ARTICLE

Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B

Ferenc Boldog1,‡, Robert M. Gemmill1,‡, James West1, Misi Robinson1, Linda Robinson1, Efang Li1, Joelle Roche1,§, Sean Todd1, Barbara Waggoner1, Ron Lundstrom2, Jan Jacobson1, Michael R. Mulkokandov3, Harold Klinger3 and Harry A. Drabkin1,*

1Division of Medical Oncology, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA, 2Eleanor Roosevelt Institute, 1899 Gaylord, Denver, CO 80206, USA and 3Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Ave., New York, NY 10461, USA

Received September 16, 1996; Revised and Accepted November 27, 1996

Loss of heterozygosity (LOH) involving 3p occurs in many carcinomas but is complicated by the identification of four distinct homozygous deletion regions. One putative target, 3p14.2, contains the common fragile site, FRA3B, a hereditary renal carcinoma-associated 3;8 translocation and the candidate tumor suppressor gene, FHIT. Using a ∼300 kb cosmid/λ contig, we identified homozygous deletions in cervix, breast, lung and colorectal carcinoma cell lines. The smallest deletion (CC19) was shown not to involve FHIT coding exons and no DNA sequence alterations were present in the transcript. We also detected discontinuous deletions as well as deletions in non-tumor DNAs, suggesting that FHIT is not a selective target. Further, we demonstrate that some reported FHIT aberrations represent normal splicing variation. DNA sequence analysis of 110 kb demonstrated that the region is high in A–T content, LINEs and MER repeats, whereas Alu elements are reduced. We note an intriguing similarity in repeat sequence composition between FRA3B and a 152 kb segment from the Fragile-X region. We also identified similarity between a FRA3B segment and a small polydispersed circular DNA. In contrast to the selective loss of a tumor suppressor gene, we propose an alternative hypothesis, that some putative targets including FRA3B may undergo loss as a consequence of genomic instability. This instability is not due to DNA mismatch repair deficiency, but may correlate in part with p53 inactivation.

INTRODUCTION

LOH involving 3p occurs frequently in carcinomas of the lung, kidney, cervix, breast and other epithelial neoplasms (1–7). However, 3p loss is complex, involving at least four distinct homozygously deleted regions (8–12). One of the most frequently lost regions is 3p14, especially in cervical carcinomas (13). This region is of interest since it contains the site of a hereditary renal carcinoma-associated translocation, t(3;8)(p14.2;q24.1) (14), and is the location of the most inducible common fragile site in the genome, FRA3B (15). We previously reported cloning of the 3;8 translocation breakpoint (16) and demonstrated by fluorescence in situ hybridization that FRA3B was further telomeric (17). These studies also suggested that breaks in FRA3B occur over a region instead of at a single site. While searching for genes in this region, we identified a homozygous deletion in the cervical carcinoma cell line, HeLa, involving marker D3S1300.

We developed a ∼300 kb cosmid/λ contig within FRA3B containing D3S1300. Probes from the region detected frequent homozygous deletions in cervical, lung, colorectal and breast carcinoma cell lines. Cervical carcinomas, which are associated with papilloma virus infection and p53 inactivation (18), were most frequently deleted and the smallest deletion occurred in cell line CC19. During our investigations, Ohta et al. reported identification of a candidate tumor suppressor gene, FHIT, which spanned the t(3;8) breakpoint and was deleted in various carcinoma cell lines (19). However, our analysis indicates that FHIT is not the target of these deletions. We also observed that 3p14 deletions tend not to occur in tumors with deficiencies in DNA mismatch repair.

‡To whom correspondence should be addressed
§Present address: CNRS URA1172, IBMIG, Universite de Poitiers, 40 Av. du Recteur Pineau, 86022 Poitiers CéDEX, France
These authors contributed equally to this work
Table 1. Results of microsatellite analysis and 3p14 deletions in selected cell lines

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Mismatch repair deficiency/ microsatellite instability</th>
<th>3p14 status</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal carcinoma&lt;br&gt;a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT-29</td>
<td>intact</td>
<td>deletion</td>
<td>this study</td>
</tr>
<tr>
<td>SW-480</td>
<td>intact</td>
<td>deletion</td>
<td>Umar et al. (25); this study</td>
</tr>
<tr>
<td>SW-403</td>
<td>intact</td>
<td>deletion</td>
<td>this study</td>
</tr>
<tr>
<td>SW-948</td>
<td>intact</td>
<td>rearranged</td>
<td>this study</td>
</tr>
<tr>
<td>SW-48</td>
<td>defective</td>
<td>no deletion</td>
<td>Boyer et al. (23); this study</td>
</tr>
<tr>
<td>DLD1/HCT15</td>
<td>defective</td>
<td>no deletion</td>
<td>da Costa et al. (24)</td>
</tr>
<tr>
<td>Cervical carcinoma&lt;br&gt;c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-41</td>
<td>intact</td>
<td>deletion</td>
<td>Larson et al. (26)d</td>
</tr>
<tr>
<td>Caski</td>
<td>intact</td>
<td>deletion</td>
<td>Larson et al. (26)d</td>
</tr>
<tr>
<td>SIHA</td>
<td>intact</td>
<td>deletion</td>
<td>Larson et al. (26)d</td>
</tr>
<tr>
<td>HeLa</td>
<td>intact</td>
<td>deletion</td>
<td>Boyer et al. (23); Umar et al. (25)</td>
</tr>
<tr>
<td>CC19</td>
<td>NT</td>
<td>deletion</td>
<td>Larson et al. (26)d</td>
</tr>
<tr>
<td>MS-751</td>
<td>intact</td>
<td>deletion</td>
<td>Larson et al. (26)d</td>
</tr>
<tr>
<td>C-33A</td>
<td>defective</td>
<td>deletion</td>
<td>Larson et al. (26)d; this study</td>
</tr>
<tr>
<td>ME-180</td>
<td>intact</td>
<td>no deletion</td>
<td>Larson et al. (26)d</td>
</tr>
</tbody>
</table>

aDefective refers to a reported deficiency in mismatch repair or microsatellite instability.
bOverall, deletions (5) or rearrangements (1) were identified in 6/12 colorectal carcinoma DNAs. In addition to those listed, deletions occurred in COLO-205 and COLO-320. No deletions were detected in SKCO-1, CaCo-2, SW-1417 and T-84.
cRepresents all cervical carcinoma samples examined.
dPersonal communication from Dr Garret M. Hampton.

DNA sequencing studies were performed to identify features that might provide insight into these breaks. Our results demonstrate that FRA3B is high in A–T content, LINES and MER repeats. In contrast, Alu elements and confirmed genes are reduced. We identified a FRA3B segment highly similar to a reported small polydispersed circular DNA, sequences which are markedly elevated in damaged or unstable genomes. FRA3B also bears an intriguing overall sequence similarity to the Fragile-X region. However, unlike the rare folate-sensitive fragile sites, no triplet repeats nor methylated CpG island was identified. These overall features may be responsible for, or contribute to, the observed instability of this region.

RESULTS

Identification of homozygous deletion

Following identification of the HeLa deletion, additional cervical carcinoma lines were examined using probes spanning a ∼300 kb cosmid/phage contig from the FRA3B region (Fig. 1A and B).

Figure 2A shows an example using cosmid c136C3. Deletions were evident in DNAs from MS751, SIHA and Caski (lanes 2, 5 and 7) whereas CC19 (lane 6) shows missing and altered bands (arrow). Altogether, homozygous deletions were detected in seven out of eight lines (87.5%). Similar hybridizations were performed using DNAs from 12 colon tumor lines, and deletions or rearrangements were seen in 50% (Fig. 2B). Other homozygous deletions (Fig. 1C) were detected in three lung (NCI-H1648, NCI-H460 and NCI-H892) and one breast carcinoma (MDA-231) cell lines. The smallest (∼40 kb) was in the cervical carcinoma line CC19. This line was characterized further by DNA sequence analysis and FHIT gene expression (see below). Two (HeLa and MDA-231) contained discontinuous deletions, which was surprising since a single deletion should have been sufficient to inactivate a tumor suppressor gene. No deletions were detected in five renal carcinoma lines (KRC/Y, CAKI-1, ACHN and KV6).

We also observed deletions in non-tumor-derived DNAs. Hybridization with c61G12 to a chromosome 3 hybrid panel revealed a partial deletion in 2A3CT (Fig. 3A). Similarly, a

Figure 1. (A) Schematic showing position of the homozygous deletion region with respect to YACs 850A6, 6SE7 and 74B2. Positions of MluI and selected XhoI restriction sites are indicated. XhoI sites were not determined for YAC 850A6. (B) Cosmid and λ clone contig spanning the homozygous deletion region within 3p14. Solid horizontal lines represent cosmid (c) and lambda (λ) inserts, as indicated, along with cleavage sites for EcoRI (short vertical bars), XhoI (X) and SalI (S) which were mapped within the central 170 kb. The MluI site corresponding to the site in YAC 74B2 is present in c55D12 and c84F12, although we have not determined its precise location. Positions of selected markers, breakpoints and integration sites for HPV-16 and pSV2neo are indicated across the top. (B) and (C) are drawn to the same scale and positions correspond exactly between the two. (C) Homozygous deletions in tumors and normal genomic DNAs. Horizontal lines denote the extent of deletion in the indicated cell lines. Parallel lines at the ends of each deletion indicate where precise boundaries were not determined. HeLa and MDA231 contain discontinuous deletions indicated by interrupted lines. Exon 5 and the direction of transcription for the FHIT gene are indicated at the bottom.
Figure 2. (A) Southern analysis of cervical carcinoma cell lines. Lanes 1–7 contain DNA from seven cervical lines, as indicated; lane 8 is a normal human DNA control. DNA samples were digested with EcoRI and hybridized with c136C3. Three cell lines were homozygously deleted for c136C3 while the arrow indicates a rearranged band present in CC19. (B) Southern analysis of colon carcinoma cell lines. Control lanes (1–4) contain respectively, λ HindIII size marker, HindIII-digested human DNA, the chromosome 3-specific hybrids 2A3CT and UCH12. Hybrid 2A3CT contains a single copy of chromosome 3 deleted for all sequences distal to 3p21.3. The chromosome retained in UCH12 is deleted for the entire short (p) arm. Lanes 5–18 contain HindIII-digested DNAs from 12 colon carcinoma cell lines (note, DLD-1 and HCT-15 are identical). The Southern blot was hybridized with c61G12 which contains D3S1300 and FHIT exon 5. Homozygous deletions were observed in three cell lines; altered bands, which may be indicative of rearrangements, were present in four additional lines.

Figure 3. Deletions in genomic DNAs from normal sources. DNA from several hybrids and YAC clones were analyzed with probes from the homozygous deletion region. In (A), DNA from seven chromosome 3 hybrids and a deletion variant of YAC clone 74B2 were digested with HindIII and hybridized with c61G12. The variant band in hybrid H3-4 represents a polymorphism identified in eight of 19 normal DNA samples (not shown). The seven hybrids all retain 3p14.2 by cytogenetic and molecular genetic analyses while missing other specific regions of chromosome 3. (B) DNA samples from YACs 74B2 and the deletion variant 74B2D were digested with EcoRI and hybridized with c31E1. Four homologous bands present in YAC 74B2 were missing in 74B2D (arrows). In (C), EcoRI-digested DNA from the hybrid A5-4 (lane 4) was compared with human (lane 1) and hybrid 3;8/4-1 (lane 3) using c31E1.

spontaneous 80 kb deletion in YAC 74B2 spanning the region was identified during single clone purification (74B2D). Figure 3B shows missing bands (arrows) in 74B2D when hybridized with c31E1. This 80 kb deleted segment encompasses the aphidicolin-induced breakpoint in hybrid A5-4 (Figs 3C and 1C), the ends of several tumor deletions (C33A, SiHa, HeLa, COLO-205 and MDA-231) and the more telomeric of two pSV2neo plasmid insertion sites which preferentially integrated into FRA3B after aphidicolin treatment (20,21).

Correlation of deletions with microsatellite instability

In cervical and colorectal carcinoma cell lines, we observed a trend of inverse correlation between the presence of a 3p14 deletion and reported microsatellite instability (22–26). However, for some lines, we were unable to discern the replication error (RER) status from the literature. Therefore, we subcloned selected lines and tested 10 clones each with up to six microsatellite loci (D3S1300, D3S1210, D3S1286, D3S1233 and AFM320yb5). Instability was accepted if new bands appeared from two or more loci, although there were no cases of only one alteration. Results for samples where we have information on both DNA mismatch repair and 3p14 deletion are shown in
Table 1. While the number of RER+ lines is small, both 3p14 deletions and the RER phenotype are usually discordant. One simple conclusion is that defects associated with microsatellite instability alone are not responsible for the observed deletions. We also observed the highest incidence of deletions in cervical carcinoma lines, where p53 alterations appear very common (27,28). This may relate to the reduced frequency of deletions observed in RER+ colorectal tumors (see Discussion).

Relative position of FHIT

To examine the role of the FHIT gene in these deletions, we derived primers (FHIT-5′ and FHIT-3′) for a one-stage RT-PCR amplification of the entire coding region (exons 5–9). An expected 638 bp product was amplified after 35 cycles from embryonal kidney 293 cells (Fig. 4A, lane 7). This was used as a probe against the contig which showed that only a single FHIT exon, located adjacent to D3S1300, was present (Fig. 1B). Comparison of our map with that described by Ohta et al. (19) allowed us to conclude that this was exon 5 (also confirmed by DNA sequencing, not shown) and demonstrated that several tumor deletions exclusively affected large introns (Fig. 1C). This occurred in CC19 and SIHA (cervix), NCI-H1648 and NCI-H460 (lung) and MDA-231 (breast).

Analysis of FHIT in the tumor lines shown in Figure 4A demonstrated that 3/6, including CC19, contained normal sized bands (arrow). To characterize further the FHIT product from CC19, four independent isolates were sequenced. Each contained only normal sequences comprising coding exons 5–9. Thus, the homozygous deletion had no apparent effect on FHIT mRNA, suggesting either a different target gene or unselected genomic instability. Additionally, a faint normal sized product was seen in MS751 (Fig. 4A, lane 5), which by Southern blot contains a homozygous deletion including exon 5. A possible explanation is that the deletion is heterogeneous within the cell population suggesting it occurred during culture.

The cervical carcinoma lines C33A and SIHA contained multiple smaller RT-PCR products with no detectable wild-type product. While C33A had an exon 5 deletion that could explain one smaller band (Fig. 1C), the multiplicity of products suggested alternative splicing. The SIHA deletion does not affect an exon although there may be non-recognized discontinuous or overlapping bi-allelic deletions. Despite these obvious differences, the RT-PCR products appeared identical (Fig. 4A, lanes 2 and 3). To examine this, we amplified FHIT coding exons from RNA or cDNAs prepared from normal (fetal brain, adult kidney), immortalized (E293) and tumor (A549, KRC/Y) samples (Fig. 4B). While normal FHIT (arrow marked N) was observed in each, a smaller product was also seen which predominated in the normal fetal brain cDNA library. When this was cloned and sequenced, exon 8 containing the conserved histidine triad motif and possible zinc-binding site (29) was missing (not shown). This variation had been reported to represent an aberration in squamous cell carcinomas of the head and neck (30). Additional larger products were identified from the adult kidney cDNA library although these have not been characterized. Thus, alternative splicing definitely occurs in normal tissues.

DNA sequence analysis

Since CC19 contained the smallest deletion, we initiated large-scale DNA sequencing studies to identify new genes or structural features that might be responsible for the genomic instability. Six genome equivalents of sequence were obtained from c81E9, c120E3, c136C3 and λ33 and the data assembled. Gaps were closed using directed primers and a 1.3 kb clone gap between c136C3 and λ33 was closed by PCR amplification using λ48 (Fig. 1B) as template. This resulted in 110.4 kb of contiguous sequence (GenBank U667272) at an accuracy of ~99.6%. Predicted restriction maps from the assembled sequence also matched those determined experimentally. Sequence analysis (shown schematically in Fig. 5) included similarity searches, repetitive sequence identification and exon prediction.

Gene search

Overall, the sequence is AT-rich (61.1%) and very depleted in CG dinucleotides. No identities to known genes were seen. GRAIL2 predicted 15 exons of which six had moderate (~0.4), five had good (0.4–0.6) and four had excellent (>0.6) scores. Four putative exons were coincident with repeat sequences and two others (positions 8.0 and 71.5 kb) were coincident with Genemark predicted coding segments. Genemark (31) was utilized in order to implement a matrix for higher A–T content regions. While 61 potential coding segments were identified, no significant similarities were observed. Many predicted coding segments clearly occurred within LINE and MER elements, and others not directly within repeats nevertheless demonstrated similarity to LINEs. Four putative exons were clustered near position 107 kb, one of which showed perfect identity with a partially sequenced cDNA, EST-N70372. However, Northern analysis failed to identify a transcript from this region (data not shown) and the cDNA sequence included a portion of the LINE element at position 108 kb. Based on end sequences from the cDNA clone and insert length, the cDNA was co-linear with genomic DNA. Further, the 3’ end of the cDNA was coincident with a poly(A) tract in genomic DNA, all suggesting this represented an unprocessed transcript, or more likely resulted from false priming of contaminating DNA. BLASTN searches determined two additional regions with similarity to non-annotated cDNAs (positions 58.2 and 64.3 kb). However, neither showed similarity to known genes nor were directly superimposed on predicted exons. That these sequences were observed adjacent to, rather than superimposed on predicted exons could be due to conservative prediction algorithms used which underestimate the extent of many exons. Alternatively, the homologies could be due to infrequent repeats. We note the presence of five remaining putative exons, denoted in Figure 5, having high probability scores and which do not overlap repeats.

Repeat sequence analysis

Analysis with Pythia (32) showed that 20.2% of the sequenced region is comprised of known repeats, a level comparable with other regions chosen for comparison (Table 2). However, in this 110.4 kb, LINE and MER elements make up the bulk of repeats. Intriguingly, the repeat composition is very similar to a 152 kb sequence from the Fragile-X region (33) and differs from arbitrarily selected segments in 3p21.3 and 4p16.3 (Table 2). In particular, λ33 (from position 96 to 110.4 kb) within the CC19 deletion is nearly identical in LINE element composition to the Fragile-X region. Similarly, both Fragile-X and FRA3B have a low level of predicted coding regions. Speculatively, these similarities may influence the observed tendency of both regions to undergo breakage. However, there are obvious differences between the two sites, notably the presence of a triplet repeat and
Figure 5. Assembled sequence data totaling 110,435 bp. GeneMark and GRAIL2 predicted exons are indicated (filled boxes = direct strand; open boxes = complementary strand). Repetitive sequences homologous to LINE (L1) and Alu elements are designated by boxes filled with diagonal lines and shading, respectively. Other repetitive elements are labeled. These largely consist of MER elements (medium reiterated repeats) and LTRs (long terminal repeats). Information about particular repetitive elements can be found in Repbase (ftp://ncbi.nlm.nih.gov/repository/repbase). Unique features of the sequence are indicated on the lowest row.

methylated CpG island in Fragile-X (34) and the common versus rare nature of FRA3B.

Other features

Position 38 kb contains a cluster of terminal deletion breakpoints induced by aphidicolin treatment (GenBank U46001) further linking our sequence to FRA3B. Flanking these aphidicolin breakpoints is a long polyurine tract and two extended variable dinucleotide repeats. The telomeric breakpoint of the CC19 deletion lies within a 1.0 kb region at position 76 kb between L1 and MER/LTR elements. Within the CC19 deletion region, near position 86 kb, is an HPV16 integration site from a cervical carcinoma (35). It is of note that this integration was associated with an interstitial deletion (35,36) which we now know would not affect FHIT coding sequences. Interestingly, a very signifi-
cant similarity \((10^{-77})\) to a small polydispersed circular \(\text{spcDNAs}\) was observed at 106 kb (GenBank X96885). Figure 6 shows a FASTA alignment over 585 bp. The region is 72.1% identical and contains interspersed blocks having up to 89% identity \((\text{e.g., spcDNAs bp 121–207})\). This spcDNA, which appears non-repetitive, was isolated from a tuberous sclerosis-associated angiofibroma \((I.\) Hinkel-Schreiner, Ph.D. Thesis). Characteristics of spcDNAs include derivation from chromosomal sequences \((37)\), association with clustered repeats \([\text{such as} \beta\text{-satellites and other clustered elements although a single family member may be predominantly involved (38)}]\) and elevation in conditions associated with genomic instability such as Fanconi’s anemia \((39)\). spcDNAs are also increased by DNA-damaging agents \((37)\) and inhibitors of DNA and protein synthesis including the fragile site inducer aphidicolin \((40)\). Given the limited number of spcDNAs that have been sequenced, this similarity may be biologically important.

**DISCUSSION**

Using a ~300 kb cosmid/\(\lambda\) contig, located ~150 kb telomeric to the 3;8 translocation breakpoint, we have identified homozygous deletions in various carcinoma cell lines that overlap the most inducible common fragile site in the genome, FRA3B. From various aphidicolin-induced breakpoints and plasmid or viral integration sites \((\text{Fig. 1})\), FRA3B represents a region rather than a single site. Studies by Wilke \textit{et al.} \((36)\) and Smith \textit{et al.} \((41)\) indicate that some clustering of breakpoints may occur. However, their rearrangements were induced by aphidicolin in a single chromosome 3-containing hybrid and, unlike the interstitial deletions we observed in tumor and non-tumor samples, appear to represent terminal breaks. Where we have defined the boundaries for the carcinoma-associated deletions accurately, one or both are contained within the FRA3B region.

Our initial hypothesis was that the smallest deletion \((CC19)\) would contain elements of a tumor suppressor gene. During our studies, Ohta \textit{et al.} \((19)\) identified the \textit{FHIT} gene with reported abnormalities in RT-PCR products. However, \textit{FHIT} has similarity to a yeast di-adenosine hydrolyase which would represent an unexpected function for a tumor suppressor gene. Our results indicate that \textit{FHIT} is not the target of these deletions. First, the CC19 deletion does not involve \textit{FHIT} coding sequences \((\text{Fig. 1B and C})\) and, based on RT-PCR and cDNA sequence analysis, the coding portion of the \textit{FHIT} transcript is normal. Second, we have observed deletions in genomic DNAs from non-tumor sources \((\text{Fig. 3A})\). The somatic cell hybrid 2A3CT was formed by spontaneous terminal deletion at 3p21.3 from a non-tumor-derived chromosome 3 hybrid, UCTP2A3 \((11)\). However, 2A3CT also acquired a 3p14 interstitial deletion overlapping an aphidicolin-induced breakpoint, a PSV2neo integration into FRA3B \((20)\) and the telomeric borders of several carcinoma-associated deletions. We also identified an overlapping deletion in an unselected subclone from YAC 74B2. While neither the hybrid nor YAC 3p14 DNA segments are in a ‘normal’ background, they clearly are unselected from a tumorigenic standpoint. We note that ‘hotspots’ of recombination in human DNA can be maintained in a yeast background \((42)\), thus it is not unreasonable that unstable regions may behave similarly or more so.

Third, discontinuous deletions appear common in this region \((\text{i.e., HeLa, MDA231})\, both from our analysis and from that reported by Ohta \textit{et al.} \((19)\). Multiple deletions might be expected if there were no common target gene and if the region was unstable. Fourth, we have observed that \textit{FHIT} undergoes alternative splicing in normal tissues \((\text{Fig. 4B})\) which explains some previously reported abnormal PCR products \((19,30,43)\). Thiagalingam \textit{et al.} \((44)\) recently reported lack of \textit{FHIT} involvement in colorectal carcinomas and suggested that PCR artifacts might be responsible for some previously observed alterations \((19,43)\). It also seems likely that many deletions would have been missed in their study since only a few markers were tested. One of the possible features that suggested \textit{FHIT} could be a tumor suppressor gene was that it crossed the hereditary renal carcinoma-associated 3;8 breakpoint \((19)\). However, we found no alterations in RT-PCR products from five renal carcinoma cell lines. Moreover, Bugert \textit{et al.} \((\text{manuscript submitted})\) have observed normal \textit{FHIT} transcripts and no point mutations in a large series of renal cancers. Thus, \textit{FHIT} does not appear to be involved in renal carcinoma. With respect to other possible target genes, from our sequencing studies we identified a 100% identity to two expressed sequence tags from a liver/spleen library. However, we determined that both clones were identical, were not expressed using a commercial Northern blot \((\text{Clontech})\), were co-linear with genomic DNA containing a poly(A) tract corresponding to their Y end, encoded no significant open reading frame and overlapped a partial LINE element.

While other tumor suppressor genes may exist within FRA3B, an alternative possibility is that the deletions are due to primary genomic instability affecting a particularly susceptible region. This hypothesis is consistent with several of the observations reported here including their discontinuity and occurrence in non-tumor cell lines. By using numerous probes, we were able to identify a high frequency of homozygous deletions, especially in cervical carcinomas where p53 inactivation is very common \((18)\). In this regard, it is interesting that we observed an inverse correlation between 3p14 deletions and microsatellite instability \((\text{RER+})\). Importantly, p53 mutations, which have been shown to destabilize the genome \((45,46)\), appear infrequent in RER+ colorectal carcinomas and gastric tumors \((47,48)\) and P. Cottu, presented at Cancer and the Cell Cycle, Lausanne, Switzerland 1996]. Thus, these findings would be consistent, at least in part, with 3p14 deletions resulting from the genomic instability which accompanies p53 inactivation. We note that although CC19 is HPV negative, it expresses a mutant p53 protein \((53)\).

What have we learned from the DNA sequence analysis to date? First, is that the region is high in A–T content with frequent LINEs and MER repeats, and is conversely low in Alu sequences and confirmed genes. In contrast to the reported rare folate-sensitive sites which are associated with expanded CGG repeats \((49,50)\), FRA3B does not contain an expanded triplet repeat or methylated CpG island; nor did we identify any telomeric repeats which have been suggested as a possible cause of breaks \((51)\). However, we do note that over the 110.4 kb region there is an overall repeat sequence similarity to 152 kb from Fragile-X. \textit{In vitro}, expanded CGG repeats have been shown to inhibit DNA replication \((52)\). Whether or not specific sequences within FRA3B may have a similar effect on replication awaits experimental testing. Our discovery of a strong spcDNA homology in FRA3B may therefore not be coincidental. While we do not know if the spcDNA site is a primary cause of FRA3B instability or simply a marker for this property, the DNA sequences reported should provide the means to test this.
Figure 6. FASTA alignment of 585 bp of spcDNA clone (X96885) and FRA3B.

Table 2. Summary of sequence features

<table>
<thead>
<tr>
<th>Sequence</th>
<th>GC content</th>
<th>Total repeats</th>
<th>Alu</th>
<th>LINE</th>
<th>MER</th>
<th>Predicted exons</th>
<th>Confirmed genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRA3B</td>
<td>38.9%</td>
<td>20.2%</td>
<td>4.9%</td>
<td>9.0%</td>
<td>3.9%</td>
<td>3.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td>(110.4 kb sequence)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRA3B</td>
<td>37.0%</td>
<td>26.6%</td>
<td>3.5%</td>
<td>17.5%</td>
<td>2.1%</td>
<td>2.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>(λ33 only)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xq27.3</td>
<td>39.0%</td>
<td>26.3%</td>
<td>5.9%</td>
<td>16.3%</td>
<td>3.0%</td>
<td>3.4%</td>
<td>2.8%</td>
</tr>
<tr>
<td>(Fragile X)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3p21.3</td>
<td>58.0%</td>
<td>19.4%</td>
<td>18.1%</td>
<td>0.0%</td>
<td>0.4%</td>
<td>9.5%</td>
<td>12.9%</td>
</tr>
<tr>
<td>(HLuca_14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4p16.3</td>
<td>54.9%</td>
<td>16.1%</td>
<td>10.8%</td>
<td>3.5%</td>
<td>1.4%</td>
<td>5.6%</td>
<td>7.6%</td>
</tr>
<tr>
<td>(Huntington’s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sequences used for comparison were derived from Xq27.3 (contains the Fragile-X region, GenBank L29074) and 4p16.3 (in the region of Huntington’s disease, GenBank Z69837). The sequence from 3p21.3 was downloaded from ftp://genome.wustl.edu/pub/gsc1/sequence/st.louis/human/shotgun/3/H_LUCA14.seq.

MATERIALS AND METHODS

Nucleic acid sources and manipulations

Cell lines. Breast, colon and cervical carcinoma lines were obtained from the American Type Culture Collection. The cervical carcinoma cell line, CC19, was established as described (53). Lung tumor lines were obtained through the Colorado Lung Cancer SPORE Tissue Culture Core Laboratory. Normal cell lines included the human lymphoblastoid cell lines TL8229 and AG4103 and the E1A-immortalized human embryonic kidney cell line, E293. The somatic cell hybrids have been described previously (54).

Libraries. The gridded chromosome 3-specific cosmid library (55) was obtained from Lawrence Livermore National Laboratories. The 850A6 YAC subclone library was previously described (16). Fetal brain and adult kidney cDNA libraries were obtained from Clontech.

DNA and RNA isolations. DNA was isolated from cell lines by standard methods. Cosmid DNAs were isolated using alkaline lysis; preparations used for DNA sequence analysis were purified further by CsCl gradient centrifugation. DNA was isolated from single or pooled phase clones by the Grossberger method (56). RNA was isolated from cell lines when the cultures reached 90% of confluency using the RNA-STAT-60 kit from Tel-Test, Inc. (Friendswood, TX).
Library screening, contig assembly and hybridization analysis. The gridded cosmid library was spotted onto filters at high density (1536 clones per filter) and hybridized using standard techniques with a 370 kb MluI fragment derived from YAC 74B2. The 850A6 YAC phage library (16) was screened using the same probe. Resulting clones were assembled into a contiguous segment by hybridization analysis using end probes and total inserts followed by analysis with the software tool SEGMAP. Complete and partial digestion analyses (57) were used to restriction map the central 170 kb of the contig prior to DNA sequence analysis.

Replication error (RER) analysis

Biotinylated primers were obtained from Research Genetics, Huntsville, AL. PCR amplifications were performed with 40 ng of template DNA utilizing hot start and touch down procedures. After separation on denaturing polyacrylamide gels, PCR products were detected using the New England BioLabs Phototope™ Detection Kit. Alleles were scored by visual inspection of band patterns.

RT-PCR analysis of FHIT gene expression

RT-PCR was performed using primers FHIT-5′ (5′-CTCGAA- TTCTTAGACCCCTATAAAGC-3′) and FHIT-3′ (5′-CTG- ATTCAGTTCTTCTGT-3′) derived from non-coding exons 4 and 10, respectively. First strand synthesis was accomplished with 1–3 mg of total RNA and the Superscript II kit (Life Technologies). Subsequent amplification utilized one-fifth of the reverse transcriptase reaction together with the FHIT primers. Standard PCR conditions of 94°C denaturation (1 min), 55°C annealing (1 min) and 72°C elongation for 35 cycles were employed. PCR products were subcloned into the EcoRI site of pBluescript II SK+ using an introduced EcoRI site present in the FHIT-5′ primer and a natural site located 21 bp downstream of the FHIT stop codon.

DNA sequence determination

Cosmid and phage clones were sequenced using a random shotgun subcloning and end-sequencing strategy. Clone DNA was sonicated and size selected by LMP-agarose gel electrophoresis. Recovered fragments of 1–2 kb were end-repaired with Klenow fragment of Escherichia coli DNA polymerase I and T4 DNA polymerase, ligated into the phosphatased EcoRV site of pBluescript II and transformed into E.coli DH10B. AmpR/β-Gal− subclones were grown in 3 ml of TB for isolation of sequencing templates. To eliminate vector sequences from the subclone library, phage inserts were amplified by long-range PCR. Inserts were gel purified, 32P labeled and hybridized to subclone libraries. Positive subclones were sequenced using an ABI 373 or 377 and the ABI Prism dye terminator cycle sequencing kit. Chromatograms were analyzed for database similarities by PHRED and PHRAP (from Dr Phil Green). Gaps were closed by primer walking. Based on independently obtained overlapping contigs, as well as analysis of cosmid vector sequences in some subclones, sequencing accuracy was at least 99.6%.

Sequence analysis

Assembled sequences were analyzed for database similarities by BLASTN and BLASTX, searching the nr and dbest databases with default parameters. Strongly similar sequences (P < e−30) were retrieved from GenBank for further analysis. Homologies to repeat sequences were found using Pythia (32). Potential exons were identified using GRAIL2 (58) and GeneMark (31) programs. GeneMark was run using 5th order matrices trained on sequences with GC content similar to the sequences checked. Sequence alignments were prepared using the FASTA program from the Wisconsin Package (GGC).

ACKNOWLEDGEMENTS

This work was supported by DAMD17-94-J-4391 (HAD, R.M.G.), CA51395 (HPK, MRM) and the DNA sequencing studies were supported by HG000358 (H.A.D., R.M.G.). We gratefully acknowledge support from the Colorado Cancer League (2531718) to F.B. Additional Core support was provided through the University of Colorado Cancer Center (CA 46934) and by Dr Wilbur Franklin through the Lung Cancer Spore.

REFERENCES


**NOTE ADDED IN PROOF**

Sequencing of the CC19 deletion region has been completed with the finalization of lambda 94, a phage proximal to lambda 33. This 18 810 bp sequence maintains the same characteristics as the balance of FRA3B. It is 63% AT and 27% is comprised of known repeats with the L1 type elements predominating.