translated through an apparent change in chromatin structure to a higher-order instability observed at the level of the chromosome. This change in chromatin structure can give rise to disease by modifying the expression of neighboring genes, as in the case of FMR1 (15), or by mediating chromosome deletion as is suggested in the 11q- (Jacobsen) syndrome (14).

Common fragile sites, on the other hand, have not been convincingly implicated in disease processes. Even though these sites make up the vast majority of fragile sites, little is known

*To whom correspondence should be addressed

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

about their structure or biological role. While the rare fragile sites arise via expansion of normally occurring CCG repeats and represent aberrations in nucleotide and chromatin structure, the large number of common fragile sites are a constant feature of the genome. This constancy of common fragile sites suggests that their overall nucleotide and/or chromatin structure is also unvarying and implies a conservation of function. Fragile sites in general are thought to represent genomic regions that replicate late in S phase of the cell cycle (16) and this has been demonstrated in the case of the FRAXA locus (17). Common fragile sites are induced by agents that perturb DNA replication. Therefore a greater understanding of common fragile sites may lead to a greater understanding of the basic processes of DNA replication and repair and how these may relate to chromatin structure.

A number of cytogenetic studies have demonstrated an association between common chromosomal fragile sites and the sites of integration of human papillomaviruses (18–20). Recently, human papillomavirus infection has been confirmed to be the causative transforming factor in over 90% of cervical cancers worldwide (21). A key step in the progression of benign cervical tumors to invasive carcinomas is the integration of the normal episomally-replicating HPV DNA into the cellular genome. It is therefore possible that common fragile sites could play an important role in the progression of cervical and other cancers, by providing targets for viral integration. However, the actual importance of common fragile sites in facilitating tumor progression through this process has not been established.

Using FISH analysis we recently identified a 1.3 Mb YAC spanning both the t(3;8) translocation associated with hereditary renal cell carcinoma (hRCC) and the common fragile site FRA3B (22,23). Here we report the further localization of FRA3B to a broad region within this YAC and show that a region of frequent breakage within FRA3B coincides with a previously characterized site of HPV16 integration in a primary cervical carcinoma (24). Furthermore, we show that the HPV16 integration event gave rise to a short chromosomal deletion limited to the local FRA3B region within 3p14.2. Interestingly, 3p14.2 lies within the smallest commonly deleted region of 3p LOH in cervical cancers (25). To our knowledge this is the first molecular characterization of an *in vivo* viral integration event within a confirmed fragile site region, supporting previous cytogenetic observations linking viral integration sites and fragile sites.

RESULTS

In our previous experiments, we used FISH analysis on metaphase spreads expressing the common fragile site, FRA3B to determine that YAC850A6 spanned this fragile site (23). Here we extended this approach to sublocalize FRA3B within YAC850A6. In this case, lambda subclones of the YAC were used as FISH probes against aphidicolin treated metaphase spreads to determine the relative positions of the probes to the fragile site.

First, YAC850A6 was subcloned into Lambda FIXII (Stratagene, La Jolla, CA). Hybridization of random lambda subclones to pulsed field blots of *Mlu*I digested YAC DNA was then used to determine the general position of several anchor subclones within YAC850A6. DNA from these subclones was then hybridized to the gridded library to extend small contigs. This 1.3 mb YAC is cleaved five times by *Mlu*I. Several lambda subclones were positioned within the two most telomeric *Mlu*I fragments, D and A (Fig. 1). The initial FISH experiments with three lambda clones from these two regions (L48, L98 and L72) indicated that the general position of FRA3B, as defined by gaps or breaks on aphidicolin treated spreads, is within the *Mlu*I D and A regions. The position of the FISH signals from lambda 48, which spans the *Mlu*I site dividing fragments F from D, was nearly always proximal to the fragile site gaps. Lambda 72, located in fragment A, generally hybridized distal to the breakage. While lambda 98 was more frequently proximal, but also crossed giving FISH signals on both sides and sometimes hybridized distal to the fragile site.

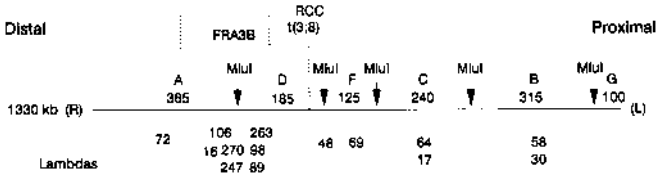

Figure 1. A schematic representation of YAC850A6. Above the map are indicated the positions *Mlu*I sites, along with the position of the t(3;8) translocation associated with hereditary renal cell carcinoma (22). The overall FRA3B region is indicated by the extent of the two stippled lines. The *Mlu*I fragments are labeled from largest to smallest A–G, with the sizes in kilobases shown below each fragment designation. The numbers below the line indicate the general position of relevant lambda clones determined as described in the text. The overall size of the YAC is shown to the left of the map and the L and R indicate the left and right pYAC4 arms, respectively. The distal (telomeric) to proximal (centromeric) orientation of the YAC is indicated.

A large region of frequent breakage is defined by FISH analysis with lambda subclones

Based on these preliminary data a 50 kb lambda contig was extended from lambda 98 (Figs 1 and 2a). This contig contains the *Mlu*I site dividing fragments D and A (Fig. 1) and therefore defines its position as being approximately 365 kb from the telomeric end of the YAC.

Table 1. FISH data for individual lambda clones relative to aphidicolin induced breakage

Clones	*N*	%Prox	%Dist	%Crossing	(%Prox - % Dist)
48	21	95.2	4.8	0	90.4
263	28	75	7.1	10.7	67.9
98	31	41.9	29	6.4	12.9
89	77	33.8	27.3	18.2	6.5
247	33	33.3	27.3	18.2	6.0
270	51	37.3	37.3	17.6	0
16	62	16.1	46.8	21	–30.7
106	60	18.3	50	21.7	–31.7
72	29	3.4	72.4	20.7	–69

N = The number of metaphase spreads scored.
The %Prox, %Dist and %Crossing refer to the percentage of the time that signals were seen proximal, distal or crossing the aphidicolin induced breakage, respectively. Note that these values do not sum to 100%, because the percentage of the time that FISH signals were observed directly in the middle of the breakage was not included in this calculation.

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

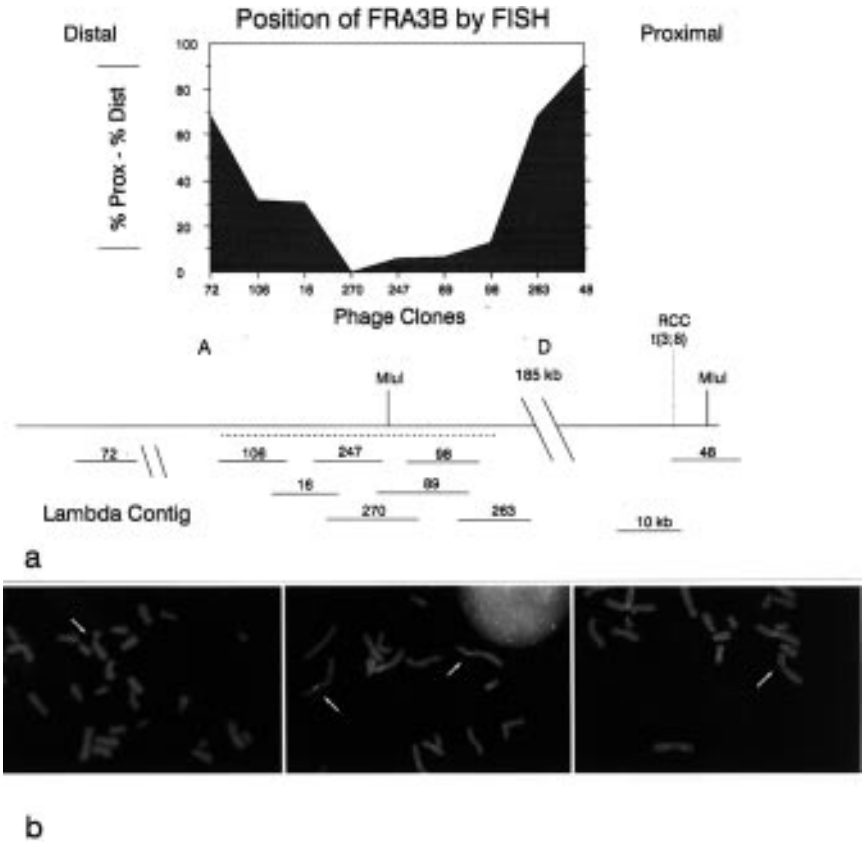

Figure 2. (**a**) The graphical results are expressed in terms of the absolute value of the difference between the percentage of time that a clone was distal or was proximal to the fragile site. The lower the value, the more often a hybridization signal was seen equally on both sides of the fragile site. The lambda contig (phage clones 106–263, represented by overlapping lines) spanning the *Mlu*I site dividing the A and D fragments and defining the region of aphidicolin induced breakage is shown below the graph. This contig spans 50 kb. The distance between this contig and clone 72 is unknown. (**b**) The three photographs show partial metaphase spreads from a single experiment in which the fragile site was induced with aphidicolin and lambda 247 used as a FISH probe. From left to right the hybridization signals indicated by the arrows can be seen to occur proximal, on both sides and distal to the fragile site, respectively.

In order to better localize the region of gaps and breaks defining FRA3B, individual members of this lambda contig were used in FISH experiments on metaphase spreads expressing the fragile site. Interestingly, we found that chromosomal breakage associated with the fragile site extended over this entire 50 kb region. More than one non-overlapping phage clone crossed the fragile site on some metaphases. That is, more than one clone in this region was found to sometimes lie proximal, sometimes distal and sometimes to cross the fragile site on different metaphase spreads from the same preparation. However, the distribution of these events differed for individual clones and generally correlated with their position within the contig. Results of the FISH experiments are summarized in Table 1 and represented graphically in Figure 2a, and an example of a single clone (lambda 247) showing variable hybridization with respect to the fragile site is shown in Figure 2b.

One possible explanation for the appearance of FISH signals on both sides of the fragile site breakage is that the lambda clones in this region contain repetitive elements capable of hybridizing throughout an extended portion of chromosome 3. The hybridization of these lambda clones to a gridded phage library did not suggest that they contain dispersed repeated elements, as the number of hybridizing clones on the grids was limited and the hybridization signals themselves were clearly positive and negative (data not shown). Throughout the course of these studies, the lambda clones from this contig were used as probes in Southern blot analyses of 1, pulsed field gels of 850A6 (Fig. 3a); 2, total human DNA and DNA from hamster–human hybrids containing only chromosome 3 (Fig. 3b); and 3, gels containing restriction digested DNA of the individual lambda clones (Fig. 3c). These Southern hybridizations were carried out under normally stringent conditions using *Cot*1 and human placental DNA as blocking agents and closely correspond to the hybridization conditions used in the FISH analyses. Furthermore, DNA blot analysis is generally a much more sensitive method than FISH analysis for discerning short regions (500 bp or less) of dispersed repetitive homology. In all of these cases, the DNA from these lambda clones showed no evidence of containing repetitive DNA that would cross-hybridize to sequences throughout the fragile site region. Taken together, the data from the FISH and Southern analyses indicate that the position of the fragile site breakage is truly variable and that we have defined a region of at least 50 kb of frequent breakage associated with FRA3B.

Lambda 247 corresponds to an HPV16 integration site from a previously characterized cervical carcinoma

In an initial effort to gain an understanding of the molecular basis of fragility of FRA3B, we selected lambda clone 247 for sequence

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

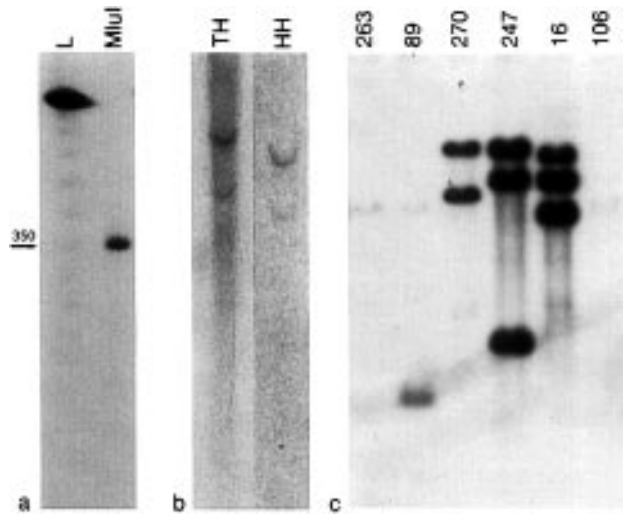

Figure 3. Hybridization of lambda 247 insert DNA to various types of Southern blots. (**a**) A pulsed field gel DNA blot. A lambda size marker ladder is shown on the left (l). The lane on the right (*Mlu*I) shows hybridization of the probe to a 365 kb *Mlu*I band in a digest of total yeast DNA from YAC strain 850A6. The 350 kb lambda oligomer marker band is indicated on the left by the arrow. (**b**) A Southern blot of *Eco*RI digested DNAs run on a 0.7% agarose gel. The left lane contains total human DNA (TH) and the right lane contains DNA from a hamster–human hybrid (HH) with only human chromosome 3. The hybridizing bands in these lanes are the 6.7 and 4.2 kb *Eco*RI restriction fragments (see Fig. 4). (**c**) This blot contains *Not*I/*Eco*RI digested DNA from the lambda clones in the lambda 247 contig. Hybridization occurs with only those bands that overlap the 247 clone.

analysis, because of its central location within the region of aphidicolin induced gaps and breaks. Lambda 247 was subcloned into pBluescript or pUC18 and subjected to automated fluorescence sequence analysis on an ABI 373A sequencer (Pharmacia). Approximately 9.5 kb of sequence from the 12.5 kb lambda clone was obtained (Fig. 4a). No large or small repeat sequences that could be potentially involved in fragile site formation were detected. Several interesting findings were revealed by BLAST (26) and FASTA (27) searches of the GenBank and EMBL sequence databases. Lambda 247 shares a large region of identity with a human papillomavirus type 16 (HPV16) integration site, characterized in the tissues of a primary cervical carcinoma (24). Bestfit analysis showed 98.03% identity of the two sequences over 936 bp starting at a *Kpn*I site at position 199 of the H705 sequences reported by Wagatsuma *et al.* (24) (Fig. 4b). The first 198 bp of the H705 sequence are derived from the M13mp18 cloning vector, so the genomic sequence initiates at the *Kpn*I site at position 199. The lack of complete identity over the 936 bp is likely to be attributable to sequencing errors, as much of our sequence was characterized on only one strand. Interestingly, the homology of lambda 247 and the HPV16 integration site sequences occurs on only the proximal side of the integration site (Fig. 4b). The sequences reported by Wagatsuma *et al.* (24) distal to the integration site do not correspond to those in the next 2000 bp of lambda 247, which is the extent of the sequence gathered here.

To confirm that the arrangement of sequences in lambda 247 represent the normal genomic structure, PCR primers were designed on either side of the integration site based on the lambda 247 sequence. These were used to successfully amplify a 386 bp sequence from total human, YAC850A6 and lambda 247 DNA (Fig. 4c). The sequence of the 386 bp PCR amplification-target in lambda 247, compared with the published sequence (24) is shown in Figure 4b and the PCR results are shown in Figure 4c. This result confirms that lambda 247 represents the normal genomic structure across the HPV16 integration site. It also indicates that the sequences proximal and distal to the HPV16 DNA at the integration site, as reported in (24), are not contiguous in normal DNA. Therefore the HPV16 integration event in this cervical carcinoma is associated with a genomic rearrangement.

To investigate the type of rearrangement, we designed an oligonucleotide probe based on the sequences distal to the integration site from the published sequence. This probe was hybridized to Southern blots containing DNA from the lambda 247 contig, in case the rearrangement consisted of a short deletion within the region, as well as to the gridded lambda FIX II subclone library of YAC850A6. No hybridization was observed with the nearby members of the lambda 247 contig, which contains approximately 20 kb of sequences distal to the integration site. However, the oligonucleotide probe identified three lambda clones from the YAC850A6 library. The insert DNA from one of these lambda clones was used as a probe on a pulsed field Southern of YAC850A6 and showed clear hybridization to the 365 kb *Mlu*IA fragment (unpublished data). These data indicate that the genomic sequences juxtaposed to the distal edge of the HPV16 DNA (24), are derived from the *Mlu*I A region. Hybridization data from a partial cosmid contig developed in the FRA3B region places the three lambda clones homologous to the distal oligonucleotide probe approximately 100 kb from lambda 247 (W. Paradee, C. M. Wilke, T. G. Glover and D. I. Smith, in preparation).

Several other features of the lambda 247 sequence were revealed by the data base searches. Two full length *Alu* repeats in the vicinity of the integration site were located using a BLAST (26) search of the *Alu* subset database within REPBASE (NCBI BLAST network server) (Fig. 4). The first has 80% identity over 290 bases to the human *Alu*-Sx subfamily consensus sequence and is located approximately 1400 bases distal to the HPV16 integration site (Fig. 4a). The second *Alu* has between 97 and 100% identity to an *Alu*-J sequence over a length of 278 bases. The 3′ end of this element ends just eight bases distal to the HPV16 integration site and also extends in a distal direction. A portion of this *Alu* homology was previously described (24).

Forty-nine bp from the 3′ end of the more telomeric *Alu* repeat and in the same transcriptional orientation is a nearly complete processed cyclophilin pseudogene. This pseudogene shows 88.1% identity to the T-cell cyclophilin cDNA (709 bp) over 488 bp by FASTA analysis and includes sequences from all five of the cyclophilin exons (28). The homology to the first exon begins at base 4 of the normal cyclophilin cDNA and the homology to the fifth exon is truncated at base 468 by cloning into Lambda FIX II (Fig. 4).

We used several simple analyses to investigate the potential role that local sequence structure may have played in the position of the HPV16 integration. The overall base content of the 3535 base sequence containing the HPV16 integration site is 57.1% A/T, with some short stretches of approximately 20 bases of up to 100% A/T. Using the Hitachi DNASIS sequence analysis program, G/C content (counting window sizes from 10 to 40 bases), we found that the HPV16 integration site falls asymmetrically within a trough of one of these regions of low G/C content (0–20%) surrounded by short regions (40–60 bases) of relatively higher G/C content (50–70%). No stemloop or direct repeat sequences greater than 7 bp are found within 50 bases of the

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

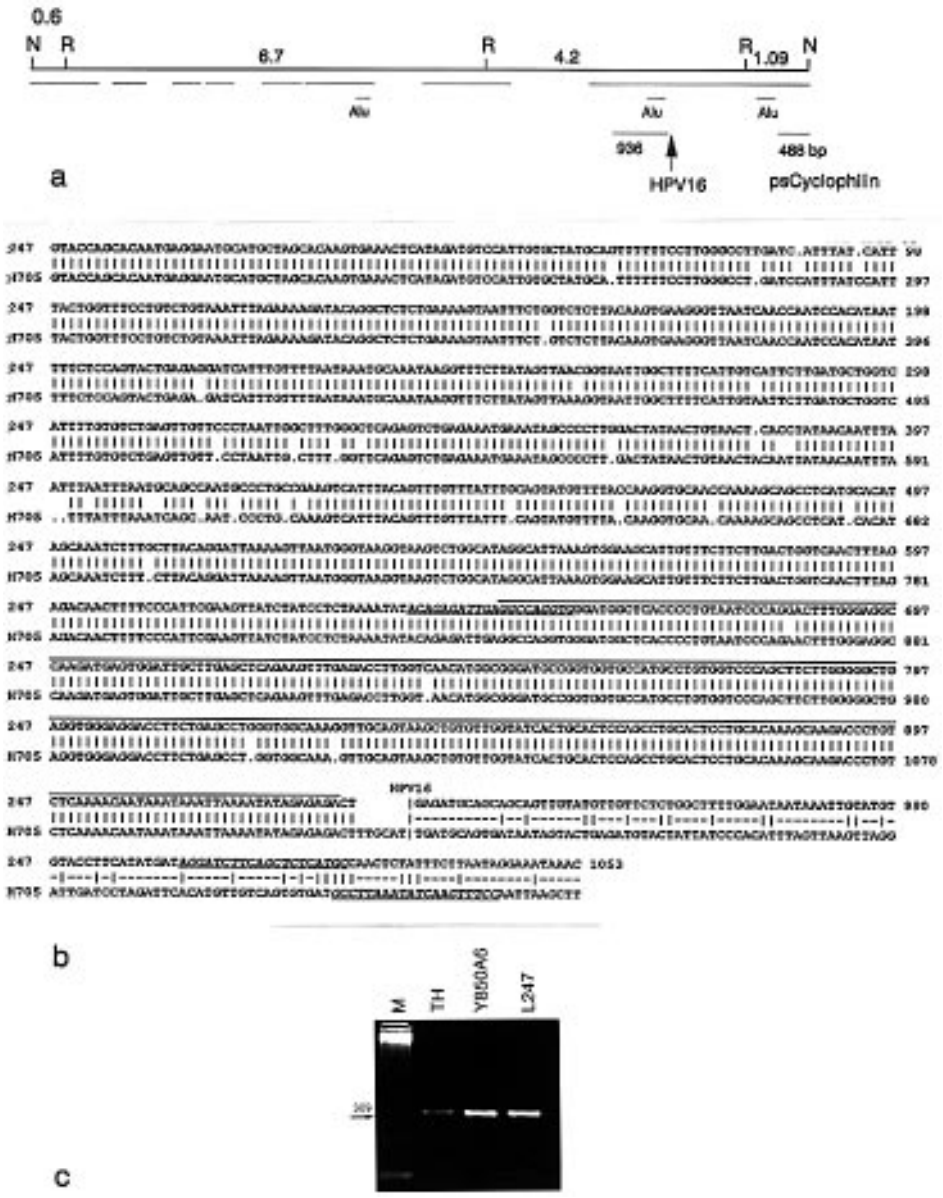

Figure 4. Sequence analysis of lambda clone 247. (**a**) The 12.5 kb of lambda 247 is shown by the bold line. The thin lines indicate the sequenced regions. The positions of *Alu* repeats, the 936 bp of homology to the HPV16 integration site (24) and the 488 bp of the cyclophilin pseudogene are indicated by further thin lines. The arrow indicates the position of the HPV16 integration. The sizes are indicated in kb above the map and N and R correspond to*Not*I and *Eco*RI sites, respectively. The 3535 bp segment discussed at length in the text is represented by the large continuous line at the right end of this figure. (**b**) Sequence homology to the H705 sequence surrounding the HPV16 integration site from Wagatsuma *et al*. (24). The alignment from the bestfit analysis is shown starting from the *Kpn*I site at the 5′ end. The *Alu* homology in this area is indicated by the line above the lambda 247 sequences. The positions of the lambda 247 PCR primers and the oligonucleotide probe from the 3′ HPV16 integration sequences are underlined and shown in italics. The site of the integration event is indicated. (**c**) PCR analysis of the integration site. The amplification of the 386 bp target site in lambda 247 is shown in total human (TH), YAC850A6 (Y850A6) and lambda 247 (L247) template DNAs. PCR products were run on a 2% agarose gel and stained with ethidium bromide. The 123 bp ladder (BRL) was used as a marker (M). The arrow indicates the position of the 369 bp marker band.

integration site. The Findpatterns searching tool from the Wisconsin GCG package (29) was used to locate a number of short sequence motifs that have been implicated in recombination including, the Chi-like octomers (GC[A/T]GG[A/T]GG) (30,31); immunoglobulin heptamers (GATAGTG) (32); alpha protein recognition sites (AAATATC) and (TTAATTC) (31); the consensus Topoisomerase II recognition sites (topoIIv: [A/G]N[T/C]NNCNNG[T/C]NG[G/T]TN[T/C]N[T/C] topoIId: (GTN[T/A]A[C/T] ATTNATNN[A/G] and topoIIi: ([T/C] [T/C] CNTA[C/G][C/T]CC[T/G][T/C][T/C]TNNC) (33); and three consensus core VNTR sequences (34). We also searched for two sequences of particular interest here, one of which is identified as a consensus common hotspot for gene deletions and translocations (T G A/G A/G G/T A/C) (35) and the second (T/A G G AG) represents a subset of the first and sometimes acts as an arrest site for DNA polymerase alpha (35).

Of the first 10 motifs, the Chi-like sequence was present only one time approximately 600 bp distal to the HPV16 integration

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

site and the other nine sequences were not present. For the last two motifs, the deletion related consensus sequence, (TG[A/G][A/G][G/T][A/C]), occurs 12 times on the upper strand and 19 times on the lower strand. The DNA polymerase alpha arrest sequence ([T/A]GGAG) occurs five times on the upper and 10 times on the lower strand. These two short motifs should occur in DNA by chance once every 256 bases (13.8× in 3535 bases) or 512 bases (6.9×), respectively. Therefore, these sequence motifs do not appear to be over represented in the 3535 bases containing the HPV16 integration site. In addition to the computer searches for repeated DNA, Southern blots containing DNA from the lambda clones of the lambda 247 contig were screened for the presence of CAG and CCG trinucleotide repeats by oligomer hybridization. No strong positive signals were observed.

DISCUSSION

By using a FISH mapping approach with available probes we have successfully localized FRA3B to a fairly broad region of YAC850A6. The findings that the fragility of FRA3B extends over a large region and does not involve the expansion of CCG trinucleotide repeats underscores the basic differences between common fragile sites and rare heritable fragile sites. The molecular mechanisms responsible for these two types of fragile sites are likely to be distinct and this is not surprising considering the differences in both the frequencies with which they occur in populations and in the methods used in their induction. Contemporary with our studies, Rassool *et al.* (7; F. V. Rassool, M. M. Le Beau M-L. Shen, M. E. Neilly, S. T. Ong, F. Boldog, H. Drabkin and T. W. McKeithan, in preparation) used a direct cloning strategy, based on integration of transfected pSV2neo DNA into the FRA3B region, to isolate sequences from FRA3B. They too have defined the same general region of YAC850A as containing the fragile site. Furthermore, the detailed mapping of aphidicolin-induced breakpoints in chromosome 3 containing hamster–human hybrids has revealed a distribution of breaks in 3p14.2 that encompasses this same region (6) (W. Paradee, C. M. Wilke, T. W. Glover and D. I. Smith, in preparation). Although the molecular basis for chromosome fragility is not yet apparent from simple sequence analysis of the clones characterized so far, the mapping of the region will allow for further studies into the mechanisms underlying common fragile site expression.

It is significant that the region we have defined as FRA3B coincides with the integration site of a human papillomavirus in a primary cervical carcinoma. The existence and position of common fragile sites have been defined through the use of stress inducing agents in an artificial cell culture environment. Their identification in this way does not point to biological significance. However, our positioning of FRA3B by *in vitro* methods and the subsequent finding of an *in vivo* generated HPV16 integration site within the region provides the first direct evidence that common fragile sites, as they have been defined, have biological relevance. These direct molecular data support the earlier statistical association of these two entities observed at a cytogenetic level.

Common chromosomal fragile sites are present in all individuals and therefore can be assumed to constitute a macromolecular manifestation of some conserved feature of chromatin structure or DNA replication and repair, or perhaps gene expression. Two characteristics of fragile sites that may be relevant to viral integration are that most occur in G-light bands, leading to suggestions that they may be associated with active gene regions (8,36) and that they are thought to be late replicating regions (16). That they are late replicating has recently been verified for FRAXA, the only fragile site in which the time of replication has been characterized (17). Several studies point to a possible mechanistic connection between the integration of DNA viruses and these two characteristics of fragile sites.

Caporossi *et al.* (37) observed that adenovirus type 5 infection and aphidicolin treatment, both of which when used alone induce common fragile sites, have a synergistic effect on the production of fragile sites when used together. Aphidicolin is an inhibitor of both DNA polymerase α and δ (38) and may induce fragile sites by further delaying DNA replication in regions that are already late replicating or slow to replicate. This synergism with aphidicolin and virus suggests that the fragile sites induced by the two different agents are the same and that the viral induction of fragile sites also involves delay in DNA replication. In another study (39), it was found that the latest replicating G-negative bands were selectively targeted by HPV16 in virus-immortalized human keratinocyte cell lines. Finally, Smith *et al.* (40), directly demonstrated that an HPV18 integration spanned a gap like a fragile site in a cell line infected *in vitro*. These *in vitro* studies have suggested a direct association between DNA tumor virus integration and fragile sites, or portions of the genome showing characteristics of fragile sites and support previous FISH studies correlating HPV integration sites in tumors or tumor cell lines with chromosomal bands known to contain fragile sites and protooncogenes (19,41). This type of data also suggests that the viral infection may play some direct role in the induction of fragile sites and the eventual site of viral integration (37).

Most DNA tumor viruses encode proteins that inhibit p53 or pRb or both. The initial infection therefore results in a relaxation of normal cell cycle controls on DNA replication and repair, establishing a favorable cellular environment for the occurrence and fixation of the multiple mutations seen in tumor progression, including increased DNA instability. Double-stranded transforming DNA viruses normally replicate extrachromosomally and only rarely integrate into the host genome. Because integration of these viruses is not a normal part of their replication cycle, it is likely that the integration events themselves are passive and signal a change in DNA metabolism to a more recombinogenic state. However, if the integration of the virus results in a selective growth advantage to the cell, then these will be the events observed in cancers and may point to regions important to tumor progression. In this way, the initial viral infection can be viewed as the first step in this series of events leading to its own integration and the integration itself is an important later step in the evolution of the cancer. Such a view is not novel and has been proposed previously (19,42,43).

Consistent with this selection perspective, is the actual small number of different chromosomal locations containing HPV16 or 18 integrations that have been characterized in primary cervical tumors or cervical cancer cell lines and the number of times that several of these locations have been repeatedly targeted in these tissues. In a summary of these integration sites Lazo *et al.* (42 and references therein) found that the 8q24 region, which contains both a fragile site and the c-*myc* gene, was targeted five of 15 times and the 2p24 region containing the N-*myc* gene was also targeted once. The 12q13 region, rearrangements of which are associated with a number of cancers was also found to be multiply targeted. By extension these authors argue that other sites found

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

to contain HPV16/18 integrations are likely to indicate the position of tumorogenic genes.

Several studies have implicated the 3p14 region in the development of cancer (44). In one study (45) loss of heterozygosity of 3p14 was found in 90% of cervical cancers and in several others (25,46) the most commonly deleted region of 3p in cervical cancer was between 3p13 and 3p21.1. The transfection of a malignant renal cell carcinoma cell line with the 3p12–3p14 region was shown to suppress tumor growth in nude mice (47). Also, both loss of heterozygosity (48) and amplification (49) of 3p14 have been observed in breast tumors. Interestingly, Wagatsuma *et al.* (24) found that H705 sequences represented the only HPV16 integration in their primary cervical tumor and that the viral and neighboring genomic DNAs were amplified five to 10 times. Similar amplifications of HPV DNAs and the c-*myc* and N-*myc* genes, sometimes associated with increased expression of these loci, have been found in a number of cervical cancer cell lines (50). The coincidence of a fragile site and an HPV16 integration/deletion event within a region commonly deleted in cervical cancers suggests that, at least in some instances, viral integration events within fragile sites may play a mechanistic role in the evolution of cancers (20). Furthermore, it suggests that the 3p14.2 region in particular may be important in cervical carcinoma progression.

MATERIALS AND METHODS

Subcloning of YAC850A6 in lambda vectors and identification of human clones

Subcloning of YAC850A6 into the Lambda FIXII vector and production of a gridded, human specific subclone library was carried out as previously described (51). Initially, random human lambda clones were hybridized to *Mlu*I digested YAC850A6 pulsed field gel blots to assign anchor clones to each of the *Mlu*I fragments. Also, the 850A6 *Mlu*I fragments themselves were ^{32}P-labeled and hybridized to the grids to assign some of the lambda clones. After the initial FISH results (below) indicating the position of the fragile region, the lambda contigs were expanded by hybridizing lambda insert DNA, to the gridded phage library.

Production of metaphase spreads and aphidicolin induction of FRA3B

PHA-stimulated whole blood lymphocyte cultures were grown in RPMI 1640 medium (Irvine Scientific) containing 10% fetal calf serum, 4 mM glutamine and penicillin/streptomycin (200 units/ml, 200 µg/ml). Fragile sites were induced with 0.4 µM aphidicolin (Sigma) for 26 h prior to harvesting. Harvesting was carried out conventionally and consisted of 45–60 min of colcemid treatment (0.07 µg/ml) followed by 15–20 min hypotonic treatment (0.075 M KCl) and multiple changes of fixative (3:1 methanol–acetic acid) before preparing slides.

FISH analysis

Total lambda DNA was biotin labeled (biotin-14-dATP) using the BioNick Labeling System (BRL, Gaithersburg, MD). One hundred to 200 ng of labeled lambda DNA and 25 µg of *Cot*1 (BRL) were denatured at 75°C in 10 µl of 2× SSC, 10% dextran sulfate, 50% formamide, 1% tween 20 and preannealed at 37°C for at least 1 h prior to application to denatured slides. Slides were denatured at 70°C for 2 min in 70% formamide, 2× SSC, pH 7.0 and put through a standard dehydrating ethanol series. Hybridization was carried out overnight. Washes consisted of 50% formamide, 2× SSC at 42°C followed by 0.1× SSC at 60°C. Signals were visualized after incubations with two layers of FITC conjugated avidin-DCS and biotinylated goat anti-avidin IgG (Vector, Burlingame, CA), followed by counterstaining with propidium iodide alone or with both DAPI and propidium iodide. The chromosomes were viewed with either a Nikon or Zeiss Axioskop epifluorescence microscope equipped with FITC and DAPI filters. Metaphase spreads were photographed on Kodak ASA 400 or 100 Gold film. For the published picture, the color negatives were digitally scanned to a Kodak PhotoCD Master Disc using a Kodak PCD Film Scanner 2000 controlled by a Kodak PCD Imaging Workstation 2200/2400. The digital image was converted to gray scale using Adobe Photoshop software and printed on a Kodak PCD Printer 200.

Subcloning and sequence analysis of lambda 247

Lambda 247 DNA was singly digested with *Not*I, *Eco*RI, *Pst*I, *Kpn*I, *Sal*I, *Xba*I, *Hin*dIII and doubly cut with *Not*I/*Eco*RI and shotgun cloned into appropriately digested pUC18 or pBluescript. None of these enzymes cleave within the Lambda Fix II vector and therefore allow the cloning of only insert DNA. DNA from individual clones from each shotgun ligation was prepared using the miniprep alkaline lysis procedure (52), digested and run on agarose gels to identify clones of various sizes. Clones with different size inserts were subjected to automated fluorescence sequence analysis using vector sequencing primers on an ABI 373A sequencer (Pharmacia). The above approach was used in an effort to maximize the amount of sequence obtained with the least number of runs. Our goal was to acquire enough sequence information to discern interesting or important characteristics of lambda 247. Most of the sequence has been obtained on one strand only. A few gaps were filled in using specific synthetic oligosequencing primers. The sequences were assembled and analyzed using the Wisconsin Genetics Computer Group DNA-analysis programs FASTA, Lineup, Bestfit, Findpatterns, Repeats and Stemloop (29), as well as the BLAST (26) and Grail servers (27) and the DNASIS program G/C Content (Hitachi). The GenBank accession numbers for the 9.5 kb of sequence characterized here are U39793–U39801 and U39804.

Southern hybridization

Southern hybridization probes were ^{32}P-labeled by the random priming technique (52). Five to 50 ng of probe was denatured and preanealed for 1 h in Church's phosphate buffer (53) in the presence of the following sheared DNAs: 10 µg/ml *Cot*1 DNA (BRL), 25 µg/ml human placental DNA, 100 µg/ml salmon sperm DNA and 2.5 µg/ml lambda DNA. Hybridizations were carried out at 65°C overnight in Church's phosphate buffer. The filters were then washed to a stringency of 0.1× SSC, 0.1% SDS at 65°C for 10–15 min. DNA oligomer probes were ^{32}P-end labeled with T4-kinase (52), hybridized for at least 4 h at 50°C in Church's phosphate buffer and washed to a stringency of at least 2× SSC, 0.1% SDS at 50°C for 30 min and sometimes to 1× SSC, 0.1% SDS at 50°C for 30 min.

PCR across the HPV16 integration site in lambda 247

The two 20-mer PCR primers SA5mfp′(GCATGAGAGCTGAAGATCCT) and SA4mfp1 (ACAGAGATTGAGGCCA GG TG) were

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

derived from the lambda 247 sequence and give rise to a 386 bp PCR product spanning the HPV16 integration site of Wagatsuma *et al*. (24). Template DNA amounts in 100 µl reactions were 100 ng total human DNA, 10 ng total yeast DNA containing YAC850A6 and 5 ng lambda DNA. Each 100 µl reaction contained 20 pmol (0.2 µM) of each primer, 200 µM of each dNTP, 1×PCR buffer (Amersham), 1.5 mM $MgCl_2$ and 2.5 U *Taq* DNA polymerase (Amersham). The reactions were subjected to 30 cycles of 1 min denaturation at 94°C, 1 min annealing at 62°C, 2 min extension at 72°C, followed by a final extension at 72°C for 10 min. The PCR products were electrophoresed on a 2% Nusieve GTG (FMC, Rockland ME) agarose gel and then stained with ethidium bromide.

ACKNOWLEDGMENTS

The authors thank Herman Dierick for critically reading the manuscript and helpful discussion, James L. Beals from the UM Biomedical Communications for help in preparing the figures and F. V. Rassool *et al*. for sharing unpublished data. We are especially grateful to Diane Miller for help in FISH experiments. This research was supported by Grant CA43222 from the National Cancer Institute. The GCG package was supported in part by Grant M01RR00042 from the National Center for Research Resources, NIH.

REFERENCES

1. Sutherland,G.R. and Hecht,F. (1985) Fragile sites on Human Chromosomes. *Oxford Monographs on Medical Genetics*, Oxford University Press, Oxford.
2. Wenger,S.L., Hennessey,J.C. and Steele,M.W. (1987) Increased sister chromatid exchange frequency at Xq27 site in affected fragile X males.*Am. J. Med. Genet.* **26,** 909–914.
3. Glover,T.W. and Stein,C.K. (1987) Induction of sister chromatid exchanges at common fragile sites. *Am. J. Hum. Genet.* **41,** 882–890.
4. Warren,S.T., Zhang,F., Licambi,G.R. and Peters,J.F. (1987) The fragile X site in somatic cell hybrids. *Science* **237,** 420–423.
5. Glover,T.W. and Stein,C.K. (1988) Chromosome breakage and recombination at fragile sites. *Am. J. Hum. Genet.* **43,** 265–273.
6. Wang,N-D., Testa,J.R. and Smith,D.I. (1993) Determination of the specificity of aphidicolin-induced breakage of the human 3p14.2 fragile site. *Genomics* **17,** 341–347.
7. Rassool,F.V., McKeithan,T.W., Neilly,M.E., van Melle,E., Espinosa III,R. and Le beau,M.M. (1991) Preferential integration of marker DNA into the chromosomal fragile site at 3p14: An approach to cloning fragile sites. *Proc. Natl Acad. Sci. USA* **88,** 6657–6661.
8. Yunis,J.J. and Soreng,A.L. (1984) Constitutive fragile sites and cancer. *Science* **226,** 1199–1204.
9. Le Beau,M.M. and Rowley,J.D. (1984) Heritable fragile sites and cancer. *Nature* **308,** 607–608.
10. 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., Oostra,B.A., Nelson,D.L. and Caskey,C.T. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell* **67,** 1047–1058.
11. Knight,S.J.L., Flannery,A.V., Hirst,M.C., Campbell,L., Christodoulou,Z., Phelps,S.R., Pointon,J., Middleton-Price,H.R., Barnicoat,A., Pembrey,M.E., Holland,J., Oostra,B.A., Bobrow,M. and Davies,K.E. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74,** 1–20.
12. Nancarrow,J.K., Kremer,E., Holman,K., Eyre,H., Dogget,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.
13. 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 CCG repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. *Nature Genet.* **8,** 229–235.
14. Jones,C., Slijepcevic,P., Marsh,S., Baker,E., Langdon,W.Y., Richards,R.I. and Tunnacliffe, A. (1994) Physical linkage of the fragile site FRA11B and the Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. *Hum. Mol. Genet* **3,** 2123–2130.
15. Pieretti,M., Zhang,F, Fu,Y-H., Warren,S.T., Oostra,B.A., Caskey,C.T. and Nelson,D.L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66,** 817–822.
16. Laird,C., Jaffe,E., Karpen,G., Lamb,M. and Nelson,R. (1987) Fragile sites in human chromosomes as regions of late replicating DNA. *Trends Genet.* **3,** 274–281.
17. Hansen,R.S., Canfield,T.K., Lamb,M.M., Gartler,S.M. and Laird,C.D. (1993) Association of Fragile X syndrome with delayed replication of the *FMR1* gene. *Cell* **73,** 1403–1409.
18. Popescu,N.C. and DiPaolo,J.A. (1989) Preferential sites for viral integration on mammalian genome. *Cancer Genet. Cytogenet.* **42,** 157–171.
19. Popescu,N.C., Zimonjic,D. and Dipaolo,J.A. (1990) Viral integration, fragile sites and proto-oncogenes in human neoplasia. *Hum. Genet.* **84,** 383–386.
20. Cannizzaro,L.A., Durst,M., Mendez,M.J., Hecht,B.K. and Hecht,F. (1988) Regional chromosome localization of human papillomavirus integration sites near fragile sites, oncogenes and cancer chromosome breakpoints. *Cancer Genet. Cytogenet.* **33,** 93–98.
21. Bosch,F.X., Manos,M.M., Munoz,N., Sherman,M., Jansen,A.M., Peto,J., Schiffman,M.H., Moreno,V., Kurman,R., Shah,K.V., International Biological Study on Cervical Cancer Study Group. (1995) Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J. Natl Cancer Inst.* **87,** 796–802.
22. Boldog,F., Gemmill,R.M., Wilke,C.M., Glover,T.W., Nilsson,A., Chandrasekharappa,S.C., Brown,R.S., Li,F.P. and Drabkin,H.A. (1993) Positional cloning of the hereditary renal carcinoma 3;8 chromosome translocation breakpoint. *Proc. Natl Acad. Sci. USA* **90,** 8509–8513.
23. Wilke,C.M., Guo,S-W., Hall,B.K., Boldog,F., Gemmill,R.M., Chandrasekharappa,S.C., Barcroft,C.L., Drabkin, H.A. and Glover,W.G. (1994) Multicolor FISH mapping of YAC clones in 3p14 and identification of a YAC spanning both FRA3B and the t(3;8) associated with hereditary renal carcinoma. *Genomics* **22,** 319–326.
24. Wagatsuma,M., Hashimoto,K. and Matsukura,T. (1990) Analysis of human Papillomavirus type 16 DNA in cervical cancers: amplification of viral sequences together with cellular flanking sequences. *J. Virol.* **64,** 813–821.
25. Kohno,T., Takayama,H., Hamaguchi,MM., Takano,H., Yamaguchi,N., Tsuda,H., Hirohashi,S., Vissing,H., Shimizu,M., Oshimura,M. and Yokota,J. (1993) Deletion mapping of chromosome 3p in human uterine cervical cancer. *Oncogene* **8,** 1825–1832.
26. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.
27. Uberbacher,E.C. and Mural,R.J. (1991) Locating protein-coding regions in human DNA by multiple sensor neural network approach. *Proc. Natl Acad. Sci. USA* **88,** 1161–1165.
28. Haendler,B., Hofer-Warbinek,R. and Hofer,E. (1987) Complementary DNA for T-cell cyclophilin. *EMBO J.* **6,** 947–950.
29. Genetics Computer Group, Program Manual for GCG Package, Version 8, April 1991, 575 Science Drive, Madison, Wisconsin, USA 53711.
30. Krowczynska,A.M., Rudders,R.A. and Krontiris,T.G. (1990) The human minisatellite consensus at breakpoints of oncogene translocations. *Nucleic Acids Res.* **18,** 1121–1127.
31. Sowerby,S.J., Kennedy,M.A., Fitzgerald,P.H. and Morris C.M. (1993) DNA sequence analysis of the major breakpoint cluster region of the BCR gene rearranged in Philadelphia-positive human leukemias. *Oncogene* **8,** 1679–1683.
32. Chen,S.J., Chen,Z., D'Auriol,L., Le Coniat,M., GrausZ,D. and Berger, R. (1989) Ph+bcr-acute leukemias: implications of Alu sequences in a chromosomal translocation occuring in the new cluster region within the BCR gene. *Oncogene* **4,** 195–202.
33. Kas,E. and Laemmli,U.K. (1992) *In vivo* topoisomerase II cleavage of the *Drosophila* histone and satellite II repeats: DNA sequence and structural characteristics. *EMBO J.* **11,** 705–716.
34. Cullen,M., Erlich,H., Klitz,W. and Carrington,M. (1995) Molecular mapping of a recombination hotspot located in the second intron of the human TAP2 locus. *Am. J. Hum. Genet.* **56,** 1350–1358.

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022

35. Krawczak,M. and Cooper,D.N. (1991) Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence environment. *Hum. Genet.* **86,** 425–441.
36. Hecht,F. (1988) Fragile sites, cancer chromosome breakpoints and oncogenes all cluster in light G-bands. *Cancer Genet. Cytogenet.* **31,** 17–24.
37. Caporossi,D., Bachetti,S. and Nicoletti,B. (1991) Synergism between aphidicolin and adenoviruses in the induction of breaks at fragile sites on human chromosomes. *Cancer Genet. Cytogenet.* **54,** 39–53.
38. Glover,T.W., Berger,C., CoyleJ. and Echo,B. (1984) DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum. Genet.* **67,** 136–142.
39. Zimonjic,D.B., Popescu,N.D. and DiPaolo,.A. (1994) Chromosome organization of viral integration sites in human Papillomavirus-immortalized human keratinocyte cell lines. *Cancer Genet. Cytogenet.* **72,** 39–43.
40. Smith,P.P., Friedman,C., Bryant,E.M. and McDougall,J.K. (1992) Viral integration and fragile sites in human papillomavirus-immortalized human keratinocyte cell lines. *Genes Chrom. Cancer* **5,** 150–157.
41. Kahn,T., Korn,B., Delius, H., Poustka,A. and zur Hausen,H. (1995) Suche nach Tumor-Suppressor-Genen in Papillomvirus-Integrationsorten am Beispiel einer HPV 6-DNA Integration in einem Tonsillen-Karzinom.Tagung der Gesellschaft fur deutsche Virologie, Giessen.
42. Lazo,P.A., Gallego, M.I., Ballester,S. and Feduchi,E. (1992) Genetic alterations by human papillomaviruses in oncogenesis. *FEBS Lett.* **300,** 109–113.
43. Gallego,M.I., Zimonjic,D.B., Popescu,N.C., DiPaolo,J.A. and Lazo,P.A. (1994) Integration site of human Papillomavirus type-18 DNA in chromosome band 8q22.1 of C4-I cervical carcinoma: DNase I hypersensitivity and methylation of cellular flanking sequences. *Genes Chrom. Cancer* **9,** 28–32.
44. Hibi,K., Takahashi,T., Yamakawa,K., Ueda,R., Sekido,Y., Ariyoshi,Y., Suyama,M., Takagi,H. and Nakamura,Y. (1992) Three distinct regions involved in 3p deletion in human lung cancer. *Oncogene* **7,** 445–449.
45. Chung,G.T., Huang,D.P., Wailo,K., Chan,M.K.M. and Wong,F.W.S. (1992) Genetic lesions in the carcinogenesis of cervical cancer. *Anticancer Res.* **12,** 1485–1490.
46. Jones,M.H. and Nakamura,Y. (1992) Deletion mapping of chromosome 3p in genital tract malignancies using microsatellite polymorphisms. *Oncogene* **7,** 1631–1634.
47. Sanchez,Y., El-Nagger,A., Pathak,S. and Killary,A.M. (1994) A tumor suppressor locus within human chromosome 3p14–p12 mediates rapid cell death of renal carcinoma *in vivo*. *Proc. Natl Acad. Sci. USA* **91,** 3383–3387.
48. Devilee,P. van den Broek,M., Kuipers-Dijkshoorn,N., Kolluri,R., Khan,P.M., Pearson,P.L. and Cornelisse,C.J. (1989) At least four different chromosomal regions are involved in loss of heterozygosity in human breast carcinoma. *Genomics* **5,** 554–560.
49. Pinkel,D. (1994) Visualizing tumor amplification. *Nature Genet.* **8,** 107–108.
50. Couturier,J., Sastre-Garau,X., Schneider-Maunoury,S., Labib, A. and Orth,G. (1991) Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression. *J. Virol.* **65,** 4534–4538.
51. Mercer,J.F.B., Livingston,J., Hall,B., Paynter,J.A., Begy,C., Chandrasekharappa,S., Lockhart,P., Grimes,A., Bhave,M., Siemieniak,D. and Glover,T.W. (1993) Isolation of a partial candidate gene for Menkes disease by positional cloning. *Nature Genet.* **3,** 20–25.
52. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular cloning, a laboratory manual*, 2nd Edition. (Cold Spring Harbor Laboratory Press, New York)
53. Church,G.M. and Gilbert,W. (1984) Genomic sequencing. *Proc. Natl Acad. Sci. USA* **81,** 1991–1995.

Downloaded from https://academic.oup.com/hmg/article/5/2/187/2568695 by guest on 20 August 2022