Human Molecular Genetics, 2000, Vol. 9, No. 11 **1651–1663**

Common chromosomal fragile site *FRA16D* sequence: identification of the *FOR* gene spanning *FRA16D* and homozygous deletions and translocation breakpoints in cancer cells

Karin Ried[1], Merran Finnis[1], Lynne Hobson[1], Marie Mangelsdorf[1], Sonia Dayan[1], Julie K. Nancarrow[1], Erica Woollatt[1], Gabriel Kremmidiotis[1], Alison Gardner[1], Deon Venter[2], Elizabeth Baker[1,3] and Robert I. Richards[1,4,+]

[1]Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide, South Australia 5006, Australia, [2]Peter MacCallum Cancer Institute, East Melbourne, Victoria 3002, Australia and [3]Department of Pediatrics and [4]Department of Genetics, The University of Adelaide, South Australia 5000, Australia

Received 13 March 2000; Revised and Accepted 27 April 2000

DDBJ/EMBL/GenBank accession nos AF217490–AF217492 and AF227526–AF227529

Fluorescence *in situ* hybridization of a tile path of DNA subclones has previously enabled the cytogenetic definition of the minimal DNA sequence which spans the *FRA16D* common chromosomal fragile site, located at 16q23.2. Homozygous deletion of the *FRA16D* locus has been reported in adenocarcinomas of stomach, colon, lung and ovary. We have sequenced the 270 kb containing the *FRA16D* fragile site and the minimal homozygously deleted region in tumour cells. This sequence enabled localization of some of the tumour cell breakpoints to regions which contain AT-rich secondary structures similar to those associated with the *FRA10B* and *FRA16B* rare fragile sites. The *FRA16D* DNA sequence also led to the identification of an alternatively spliced gene, named *FOR* (fragile site *FRA16D* oxidoreductase), exons of which span both the fragile site and the minimal region of homozygous deletion. In addition, the complete DNA sequence of the *FRA16D*-containing *FOR* intron reveals no evidence of additional authentic transcripts. Alternatively spliced *FOR* transcripts (*FOR I*, *FOR II* and *FOR III*) encode proteins which share N-terminal WW domains and differ at their C-terminus, with *FOR III* having a truncated oxidoreductase domain. *FRA16D*-associated deletions selectively affect the *FOR* gene transcripts. Three out of five previously mapped translocation breakpoints in multiple myeloma are also located within the *FOR* gene. *FOR* is therefore the principle genetic target for DNA instability at 16q23.2 and perturbation of *FOR* function is likely to contribute to the biological consequences of DNA instability at *FRA16D* in cancer cells.

INTRODUCTION

Chromosomal instability is a feature of certain types of cancer. It is not yet clear whether such instability represents the outcome of a selection process involving the gain, loss or alteration of specific genetic material or whether certain regions of the genome are predisposed to instability. Fragile sites are chromosomal structures which have been proposed to have a determining role in cancer-associated chromosomal instability (1). There are in excess of 100 fragile sites in the human genome (2), of which the fragile site *FRA11B* is located within the *CBL2* proto-oncogene (3,4) and the *FRA3B*, *FRA7G* and *FRA16D* sites have been located within or adjacent to regions of instability in cancer cells (5–11).

There are two distinct forms of chromosomal anomaly referred to as fragile sites (12). The 'rare' form is polymorphic in the population and is accounted for by the expansion of repeat DNA sequences beyond a copy number limit. The 'common' form is present at many loci in all individuals. Despite determination of the complete sequence of the common fragile site *FRA3B* (13–15) and partial sequence analysis of the common fragile sites *FRA7G* (8,9) and *FRA7H* (16), the molecular basis for common fragile sites is not yet understood. Fragile sites are also distinguished by the culture conditions required for their induction. Common fragile sites are (mainly) induced by aphidicolin, whereas the rare fragile sites are induced by either high or low concentrations of folate, AT-rich binding chemicals such as distamycin A or by bromodeoxyuridine. The role of chromosomal fragile sites in human genetic disease was thought to be restricted to fragile X syndrome caused by the *FRAXA* fragile site; however, a mild form of mental retardation has been associated with *FRAXE*

[+]To whom correspondence should be addressed at: Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide, South Australia 5006, Australia. Tel: +61 8 8204 7111; Fax: +61 8 8204 7342; Email: rrichard@medicine.adelaide.edu.au

(17). In addition, the *FRA11B* fragile site appears to predispose to 11q breakage leading to some cases of Jacobsen syndrome (3,4), raising the possibility of other diseases of diverse phenotypes being associated with other fragile sites.

Recent detailed molecular analysis of fragile site loci has demonstrated that the common fragile site *FRA3B* is located within a region subject to localized deletion and that this deletion is frequently observed in certain forms of cancer (5,6). *FRA3B* lies proximal to the major region of loss of heterozygosity (LOH) on chromosome 3p previously shown to be responsible for deletion of the *VHL* tumour suppressor (18). The cancer-associated *FRA3B* deletions can result in inactivation of a gene (*FHIT* for fragile histidine triad) which spans the fragile site. The *FHIT* gene product has been shown to have a role in tumour growth (19); however, there has been a great deal of controversy over the functional significance of *FHIT* loss to the tumorigenic process (20,21).

The sequence of the region containing *FRA3B* has been determined; however, neither the molecular basis for expression of the fragile site nor the cancer-associated instability is clearly resolved (13–15). The *FRA3B* region sequence has an abundance of the L1 type of long interspersed nuclear element (LINE) DNA repeats. It may be that these elements constitute the molecular basis for fragile site expression, as the expansion of repeat elements is the only known cause for 'rare' fragile sites. The L1 elements are frequently found at or near the boundaries of *FRA3B*/*FHIT* deletions in cancer cell DNA (14,15) and have therefore been proposed to facilitate the deletion process.

Analysis of the common fragile site *FRA7G* has also demonstrated that this fragile site is located within a region of frequent deletion in breast and prostate cancer (8,9). This region contains two members of the caveolin gene family the deletion of which may play an important role in tumorigenesis (7,22). Only a partial sequence for this fragile site is available and so it is not clear whether other genes may be affected by *FRA7G*-associated chromosomal instability. Molecular studies have also been undertaken at the *FRA7H* fragile site and the region around an SV40 integration site at this locus has been sequenced (16). No genes were identified in this 161 kb sequence; however, this sequence does not appear to span the entire fragile site and further analysis may reveal associated genes.

We have previously localized the minimum region required for cytogenetic expression of *FRA16D* by establishing a contig of subclones across the region and determining their position with respect to *FRA16D* by fluorescence *in situ* hybridization (10). DNA markers from within this region were used to detect instability in tumour cell lines that had previously been shown to exhibit instability at the *FRA3B* locus. One of these cell lines, the gastric adenocarcinoma AGS, was shown to have a homozygously deleted region that spanned the *FRA16D* fragile site. Using representational difference and PCR deletion analysis Paige *et al.* (11) have identified the *FRA16D* region as a site of homozygous deletion in three additional tumours (from ovary, lung and colon).

We have determined the complete DNA sequence of a minimal tile path of subclones spanning *FRA16D* to define the molecular limits of this common fragile site region. Within this sequence genetic and/or physical markers have been identified that delineate the boundaries of the observed DNA instability in cancer cells. The *FRA16D* region sequence also allowed identification of potential genes as expressed sequence tags (ESTs) as likely candidate targets for instability and found no evidence for a transcript within the 270 kb containing *FRA16D*. Instead, a transcript (*FOR* for fragile site *FRA16D* oxidoreductase) was identified which spans the *FRA16D* fragile site, the common minimal region of homozygous deletion found in adenocarcinomas and three out of five translocation breakpoints in multiple myeloma. Transcripts from the alternatively spliced *FOR* gene encode proteins with common N-terminal WW domains and variable homology to the oxidoreductase family of proteins. *FOR* is therefore the most likely gene to convey any biological consequences of DNA instability at the *FRA16D* locus.

RESULTS

DNA sequence spanning *FRA16D*

The DNA sequence spanning *FRA16D* was determined by a combination of approaches. Firstly, a tile path of λ subclones of YAC My801B6 and BAC 325M3 was restriction mapped with restriction endonucleases *Eco*RI, *Hin*dIII, *Bam*HI and *Sac*I in order to provide a reference framework with which to anchor the DNA sequence (10). Secondly, either whole BAC DNA preparations of BAC325M3 or BAC353B15 or specific restriction fragments from the λ subclone tile path were used as feedstock DNA for construction of random insert plasmid libraries. Sequences from the region between BAC325M3 and BAC353B15 [λ subclone tile path λ32–λ191 (10)] were subjected to long range PCR and restriction digest analysis in order to verify the integrity of this sequence. Sequenced subclones were also ordered by hybridization with individual λ subclones from the minimal tile path. The DNA sequences were therefore assembled in a directed rather than random manner. This approach greatly assisted in the assembly of those regions that were rich in DNA repeats. The 270 kb contiguous sequence spanning *FRA16D*, with an average 4-fold sequence coverage, has been deposited in GenBank (accession no. AF217490) (Fig. 1).

Relationship between deletion and translocation breakpoints and *FRA16D*

PCR analysis of sequence tags across the *FRA16D* region was used to refine the location of deletion breakpoints in the AGS and HCT116 tumour cell lines (Fig. 1 and Table 1). Both cell lines showed at least two distinct regions of homozygous deletion, indicating a minimum of three deletion events on the two chromosomes 16 in each cell line. Four regions of the *FRA16D*-spanning sequence were particularly difficult to determine because of their composition (as evident by DNA polymerase pausing in sequencing). Each of these sequences coincided with breakpoint regions in HCT116 or AGS tumour cell lines (Fig. 1, and referred to as 'pause sites' in Table 1). The unstable regions consisted of: (i) a poly(A) homopolymer region at 144–145 kb of DNA sequence AF217490; (ii) an imperfect CT repeat of 320 bp at positions 177–178 kb; (iii) an 8 kb region at positions 191–199 kb encompassing a poly(A) homopolymer region followed by an AT repeat, a poly(T) homopolymer repeat and two inverted (hairpin-forming)

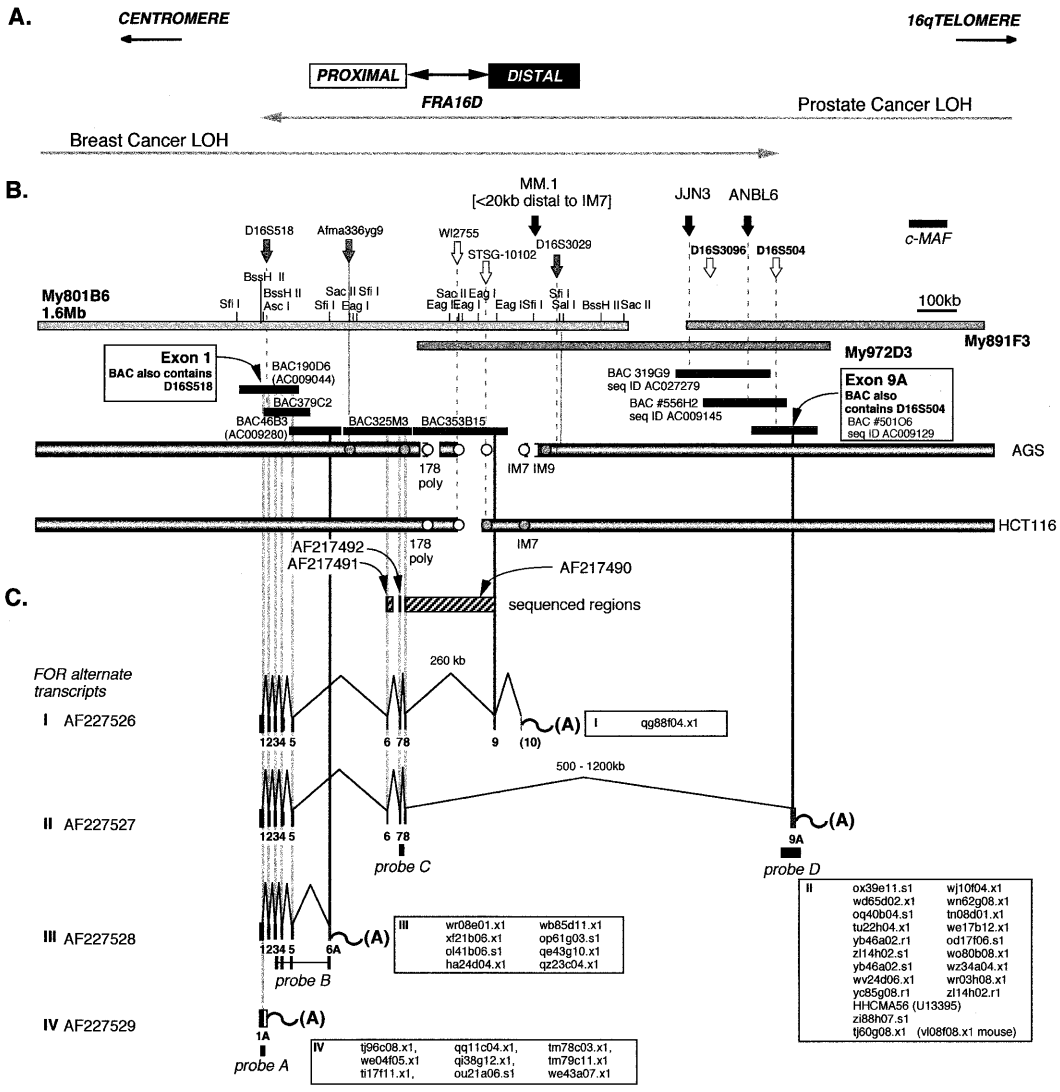


Figure 1. Map of the *FOR* transcripts and deletion breakpoints in 16q23.2 with respect to *FRA16D*. (**A**) Extent of loss of heterozygosity regions in breast (28) and prostate cancer (29) in relation to the cytogenetic position of the *FRA16D* fragile site as determined by fluorescence *in situ* hybridization of a tile path of subclones (10). (**B**) Map of YACs which span the *FRA16D* region showing approximate location of multiple myeloma breakpoints (MM.1, JJN3 and ANBL6) determined by Chesi *et al.* (23). Location of homozygously deleted regions in AGS and HCT116 tumour cell lines as determined by STS content. The locations of various partial BAC sequences (as evidenced by STS content) are indicated. (**C**) Location of the DNA sequences determined in this study including those spanning *FRA16D* (striped boxes with GenBank accession nos) and the respective exons of the alternatively spliced *FOR* gene transcripts (numbered black boxes). Clusters of EST sequences representative of each of the alternative mRNA 3′ ends are given.

repeats; (iv) a TG repeat followed by a homopolymer region [poly(T)] at 212–213 kb. This fourth sequence is located within a common breakpoint region for the AGS and HCT116 cell lines at 211.7–219.9 kb of AF217490. PCR across each of the breakpoint regions in the AGS and HCT116 cell lines using primers from positive flanking STSs failed to produce products, suggesting that additional cryptic instability (e.g. inversions or amplifications) may also be present.

The locations of three previously identified multiple myeloma breakpoints (23) were determined by either BLAST scanning of partial database sequences or by PCR of STSs on the tile path of λ subclones spanning *FRA16D*. The location of MM1 was verified by PCR of λ subclone λ131 (which also contained the STS IM7). The locations of JJN3 and ANBL6 are based on their presence or absence in the BAC sequences GenBank accession nos AC027279 and AC009145 and the relative position of these BAC sequences with respect to the marker D16S504. The location of the translocation breakpoints (Fig. 1) is in agreement with that published by Chesi *et al.* (23).

Repeat elements and DNA flexibility in the *FRA16D*-spanning sequence

The DNA sequence spanning *FRA16D* was analysed for type and quantity of DNA repeats since DNA repeats are known to be the molecular basis for the rare fragile sites and have been proposed to have a role in common fragile site instability (15). Comparison of the *FRA16D*-spanning sequence with that of other common fragile sites, *FRA3B* and the partially sequenced *FRA7H* and *FRA6E* regions (Fig. 2A), gave a surprising degree

Table 1

BREAKPOINTS	STS name	ACCESSION NUMBER	AF217490 sequence #	Pause Sites (kb)	FlexStab Peaks (kb)	BB (duplex)	AGS (duplex)	HCTII6 (duplex)	PRIMER	PRIMER SEQUENCE
proximal	coxido3	prox/dis	4220-4582			+(+)	+(+)	+(+)	8052 / 8053	ACAGATATTAACGACTCCTTGGGA / ATGGGCTGGACAATGTTATTACTG
	AA368108		77046-77276		*38.5-39.5*	+(+)	+(+)	+(+)	7125 / 7126	TAATCCTCAGCCTCTAGAATGCCT / GTATGATGATTTTCAGGGAGAAAC
	16D-15/16D-16		89367-89644			+(+)	+(+)	+(+)	7538 / 7539	CTTCGCATATGTATTTATGGC / CAGCTCCATAACGTGACATGT
AGS 654bp	16D-15/16D-36		89367-90681			+(+)	+(+)	+(+)	7538 / 8645	CTTCGCATATGTATTTATGGC / TGTCTTGAAGCCAGACACCT
	16D-1/16D-60		91184-91315		*91 - 91.5*	+(+)	-(+)	+(+)	6069 / 8739	TGAGGGCCACAGAGCTTGCT / ACCTTCCATGGAAAGGTCTG
	16D-17/16D-18		91929-92297			+(+)	-(+)	+(+)	7540 / 7541	AACATACTGTGACCATGCAGT / TCAGAAGTCTTCGAGACATG
	mouse cDNA	CC77236	94143-94604			+(+)	-(+)	+(+)	6613 / 6614	GACTCATGAAACCTAGAAGG / CCCGACTTTTTACCAAAGG
	z20768	Z20768	109423-109719			+(+)	-(+)	+(+)	7059 / 7060	AGGTGGAACTGATTTTCTCACTG / TACATGAACTGTGGCAGCTGG
	ye25b03	T91653	111666-111817			+(+)	-(+)	+(+)	7021 / 7022	AAGAATTTCCCGATTCTCTTTCCAC / ATTGCATCTCACCCCGACATTAGC
	zt58b03	AA3988024	113855-114010			+(+)	-(+)	+(+)	7020 / 7019	TCAATGGGTTAGGCACAGACC / TGTCCTCAACTGATTCTTACAAAC
	16D-21		114405-114635			+(+)	-(+)	+(+)	8046 / 8047	TGATGTACCCTGTGTTCTGC / TCTGCACATCGGCTTCTGTT
	16D-20		115815-116145			+(+)	-(+)	+(+)	8044 / 8045	GTGGAGGGAACATCCATCCT / CTAGGCTTTTCCTACTCTAG
Hct 1835bp	16D-19		121569-121959			+(+)	-(+)	+(+)	8042 / 8043	TCCCATTCAGTCATGATATGC / CCCAATTCTGATGGGGTTCC
	16D-61		123569-123794			+(+)	-(+)	-(+)	8740 / 8741	GAAATTGGGACTCCCATCCT / GCCTGTGTGATCTGAGAAAC
	16D-24		124570-124714			+(+)	-(+)	-(+)	8123 / 8124	GTAATCCAGCTCCATGCTGG / GTTCTGAATGCGGAAATCTG
	16D-25		126129-126324			+(+)	-(+)	-(+)	8125 / 8126	ATGCTCTACTGAGGAGTGGA / TGCTACTGCCTTAAGGCAGG
	zf68e10	AA063112	126806-126886			+(+)	-(+)	-(+)	7023 / 7024	AATCCCCAGGACAGTCTCATTTGG / GGAACCAAATCTGCGTGCCCTGAA
	178POLY		135538-135920			+(+)	-(+)	-(+)	7253 / 7254	TACGCTGAATATCACTGGAG / TGCCAACCAGTACACAAATA
	16D-23		138601-138901			+(+)	-(+)	-(+)	8080 / 8081	GGAGAACATGGTATCAGAAC / CAATCTCCTGACTGCACTGT
	16D-26		140081-140231			+(+)	-(+)	-(+)	8182 / 8183	AGTGACAGAGTGTGTCAAGC / GTAGTGAACAGAGGCTCTAG
	16D-27		141181-141301			+(+)	-(+)	?	8184 / 8185	CCACTTCTACCTTAGCTATG / ACAGAGTGGCCTGTAATGAG
	16D-22		142211-142451			+(+)	-(+)	-(+)	8078 / 8079	GGTTTCCATCCTGAAGAGCT / AGACCCGTAGTCTAGCATTC
	16D-28A		143049-143141			+(+)	-(+)	-(+)	8335 / 8336	GCACTGGGCTGGAAATGAAT / TATTCACACTGCCAGCTCCT
	16D-28		143191-143301			+(+)	-(+)	-(+)	8200 / 8201	ACAGAAAGTGAAACGGTGCC / GCAATACTTCCCTTCAAACC
	ab10e01	AA485822	144337-144529	144 - 145		+(+)	-(+)	?	7055 / 7056	GAAACTGTTGAGAACTAGGAAAAG / AGCCTGTTTTAATCCTGAATTATG
Hct 1549bp	16D-62		145841-145971			+(+)	-(+)	-(+)	8742 / 8743	TTGCCCAGTGTACTCTCTGT / GACAGGAAGCAAGTGAACCT
	qz19h11	AI361849	147540-147687		*147 -147.5*	+(+)	-(+)	+(+)	7026 / 7025	CAAAGTTGTTAAGACTATCTAAGA / GAGTTTATAGACTAATGGGTAGG
	nk32e05	AA555206	151920-152728		*150 -151*	+(+)	-(+)	+(+)	7141 / 7142	GCCAGGAAACTCCTGTACTTCAAG / GCAATCAATCTGTAATTGGTGCTC
	nk32e05ex1		152228-152373			+(+)	-(+)	+(+)	7047 / 7046	TTCACTCCGAGCCATCGACTTCTG / TGTCAAAGAAAACAGCAAACATGG
	AA349717	AA349717	161704-162066			+(+)	-(+)	+(+)	7016 / 7015	CATCCTGTATTAGTCCTCCTTG / GATGCAATCCAGGTATTTTATCAC
	zh97a01	AA007376	173134-173589		*170 - 171*	+(+)	-(+)	+(+)	7013 / 7014	ATACCACTTCTCCTAGGGATC / CAGCAATGCTTTAATAGTGACAG
	yd85h12	T80206	173728-174001			+(+)	-(+)	+(+)	7012 / 7011	ACAGCAGGGATGGAAGAGGCA / CATGCACCAGTCAAGGAAGGC
	16D-46		175629-175819			+	-(+)	+	8654 / 8655	TCCACCTTGCCAGTATCTAG / GATATTGGCACTTCTCCAGG
AGS 3962bp	16D-70		177861-178091	177 - 178		+(+)	- ?(+)	+(+)	8791 / 8792	TCGCTCTTCTACTCTCAGTG / GAGGCGGAGTTGTAATTCTG
	16D-47		179801-180011			+	+(+)	+ neg +	8656 / 8657	GTTTGTGAGTCCAGAGACGA / TTCTTTGGAGACTTCTCCGG
AGS 3030bp	16D-57		181012-181151			+(+)	v.faint + (+)	+(+)	8723 / 8724	GTGAGGTAAGTTCGTCAGAC / AGCAGCTGGGAACTGTAAAG
	16D-67		182422-182562			+(+)	-(+)	+(+)	8785 / 8786	GTAACAGTGACGCAAACGAC / CCAAGTGAGGACACCTTGAA
	16D-48		184201-184356			+	-(+)	+(+)	8658 / 8659	GATCTCACTCCAGTATCAGC / AGCTCATTGGCTTCTGGCTT
	16D-49		187440-187701			+(+)	-(+)	+(+)	8660 / 8661	ACAGCTCCCTCCCTAATCTT / GTCTGTAGCAGAGCCTGAAA
AGS 1720bp	16D-68		188559-188610			+(+)	-(+)	+(+)	8787 / 8788	TGTGTCTTTGGCCTGGAAAG / GTGGGAATGCATACACATGG
	16D-54		189441-189562			+(+)	+(+)	+(+) mult.	8715 / 8716	ACAAAACCCAGCGTGGTGAT / GACAACAGGCCTATCACCTT
	16d190		190151-190340	191 - 199	*191 - 192*	+(+)	+(+)	+(+)	8166 / 8167	CAGGAAATCTTTGTGCATGGCTGT / ATAAAGATGGGCAATGTGCACATG
	16D-58		192401-192559			+(+)	+(+)	+(+)	8725 / 8726	CTGAGTTCCTGTAGCATGTG / AGCTACTAGGTACACCATCC
	16D-50		193311-193478			+(+)	+(+)	+(+) neg +	8662 / 8663	CAGGTCACTACGTATTGCGT / CTGCCACAGGCTTTACCTTA
	16D-59		196321-196504			+(+)	v.faint + (+)	+(+)	8727 / 8728	ACACACTTCCACCTACAGAG / TCTACGTGCAAGATGCTGTG
AGS 209bp	16D-51		197361-197531			+(+)	+(+)	+(+) neg +	8664 / 8665	TGTTCACCTGGAGTCTGTAC / ATGTGTGGCGGCTACAAATC
	16D-55		197760-197847			+(+)	-(+)	+(+)	8717 / 8718	CTCACTAAAGGGGAGGAATC / TAGGTTGTGGATTGTGTGGG
	16D-56		199482-199629			+(+)	-(+)	+(+)	8719 / 8720	GCTTTATGGGCCTTCTCTTC / GTAAAGTGGGAGAACTCACC
	16D-52		200171-200339			+(+)	-(+)	+(+)	8666 / 8667	CAAATCCTGGGAAGACTCTG / TCTTGCCTCTCTTGATCCAG
AGS 5690bp	16d202		201624-201851			+(+)	-(+)	+(+)	8164 / 8165	ACCACTGCTCGCATTCCCATGTTG / GAGAAATCTTACTGACAACACTGCTA
	16D-53		206401-206579			?	?	?	8668 / 8669	ATGGAGGTCAACTACGGGAA / GAATAACAGGGTTCAACCCC
	16D-69		207561-207701			+(+)	+(+)	+(+)	8789 / 8790	TCGTTTGGGATTCCTCATCC / GTTGAGTCGCAAGAGGTACA
	16D-29		208901-209111			+(+)	+(+)	+(+)	8266 / 8267	ACCCAAGTTTGGCCTCTCTG / AAAGGACTGAGAATGGCGGA
	16D-31		210401-210581			+(+)	?	?	8337 / 8338	CAGGGCCTATTGTGTTGTGT / ACAGACAGGTGAGCAAGTTC
	16D-32		211701-211811	212 -213		+(+)	+(+)	+(+)	8339 / 8340	CAGTCTCCTGTCCCATACTT / CTTTGCGCACATGCAGAAGT
	16D-63		213801-213899			+(+)	+	+(+)	8744 / 8745	CAACAAAGCCTTTCACGCTG / TGCATTTGATTGTGGCTCCG
Hct116 422bp	16D-35		214171-214271			?	?	?	8426 / 8427	TCATTCATTCTGCTGGCCCT / GGCCGTTTAACTACCTCTTG
	16D-30		214341-214435			?	+(+)	-(+)	8268 / 8269	CACAGTGATAGACCCTGAGT / CTGCTTTCTGCCAAGCTTGA
AGS 5179bp	16D-30/16D-44		214341-214590			+(+)		-(+)	8268 / 8652	CACAGTGATAGACCCTGAGT / AAGGACTTTCGGGCTCTTGA
	ETA1	AA010088	215802-215892			+(+)	?	-(+)	7247 / 7248	TATTTGCAGCATGTCTTGCC / TGCTTGGCACATGGTAAAAA
	16D-34		216401-216511			?	?	?	8428 / 8429	ACTCTTTGCTTGCTCAGCAG / CAAGATATCAAGGGAGCTGC
	WI2755	M78830	219798-219880			+(+)	-(+)	?faint+	1552 / 1553	TCCACCATCCTGGGTCTAGCT / CCTGATCCATGTGTTTTGAGATCTC
	16D-64		222530-222751			+(+)	-(+)	-(+)	8746 / 8747	GATTGATGGGCATACAGAGG / CTACTAGCCTCTAGCTGCTA
	16D-65		227721-227921			+(+)	-(+)	-(+)	8748 / 8749	CACCACCACATCCAAATGTC / CAGTTTCACTCTCAGACTGC
	AA297515	AA297515	234928-235125			+(+)	-(+)	?faint+	7061 / 7062	GTTTGCAGTAAGCTAGGATTAAG / AGTCCTGTATATGCAGATGAAATG
	51A6		235378-235593			+(+)	-(+)	-(+)	7256 / 7255	AAACAGAAGGAGCACCAG / CCAGGAACACCAAAACAC
	R81094	R81094	241224-241574			+(+)	-(+)	-(+)	7018 / 7017	TCGTACTTCGAGGGCAAGGAC / CAACTTCCAGGATGGTCTACC
	RD69		260790-260920			+(+)	-(+)	-(+)	7249 / 7250	GTGGAACCTTTAGGGCTTAT / TTAGCAAAAGAAAGGGAGTG
Hct116 1513	16D-66		262231-262371			+(+)	-(+)	-(+)	8750 / 8751	GTTCAAGCCATACGTCACCA / ATTATGGGACTCAGCCAGAG
	801A		263904-264113			+(+)	-(+)	+(+)	8038 / 8039	CAAATATTGCAGGTGGCCGAAGCC / CTTCTAAAGCAATAACGACTGGTATTC
spans exon9 FORI	oxido4		264113-264357			+(+)	-(+)	+(+)	8040 / 8041	GAATACCAGTCGTTATTGCTTTAGAAG / GTCACCTTTGTCATAAGATAAAGC
	294A		264357-265251			+(+)	-(+)	+(+)	8034 / 8035	GCTTTATCTTATGACAAAGGTGAC / ACTATTAGATTGAGAGAAATGGACTAG
	294B		265251-265561			+(+)	-(+)	+(+)	8036 / 8037	CTAGTCCATTTCTCTCAATCTAATAGT / GTCTATAAATCTACAGGTGTTACCC
	sts123		not in AF217490			+(+)	-(+)	+(+)	7968 / 7969	ACCAAGAGTCTATGAGCTCGAAGG / CTTTACGCTGAGAGCTGCGCCATC
	sts234		not in AF217490			+(+)	-(+)	+(+)	7970 / 7971	GCTGATGTTGTAGCTGTAATCACC / AGATGCTTATATATCTAGCAGGTC
	STSG10102	Z23147	not in AF217490			+(+)	-(+)	+(+)	2984 / 2985	CAGACCACCATATTTGTCCTACTAGG / TGCAGAATCATAAAACTCGGG
AGS ~1500bp	IM7	H54363	not in AF217490			+(+)	-(+)	+(+)	7259 / 7260	ACAGCCATGTGAGTGAACCA / TGGGCAGTTATCCACTGTTG
	410S1A	I410	not in AF217490			+(+)	+(+)	+(+)	8311 / 8312	CACACCTCTGACCACAAATC / TCCAGGAAATACAGCTGGAG
	410S1B	I410	not in AF217490			+(+)	+(+)	+(+)	8341 / 8342	CCTGAATAGCTGGGACTACA / TGCTTGAGGCTATGGCTTCA
	410S3	I410	not in AF217490			+(+)	+(+)	+(+)	8160 / 8161	AGCTCCAAAGACAAGGGCAA / CCCAATCACACAGTCAGGAGA
	330S1	I330	not in AF217490			+(+)	+(+)	+(+) neg faint	8162 / 8163	GCTTGTCTCAAGCTGCTGAT / CTGTAATGCTCTCCAGGCAT
	330B1	I330	not in AF217490			+(+)	+(+)	+(+)	8158 / 8159	GTGTTCAAGACACTGTGCTA / AAGCCATGCTGCTATGGTTT
	330B4		not in AF217490			+(+)	+(+)	+(+)	7950 / 7951	GATCACAGCAAAGGTGAGCT / CTCACACCAATGGGCATGGA
distal	IM2	AA229508	not in AF217490			+(+)	+(+)	+(+)	7438 / 7437	GGCATGATGGAGGAATTTTA / TGTTTGTCTGGTGATGGTTA

of variation in the type and quantity of repeat DNA sequences. LINE-1 elements, which have been proposed to play a major role in the DNA instability seen at the *FRA3B* locus, were significantly under-represented at the *FRA16D* locus when compared with not only the other fragile site loci but also the human genome in general (Fig. 2A).

Common fragile site DNA sequences have been identified as containing regions of flexibility as determined by the computer program FlexStab (16). This program was obtained from the website of Dr B. Kerem (http://leonardo.ls.huji.ac.il/departments/genesite/faculty/bkerem.htm) and was used to analyse the *FRA16D* DNA sequence. Six regions of flexibility

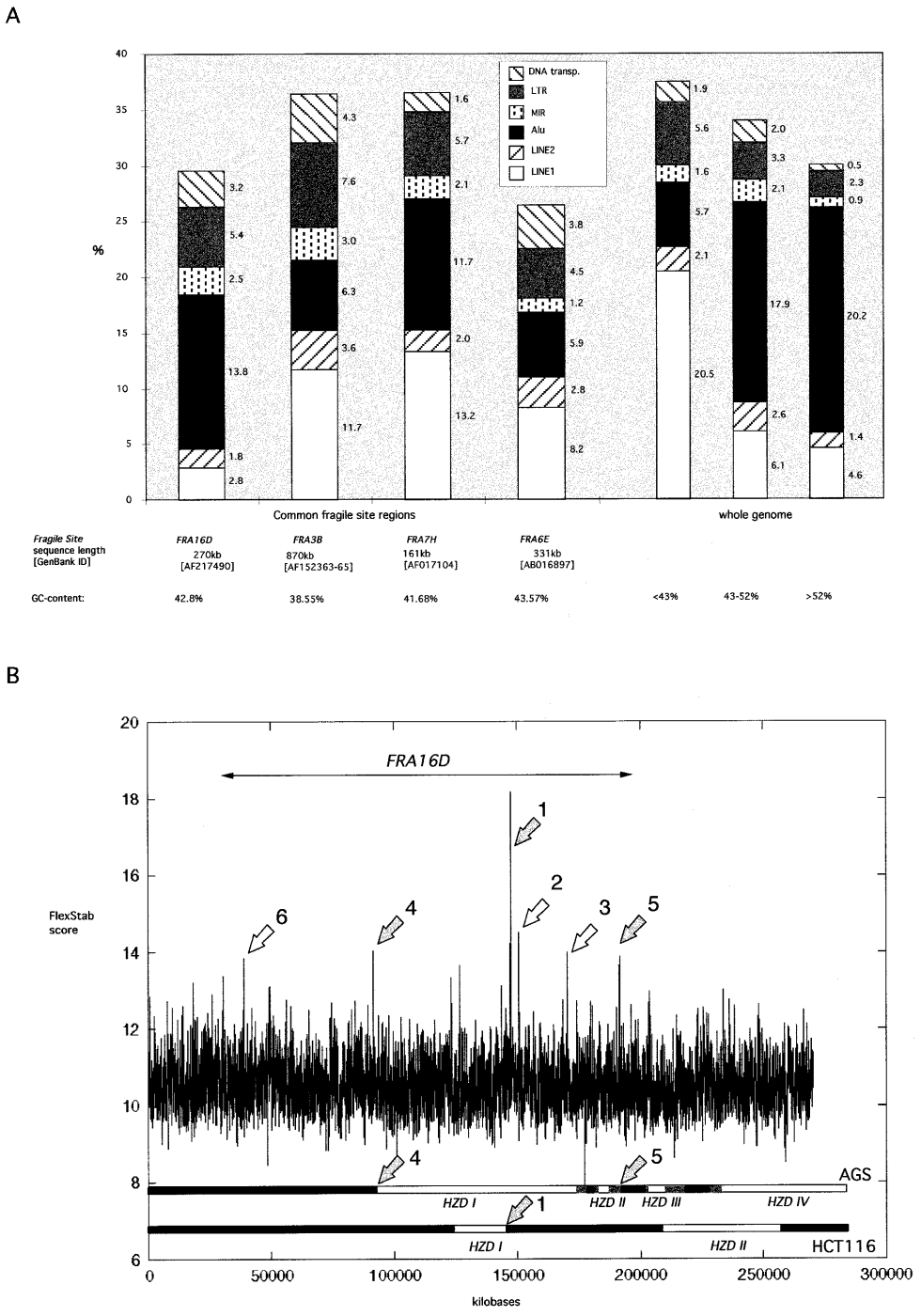


Figure 2. Physical properties of the *FRA16D*-spanning DNA sequence. (**A**) Comparison of the DNA repeat composition of the *FRA16D*-spanning sequence with those of other common fragile sites and regions of the whole genome. Values for various %GC content fractions of the human genome are from Smit (30). (**B**) FlexStab analysis of DNA 'flexibility' for the *FRA16D* DNA sequence. The location of *FRA16D* is indicated as are the six highest peaks of flexibility score. The position of these peaks with respect to breakpoint locations in the AGS and HCT116 tumour cell lines is also shown. *HZD I–IV* refer to homozygously deleted regions. Black bars represent DNA sequences that are present; white bars represent DNA sequences that are absent; grey bars represent DNA sequences that are of unknown deletion status.

were identified (Fig. 2B), two of which (peaks 1 and 4) coincide with breakpoints of homozygously deleted regions in the HCT116 and AGS cell lines, respectively. A third peak (peak 5) maps in the vicinity of another breakpoint in the AGS cell line (Fig. 2B).

The alternatively spliced *FOR* gene spans fragile site *FRA16D*

Scanning of the 270 kb sequence spanning *FRA16D* by BLAST homology searches revealed a paucity of EST homologies, almost all of which could be accounted for as illegiti-

mate (Table 2). The exceptions were consecutive exons corresponding to sequences from the EST qg88f04.x1 (Fig. 1). These exons therefore locate *FRA16D* within a 260 kb intron. BLAST searches with the qg88f04.x1 EST sequence revealed considerable overlap with clusters of ESTs, the longest available sequence of which was HHCMA56 (U13395). Part of the HHCMA56 sequence has been mapped previously ~700 kb distal to *FRA16D* (11). ESTs qg88f04 and HHCMA56 clearly have distinct 3′ end sequences and were therefore referred to as transcripts I and II. Another cluster of ESTs (transcript III) was found to share 5′ but not 3′ end sequences with transcripts I and II. A fourth cluster of ESTs (transcript IV) was found to share sequence homology; however, this overlap is between the 5′ most sequences of transcripts I–III and the 3′ end of the EST cluster, suggesting that it represents an overlapping gene rather than another alternatively spliced transcript.

All other ESTs with homology to the *FRA16D*-spanning sequence exhibit features which discount them as likely to represent authentic transcripts. In most cases the homology between the EST and the chromosomal DNA sequence was collinear and therefore not interrupted by introns. The exceptions (Table 2) were chimeric, having either additional unique sequences from elsewhere in the genome than 16q23.2 or Alu sequences of unattributable origin. The 3′ end sequences that had poly(A) tails also had poly(A) sequences in the chromosomal DNA, indicating that the poly(A) sequence in the EST was not a post-transcriptional addition. In addition, there were no expected polyadenylation signal sequences (AATAAA) located ~20 bp upstream of the poly(A) tracts. A single exception was qz19h11, where the EST has a poly(A) segment while the genomic DNA does not; however, there is no poly(A) signal and no intron boundaries are evident. The majority of the ESTs were singletons indicative of rare events and are most likely due to rare oligo(dT) priming from a chromosomal DNA contaminant in the cDNA library preparations. ESTs that are in the same transcriptional orientation as the *FOR* gene could represent 3′ exons of yet additional alternatively spliced forms of *FOR* gene transcripts.

The 270 kb *FRA16D*-spanning sequence was also subjected to GenScan gene prediction analysis. This analysis revealed only three likely exons, one of which (at positions 4216–4621) was exon 8 of the *FOR* gene. The others (at 8134–7818 and 225 339–225 270) were both in the opposite orientation to the *FOR* gene and neither corresponded to any of the EST homologies that were detected in the BLAST searches. The GenScan analysis did not identify exon 9 of the *FOR I* transcript. This is consistent with this exon (which contains mainly 3′ untranslated region) being a poorly utilized alternative splice pathway giving rise to the low abundance *FOR I* mRNA. The remainder of the *FOR I* mRNA 3′ untranslated region (consisting of an AT-rich sequence of only 30 bases) is thought to be encoded by an additional exon (exon 10) (Fig. 1); however, this exon is yet to be located in the chromosomal DNA sequence.

5′-RACE experiments using mRNA from normal (HS578BST) and tumour (T47D) cells were utilized to extend and confirm the sequences of the clusters of GenBank EST sequences of transcripts I–IV and to determine the organization of the alternatively spliced mRNAs which they represent (Fig. 3). Transcripts I–III were found to have a common 5′ end, indicating a common promoter. The exons shared and utilized in the alternatively spliced mRNAs were identified in BAC sequences AF217491, AF217492, AC009044, AC009280 and AC009129 (Fig. 1). The confinement of distribution of EST sequences amongst exons confirmed that the different transcripts were due to alternative splicing. Transcripts I–III share

Table 2. Location of singleton EST homologies in the *FRA16D*-spanning DNA sequence

EST	GenBank accession no.	Location within *FRA16D* (kb)
qg88f04.x1	AI219858	4 (exon 8 of *FOR I*)
EST03811	T05922	25
tr01e09.x1	AI914748	44
41d9	W28001	51
tx82e01.x1	AI689243	61
tj44g06.x1[a]	AI868292	63
yq75a06.r1	R98219	68
ow72g03.x1	AI022399	72
yd85d04.r1	T87470	74
EST79307	AA368108	77
ye42h09.s1	T96062	80
yd31b08.r1	T85134	86
tq14e08.x1	AI697096	93
C77236 (mouse)	C77236	94
xr51f08.x1	AW302008	95
40.H12R	AA447280	97
Z20768	Z20768	109
ye25b03.r1	T91653	111
zt58b05.s1	AA398024	113
np61g05.s1	AA633361	119
ti22c10.x1[b]	AI432669	120
EST214855 (rat)	AI169027	120
zf68e10.r1	AA063112	126
za57h10.s1	N70268	133
ab10e01.r1	AA485822	144
qz19h11.x1[c]	AI361849	147
nk32e05.s1[b]	AA555206	152
EST56568	AA349718	160
EST56567	AA349717	161
zh97a01r1	AA007376	173
yd85h12.r1	T80206	173
am38c03.s1	AA884576	178
43b3	W28177	190
ye76g02.r1	R02647	199
ye90d10.s1	R18122	195
EST37087	AA333054	213
ETA1	AA010088	215
EST00978	M78830	220
EST113064	AA297515	235
EST15m13	R81094	241
ua24g03.r1	AA982977	249
wc43f06.x1	AI799856	258
qg88f04.x1	AI219858	264 (exon 9 of *FOR I*)

[a]Spliced to Alu sequence of unknown origin.
[b]Chimeric, with sequences from other chromosomes.
[c]No AATAAA preceding poly(A) tail, no intron splice site.

a common initiation methionine with an adjacent 5′ Kozak translation initiation sequence and an upstream in-phase termination codon. The open reading frames code for proteins of 41.2, 46.7 and 21.5 kDa, respectively. Each of these open reading frames shares homology with the oxidoreductase family of proteins and therefore the gene has been named *FOR* (fragile site *FRA16D* oxidoreductase), with the alternatively spliced transcripts I–III referred to as *FOR I*, *FOR II* and *FOR III*, respectively.

Northern blot analysis with various *FOR* exon probes identified the 2.3 kb *FOR II* transcript as the predominant and ubiquitously expressed mRNA, with *FOR I* and *FOR II* mRNAs showing a similar pattern of expression (Fig. 4). A DNA probe spanning the 5′ exons detected additional RNAs with a different tissue-specific pattern. A cluster of ESTs (Fig. 1) with homology limited to exon 1 of the *FOR* gene was found from a BLAST search of the databases. This suggests that these transcripts (referred to as *FOR IV*) might arise from a different promoter and may well constitute a different gene, the 3′ end of which overlaps with the 5′ end of *FOR* (Fig. 1). The 3′ end sequences of these ESTs contain a very short open reading frame (4.1 kDa) which is truncated with respect to that seen in the *FOR* transcripts. The complete *FOR I–FOR III* mRNA and partial related transcript sequence (*FOR IV*) were determined from 5′-RACE and RT–PCR products and deposited in GenBank (accession nos AF227526, AF227527, AF227528 and AF227529). An mRNA of ~2.7 kb in length of unknown identity was detected with probe B in the kidney and three RNAs (~5.5, ~8 and ~11 kb) were found in various tissues (Fig. 4), indicating that additional alternatively spliced *FOR* transcripts are likely. The complete length of the *FOR IV* mRNA is yet to be determined and may account for one or more unidentified transcripts.

FOR mRNA in normal and tumour cells

RT–PCR and 5′-RACE were used to detect the various *FOR* transcripts in normal and tumour cells. Striking differences between the presence/absence of *FOR I* and *FOR III* transcripts was noted for the 'normal' fibroblast-like cell line HS578BST and various tumour cell lines (Fig. 5). 5′-RACE and RT–PCR products for transcript-specific PCR were sequenced to confirm the identity of the respective products. The sequence of the aberrant RT–PCR product from cell line MDA-MB-453 generated using a *FOR III*-specific primer contains a retroviral element (HERV-H) 5′ of exons 5 and 6A of *FOR* (GenBank accession no. AF239665). In addition, one EST (qz23c04.x1) identified in database BLAST searches contains exons 1–3 of *FOR* spliced at the 3′ end to another retroviral element, LTR13.

Homozygous deletion of *FOR I* exon 9 detected in AGS tumour cells suggests that the gain of *FOR I* transcript will not be a common event in tumour cells. Similarly, the loss of *FOR III* transcript is not common to all tumour cells as *FOR III*-specific RT–PCR products were readily detected in both AGS and HCT116 cells (Fig. 5).

FOR-encoded proteins

The alternatively spliced mRNAs transcribed from the gene each show homology to the oxidoreductase superfamily of proteins. The open reading frames of the alternatively spliced *FOR* gene mRNAs I–III have a common N-terminus which contains two WW domains (Fig. 3). The first WW domain is truncated in the *FOR IV* open reading frame; however, since this mRNA appears to originate from a distinct promoter it may well be that an upstream reading frame is utilized in this mRNA. The open reading frame from the *FOR III* transcript retains the WW domains; however, it is truncated for approximately half the length of the oxidoreductase homology (Fig. 3).

DISCUSSION

Structure/composition of common fragile sites

A principle aim of the current study was to compare the DNA sequence spanning the *FRA16D* fragile site with that of other common fragile site sequences in order to identify possible shared sequence elements which might account for common properties. Given that it is DNA repeat sequences that constitute the molecular basis of rare chromosomal fragile sites, the overall repeat composition of the *FRA16D*-spanning DNA sequence was strikingly different to other common fragile site sequences (Fig. 2), although those from the vicinity of *FRA7H* and *FRA6E* have not been shown to span the fragile site. A further difference was noted with respect to LINE elements, which have been proposed to play a role in the DNA instability seen at the *FRA3B* common fragile site (15). No consistent association was seen between LINE elements and the *FRA16D* breakpoints in AGS and HCT116 tumour cell DNAs, suggesting that LINE elements are not an essential feature of common fragile site-associated DNA instability in cancer. Instead, Alu repeats were more common in *FRA16D* than LINE elements. These differences suggest that it is not the broad composition of the region which gives rise to chromosomal fragility and, therefore, it is more likely that localized sequence elements (such as DNA polymerase pause sites) play the determining role.

Analysis of the *FRA16D* sequence using the FlexStab program supports the proposal that regions of flexibility contribute to DNA instability in the vicinity of fragile sites (16) as two peaks of flexibility correspond to homozygous deletion breakpoints in tumour cells. These DNA sequences contain the simple AT dinucleotide repeat which is capable of forming secondary structures. The AT-rich DNA sequences at nucleotide positions 191.6–198.4 kb of AF217490 are also capable of forming hairpin structures *in vitro* and are therefore similar to the AT-rich DNA repeats associated with the *FRA10B* and *FRA16B* rare fragile sites. These DNA polymerase pause sites coincide with some of the breakpoint locations in AGS and HCT116 tumour cells. Given that DNA breakage is a form of chromosome fragility, then the coincidence of cancer cell breakpoints with these AT-rich secondary structures suggests that the latter may have a role in cytogenetic expression of the *FRA16D* fragile site. While *FRA16D* is a common fragile site and is therefore expressed in all individuals, there is a possibility that polyporphism in the vicinity of *FRA16D* may contribute to differences in the percentage of metaphases which exhibit cytogenetic expression of the fragile site.

Fragile site DNA sequences have been found to act as regions of delayed replication (24–26) and in so doing may contribute to the coupling of DNA replication with cell division. Deletion of fragile site sequences in cancer cells may therefore confer a selective advantage to the tumour cell in a manner which is otherwise independent of the gene in which

A. *FOR I* mRNA open reading frame

MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLPYGWEQETDENG
QVFFVDHINKRTTYLDPRLAFTVDDNPTKPTTRQRYDGSTTAMEILQGRDFTGKVVVVTGANSGIGFETAKSFAL
HGAHVILACRNMARASEAVSRILEEWHKAKVETMTLDLALLRSVQHFAEAFKAKNVPLHVLVCNAATFALPW
SLTKDGLETTFQVNHLGHFYLVQLLQDVLCRSAPARVIVVSSESHRFTDINDSLGKLDFSRLSPTKNDYWAMLAY
NRSKLCNILFSNELHRRLSPRGVTSNAVHPGNMMYSNIHRSWWVYTLLFTLARPFTKSMVSDCLVEGGHF

B. *FOR II* mRNA open reading frame

MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLPYGWEQETDENG
QVFFVDHINKRTTYLDPRLAFTVDDNPTKPTTRQRYDGSTTAMEILQGRDFTGKVVVVTGANSGIGFETAKSFAL
HGAHVILACRNMARASEAVSRILEEWHKAKVETMTLDLALLRSVQHFAEAFKAKNVPLHVLVCNAATFALPW
SLTKDGLETTFQVNHLGHFYLVQLLQDVLCRSAPARVIVVSSESHRFTDINDSLGKLDFSRLSATKNDYWAMLA
YNRSKLCNILFSNELHRRLSPRGVTSNAVHPGNMMYSNIHRSWWVYTLLFTLARPFTKSMQQGAATTVYCAAVP
ELEGLGGMYFNNCCRCMPSPEAQSEETARTLWALSERLIQERLGSQSG

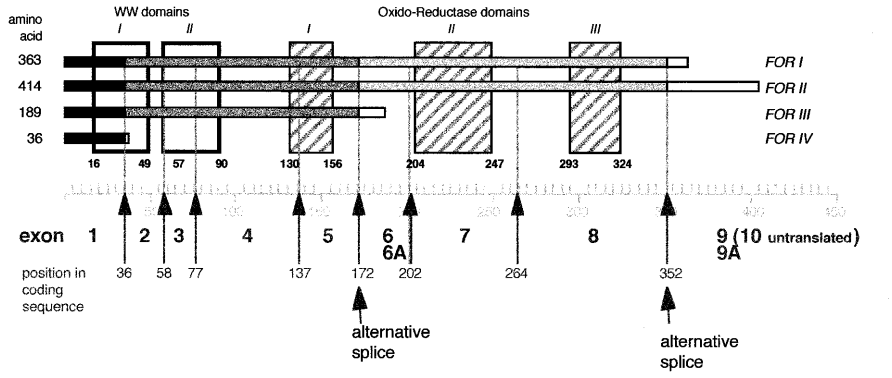
C. *FOR III* mRNA open reading frame

MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLPYGWEQETDENG
QVFFVDHINKRTTYLDPRLAFTVDDNPTKPTTRQRYDGSTTAMEILQGRDFTGKVVVVTGANSGIGFETAKSFAL
HGAHVILACRNMARASEAVSRILEEWKTKYHPPPEKCRIKIFH

D. *FOR IV* mRNA open reading frame

MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYAK

E. Comparison of *FOR* open reading frames and location of WW and Oxido-Reductase domains



F. *FOR* WW domains

FOR sequence (WW1)	DELPPGWEERTTKDGWVYYANHTEEKTQWEHPKT
FOR sequence (WW2)	GDLPYGWEQETDENGQVFFVDHINKRTTYLDPRL
WW domain consensus	--LP-GWE---tttGt-YYh-HNTtTTtW-tPt-

Figure 3. Comparison of *FOR* open reading frames and location of WW and oxidoreductase domains. (**A**–**D**) Open reading frames from *FOR I*–*FOR IV* mRNAs, respectively. (**E**) Comparison of *FOR* open reading frames and location of WW (white boxes) and oxidoreductase domains (striped boxes). Oxidoreductase domains were identified by alignment of protochlorophyllide reductase precursor proteins from *Arabidopsis thaliana* (accession no. P21218), *Pisum sativum* (Q01289), barley (P13653) and *Avena sativa* (P15904), oxidoreductases from *Streptomyces antibioticus* (Q03326) and *Neurospora crassa* (Q92247) and FIXR protein from *Bradyrhizobium japonicum* (P05406) with the *FOR II* open reading frame. The three homologous oxidoreductase domains contained: (i) five completely conserved amino acids over a 28 amino acid sequence; (ii) seven completely conserved amino acids over a 47 amino acid sequence; (iii) five completely conserved amino acids over a 33 amino acid sequence. (**F**) Comparison of *FOR* WW domains with the WW domain consensus sequence.

the fragile site is located. Alternatively, given the coupling of replication with transcription, the deletion of a sequence which normally delays replication may have an effect on the expression of genes in the vicinity.

Identification of the *FOR* gene spanning *FRA16D*

Given the proposed role of the *FHIT* gene in mediating the biological consequences of *FRA3B*-associated DNA instability in cancer cells we sought to identify the closest gene to *FRA16D* which might mediate the biological effects of *FRA16D*-associated DNA instability in cancer (10,11). Sequence analysis of the *FRA16D*-spanning DNA sequence revealed the *FOR* gene as the sole transcript in the immediate vicinity of the minimal region of homozygous deletion in cancer cells. Alternative exons of this gene were found to flank both the *FRA16D* fragile site and the tumour cell deleted regions, the alternative exon 9 being deleted in the AGS cell line. No additional authentic transcripts from within the *FOR*

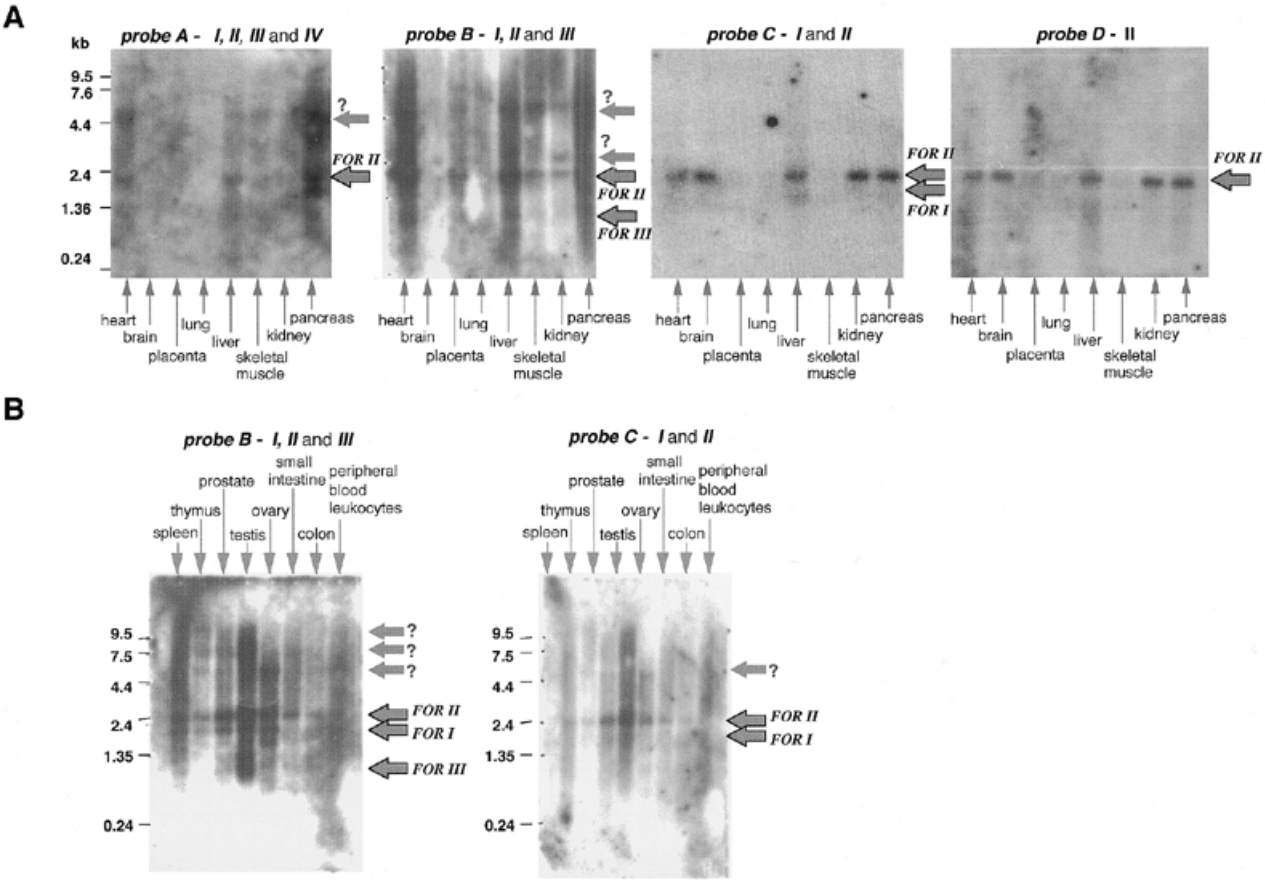


Figure 4. Northern blots of RNA from various human tissues. Expected *FOR* mRNAs (I–III) are indicated for the respective DNA probes which span various exons of the *FOR* gene. Large arrows indicate *FOR* mRNAs (*FOR I* ~1.8 kb, *FOR II* ~2.2 kb, *FOR III* ~ 0.74 kb). Other unidentified transcripts (possibly including *FOR IV*) are indicated by small arrows. (**A**) Blots of mRNA from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas hybridized to probes A, B, C and D, respectively. (**B**) Blots of mRNA from spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocytes hybridized to probes B and C, respectively.

gene intron were evident. In an analogous situation to that seen at the *FRA3B* locus, deletions at *FRA16D* appear to be largely intronic to the major alternatively spliced transcript (*FOR II*). Studies on the *FHIT* gene spanning *FRA3B* have given contradictory results in terms of the possible role of *FHIT* in neoplasia. Therefore, it will be of interest to determine the effect of deletions at *FRA16D* on the *FOR* gene transcripts and whether any of the *FOR*-encoded proteins and their dysfunction brought about by *FRA16D*-associated deletion have a role to play in cancer.

The finding that the *FOR* gene spans *FRA16D* represents a shared feature of 'common' fragile sites *FRA3B* and *FRA16D* in that they are both unstable regions located within large introns of large genes; *FHIT* (>1 Mb) and *FOR* (>1 Mb). The sequences determined at the *FRA7G*, *FRA7H* and *FRA6E* loci may be of insufficient length to identify the respective spanning genes.

Differential expression of alternatively spliced and aberrant *FOR* transcripts in normal and tumour cells

RT–PCR and 5′-RACE gave differing patterns of *FOR* transcript expression in various normal and tumour cell lines. It will be of interest to determine whether there are differences in the ratio of *FOR* transcripts which are consistent with the biological characteristics of various cell types, e.g. neoplastic state or metastatic potential. It is unlikely that the presence of *FOR I* transcripts will be a common property of tumour cells since at least the AGS cell line is homozygously deleted for *FOR I* exon 9. Additional aberrant *FOR* transcripts, including sequences fused to retroviral LTRs, were detected in tumour cells. Extensive studies on the *FHIT* gene at the *FRA3B* locus have yet to resolve the role (if any) of aberrant transcripts of this gene in neoplasia. It is likely that extensive functional studies of alternative *FOR* transcript expression will be needed in order to determine what contribution (if any) this phenomenon contributes to the biological characteristics of the cells where it is observed.

It may well be that the ratio of the various *FOR* transcripts is perturbed by DNA instability in the region and that it is the resultant alteration in relative abundance of the various *FOR*-encoded proteins which mediates the biological consequences of DNA instability at *FRA16D*. For example, the homozygous deletion in AGS cells deletes exon 9 of the *FOR I* transcript and may have an effect on the stability of the *FOR II* transcript; however, this deletion is unlikely to have any effect on the *FOR III* transcript, which terminates well outside the homozygously deleted region.

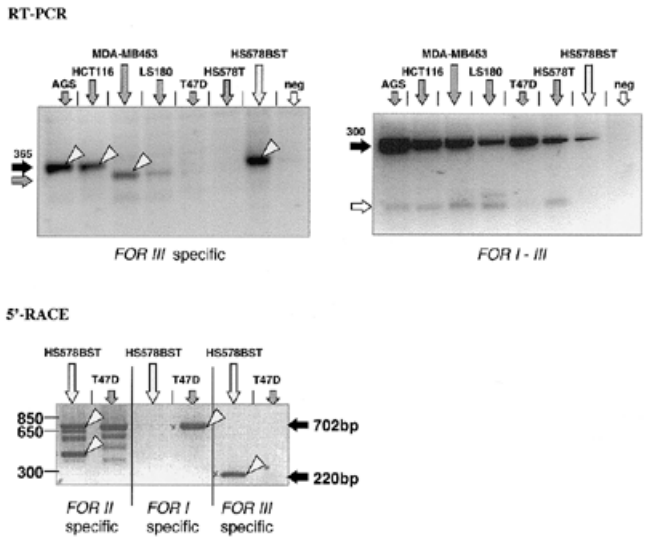


Figure 5. *FOR* transcripts in normal and tumour cells. Products that were subjected to sequence analysis are indicated by arrowheads. (**A**) RT–PCR were either 'specific' for the *FOR III* transcript or 'general', being able to detect *FOR I–III* mRNAs. (**B**) 5′-RACE specific for the *FOR I*, *FOR II* and *FOR III* transcripts in 'normal' HS578BST cells and T47D tumour cells.

Possible function of *FOR* and role in neoplasia

The *FOR*-encoded proteins show sequence homology to the oxidoreductase family of proteins and contain WW domains. Other members of this family of proteins include the *YES* proto-oncogene-associated proteins and *NEDD* ubiquitin ligases.

The open reading frame from the *FOR III* transcript retains the WW domains; however, it is truncated for approximately half the length of the oxidoreductase/ubiquitin ligase homology (Fig. 3). The *FOR III* protein is therefore likely to be able to bind proteins that recognize the common *FOR I* and *FOR II* WW domains but not be able to perform the enzymatic function encoded by the *FOR I* and *FOR II* proteins (possibly ubiquitination). Such characteristics make the *FOR III* protein a likely competitor of *FOR I* and/or *FOR II*. Since ubiquitination facilitates the process of specific protein turnover, *FOR III* could therefore act to prolong the half-life of its substrate by competing with *FOR I* and/or *FOR II*.

WW domains are regions of protein–protein interaction that bind polyproline-rich motifs (PY domains) in specific partner proteins. Specificity in this interaction is determined by differences in particular amino acids in the various WW domains. Proteins known to bind to WW domains include the *YES* proto-oncogene product and p53 binding protein-2 (27). Alteration in the relative levels of the *FOR*-encoded proteins as a consequence of *FRA16D*-associated instability is therefore likely to influence the biological function of the PY motif-containing protein(s), which is (are) the normal binding partner that the *FOR* proteins share through their WW domain.

The majority of deletions in the 16q23.2 region are heterozygous, with the homozygous deletions being confined and limited in number (11). Cells which still have the capacity to produce *FOR II* protein (from a normal chromosome 16 *FOR* allele) might have an elevated level of *FOR III* (through *FRA16D*-associated deletion of the other chromosome 16 allele) and therefore have a selective 'heterozygote' advantage.

The finding of aberrant *FOR*-related transcripts spliced to retroviral RNA sequences in tumour cells that do not necessarily exhibit *FRA16D* homozygous deletion (e.g. MDA-MB-453, Fig. 4) suggests that dysfunction of the pathway involving the *FOR* WW domain could be a common event in neoplasia, perhaps through other forms of *FRA16D*-related DNA instability such as DNA insertion or translocation. Three out of five previously mapped multiple myeloma translocations (23) map within the *FOR* gene, suggesting that DNA instability at the *FRA16D* locus and aberrant expression of the *FOR* gene may have a variety of roles to play in various forms of cancer.

The c-*MAF* oncogene is also located in the vicinity of *FRA16D* (Fig. 1) and has been identified as a target for multiple myeloma translocations, exhibiting dysregulation of expression of the translocated allele (23). It is not yet clear whether deletions in the *FRA16D* region, such as the LOH often seen in breast cancer or the homozygous deletions seen in certain adenocarcinomas, have an effect on c-*MAF* expression. The location of the various forms of DNA instability in this region with respect to the *FOR* gene suggest that this gene may also be affected; it is a distinct possibility that both c-*MAF* and *FOR* have a role to play in mediating the biological consequences of *FRA16D*-associated DNA instability.

It will be of interest to determine whether there is a general relationship between common chromosomal fragile sites and DNA instability in cancer. In this regard it is noteworthy that there are two other common regions of cancer cell LOH on 16q in addition to that involving *FRA16D* (28). One of these at 16q24.3 maps near the telomere while the other at 16q22.1 maps near another common fragile site, *FRA16C*.

MATERIALS AND METHODS

Cell lines

Cell lines AGS, HCT116, HS578BST, HS578T, LS180, MDA-MB-453 and T47D are from the Department of Cytogenetics and Molecular Genetics, WCH collection, and were originally obtained from the American Type Culture Collection or the European Collection of Cell Cultures. AGS and LS180 cells were grown as described previously (10). HS578BST cells were grown in OPTI-MEM with L-glutamine, 0.01 µg/ml epidermal growth factor, 0.5 µg/ml hydrocortisone, 8% fetal calf serum in 5% CO_2. T47D, MDA-MB-453 and HS578T cells were grown in RPMI 1640 with L-glutamine, 10% fetal calf serum in 5% CO_2.

Large-scale sequencing of *FRA16D*

Sequencing of the 270 kb region spanning *FRA16D* consisted of: (i) sonication libraries; (ii) nebulization libraries of BAC clones 325M3 and 353B15; and (iii) restriction fragments of λ clones (for sequencing between BAC 325M3 and BAC 353B15).

Construction of sonication libraries. For DNA sonication and cloning we modified the protocol from the Sanger Centre (http://www.sanger.ac.uk/Teams/Team53/sonication.shtml). Aliquots of 1 µg of each BAC DNA were sonicated in 300 µl

H_2O and 8 µl 10× mung bean buffer (500 mM Na acetate, 300 mM NaCl, 10 mM $ZnSO_4$, pH 5.0) on ice for 20 s using a Heat Systems Sonicator W-225 (50% duty, 3.5 power) (Ultrasonic, Farmingdale, NY). After reducing the volume to 80 µl, blunt ends were created by adding 40 U of mung bean nucleases (Biolabs, Beverly, MA) and incubating the mixture at 30°C for 25 min. The products were size fractioned on a 1% agarose gel and fragments ranging from 0.7 to 2 kb were extracted with the QiaQuick Gel Extraction kit (Qiagen, Hilden, Germany). Samples of 1500 ng sonicated DNA (used in 500 ng aliquots) were ligated into vector pUC18-Sma (Amersham Pharmacia, Uppsala, Sweden) at 16°C overnight and transformed into electroporation-competent Sure cells (Stratagene, La Jolla, CA). Samples of 600 and 1500 clones, respectively, of the sonication libraries of BAC 325M3 and 353B15 were gridded on 96-well plates and sequenced in one direction using the M13-forward primer. Sequences were assembled into contigs using the Staden Package (MRC, Cambridge, UK) on a UNIX computer and edited in LASERGENE (Macintosh). For a selected number of clones additional sequences with the M13-reverse primer were retrieved and assembled. Additional sequencing primers were designed and PCR products sequenced to close the gaps between contigs.

Construction of nebulization libraries. Aliquots of 10 µg of each BAC DNA were mixed with 200 µl 10× TM buffer (500 mM Tris–HCl, pH 7.5, 150 mM $MgCl_2$) and 1 ml sterile glycerol and H_2O added to 2 ml. The mixture was pipetted into an IPI nebulizer and nebulized at 10 p.s.i. for 45 s. The nebulized DNA was then precipitated, end repaired, size fractionated and cloned as described for the sonicated DNA. Samples of 300 and 500 nebulized clones, respectively, of BAC 325M3 and 353B15 were sequenced as described above and included in the assemblies. Subclones for sequencing of BAC 353B15 were picked randomly, whereas BAC 325M3 subclones were selected after hybridization of specific λ clones of the tile path made from BAC 325M3 (10).

Subcloning of restriction fragments of λ -clones between λ-32 and λ-191 in vector pUC19. Clones were sequenced with M13-forward and M13-reverse primers as well as with sequence-specific primers. In some cases subclones derived from specific restriction fragments were also subject to sonication, shotgun cloning and sequencing.

Sequencing was performed with the ABI Big Dye Terminator kit from Perkin Elmer (Foster City, CA). In cases where sequencing with the Big Dye Terminator kit failed, a dRhodamine Terminator kit (ABI) was used, as recommended for GT-rich and homopolymeric regions by the ABI DNA sequencing guide.

Long range PCRs were performed on ~70 kb of DNA sequence in the region between BACs 325M3 and 353B15. Blood bank DNA was used as template. The remaining tile path of λ subclones (10) was subjected to restriction analysis to give a redundant restriction map of the *FRA16D* region.

The final sequence was analysed using BLAST (http://www.ncbi.nlm.nih.gov/BLAST), REPEATMASKER (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker), GENSCAN (http://CCR-081.mit.edu/GENSCAN.html) and FLEXSTAB (http://leonardo.ls.huji.ac.il/departments/genesite/faculty/bkerem.htm).

Northern blot hybridization

Probes for hybridization on multiple tissue northern blots from Clontech were: probe A, part of exon 1A (163 bp), positions 298–461 of AF227529; probe B, exons 3–6A (366 bp), positions 291–657 of AF227528; probe C, exon 7 (186 bp), positions 690–876 of AF227526; probe D, part of exon 9A (779 bp), positions 1182–1961 of AF227527.

RNA extraction

RNA was extracted from 1×10^7 cells for each of the cell lines using the RNeasy Mini Kit from Qiagen. The cells were disrupted by addition of 600 µl of lysis buffer RLT (supplied with the kit). The lysed cells were homogenized by passing 5–10 times through a 21G (0.8 × 38 mm) needle attached to a 5 ml syringe. Aliquots of 600 µl of 70% ethanol were added and the samples were applied to RNeasy Mini Spin columns. Purification and elution of the samples were carried out according to the kit manual. A total of 35–98 µg of RNA was obtained.

RT–PCR

Reverse transcription was carried out in a 40 µl reaction volume using 12–33 µg of total RNA from cell lines AGS, HCT116, MDA.MB.453, LS180, T47D, HS578T and HS578BST, respectively, according to the product sheet of the Gibco BRL Superscript RNAse H^- Reverse Transcriptase kit (Gibco BRL, Gaithersburg, MD) except for the addition of 20 U RNase inhibitor (Rnasin; Promega, Madison, WI) to the mixture.

Aliquots of 100 ng of cDNA were amplified in PCRs using various cDNA primer combinations under standard PCR conditions (10 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 30 s; then 25 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 30 s).

Primers (5′→3′) used in RT–PCR were: HHCMA-F (ATCTTGGCCTGCAGGAACATGGCA) and wb85-F (TTATTCTGCACTTTTCTGGCGGAG), *FOR III*-specific; FOR-ex3 (GAACAAGAAACTGATGAGAACGGA) and wb85-F, *FOR III*-specific; wb85-E12 (TTACTACGCCAATCACACCGAGGA) and wb85-A (TGAATTAGCTCCAGTGACCACAAC), common for *FORI*, *FOR II* and *FOR III*.

5′-RACE

Complete 5′ ends of transcripts *FOR I*, *FOR II* and *FOR III* were determined by 5′-RACE experiments including first strand cDNA synthesis, purification, TdT tailing of the cDNA, PCR of dC-tailed cDNA and nested amplification according to the Gibco BRL instruction manual.

Aliquots of 1 µg of total RNA of cell lines HS578BST (normal) and T47D (tumour) were taken as templates. First strand cDNA synthesis was conducted with the following specific GSP1 primers: *FOR I*, coxido-R, 5′-TTATTTCAGCACTCAGCTCAAAGTCAC-3′; *FOR II*, HHCMA-B, 5′-AGCAAAGAGACCTATGCCTAGCCCA-3′; *FOR III*, wb85-F, 5′-TTATTCTGCACTTTTCTGGCGGAG-3′.

PCR of the dC-tailed cDNA was carried out with the GSP2 primers: *FOR I* and *FOR II*, coxido-32, 5′-ATATCTG-

TAAATCGATGGGACTCTG-3′; *FOR III*, wb85-A, 5′-TGAATTAGCTCCAGTGACCACAAC-3′.

Nested amplification was done with 5 µl of a 1:100 dilution of GSP2 PCR products and the GSP3 primers: *FOR I* and *FOR II*, coxido-21, 5′-ACATGAAGAGGCACATTCTTGGCCT-3′; *FOR III*, wb85-E, 5′-TCCTCGGTGTGATTGGCG-TAGTAA-3′ in combination with the AUAP primer (Gibco BRL).

PCR products were extracted with a QiaQuick kit from agarose gels after electrophoresis and sequenced directly with GSP3 primers and primer tj96-C (5′-GGAGGCAGCTCGTC-CTCACTG-3′).

cDNA sequence of *FOR IV* (GenBank accession no. AF227529)

The preliminary cDNA sequence of the *FOR IV* transcript is incomplete at its 5′ end at this stage. The sequence determined so far derives from overlapping EST clones qf42f03.x1 (AI149681) and tm79c11.x1 (AI570665). The latter was sequenced additionally with the internal primer tj96-C (5′-GGAGGCAGCTCGTCCTCACTG-3′).

Determination of breakpoints in cell lines AGS and HCT116

Deletions in cell lines AGS and HCT116 were determined in duplex STS–PCRs as described in Mangelsdorf *et al.* (10). All primers are listed 5′→3′ in Table 1.

Four regions of homozygous deletion (referred to as *HZD I–HZD IV*) were detected in the AGS cell line. The proximal breakpoint for *HZD I* in AGS was narrowed down to 654 bp between STSs 16D-15/16D-36 (+) and 16D-1/16D-60 (–); the distal breakpoint of *HZD I* of 3962 bp is between STS 16D-70 (–) and 16D-47 (+). The proximal breakpoint for *HZD II* in AGS was narrowed down to 3030 bp between STSs 16D-57 (+) and 16D-67 (–); the distal breakpoint of *HZD II* of 1720 bp is between STS 16D-68 (–) and 16D-54 (+). The proximal breakpoint for *HZD III* in AGS was narrowed down to 209 bp between STSs 16D-51 (+) and 16D-55 (–); the distal breakpoint of *HZD III* of 5690 bp is between STS 16D-202 (–) and 16D-69 (+). The proximal breakpoint for *HZD IV* in AGS was narrowed down to 5179 bp between STSs 16D-30/16D-44 (+) and ETA1 (–); the distal breakpoint of *HZD IV* of ~1500 bp is between STS IM7 (–) and 410S1A (+).

Two regions of homozygous deletion (referred to as *HZD I* and *HZD II*) were detected in the HCT116 cell line. The proximal breakpoint for *HZD I* in HCT116 was narrowed down to 1835 bp between STSs 16D-19 (+) and 16D-61 (–); the distal breakpoint of *HZD I* of 1549 bp is between STS 16D-62 (–) and qz19h11 (+). The proximal breakpoint for *HZD II* in HCT116 was narrowed down to 422 bp between STSs 16D-63 (+) and 16D-30 (–); the distal breakpoint of *HZD II* of 1513 bp is between STS 16D-66 (–) and 801A (+).

For determining the presence of exon 9 of *FOR I* (51 bp) in the AGS cell line a duplex PCR with genomic primers from the dystrophin gene (*DMD*) as described in example 1 was carried out with primers 8040/8041 (Table 1).

ACKNOWLEDGEMENTS

We thank Oliva Handt for comments on the manuscript and Shirley Richardson for excellent technical assistance. R.I.R. thanks Shelley Richards for support and encouragement. K.R. was supported by a post-doctoral fellowship from Deutscher Akademischer Austausch Dienst (Germany). This work was supported by grants from the National Health and Medical Research Council of Australia, the Women's and Children's Hospital Research Foundation and the Anti-Cancer Foundation of the Universities of South Australia.

REFERENCES

1. Yunis, J.J. and Soreng, A.L. (1984) Constitutive fragile sites and cancer. *Science*, **226**, 1199–1204.
2. Sutherland, G.R. and Richards, R.I. (1995) The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet. Dev.*, **5**, 323–327.
3. 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 a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. *Hum. Mol. Genet.*, **3**, 2123–2130.
4. Jones, C., Penny, L., Mattina, T., Yu, S., Baker, E., Voullaire, L., Langdon, W.Y., Sutherland, G.R., Richards, R.I. and Tunnacliffe, A. (1995) Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene *CBL2*. *Nature*, **376**, 145–149.
5. Ohta, M., Inoue, H., Cotticelli, M.G., Kastury, K., Baffa, R., Palazzo, J., Siprashvilli, Z., Mori, M., McCue, P., Druck, T., Croce, C.M. and Huebner, K. (1996) The *FHIT* gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell*, **84**, 587–597.
6. Sozzi, G., Veronese, M.L., Negrini, M., Baffa, R., Cotticelli, M.G., Inoue, H., Tornielli, S., Pilotti, S., De Gregorio, L., Pastorino, U. *et al.* (1996) The *FHIT* gene at 3p14.2 is abnormal in lung cancer. *Cell*, **85**, 17–26.
7. Engelman, J.A., Zhang, X.L. and Lisanti, M.P. (1998) Genes encoding human caveolin-1 and -2 are co-localized to the D7S522 locus (7q31.1), a known fragile site (FRA7G) that is frequently deleted in human cancers. *FEBS Lett.*, **436**, 403–410.
8. Huang, H., Qian, C., Jenkins, R.B. and Smith, D.I. (1998) Fish mapping of YAC clones at human chromosomal band 7q31.2: identification of YACs spanning *FRA7G* within the common region of LOH in breast and prostate cancer. *Genes Chromosomes Cancer*, **21**, 152–159.
9. Huang, H., Qian, C., Profit, J., Wilber, K., Jenkins, R.B. and Smith, D.I. (1998) *FRA7G* extends over a broad region: coincidence of human endogenous retroviral sequences (HERV-H) and small polydispersed circular DNAs (spcDNA) and fragile sites. *Oncogene*, **16**, 2311–2319.
10. Mangelsdorf, M., Ried, K., Woollatt, E., Dayan, S., Eyre, H., Finnis, M., Hobson, L., Nancarrow, J., Venter, D., Baker, E. and Richards, R.I. (2000) Chromosomal fragile site *FRA16D* and DNA instability in cancer. *Cancer Res.*, **60**, 1683–1689.
11. Paige, A.J.W., Taylor, K.J., Stewart, A., Sgouros, J., Gabra, H., Smyth, J.F., Porteous, D.J. and Watson, J.E.V. (2000) A 700kb physical map of a region of 16q23.2 homozygously deleted in multiple cancers and spanning the common fragile site FRA16D. *Cancer Res.*, **60**, 1690–1696.
12. Sutherland, G.R., Baker, E. and Richards, R.I. (1998) Fragile sites still breaking. *Trends Genet.*, **14**, 501–506.
13. Boldog, F., Gemmill, R.M., West, J., Robinson, M., Robinson, L., Li, E., Roche, J., Todd, S., Waggoner, B., Lundstrom, R. *et al.* (1997) Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B. *Hum. Mol. Genet.*, **6**, 193–203.
14. Inoue, H., Ishii, H., Alder, H., Snyder, E., Druck, T., Huebner, K. and Croce, C.M. (1997) Sequence of the *FRA3B* common fragile site: implications for the mechanism of *FHIT* deletion. *Proc. Natl Acad. Sci. USA*, **94**, 14584–14589.
15. Mimori, K., Druck, T., Inoue, H., Alder, H., Berk, L., Mori, M., Huebner, K. and Croce, C.M. (1999) Cancer-specific chromosome alterations in the constitutive fragile region *FRA3B*. *Proc. Natl Acad. Sci. USA*, **96**, 7456–7461.
16. Mishmar, D., Rahat, A., Scherer, S.W., Nyakatura, G., Hinzman, B., Kohwi, Y., Mandel-Gutfroind, Y., Lee, J.R., Drescher, B., Sas, D.E., Margalit, H., Platzer, M., Weiss, A., Tsui, L.-C., Rosenthal, A. and Kerem, B. (1998) Molecular characterisation of a common fragile site

(*FRA7H*) on human chromosome 7 by the cloning of an SV40 integration site. *Proc. Natl Acad. Sci. USA*, **95**, 8141–8146.
17. 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. *et al.* (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in *FRAXE* mental retardation. *Cell*, **74**, 127–134.
18. Gnarra, J.R., Tory, K., Weng, Y., Schmidt, L., Wei, M.H., Li, H., Latif, F., Liu, S., Chen, F., Duh, F.M. *et al.* (1994) Mutations in the VHL tumour suppressor gene in renal carcinoma. *Nature Genet.*, **7**, 85–90.
19. Siprashvilli, Z., Sozzi, G., Barnes, L.D., McCue, P., Robinson, A.K., Eryomin, V., Sard, L., Tagliabue, E., Greco, A., Fusetti, L. *et al.* (1997) Replacement of Fhit in cancer cells suppresses tumorigenicity. *Proc. Natl Acad. Sci. USA*, **94**, 13771–13776.
20. Huebner, K., Garrison, P.N., Barnes, L.D. and Croce, C.M. (1998) The role of the *FHIT/FRA3B* locus in cancer. *Annu. Rev. Genet.*, **32**, 7–31.
21. Le Beau, M.M., Drabkin, H., Glover, T.W., Gemmill, R., Rassool, F.V., McKeithen, T.W. and Smith, D.I. (1998) An FHIT tumour suppressor gene? *Genes Chromosomes Cancer*, **21**, 281–289.
22. Galbiati, F., Volonte, D., Engelman, J.A., Watanabe, G., Burk, R., Pestell, R.G. and Lisanti, M.P. (1998) Targeted downregulation of caveolin-1 is sufficient to drive cell transformation and hyperactivate the p42/44 MAP kinase cascade. *EMBO J.*, **17**, 6633–6648.
23. Chesi, M., Bergsagel, P.L., Shonukan, O.O., Martelli, M.L., Brents, L.A., Chen, T., Schrok, E., Ried, T. and Kuehl, W.M. (1998) Frequent dysregulation of the c-*MAF* proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. *Blood*, **91**, 4457–4463.
24. Hansen, R.S., Canfield, T.K., Field, A.D., Mumm, S., Laird, C.D. and Gartler, S.M. (1997) A variable domain of delayed replication in *FRAXA* fragile X chromosomes: X inactivation-like spread of late replication. *Proc. Natl Acad. Sci. USA*, **94**, 4587–4592.
25. Subramanian, P.S., Nelson, D.L. and Chinault, A.C. (1996) Large domains of apparent delayed replication timing associated with triplet repeat expansion at *FRAXA* and *FRAXE*. *Am. J. Hum. Genet.*, **59**, 407–416.
26. Le Beau, M.M., Rassool, F.V., Neilly, M.E., Espinosa III, R., Glover, T.W., Smith, D.I. and McKeithan, T.W (1998) Replication of a common fragile site, *FRA3B*, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum. Mol. Genet.*, **7**, 755–761.
27. Pirozzi, G., McConnell, S.J., Uveges, A.J., Carter, J.M., Sparks, A.B., Kay, B.K. and Fowlkes, D.M. (1997) Identification of novel human WW domain-containing proteins by cloning of ligand targets. *J. Biol. Chem.*, **272**, 14611–14616.
28. Latil, A., Cussenot, O., Fournier, G., Driouch, K. and Lidereau, R. (1997) Loss of heterozygosity at chromosome 16q in prostate adenocarcinoma: identification of three independent regions. *Cancer Res.*, **57**, 1058–1062.
29. Chen, T., Sahin, A. and Aldaz, C.M. (1996) Deletion mapping of chromosome 16q in ductal carcinoma in situ of the breast: refining a putative tumor suppressor gene region. *Cancer Res.*, **56**, 5605–5609.
30. Smit, A. (1996) The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.*, **6**, 743–748.