

Evidence that instability within the FRA3B region extends four megabases

Nicole A Becker¹, Erik C Thorland², Stacy R Denison¹, Leslie A Phillips¹ and David I Smith^{*,1}

¹Department of Experimental Pathology, Mayo Foundation, Rochester, Minnesota, MN 55905, USA; ²Departments of Biochemistry and Molecular Biology, Mayo Foundation, Rochester, Minnesota, MN 55905, USA

FRA3B is the most frequently expressed common fragile site localized within human chromosomal band 3p14.2, which is frequently deleted in many different cancers, including cervical cancer. Previous reports indicate aphidicolin-induced FRA3B instability occurs over ~500 kb which is spanned by the 1.5 Mb fragile histidine triad (FHIT) gene. Recently an HPV16 cervical tumor integration, 2 Mb centromeric to the published FRA3B region, has been identified. FISH-based analysis with a BAC spanning the integration has demonstrated this integration occurs within the FRA3B region of instability. These data suggest that the unstable FRA3B region is much larger than previously reported. FISH-based analysis of aphidicolin-induced metaphase chromosomes allowed for a complete characterization of instability associated with FRA3B. This analysis indicates that fragility extends for 4 Mb. Within this region are a total of five genes, including FHIT. FRA3B gene expression analysis on a panel of cervical tumor-derived cell lines revealed that three of the five genes within FRA3B were aberrantly regulated. A similar analysis of genes outside of FRA3B indicated that the surrounding genes were not aberrantly expressed. These data provide additional support that regions of instability associated with CFSs and the genes contained within them, may play an important role in cancer development.

Oncogene (2002) 21, 8713–8722. doi:10.1038/sj.onc.1205950

Keywords: common fragile sites; FRA3B; human papillomavirus; cervical cancer; viral integration

Introduction

The development of cervical cancer is highly associated with HPV infection (Choo, 1998). HPV sequences have been identified within more than 95% of analysed cervical tumors, with high-risk subtypes HPV16 and HPV18 being the most prevalently recovered sub-types (Walboomers *et al.*, 1999).

In premalignant cervical lesions, the HPV genome is typically maintained in its episomal form. However, in the majority of invasive cervical carcinomas the HPV genomic DNA has been integrated into the host cell genome. Initial analysis involving cervical integration sites proposed that the integration events occurred randomly, but later cytogenetic analysis suggested a correlation between the sites of HPV integration and chromosomal bands containing common fragile sites (CFSs; Ref. Popescu *et al.*, 1990). We rescued the sites of HPV integration from a number of HPV16-positive cervical tumors and confirmed at the molecular level that HPV16 preferentially integrates into different CFS regions in ~50% of cervical tumors analysed (Thorland *et al.*, 2000, 2002; Denison *et al.*, submitted).

Fragile sites (FSs) are reproducible, non-random regions of chromosomal instability that are observed when cells are cultured under appropriate tissue culture conditions. FSs are divided into two classifications, rare fragile sites (RFSs) and common fragile sites (CFSs). RFSs are defined as sites that are observed in less than 5% of the population. CFSs appear to be present in all individuals with varying levels of expression. Molecular analysis of four CFSs (FRA6E, FRA7G, FRA7H, and FRA16D) has revealed that aphidicolin-induced decondensation/breakage within these CFSs occurs over broad genomic regions ranging from 200 kb to greater than 2 Mb (Huang *et al.*, 1998; Krummel *et al.*, 2000; Mishmar *et al.*, 1998; Tatarelli *et al.*, 2000). Additionally, CFSs have been demonstrated to be sites of elevated sister chromatid exchange (Glover and Stein, 1987), translocations and deletions in tumors (Boldog *et al.*, 1997; Fang *et al.*, 2001; Glover *et al.*, 1988; Glover and Stein, 1988), intrachromosomal gene amplification (Coquelle *et al.*, 1997), and *in vivo* tumor-associated viral integration (Mishmar *et al.*, 1998; Rassool *et al.*, 1991; Thorland *et al.*, 2000; Wilke *et al.*, 1996). These observations suggest that the inherently unstable regions within CFSs may be predisposed to chromosomal breakage and rearrangement during cancer development (Popescu *et al.*, 1990).

The most commonly expressed CFS in humans is FRA3B, cytogenetically located at 3p14.2 (Smeets *et al.*, 1986). The region 3p14.2 is deleted in a variety of histologically different cancers including renal cell carcinoma, and cancers of the lung, pancreas, and cervix (Connolly *et al.*, 2000; Herzog *et al.*, 2001; Rabbitts, 1994; Sandberg, 1990; Solomon *et al.*, 1991).

*Correspondence: DI Smith, Department of Experimental Pathology, Mayo Foundation, Hilton 800, 200 First Street S.W., Rochester, MN 55905, USA; E-mail: smith.david@mayo.edu

Received 10 April 2002; revised 31 July 2002; accepted 7 August 2002

Previous data mapped a HPV16 viral integration from a cervical tumor within FRA3B (Wilke *et al.*, 1996). Both ends of this integration were defined, revealing the integration is associated with a 96 kb deletion within intron 4 of the FHIT gene. Recently an additional HPV16 cervical integration has been mapped to the 3p14.1 region with FISH-based analysis indicating the position of the integration is contained within the region of FRA3B fragility (Thorland *et al.*, 2002). Previous analyses had defined FRA3B as a large region of genomic instability covering ~500 kb (Paradee *et al.*, 1996; Rassool *et al.*, 1996; Wilke *et al.*, 1996; Zimonjic *et al.*, 1997). Unfortunately, the original data describing the instability at FRA3B did not fully define the boundaries of the fragility, with the telomeric boundary being defined by a cosmid that hybridizes centromeric to breakage at FRA3B in 10% of the metaphases. Similarly, the centromeric boundary is defined by a cosmid that hybridizes telomeric to breakage at FRA3B in 8% of the metaphases (Zimonjic *et al.*, 1997). The new HPV16 integration is located 2 Mb centromeric to the previously defined FRA3B region of instability therefore this led us to re-investigate the entire FRA3B region to completely characterize the 'center' as well as the 'ends' of the fragility. FISH and sequence-based analysis reveals fragility at FRA3B extends over a 4 Mb region containing five genes. In addition to FHIT, FRA3B contains protein tyrosine phosphatase receptor, gamma (PTPRG), HT021, calcium dependent activators for protein secretion (CADPS), spinocerebellar ataxia 7 (SCA7, Table 1). Gene expression analysis using cervical tumor-derived cell lines reveal that the majority of the genes in FRA3B are down-regulated whereas the surrounding genes do not demonstrate aberrant gene expression. Our data indicate that the region of genomic instability associated with FRA3B, is much larger than previously believed. This large region of defined genomic instability offers a possible explanation for the frequent deletions and alterations 3p14.2 in a variety of human cancers.

Results

Cervical tumor HPV16 integration at 3p14.2

RS-PCR, was previously employed to identify HPV16 genomic integrations into cervical tumor DNA (Thorland *et al.*, 2000, 2002). Initial sequence data revealed an HPV16 integration into a specific, non-repetitive region of the human genome localizing to 3p14.2.

We sought to confirm and further characterize this new HPV16 integration relative to the published FRA3B region. PCR was performed to validate the HPV16 integration site. A specific PCR primer in the cellular integration flanking sequence was designed. PCR amplification on cervical tumor DNA was performed using the cellular primer and HPV16 RS-PCR primers (Figure 1a). PCR products of the expected sizes were obtained for all three of the HPV16/cellular primer combinations (Figure 1b).

These HPV16/cellular primer combinations did not amplify from normal human genomic DNA (data not shown). These data validate the viral integration and eliminate the possibility of a RS-PCR artifact. RS-PCR hybrid products were BLASTN searched in the NCBI non-redundant complete sequences (nr) and high throughput genomic sequences (htgs) databases to determine what similarities existed to the HPV16 and human genomes (Figure 1a). The BLASTN searches of the HPV16/human sequence hybrid reveals that the HPV16 portion of the PCR-rescued DNA fragment mapped to the HPV16 regulatory protein E2, whereas the human genomic portion mapped to unordered BAC clone RP11129_K_20 localized to chromosome 3. Analysis of the integration site sequence reveals a 15 bp orphan sequence that can not be aligned with either the HPV16 or the human genomic sequence and creates a 12 bp direct repeat (Figure 1a). These results confirm the initial RS-PCR results and validate the site of viral integration within cervical tumor, CC61. FISH analysis on aphidicolin-treated metaphase chromosomes further positioned clone 129_K_20 cytogenetically to 3p14.2. As the ends of FRA3B had only been loosely defined, BAC 129_K_20 was used as a FISH-based probe to determine whether it was within the FRA3B region of instability.

FISH-based analysis of 50 aphidicolin-induced metaphase breaks at 3p14.2 revealed that clone 129_K_20 was contained within FRA3B. This BAC hybridized proximal to the region of aphidicolin-induced breakpoints in 42 out of 50 metaphases and distal in eight out of 50 metaphases (Figure 2; Table 2). Since FRA3B is cytogenetically localized to 3p14.2, we sought to further localize clone 129_K_20 relative to the published FRA3B site. Using the NCBI, the Santa Cruz Genome Center, and the San Antonio Genome Center public databases, we assembled a BAC contig around the published FRA3B fragile site (Paradee *et al.*, 1996; Rassool *et al.*, 1996). Unassembled BACs were aligned with the use of Sequencher 4.1.2 (Figure 3a). This revealed the cervical tumor CC61 integration to be ~2 Mb centromeric (proximal) to the previously defined HPV16 integration (Figure 3b). Only one end of this novel HPV16 integration has been localized at this time. This end localizes between two known genes, CADPS and SCA7. Therefore, unlike the previously identified HPV16 integration into FRA3B, in which a 96 kb intronic deletion between FHIT exons 4 and 5 was identified, it is not known whether this novel integration is also associated with a deletion or a more complex rearrangement.

Mapping the complete region of instability within FRA3B

Previous published reports estimated the size of FRA3B to be ~500 kb and completely contained within the FHIT locus (Paradee *et al.*, 1996; Rassool *et al.*, 1996). These previous data did not define the boundaries of fragility, but instead focused on the 'active' region which contains the previous HPV16 integration, a hereditary

Table 1 Genes contained within and surrounding common fragile site FRA3B (3p14.2)

Gene	UniGene	Gene definition	Genomic size ^a	Approx. position ^b	RT oligonucleotide primers ^c	Produce size ^d
ARHGEF3	Hs.6066	Rho guanine nucleotide exchange factor 4	120 kb	1.1 Mb	GTTCCTAAATCCCACCACC TGCTTCTCCAAACCGTTC	269
APPL	Hs.27413	Adaptor protein containing pH domain	45 kb	750 kb	ATGTGATTCTGTGGACTG CCAAGGGGGAATATCTAC	317
ARF4	Hs.75290	ADP-ribosylation factor 4	20 kb	650 kb	TGCTACTTTTTGCAAACAAAC CCAAACCAGTCCCAGATAC	291
TU3A	Hs.8022		16 kb	500 kb	CCAGCTCATCAAGAAGAAG GTACAGAAGGGCTGAAGG	271
FHIT	Hs.77252	Fragile histidine triad gene	1.4 Mb	within FRA3B	GGCCAACATCTCATCAAG TTTCCTCCTCTGATCTCC	388
PTPRG	Hs.89627	Protein tyrosine phosphatase, receptor type, G	680 kb	within FRA3B	CAACTGGAGAATGAAAATG CTAGAGTCTGGCAAAAAAG	322
HT021	Hs.47166		15 kb	within FRA3B	AAAGATGACTTCCTGTTG GCTCTTCTAAAAGAACTG	321
CADPS	Hs.151301	Calcium dependent activator protein for secretion	70 kb	within FRA3B	TACAAGCAATATGGAGCAC GGTTCATTTTTGGTTTTAG	335
SCA7	Hs.108447	Spinocerebellar ataxia 7	130 kb	within FRA3B	GTCTGTTTTCCCAACCTC GAAAGGTCTACAGTAACG	300
BAIAP1	Hs.169441	Brain-specific angiogenesis inhibitor-assoc. protein	200 kb	500 kb	AATTTTAGAGATCAATGGTGAG CACGTAAGCAAGCAAAAAG	319
UBE1C	Hs.154320	Ubiquitin-activating enzyme E1C	40 kb	750 kb	ATTCCCTTGAATAATTACTTG GTTACCGACTGTAAGTAAAGTG	248
MITF	Hs.166017	Microphthalmal-associated transcription factor	286 kb	1 Mb	GCATCATGCAGACCTAAC GTCTCTCCATGCTCATAC	255
β -tubulin	Hs.336780	Tubulin, beta polypeptide		Control gene	GCATCAACGTGTACTIONACAA TACGAGCTGGTGGACTGAGA	454
GAPDH	Hs.169476	glyceraldehyde-3-phosphate dehydrogenase		Control gene	ACCACAGTCCATGCCATCAC TCCACCACCCTGTTGCTTGTA	450
β 2	Hs.75415	Beta-2-microglobulin		Control gene	AGCTGTGCTCGCGCTACTCTCTC GTGTCCGATTGATGAAACCCAGACAC	140

^aEstimated genomic size as determined by sequence alignment and Santa Cruz genomic database. ^bApproximate gene distance relative to the proximal or distal ends of FRA3B as determined by sequence alignment and Santa Cruz genomic database. ^cSemi-quantitative RT-PCR oligonucleotide primers upper (forward) and lower (reverse) are shown 5' to 3'. ^dSemi-quantitative RT-PCR product size (bp)

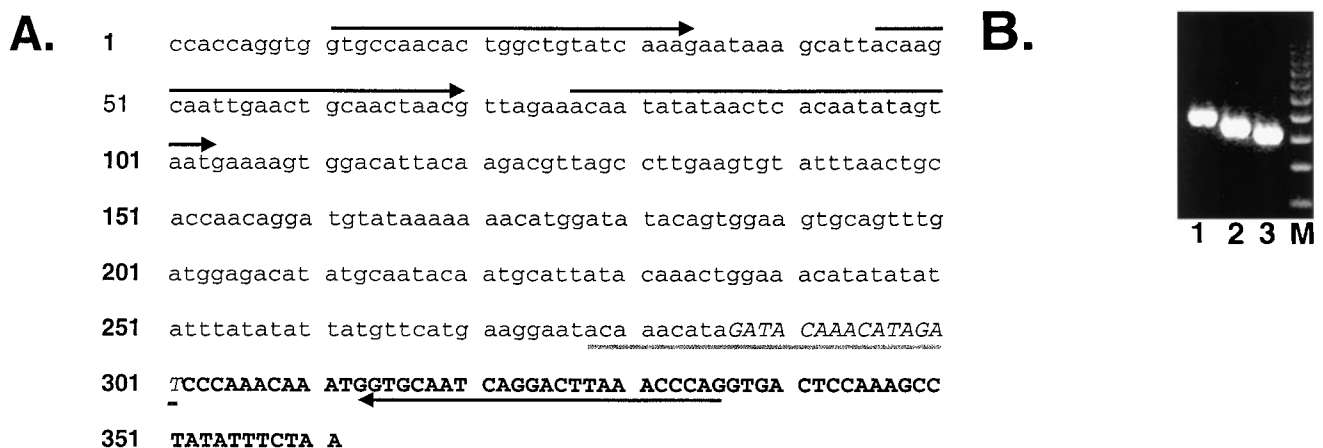


Figure 1 RS-PCR sequence results and integration confirmation for tumor CC61. (a) Lowercase letters indicate HPV16 sequence. Uppercase bold letters indicated human genomic sequence corresponding to BAC clone RP11 129_K_20 flanking the site of HPV16 integration. Uppercase italicized letters indicate bases that match neither HPV16 nor 129_K_20 sequences. The gray underlined letters indicate a 12 bp direct repeat at the site of viral integration. (b) Integration verification of the HPV16 integration in CC61. PCR was performed with three different oligonucleotides derived from HPV16 sequence (Thorland *et al.*, 2000) (Lane 1: HPV16-2929-24D; Lane 2: HPV16-2964-25D; Lane 3: HPV16-2995-28D) and an oligonucleotide derived from the sequence flanking the site of HPV16 integration (5'-CTGGGTTAAGTCCTGATGCACC) corresponding to BAC 129_K_20. These primer combinations produce products of 326, 291, and 260 bp, respectively. Forward HPV16 specific (black arrows above sequence) and reverse 3p14.2 specific (black arrow below sequence) primers are indicated. Fifty bp ladder (M) is shown (right)

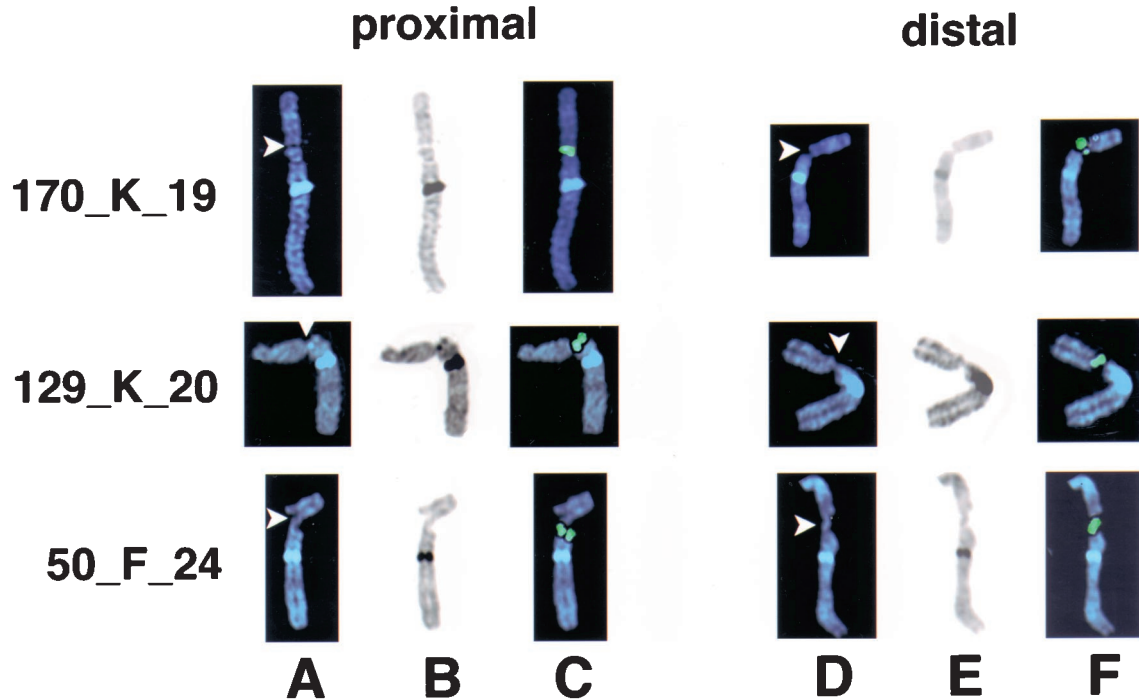


Figure 2 FISH-based analysis depicting aphidicolin-induced breakage at proximal and distal end of FRA3B (3p14.2). Breakage at 3p14.2 (white arrows) was initially identified using a DAPI image (a,d). Inverted DAPI banding confirmed breakage at 3p14.2 (b,e). Hybridization signal (green) of biotin-labeled specific BACs on the same aphidicolin-induced metaphase (c,f). Proximal (a–c) and distal (d,e) are shown for three BACs spanning the region of fragility. A summary of FISH results are shown in Table 2. BAC 170_K_19 (distal), 129_K_20 (HPV16 integration), 50_F_24 (proximal) are listed (left of panel)

Table 2 FISH analysis for BAC clones spanning the fragile site 3p14.2 (FRA3B)

Clones ^a	# Breaks ^b	Location of fluorescent signal relative to chromosomal location 3p14.2		
		Proximal (%)	Crossing (%)	Distal (%)
719_N_22	51	0 (0)	0 (0)	51 (100)
70_P_20	50	0 (0)	2 (4)	48 (96)
170_K_19	50	3 (6)	0 (0)	47 (94)
29_O_20	21	2 (9.5)	2 (9.5)	17 (81)
1_A_6	50	32 (64)	7 (14)	11 (22)
137_N_22	20	15 (75)	3 (15)	2 (10)
154_D_3	20	13 (65)	2 (10)	5 (25)
204_J_18	21	17 (80.9)	0 (0)	4 (19.1)
129_K_20	50	42 (80)	0 (0)	8 (20)
284_K_18	53	47 (88.7)	4 (7.5)	2 (3.7)
585_O_21	50	45 (90)	0 (0)	5 (10)
50_F_24	20	18 (90)	0 (0)	2 (10)
152_N_21	50	50 (100)	0 (0)	0 (0)

^aResearch Genetics RP-11 BAC clones hybridized to APC treated metaphase spreads. ^bNumber of APC induced metaphase breaks analysed for breakage at 3p14.2. The breaks were scored as either proximal (centromeric), crossing the break, or distal (telomeric)

renal cell carcinoma translocation breakpoint (hRCC t(3:8) translocation), and aphidicolin induced breakpoint clusters (Wang *et al.*, 1997; Zimonjic *et al.*, 1997). Since clone 129_K_20, which is 2 Mb proximal to the previous HPV16 integration, still localizes within the region of aphidicolin-induced instability, this suggests that instability at FRA3B is significantly larger than previously published results indicate. To define the entire

region of instability, we chose BACs across the region for FISH analysis on aphidicolin-induced metaphase chromosomes (Figure 3a). Using a single individuals blood culture, 50 metaphase breaks at FRA3B (3p14.2) were used to define the boundaries of fragility. A minimum of 20 breaks was counted when both proximal and distal breaks were observed for an individual BAC (Table 2). Our data indicates that RP11 152_N_21 and RP11 719_N_22 reside outside of FRA3B at its proximal and distal ends, respectively. Contig assembly across the region reveals that FRA3B fragility extends for approximately 4 Mb with the distal end contained within 3p14.2 and the proximal end extending into 3p14.1 (Figure 3a). One sequence gap is present that could not be closed by available sequence from the public databases (Figure 3a, double vertical black lines). This gap is located between FHIT exons 3 and 4.

Individual BAC clones hybridizing with approximately equal frequency proximally and distally to decondensation/breakage are observed within the FHIT locus between RP11 clones 137_N_22 and 29_O_20 (Figure 3a, Table 2). This localized region is considered the 'active' region, containing the previously published HPV16 integration site, the hRCC t(3:8) translocation, and the aphidicolin breakpoint clusters (Boldog *et al.*, 1993; Cohen *et al.*, 1979; Wang *et al.*, 1997; Wilke *et al.*, 1996; Zimonjic *et al.*, 1997) (Figure 3b). Our assembled BAC contig for FRA3B contains five known genes: SCA7, CADPS, HT021, PTPRG, and FHIT. The first four genes are completely

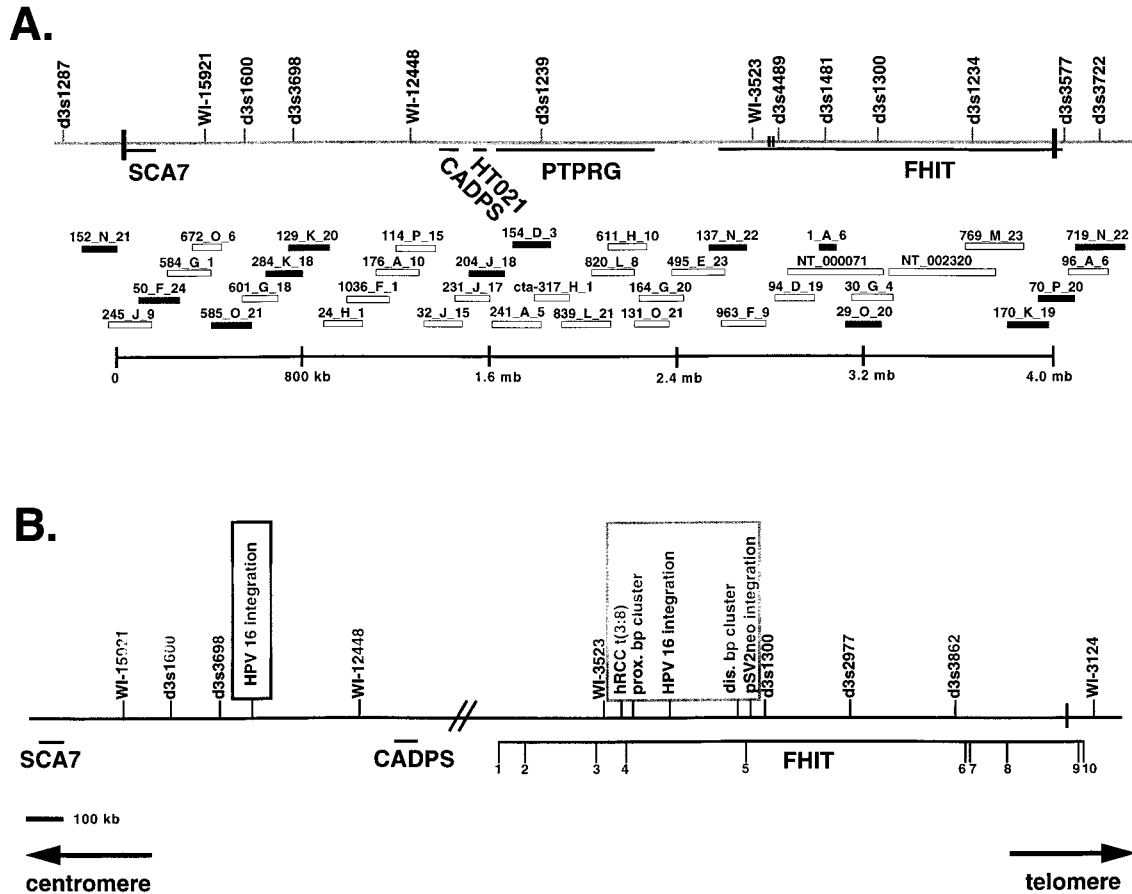


Figure 3 FRA3B BAC contig. (a) FRA3B BAC contig defining the proximal (centromeric) and distal (telomeric) ends. Markers across the region are indicated above the light gray line. The black vertical lines represent the proximal and distal ends for FRA3B. The double black vertical lines indicate sequence gaps of unknown size. The FRA3B genes and their relative genomic size (horizontal black lines) are indicated below the solid gray line. The NT numbers refer to complete NCBI sequences, 1_A_6 is a P1 cosmid clone (Paradee *et al.*, 1996), and cta-317_H_1 is a BAC clone sequenced by Washington University. All other clones are Research Genetics RP-11 BACs. BACs analysed by FISH (gray boxes) and BACs spanning the fragile site (white boxes) are indicated. The estimated genomic bp size of the region is indicated. (b) FRA3B ‘active’ site and new HPV16 integration. FRA3B ‘active region’ (gray box) depicts sites of integration, aphidicolin break points, a hRCC t(3:8) translocation previously mapped to the region (Paradee *et al.*, 1996; Wilke *et al.*, 1996). New HPV16 integration (black box) from tumor CC61 is indicated. FRA3B genes and their genomic size (horizontal black lines) are indicated with FHIT exons marked. Markers across the region are indicated above the black line. The black vertical line (telomeric end) represents the FRA3B proximal end. Double black vertical lines indicate sequence break FRA3B genes CADPS and FHIT. Directions of centromere and telomere are shown (bottom). A 100 kb reference size is also provided

contained within the fragile site with SCA7 being on the very proximal end of FRA3B. Sequence analysis indicates that the 3’ end of SCA7 is approximately 10 kb from BAC 152_N_21, which is outside the region of FRA3B fragility (Figure 3a). Only FHIT has a portion of the gene outside of the fragile site, with exons 9 and 10 localizing distal to the region of fragility (Figure 3).

Expression analysis of genes in and surrounding FRA3B

We assessed the expression of all genes within and surrounding the FRA3B region of instability in nine tumor-derived cervical cell lines. Previous work has demonstrated aberrant transcripts and loss of FHIT expression in cervical cell lines (Connolly *et al.*, 2000; Greenspan *et al.*, 1997; Hendricks *et al.*, 1997). In addition, FHIT and PTPRG, two genes contained within

FRA3B, have been proposed to be tumor suppressor genes (Cool and Fischer, 1993; Druck *et al.*, 1998).

For this analysis, we performed semi-quantitative RT-PCR on tumor-derived cervical cell lines using cultured keratinocytes as our normal control. Oligonucleotide DNA primers were designed to specifically amplify each of the five FRA3B genes (Table 1). These primers were optimized and multiplex RT-PCR was performed with either GAPDH primers or β -tubulin primers as controls. CADPS was not expressed in normal keratinocytes and was not further analysed for expression in the cervical cell lines. Of the remaining four genes, three (FHIT, PTPRG, and HT021) demonstrated aberrant gene expression (Figure 4). SCA7 was expressed in all cell lines analysed with intensities similar to normal keratinocytes. FHIT, however, had a complete loss of detectable expression in 66% (six out of nine) of the cell lines with apparent down regulation in the remaining three cell lines (Figure 4).

HT021 and PTPRG had complete LOE in 44.4% (four out of nine) and 22.2% (two out of nine), respectively. PTPRG was up regulated in two of the cell lines (Figure 4, lanes 7 and 9) and down regulated in an additional two cell lines (Figure 4, lanes 4 and 5). By comparison, HT021 had down regulation in one cell line (Figure 4, lane 4) and normal expression in the remaining four cell lines (Figure 4, lanes 6, 9, 10, and 11). These data reveal that three out of five genes in FRA3B have either complete loss of expression or aberrant gene expression in the analysed cervical cell lines.

To determine if only genes contained within FRA3B demonstrated aberrant gene expression, we also analysed genes outside the region of fragility (Figure 4, Table 1). Genes outside of FRA3B were chosen at random for analysis. The genes chosen localized between 500 kb and 1 Mb proximal or distal to the BACs, which defined the proximal and distal ends of the FRA3B region of instability (Table 1). Four distal (telomeric) genes were analysed, ARHGEF3, APPL, ARF4, and TU3A (Table 1). TU3A was not expressed in normal keratinocytes and was not further analysed for expression in the cervical cell lines. The remaining three genes (ARHGEF3, APPL, ARF4) all demonstrated expression levels similar to the normal control (Figure 4). Only APPL indicated lower expression levels, but no complete loss in any of the cell lines analysed (Figure 4). Three proximal (centromeric) genes were analysed, BAIAP1, UBE1C, and MITF (Table 1). Two of the three (UBE1C and BAIAP1)

proximal genes did not demonstrate aberrant gene expression (Figure 4). MITF was up-regulated in two of the examined cervical cell lines (Figure 4, lanes 3 and 6), but did not demonstrate loss of expression in any of the remaining cell lines (Figure 4).

Discussion

The integration of high-risk HPV subtypes is temporally associated with the development and/or progression of cervical cancer (Choo, 1998). It remains to be determined, however, if these integration events are pivotal in the development of cervical cancer. The genomic positions of HPV integrations have been of great interest to determine the significance of the integration site in cervical cancer development. Initial cytogenetic studies localized a high percentage of HPV genomic integrations to bands containing CFSs (Popescu *et al.*, 1990). We previously demonstrated at the molecular level that HPV16 integration events preferentially occur (~ 50% of integrations analysed) within CFSs (Thorland *et al.*, 2000, 2002). Preferential viral integration suggests that a unique feature exists within these unstable regions, but it is not known if this feature is solely the fragility within CFSs or the targeting of specific genes. Interestingly, many of the characterized integration events occurred within or immediately adjacent to genes proposed to play a role in cancer development. This not only supports the

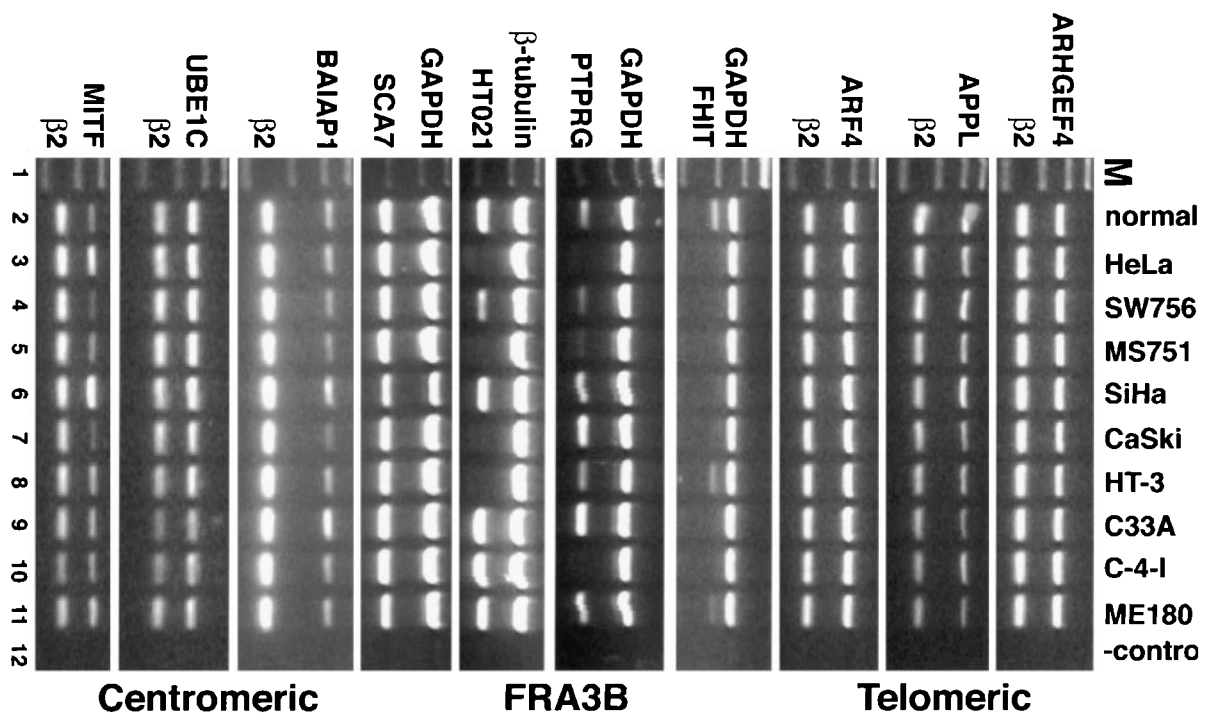


Figure 4 Semi-quantitative RT-PCR for genes in and surrounding FRA3B. Genes analysed along with internal control are indicated (left of panel). The normal control (cultured keratinocytes), tumor-derived cell lines, and negative control (H₂O) are shown (top of panel). Relative gene position is indicated (right of panel). Expected gene product size is described in Table 1. One hundred bp ladder (M, lane 1)

hypothesis that the integration events may be targeting regions of instability, such as the CFSs, but also indicates that genes at the site of integration may be important in the development of cervical cancer.

In this report, we further characterized a previously defined (Thorland *et al.*, 2002) single HPV16 integration event that occurred within FRA3B at 3p14.2. FISH analysis revealed that a BAC clone which spanned this integration mapped within the proximal end of FRA3B even though the integration occurred more than 2 Mb centromeric to the HPV16 integration originally described by Wilke *et al.*, 1996 (Figure 3a). The new HPV16 integration did not occur within any known gene or EST, but instead localizes between the two known genes SCA7 and CADPS (Figure 3). The other end of this HPV16 integration has not been identified, so it is not known if this integration is associated with a deletion at 3p14.2 or a more complex translocation. In addition, sequence analysis at the HPV16 integration site identified a short 15 bp sequence that is not associated with either HPV16 or human sequences at 3p14.2 (Figure 1a). This unknown sequence creates a 12 bp direct repeat that is located at the junction between the HPV and human sequences (Figure 1a). Orphan sequences such as this have been observed at the junctions of other HPV integrations (Gallego *et al.*, 1997). These sequences may be derived from cellular sequences involved in the integration process or may be due to illegitimate repair after the integration event. However, this novel sequence was not observed in any of the other HPV16 integrations described by Thorland *et al.* (2000, 2002).

The identification of this new HPV16 integration led to the complete characterization of the region of instability surrounding FRA3B. In the previous analysis of FRA3B, a 500 kb cosmid contig around the HPV16 integration described by Wilke *et al.*, 1996 was constructed with cosmids derived from the original YAC (850A6) that was found to span the hRCC (Paradee *et al.*, 1996; Rassool *et al.*, 1996). These cosmid clones were used as FISH-based probes to characterize the FRA3B region and this defined the 'center' of FRA3B instability as localizing between FHIT exons 4 and 5. However, these data did not define the 'ends' or boundaries of fragility in FRA3B. Based upon the HPV16 integration which is 2 Mb centromeric of the 'center' of FRA3B, but which still occurs within the region of instability, we sought to completely characterize the region of instability at FRA3B. To define the boundaries of fragility we assessed 50 metaphase breaks for each BAC clone at 3p14.1-3p14.2, offering a more comprehensive definition of the ends of FRA3B (Table 2).

In addition to defining the boundaries of fragility, we also wanted to determine the FRA3B 'active' site and compare this with results with previous reports (Paradee *et al.*, 1996; Rassool *et al.*, 1996). The 'active' site for a CFS can be defined as the region of instability where breakage, determined by FISH-based analysis, occurs with a relatively equal frequency both

proximal and distal to the large insert clone tested. Our data indicate the FRA3B 'active' site remains between exons 4 and 5 of the FHIT locus, which agrees with previous reports (Paradee *et al.*, 1996; Rassool *et al.*, 1996). However, the 'active' site is not in the physical center of the fragile site. The proximal end for FRA3B extends ~3 Mb beyond the center of instability, whereas the distal end only extends ~700 kb (Figure 3). Interestingly, the most proximal gene, SCA7, contains an unstable (CAG) trinucleotide repeat in the coding sequence. In RFSs a link has been made between unstable repeats, either trinucleotide (CCG) repeats or mini-satellite (AT) expansions, and the mechanism of RFS fragility (Usdin and Woodford, 1995; Verkerk *et al.*, 1991). Although the trinucleotide expansion in SCA7 patients is different (CAG *vs* CCG), both of these expansions result in unusual structures *in vitro* that could inhibit replication machinery (Kang *et al.*, 1995; Parniewski *et al.*, 1999; Usdin and Woodford, 1995). Since the mechanism for RFS fragility is not completely understood, but known to be linked to unstable expansions associated with a disease-state, it is possible that the variable, unstable SCA7 CAG repeat found even in normal individuals is sufficient to affect fragility without having the full disease state expansion. The presence of SCA7 in FRA3B is the first unstable trinucleotide repeat that has been shown to be contained within a CFS and may offer a potential link between CFSs and RFSs in terms of their mechanism of fragility.

To determine if any of the five known genes within FRA3B are targeted for gene inactivation in cervical cancer, we assessed gene expression in nine tumor-derived cervical cell lines. Expression analysis revealed that the very large PTPRG and FHIT genes were down-regulated or showed LOE in cervical tumor cell lines. This indicates that there may be a targeting of the 'large' genes in FRA3B, but HT021 (an average-sized gene covering ~60 kb) also had marked inactivation. The most proximal gene, SCA7, did not reveal any down regulation (Figure 4). Since this gene is on the very end proximal of FRA3B, the instability associated with this site may not have an effect on gene regulation. In order to confirm that the genes within FRA3B may be specifically targeted for gene inactivation in cervical tumor-derived cell lines, we also assessed the expression of genes outside the region of FRA3B instability (Figure 4). The surrounding genes did not indicate loss of expression or down-regulation (Figure 4). These data support our FISH-based observations of a 4 Mb region of FRA3B instability and demonstrates that the genes within the region of fragility are being targeted for gene inactivation. Several previous studies, using different tumor types, have determined that the aberrant and alternative transcripts associated with FHIT expression are a complex combination of methylation changes and genomic deletions (Gonzalez *et al.*, 1998; Huebner *et al.*, 1998; Tanaka *et al.*, 1998; Vertino *et al.*, 1993; Zochbauer-Muller *et al.*, 2001). In addition, a recent publication studied alterations in the FHIT/FRA3B

region by generating somatic cell hybrid that separated chromosome 3 homologs from each. The data reveal multiple different chromosome 3 homologs isolated from each individual cell line (Corbin *et al.*, 2002). The net effect is that the instability within this region generates a great deal of heterogeneity and complexity.

The data generated in this paper is in complete agreement with previously published reports characterizing FRA3B (Kastury *et al.*, 1996; Paradee *et al.*, 1996). The key distinction is that we have now more comprehensively analysed clones from this region to define the ends of decondensation/breakage in the FRA3B region. It is interesting that the 'active' center of this region is not in the physical center of this fragile site. The proximal end for FRA3B extends for ~3 Mb beyond the center of instability, whereas the distal end extends ~700 kb. FISH and sequence analysis reveals the region of instability for FRA3B fully extends over a 4 Mb region containing five genes. Our data indicates that the regions of genomic instability associated with CFSs, specifically FRA3B, are much larger than previously believed. This large region of defined genomic instability offers an explanation for why 3p14.2 has been shown to be involved in a variety of cancers including cervical cancer. The data generated defining the FRA3B ends will allow for further analysis of this region of instability.

Materials and methods

Cervical tumor samples, RS-PCR, and HPV16 integration identification

Cervical tumors, grade 2–4 squamous cell carcinomas, were obtained from 26 patients. DNA was extracted and HPV typing was performed as described in Gostout *et al.* (1998). RS-PCR (Sarkar *et al.*, 1993) was performed on samples using oligonucleotide primers and conditions previously described (Thorland *et al.*, 2000, 2002).

BAC clone selection and FRA3B contig assembly

Sequenced RS-PCR products were BLASTN searched in both the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) non-redundant complete sequences (nr) and high throughput genomic sequences (htgs) databases to determine what similarities existed to the HPV16 and human genomes. BACs containing the human genomic sequence were obtained from Research Genetics and extracted BAC DNA was used as a probe for FISH-based cytogenetic localization. A BAC contig surrounding the published FRA3B region was assembled using the NCBI, the Santa Cruz Genome Center (<http://genome.ucsc.edu>), and the San Antonio Genome Center (<http://apollo.uthscsa.edu>) public databases. Unassembled BACs were aligned using Sequencher 4.1.2 (Gene Codes Corp., Ann Arbor, MI, USA).

Cytogenetic localization of clones

Metaphase preparations were obtained from blood cultures established from 1 ml of peripheral whole blood and 9 ml of Chang Media PB (Irvine Scientific, Santa Ana, CA, USA).

Cultures were incubated at 37°C in 5% CO₂ for 72 h. Twenty-four hours prior to harvest, the cultures were induced with 0.4 μM aphidicolin (Sigma-Aldrich, St. Louis, MO, USA) solution. Cell harvest and metaphase preparations followed routine cytogenetic techniques.

For each BAC clone, standard nick translation was used to incorporate biotin-16-dUTP (Boehringer/Roche, Indianapolis, IN, USA) into 1 μg of purified BAC DNA followed by precipitation and hybridization to aphidicolin-treated metaphase chromosomes according to the protocol described by Verma and Babu (1989). Probe detection and amplification followed the manufacturer's protocols (Ventana Medical Systems, Tucson, AZ, USA) with minor modifications. Chromosomes were counterstained with DAPI. Photomicroscopy was performed using a Zeiss Axioplan 2 fluorescence microscope and Powergene MacProbe software (Applied Imaging, Santa Clara, CA, USA). The position of each individual BAC clone relative to FRA3B was determined by the analysis of aphidicolin-treated metaphases with breakage at 3p14.2. A minimum of 50 breaks at 3p14.2 was used to define the proximal and distal ends. Breakage at 3p14.2 was identified by band location as established by DAPI banding. A clone was determined to be within the fragile site region if hybridization signal was observed on both sides of decondensation/breakage or if signal was observed as occurring proximal (centromeric) in one metaphase and distal (telomeric) in a separate metaphase.

Cell culture and RNA isolation

Primary keratinocytes were cultured as previously described and used as a normal control (Poumay and Pittelkow, 1995). Nine cervical tumor-derived cell lines (HeLa, SW756, MS751, SiHa, CaSki, HT-3, C-33-A, C-4-I, and ME-180) were obtained from the American Type Culture Collection. The cells were maintained at 37°C in 5% CO₂ in their recommended media. Total RNA was extracted from normal cultured keratinocytes and tumor-derived cervical cell lines using TRIZOL reagent (GibcoBRL, Rockville, MD, USA) following the manufacturer's instructions.

Semi-quantitative RT-PCR

For each reverse transcription (RT) reaction, DNA was eliminated by treating 5 μg of total RNA with RNase-free DNase I for 15 min at 25°C followed by DNase I inactivation by addition of 2 mM MgCl₂ and incubation at 65°C for 10 min. DNase-treated RNA RT was performed using MMLV-RT (GibcoBRL, Rockville, MD, USA).

Gene specific DNA oligonucleotide primers were designed for the genes contained within and surrounding FRA3B using Oligo 6.4 software (Molecular Biology Insights, Cascade, CO, USA) and multiplexed for semi-quantitative RT-PCR with either β-tubulin, GAPDH, or β2 control primers (Table 1). Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA, USA). The semi-quantitative RT-PCR reactions (12.5 μl total volume) contained 50 ng of reverse-transcribed cDNAs, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.2 mM dNTPs, 1.5 mM MgCl₂, 5 pmol of each gene specific primer, 0.5 pmol of the control primer and 0.1 U of Taq polymerase (Promega, Madison, WI, USA). The conditions for amplification were: 98°C for 3 min, then two cycles of 98°C for 30 s, 55–59°C for 30 s, and 72°C for 30 s, followed by 28 cycles of 94°C for 30 s, 55–59°C for 30 s, and 72°C for 30 s with a final extension of 72°C for 10 min. Samples were analysed by ethidium bromide agarose gel electrophoresis.

Abbreviations

BAC, bacterial artificial chromosome; CFS, common fragile site; FISH, fluorescence *in situ* hybridization; FS, fragile site; HPV, human papillomavirus; Mb, megabase pairs

References

- Boldog FL, Gemmill RM, West J, Robinson M, Robinson L, Li E, Roche J, Todd S, Waggoner B, Lundstrom R, Jacobson J, Mullokandov MR, Klinger H and Drabkin HA. (1997). *Hum. Mol. Genet.*, **6**, 193–203.
- Boldog FL, Gemmill RM, Wilke CM, Glover TW, Nilsson AS, Chandrasekharappa SC, Brown RS, Li FP and Drabkin HA. (1993). *Proc. Natl. Acad. Sci. USA*, **90**, 8509–8513.
- Choo KR. (1998). *The genetic basis of human cancer*. Vogelstein B and Kinzler KW (ed). New York: McGraw-Hill, Inc., pp. 631–637.
- Cohen AJ, Li FP, Berg S, Marchetto DJ, Tsai S, Jacobs SC and Brown RS. (1979). *N. Engl. J. Med.*, **301**, 592–595.
- Connolly DC, Greenspan DL, Wu R, Ren X, Dunn RL, Shah KV, Jones RW, Bosch FX, Munoz N and Cho KR. (2000). *Clin. Cancer Res.*, **6**, 3505–3510.
- Cool DE and Fischer EH. (1993). *Semin. Cell Biol.*, **4**, 443–453.
- Coquelle A, Pipiras E, Toledo F, Buttin G and Debatisse M. (1997). *Cell*, **89**, 215–225.
- Corbin S, Neilly ME, Espinosa III R, Davis EM, McKeithan TW and LeBeau MM. (2002). *Cancer Res.*, **62**, 3477–3484.
- Druck T, Berk L and Huebner K. (1998). *Oncol. Res.*, **10**, 341–345.
- Fang JM, Arlt MF, Burgess AC, Dagenais SL, Beer DG and Glover TW. (2001). *Genes Chromosomes Cancer*, **30**, 292–298.
- Gallego MI, Schoenmakers EFPM, Van de Ven WJM and Lazo PA. (1997). *Mol. Carcinog.*, **19**, 114–121.
- Glover TW, Coyle-Morris JF, Li FP, Brown RS, Berger CS, Gemmill RM and Hecht F. (1988). *Cancer Genet. Cytogenet.*, **31**, 69–73.
- Glover TW and Stein CK. (1987). *Am. J. Hum. Genet.*, **41**, 882–890.
- Glover TW and Stein CK. (1988). *Hum. Genet.*, **43**, 265–273.
- Gonzalez MV, Pello MF, Ablanedo P, Suarez C, Alvarez V and Coto E. (1998). *J. Clin. Pathol.*, **51**, 520–524.
- Gostout BS, Podratz KC, McGovern RM and Persing DH. (1998). *Am. J. Obstet. Gynecol.*, **179**, 56–61.
- Greenspan DL, Connolly DC, Wu R, Lei RY, Vogelstein JT, Kim YT, Mok JE, Munoz N, Bosch FX, Shah KV and Cho KR. (1997). *Cancer Res.*, **57**, 4692–4698.
- Hendricks DT, Taylor R, Reed M and Birrer MJ. (1997). *Cancer Res.*, **57**, 2112–2115.
- Herzog C, Crist K, Sabourin C, Kelloff G, Boone C, Stoner G and You M. (2001). *Mol. Carcinog.*, **30**, 159–168.
- Huang H, Qian JH, Proffitt J, Wilber K, Jenkins RB and Smith DI. (1998). *Oncogene*, **16**, 2311–2319.
- Huebner K, Garrison PN, Barnes LD and Croce CM. (1998). *Ann. Rev. Genet.*, **32**, 7–31.
- Kang S, Oshshima K, Shimizu M, Amirhaeri S and Wells RD. (1995). *J. Biol. Chem.*, **270**, 27014–27021.
- Kastury K, Ohta M, Lasota J, Moir D, Dorman T, LaForgia S, Druck T and Huebner K. (1996). *Genomics*, **32**, 225–235.
- Krummel KA, Roberts LR, Kawakami M, Glover TW and Smith DI. (2000). *Genomics*, **69**, 37–46.
- Mishmar D, Rahat A, Scherer SW, Nyakatura G, Nihzmann B, Kohwi Y, Mandel-Gutfroind Y, Lee JR, Drescher B, Sas DE, Margalit H, Platzner M, Weiss A, Tsui LC, Rosenthal A and Kerem B. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 8141–8146.
- Paradee W, Wilke CM, Wang L, Shridhar R, Mullins CM, Hoge AW, Glover TW and Smith DI. (1996). *Genomics*, **35**, 87–93.
- Parniewski P, Bacolla A, Jaworski A and Wells RD. (1999). *Nucleic Acids Res.*, **27**, 616–623.
- Popescu NC, Zimonjic DB and DiPaolo JA. (1990). *Hum. Genet.*, **84**, 383–386.
- Poumay Y and Pittelkow MR. (1995). *J. Invest. Dermatol.*, **104**, 271–276.
- Rabbitts TH. (1994). *Nature*, **372**, 143–149.
- Rassool FV, Le Beau MM, Shen ML, Neilly ME, Espinosa III R, Ong ST, Boldog FL, Drabkin H, McCarroll R and McKeithan TW. (1996). *Genomics*, **35**, 109–117.
- Rassool FV, McKeithan TW, Neilly ME, van Melle E, Espinosa R and Le Beau MM. (1991). *Proc. Natl. Acad. Sci. USA*, **88**, 6657–6661.
- Sandberg AA. (1990). *The chromosome in human cancer and leukemia*. 2nd edn. New York: Elsevier Science Publishing Co., Inc, 1315 pp.
- Sarkar G, Turner RT and Bolander ME. (1993). *PCR Methods Appl.*, **2**, 318–322.
- Smeets DF, Scheres JM and Hustinx TW. (1986). *Hum. Genet.*, **72**, 215–220.
- Solomon E, Borrow J and Goddard AD. (1991). *Science*, **254**, 1153–1160.
- Tanaka H, Shimada Y, Harada H, Shinoda M, Hatooka S, Imamura M and Ishizaki K. (1998). *Cancer Res.*, **58**, 3429–3434.
- Tatarelli C, Linnenbach A, Mimori K and Croce CM. (2000). *Genomics*, **68**, 1–12.
- Thorland EC, Myers SL, Persing D, Sarkar G, McGovern RM, Gostout BS and Smith DI. (2000). *Cancer Res.*, **60**, 5916–5921.
- Usdin K and Woodford KJ. (1995). *Nucleic Acids Res.*, **23**, 4202–4209.
- Verkerk A, Pieretti M, Sutcliffe JS, Fu YH, Kuhl D, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, van Ommen G-JB, Blonden L, Riggins GJ, Chastain JL, Kunst GH, Caskey CT, Nelson D, Oostra BA and Warren ST. (1991). *Cell*, **65**, 905–914.
- Verma RS and Babu A. (1989). *Human Chromosomes: Manual of Basic Techniques*. Verma RS and Babu A (ed). New York: Pergamon Press, Inc., 240 pp.
- Vertino PM, Spillare EA, Harris CC and Baylin SB. (1993). *Cancer Res.*, **53**, 1684–1689.
- Walboomers JMM, Jacobs MV, Manos MM, Bosch FX, Kummer A, Shah KV, Sniders PJJ, Peto J, Meuer CJLM and Munoz N. (1999). *J. Pathol.*, **189**, 12–19.
- Wang L, Paradee W, Mullins C, Shridhar R, Rosati R, Wilke CM, Glover TW and Smith DI. (1997). *Genomics*, **41**, 485–488.
- Wilke CM, Hall BK, Hoge AW, Paradee W, Smith DI and Glover TW. (1996). *Hum. Mol. Genet.*, **5**, 187–195.

Zimonjic DB, Druck T, Ohta M, Kastury K, Croce CM, Popescu NC and Huebner K. (1997). *Cancer Res.*, **57**, 1166–1170.

Zochbauer-Muller S, Fong KM, Maitra A, Lam S, Geradts J, Ashfaq R, Virmani AK, Milchgrub S, Gazdar AF and Minna JD. (2001). *Cancer Res.*, **61**, 3581–3585.