The FRA2C common fragile site maps to the borders of MYCN amplicons in neuroblastoma and is associated with gross chromosomal rearrangements in different cancers

Anne Blumrich¹, Marc Zapatka², Lena M. Brueckner¹, Diana Zheglo¹, Manfred Schwab¹ and Larissa Savelyeva¹,*

¹Department of Tumor Genetics and ²Department of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg 69121, Germany

Received September 24, 2010; Revised and Accepted January 18, 2011

INTRODUCTION

Common fragile sites (cFS) represent chromosomal regions that are prone to breakage after partial inhibition of DNA synthesis. Activation of cFS is associated with various forms of DNA instability in cancer cells, and is thought to be an initiating event in the generation of DNA damage in early-stage tumorigenesis. Only a few cFS have been fully characterized despite the growing interest in cFS instability in cancer genomes. In this study, six-color fluorescence in situ hybridization revealed that FRA2C consists of two cFS spanning 747 kb FRA2Ctel and 746 kb FRA2Ccen at 2p24.3 and 2p24.2, respectively. Both cFS are separated by a 2.8 Mb non-fragile region containing MYCN. Fine-tiling array comparative genomic hybridization of MYCN amplicons from neuroblastoma (NB) cell lines and primary tumors revealed that 56.5% of the amplicons cluster in FRA2C. MYCN amplicons are either organized as double minutes or as homogeneously stained regions in addition to the single copy of MYCN retained at 2p24. We suggest that MYCN amplicons arise from extra replication rounds of unbroken DNA secondary structures that accumulate at FRA2C. This hypothesis implicates cFS in high-level gene amplification in cancer cells. Complex genomic rearrangements, including deletions, duplications and translocations, which originate from double-strand breaks, were detected at FRA2C in different cancers. These data propose a dual role for cFS in the generation of gross chromosomal rearrangements either after DNA breakage or by inducing extra replication rounds, and provide new insights into the highly recombinogenic nature of cFS in the human cancer genome.

*To whom correspondence should be addressed. Tel: +49 6221423274; Fax: +49 6221423277; Email: l.savelyeva@dkfz.de

© The Author 2011. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
be common features of cFS (12,13). A genetic assay in yeast showed that an AT-rich region of ~500 bp, which spans the peak of highest flexibility in FRA16D and contains a polymorphically perfect AT repeat, may stall replication fork progression and increase chromosome fragility in an AT length-dependent manner (14). It has been suggested that AT stretches lead to an accumulation of DNA secondary structures that may cause double-strand breaks (DSB) at a fragile site. However, such sequence motifs are also widely dispersed in non-fragile regions (15), indicating that a combination of DNA characteristics rather than a single consensus sequence accounts for the increased genomic instability at cFS.

The cell cycle checkpoint protein ATR, and its downstream targets (16) as well as key components of the homologous recombination (HR) and non-homologous end-joining (NHEJ) pathways (17) are essential to maintain cFS stability. Replication of fragile site regions is delayed (18), and it is hypothesized that their breaks arise from regions of unreplicated single-stranded DNA accumulated at the stalled or collapsed replication forks (1).

cFS have been associated with genomic instability during tumorigenesis (19–22). Expression of FRA3B, FRA6E and FRA16D results predominantly in deletions of the tumor suppressor genes FHIT (23,24), PARK2 (25,26) and WWOX (6,27), respectively. Allelic imbalances at fragile site genes are associated with early-stage cancer development, and have been detected within many types of tumors (28,29).

Recently, large-scale screening of 746 cancer cell lines revealed that at least half of the homozygous deletions attributed to the likely underlying cause occurred in fragile site regions (30).

Amplification is one of the most frequent genomic alterations in cancer. Amplified genes represent a major driving force for deregulation of cellular growth, and promoting tumorigenesis (31). Gene amplification was initially described in drug-resistant rodent cell lines (32,33). Analyses of amplicon structures indicated that both intrachromosomal and extrachromosomal mechanisms may lead to amplification (34). Intrachromosomal amplification via breakage–fusion–bridge (BFB) cycles results in inverted, amplified structures up to several megabases in length accumulated at the resident chromosomal loci (35,36). BFB cycle amplification can be initiated in vivo by fragile site activation in rodent model systems (37). In human esophageal adenocarcinoma, gastric and breast carcinoma cell lines, the amplification of MET and PIP via BFB cycles resulted in amplicons with boundaries clustering in cFS (38–40). The mechanism of extrachromosomal amplification is more poorly understood. After extrachromosomal amplification occurs, the amplified DNA can be visualized as either circular, extrachromosomal double minutes (DM) or homogeneously staining regions (HSR) located distant from the resident site. Occurrence of high-level copy number DM has been explained by the replication–excision model (32) based on chromosomal regions looping out during DNA replication. Recently, chromosome fragmentation as a consequence of replication stress was also suggested to trigger the formation of DM (41). This model connects replication stress and amplification, implying that cFS activation may be involved in high copy number gene amplification. However, whether cFS are responsible for initiating extrachromosomal amplification in cancer cells remains unknown.

The genomic location of FRA2C was determined in this study. FRA2C consists of FRA2Ctel (2p24.3) and FRA2Ccen (2p24.2), and both cFS flank a 2.8 Mb non-fragile region containing MYCN. FRA2Ctel and FRA2Ccen encompass NBS1, NT3C1B and RDH14, and both cFS are enriched in flexibility peaks compared with the non-fragile regions. Complex genomic rearrangements, including deletions, duplications and translocations, were detected in different tumors at both cFS. We showed that MYCN amplicons and gains map to FRA2Ctel or FRA2Ccen in 56.5% of neuroblastoma (NB) cell lines and primary tumors. These results raise the possibility that initiation of MYCN amplification may be caused by extra replication rounds of secondary DNA structures accumulated at FRA2Ctel and/or FRA2Ccen. These data propose a dual role for cFS in the generation of gross chromosomal rearrangements either after DNA breakage or by producing extra replication rounds, and provide new insights into the highly recombinogenic nature of cFS in the human cancer genome.

RESULTS

FRA2C is composed of two distinct cFS

FRA2C is assigned to chromosomal band 2p24 and is the most telomeric fragile site on the short arm of chromosome 2. We mapped FRA2C by inducing cFS breaks using aphidicolin, then hybridizing six BAC clones (RP11-532A9, RP11-473K18, RP11-355H10, RP11-723P04, RP11-513P11 and RP11-498022), spaced at 2–3 Mb intervals across a broad region encompassing 2p24, to metaphase spreads (Table 1). About 4% exhibited breaks in two regions at 2p24. Some displayed breaks in either one of the two regions in each chromosome 2 homolog (Fig. 1A, left panel), and some displayed two distinct breaks on the same chromosome (Fig. 1A, right panel). One region of breaks occurred more distally (closer to the telomere) on chromosome 2, between RP11-473K18 and RP11-355H10, and a second more proximally (closer to the centromere) between RP11-723P04 and RP11-513P11, where the majority of breaks (17/25) were observed. Breaks were not observed within the 2.8 Mb region between RP11-355H10 and RP11-723P04. These data demonstrate the presence of two cFS at 2p24, which we refer to as FRA2Ctel and FRA2Ccen.

Four additional BAC clones (RP11-59C11, RP11-314E10, RP11-247H16 and RP11-422A6) spaced at 300–400 kb intervals between RP11-473K18 and RP11-355H10 were used for a more detailed mapping of FRA2Ctel. RP11-59C11 and RP11-314E10 hybridized distally to all breaks, and RP11-422A6 hybridized proximally to all breaks, while RP11-247H16 was proximal to, crossed or distal to FRA2Ctel breaks. Thus, all observed breaks were located within ~800 kb between RP11-314E10 and RP11-422A6. We determined the boundaries of the FRA2Ctel sequence by examining the hybridization pattern of seven contiguous BAC clones (Fig. 1B, Table 1). Five BAC clones covered the entire FRA2Ctel region, encompassing 747 kb from RP11-526G2 to RP11-32P22.
Table 1. Breaks and gaps at FRA2Ctel and FRA2Ccen

<table>
<thead>
<tr>
<th>Clone ID</th>
<th>Location on chr 2 (bp)</th>
<th>Distal (D)</th>
<th>Within (W)</th>
<th>Proximal (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRA2Ctel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-532A9</td>
<td>12424210–12632086</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RP11-473K18</td>
<td>14262498–14237009</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RP11-59C11</td>
<td>14399003–14519890</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RP11-314E10</td>
<td>14647806–14849828</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RP11-526G2</td>
<td>14851340–14990770</td>
<td>16</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>RP11-570F4</td>
<td>14988771–15120639</td>
<td>10</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>RP11-247H16</td>
<td>15118660–15209710</td>
<td>13</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>RP11-619O15</td>
<td>15181883–15375591</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>RP11-32P22</td>
<td>15414553–15599790</td>
<td>0</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>RP11-422A6</td>
<td>15597791–15736468</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>RP11-355H10</td>
<td>16016777–16211737</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>FRA2Ccen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-723P4</td>
<td>18141279–18318288</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RP11-101M1</td>
<td>18180770–18344726</td>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RP11-405K16</td>
<td>18342727–18389240</td>
<td>29</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RP11-720L11</td>
<td>18387241–18491084</td>
<td>28</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>RP11-120J28</td>
<td>18499085–18667774</td>
<td>25</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>RP11-30H15</td>
<td>18663775–18794410</td>
<td>18</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>RP11-185BP6</td>
<td>18793411–18970221</td>
<td>5</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>RP11-59BP8</td>
<td>18970222–18976440</td>
<td>2</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>RP11-7822</td>
<td>18976461–19135869</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>RP11-806C10</td>
<td>19014101–19278609</td>
<td>0</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>RP11-101A12</td>
<td>19186962–19361711</td>
<td>0</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>RP11-42N12</td>
<td>20112988–20248857</td>
<td>0</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>RP11-513P11</td>
<td>20609311–20862075</td>
<td>0</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>RP11-498O22</td>
<td>23485119–25680421</td>
<td>0</td>
<td>0</td>
<td>60</td>
</tr>
</tbody>
</table>

aNumber of gaps/breaks occurring distal to the hybridization signal.
bNumber of gaps/breaks occurring within the hybridization signal, resulting in a split signal.
cNumber of gaps/breaks occurring proximal to the hybridization signal.

Four BAC clones (RP11-111J6, RP11-818B6, RP11-101A12 and RP11-42N12) from the region between RP11-723P04 and RP11-513P11 were used to map FRA2Ccen. Breakage always occurred between the RP11-723P04 (centromeric) and RP11-101A12 (telomeric) BAC clones, and covered a 1 Mb large region. Ten partially overlapping BAC clones were used to further resolve this region (Fig. 1B, Table 1). RP11-723P04, RP11-101M1 and RP11-405K16 were distal to all of the 29 FRA2Ccen breaks. RP11-806C1 and RP11-101A12 were always proximal to the FRA2Ccen breaks. The remaining six BAC clones were either proximal, crossed or distal to the FRA2Ccen breaks (Fig. 1B, Table 1). These six BAC probes spanned the entire genomic region of FRA2Ccen, including more than 746 kb. Overall, the examination of both the telomeric and centromeric borders of all breaks confirmed the presence of the two distinct cFS at 2p24. Breaks at FRA2Ccen were 3-fold more frequent than at FRA2Ctel.

To construct a physical and transcriptional map of the region containing both FRA2Ctel and FRA2Ccen, the location of BAC clones covering the cFS was verified by end sequencing. Alignment of the BAC contigs with the genomic sequence from public databases revealed that FRA2Ctel and FRA2Ccen are located in the 2p24.3 G-band and 2p24.2 R-band, respectively, and contain the NBAS, NT5C1B and RDH14 genes (Fig. 1C). At least 94% of the genomic NBAS sequence (synonymous NAG) maps within FRA2Ctel. NBAS consists of 52 exons spanning 395 kb, and has a predicted transcript size of 7.3 kb. FRA2Ccen contains NT5C1B, a cytosolic 5-prime nucleotidase, and RDH14, retinol dehydrogenase 14 (all-trans/9-cis/11-cis). Nine genes are located within the 2.8 Mb non-fragile region between FRA2Ctel and FRA2Ccen, including MYCN, which is amplified in 20% of all NBs. FISH mapping with contiguous BAC probes revealed that FRA2C is composed of two cFS, FRA2Ctel and FRA2Ccen, which contain the NBAS, NT5C1B and RDH14 genes. Each fragile site spans ~750 kb, and FRA2Ctel and FRA2Ccen are separated by a non-fragile DNA region of ~2.8 Mb.

FRA2Ccen and FRA2Ctel are enriched in DNA-flexibility peaks and AT repeats

The molecular basis contributing to the inherent fragility of cFS remains poorly understood. Therefore, fragile site sequences were analyzed for repeat motifs, DNA flexibility and AT stretches since these features are suggested to promote instability at cFS. The DNA sequences spanning FRA2Ctel and FRA2Ccen were analyzed using the RepeatMasker program to assess DNA repeat composition. The 747 kb FRA2Ctel sequence has a GC content of 39.48% and is composed of 43.16% repetitive elements, including LINE1 (17.85%), LINE2 (4.84%), Alu (5.86%), MIR (5.00%), LTR (9.34%) and DNA transposons (3.78%) (Table 2). FRA2Ccen contains a similar distribution of different types of repetitive elements (Table 2). These values correspond to genomic regions having an AT content in the range of 60–62% (42). The flexibility of the FRA2Ctel and FRA2Ccen sequences was analyzed using the TwistFlex software. Both cFS are highly enriched in flexibility peaks relative to non-fragile regions, which exhibit on average 3.3 peaks/100 kb in a G-band and 1.8 peaks/100 in an R-band kb (43). In particular, FRA2Ctel is 3.18-fold more enriched (10.8 peaks per 100 kb, Table 2) in predicted flexibility peaks even relative to other characterized cFS sequences, which on average hold 3.4 peaks/100 kb in an R-band. For FRA2Ctel, situated in a G-band, has a predicted flexibility of 6.16 peaks per 100 kb. These data are in line with the observation that enrichment of DNA flexibility is characteristic of cFS. We analyzed FRA2Ctel and FRA2Ccen for (AT/TA)20 perfect repeats using the Perfect Microsatellite Repeat Finder software. A perfect (AT/TA)20 repeat was seen in FRA2Ccen. This perfect AT stretch is located in the maximum DNA-flexibility peak having 17.6 repeat of twist in FRA2Ccen, which is reminiscent with previous findings for FRA16D (14). In contrast, FRA2Ctel harbored no expanded perfect (AT/TA)20 repeats, demonstrating that long stretches of perfect AT repeats are not a ubiquitous characteristic of cFS sequences. Taken together, sequence analyses of the 1.5 Mb region holding both cFS showed that both parts of FRA2C are AT-rich and abundant in peaks of enhanced DNA flexibility.

FRA2Ccen and FRA2Ctel are targeted by genomic rearrangements in different tumors

To ascertain whether FRA2Ctel or FRA2Ccen are generally associated with genomic rearrangements in breast
and colon cancer cell lines as well as primary NBs without MYCN amplification were examined using array comparative genomic hybridization (CGH) and FISH. Genomic gains of up to a 6-fold copy number change were detected in four of seven NB cell lines. CHLA-90, LA-N-6, GI-M-EN and NB-69 displayed large gains of ~30–40 Mb on the short arm of chromosome 2. Figure 1. FRA2C consists of the two fragile sites FRA2Ctel and FRA2Ccen. (A) Breaks at distal 2p occur within two different chromosomal regions. Each of the chromosome homologues displayed breaks in either one of the two regions on distal 2p (left panel) or two distinct breaks occurred simultaneously on the same chromosome (right panel). Arrows indicate breaks in FRA2Ctel and arrowheads in FRA2Ccen. BAC probes are represented by RP11-473K8 (green), RP11-3a5H10 (red) and RP11-513P11 (yellow). (B) FISH-mapping determined borders of FRA2Ctel and FRA2Ccen on metaphases of aphidicolin-treated human lymphocytes. The ideogram of the short arm of chromosome 2 with the relative position of FRA2C is shown. FISH signals of contiguous BAC clones are shown for FRA2Ctel (left) and FRA2Ccen (right) on 2p. BAC clones displayed crossing signals in different metaphases (black arrows below), indicating that these probes span the regions of breakage at 2p24. The BAC probes RP11-314E10 and RP11-405K16 produced consistently signals distal to the breaks at FRA2Ctel and FRA2Ccen, respectively. RP11-422A6 and RP11-806C10 showed signals exclusively proximal to the breaks at FRA2Ctel and FRA2Ccen, respectively. Coloured bars above the panels indicate the relative position of the DNA probes and dashed boxes indicate their overlapping regions. A summary of all FISH-mapping results is shown in Table 1. (C) Genetic and physical map of the entire FRA2C region. FRA2Ccen and FRA2Ctel are located on the chromosomal bands 2p24.3 and 2p24.2, spanning 746 and 747 kb, respectively. FRA2Ctel consists of the 394 kb cFS gene NBAS and FRA2Ccen encompasses RH14 and NT5C1B. Both fragile regions flank the 2.8 Mb non-fragile region containing MYCN. Names of BAC probes showing a break or gap within the hybridization signal are bold. All data in this figure are illustrated according to the NCBI build 36.1/March 2006.
Multiple chromosomal alterations present in breast cancer cell lines (Fig. 2C and D). The chromosomal alterations include duplications, insertions, deletions, and translocations clustered in either the proximal or distal portion of 2p. For example, a gain was detected within the large duplicated region in CHLA-90 (RP11-459K11 and G248P8046H2), which is centromerically to its centromeric border. Such rearrangements, including deletions, duplications, and translocations, were associated with the chromosomal regions of FRA2Ctel and FRA2Ccen in more than 30% of examined cell lines derived from different cancer entities. The chromosomal alterations present in FRA2Ctel and FRA2Ccen are typical those known to originate from faulty DSB repair in cancer cells, and are characteristic of the instability at cFS.

### Table 2. In silico analyses for FRA2Ctel and FRA2Ccen

<table>
<thead>
<tr>
<th>RepeatMasker</th>
<th>FRA2Ctel Percentage of sequence (%)</th>
<th>FRA2Ccen Percentage of sequence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINEs</td>
<td>10.86</td>
<td>7.90</td>
</tr>
<tr>
<td>ALUs</td>
<td>5.86</td>
<td>3.79</td>
</tr>
<tr>
<td>MIRs</td>
<td>5.00</td>
<td>4.06</td>
</tr>
<tr>
<td>LINEs</td>
<td>23.09</td>
<td>22.98</td>
</tr>
<tr>
<td>LINE1</td>
<td>17.85</td>
<td>18.67</td>
</tr>
<tr>
<td>LINE2</td>
<td>4.84</td>
<td>4.21</td>
</tr>
<tr>
<td>LTR elements</td>
<td>9.34</td>
<td>8.89</td>
</tr>
<tr>
<td>DNA elements</td>
<td>3.87</td>
<td>3.37</td>
</tr>
<tr>
<td>Unclassified</td>
<td>0.26</td>
<td>0.03</td>
</tr>
<tr>
<td>Total interspersed repeats</td>
<td>47.42</td>
<td>43.16</td>
</tr>
<tr>
<td>Small RNA</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>Satellites</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Low complexity</td>
<td>0.44</td>
<td>0.56</td>
</tr>
<tr>
<td>Simple repeats</td>
<td>0.71</td>
<td>1.22</td>
</tr>
<tr>
<td>GC level</td>
<td>39.48</td>
<td>38.54</td>
</tr>
<tr>
<td>Number of unit (N)</td>
<td>Number of unit (N)</td>
<td></td>
</tr>
<tr>
<td>Microsatellite repeats</td>
<td>AT</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>TA</td>
<td>0</td>
</tr>
<tr>
<td>TwistFlex</td>
<td>Flexibility peaks/ 100 kb</td>
<td>6.16</td>
</tr>
</tbody>
</table>

DNA sequence was downloaded from the Ensemble database (hg18) and analyzed using the RepeatMasker-, Perfect Microsatellite Repeat Finder- and TwistFlex-software.

FRA2Ctel and FRA2Ccen map to the boundaries of MYCN amplicons

The genomic location of MYCN raises the hypothesis that MYCN amplicons may depart from recombinogenic events at FRA2Ctel and FRA2Ccen. To determine a possible relationship between amplified MYCN and the FRA2Ctel and FRA2Ccen sites, MYCN amplicon borders were defined using high resolution array CGH for the entire chromosome 2 as well as the 2p arm (average resolution of 600 and 200 bp, respectively). Genomic DNA from 15 MYCN-amplified NB cell lines and 9 MYCN-amplified localized primary NB tumors were hybridized to the array. Additionally, we assessed the genes and pseudogenes co-amplified in the amplicons. The copy number rearrangements, including amplifications, gains, and deletions on chromosome 2 are shown in Figure 3A and in Supplementary Material, Figure S2. We also detected other sequences amplified at or in the vicinity of FRA2Ctel and FRA2Ccen, which did not include MYCN. Non-contiguous and contiguous MYCN amplicons ranged between 0.1 and 2.8 Mb (Fig. 3B, blue bars). The smallest contiguous MYCN amplicons were detected in the CHP-126 cell line and NB-S498 tumor (100 and 252 kb, respectively), while the largest were 1570 and 1368 kb in the LA-N-5 cell line and NB-S413 tumor, respectively. MYCN maintained a centric position within 29% of the amplicons from both NB cell lines and tumors, and was localized telomERICally in 54% and centromERICally in 17%. Gains spanning 116 kb to 31.38 Mb in size were detected in eight NB cell lines and five primary NBs on the 2p arm (Fig. 3B, red bars). Approximately half of all MYCN amplicons detected also included NBAS in its entirety or in part. The DX1 and FAM49A genes were co-amplified with MYCN in 54.2 and 33.3% of the cases, respectively. In ~85% of the cases, the pseudogenes, USE- snRNA, U13-like scRNA, SRP RNA- and ACA 40-like both snoRNA, were included in the genomic amplification or gain (Supplementary Material, Table S1).

The proximal or distal boundaries of amplicons or gains clustered in either FRA2Ccen or FRA2Ctel or in both cFS in ~80% of NB cell lines and 33% of NB primary tumors (Fig. 3B). On the whole, 41.6% of all amplicon boundaries represent a proximal or distal rearrangement border mapping to FRA2Ctel or FRA2Ccen. The remaining borders of the amplicons or gains were near to the cFS, mapping within 332.4 and 394.5 kb proximal to FRA2Ctel in NB cell lines and primary tumors, respectively. Amplicon borders or gains also clustered within FRA2Ccen in 25% of NB cell lines and tumors and within FRA2Ctel in 58%. Most amplicon borders were located in the region spanning 15.5–16.5 Mb on 2p in NB cell lines and NB primary tumors (Fig. 3C). In total, 52 and 21 MYCN amplicon borders were located in the 14.0–20.0 Mb region on 2p in NB cell lines and primary tumors, respectively. These results may reflect a high frequency of recombinogenic events at replication forks arising during perturbations of DNA replication in FRA2C.
Different architectures of DM exist in NB cell lines

MYCN amplicons and gains were investigated using FISH and array CGH to gain insight into the architecture of genomic rearrangements creating DM involving FRA2Ctel and FRA2Ccen. DM are the predominant amplicon structure in primary NB. The NB cell lines, Vi-856 and TR-14, were...
Figure 3. MYCN amplicon borders cluster in FRA2Ctel in NB cell lines and primary tumors. (A) Fine-tilling array CGH determined MYCN amplicons represented as an increased log2 ratio (y-axis) for CHP-126 and LA-N-2. The x-axis shows the genomic location of the amplicon or gain on 2p including FRA2Ctel and FRA2Ccen borders. (B) Amplicons are illustrated as blue bars on 2p for NB cell lines and primary tumors. Chromosomal gains are indicated in red bars, losses in green bars and no alteration is left blank. Analyses were performed for 15 NB cell lines containing MYCN amplification either as DM or HSR and nine MYCN amplified primary tumors. Amplicons appeared either continuous or discontinuous. (C) A schematic overview shows DSB for gains and amplicon borders located on 2p25.1 to 2p24.1 in NB cell lines and NB primary tumors. Breakpoint intervals for gains are indicated as red bars and borders for amplified regions as blue bars for each individual cell line and tumor. All data in this figure are illustrated according to the NCBI build 36.1/March 2006.
chosen because they harbor MYCN amplicons as DM. Array CGH revealed complex genomic rearrangements in a 20 Mb region from the telomere on 2p in Vi-856 and TR-14, which comprised gains and amplicons (Fig. 4A). In Vi-856, four distinct regions of high-level amplifications (≏100 copies) and three regions of gain were detected. The region spanning 16.31–16.88 Mb showed an intermediate copy number state of a 30-fold gain. TR-14 harbors six highly amplified regions (50–150 copies) within the 20 Mb from the telomere on 2p. The most telomeric amplicon spanned 2.27–2.38 Mb, and the most centromeric spanned 16.98–17.98 Mb on 2p. FISH was performed for both cell lines using the corresponding DNA probes for each genomic region exhibiting a copy number change by array CGH (Fig. 4B). Gains in Vi-856 were visualized by hybridizing five BAC probes, detecting three different levels of copy number changes. FISH detected seven copies of RP11-342E23 (I, Fig. 4B) and RP11-299D11, eight copies of RP11-314E10 (II, Fig. 4B) and RP11-102G8 and nine copies of RP11-120J4 (III, Fig. 4B). The 7-fold amplification was the lowest level of gain detected on 2p for the Vi-856 cell line. The sequences were organized as one normal copy on 2p, two derivative chromosomes carried a 60 Mb duplication and two other derivative chromosomes harbored portions of 2p (I, Fig. 4B). The RP11-314E10, RP11-102G8 and RP11-120J4 probes also detected two additional copies located on other chromosomes (II and III, Fig. 4B). The intermediately amplified region (16.31–16.88 Mb) was integrated on other derivative chromosomes and existed as exogenous DM (RP11-65N17, VIII, Fig. 4B). The intermediate amplification of this region can be explained by the occurrence of this sequence only in a portion of MYCN-DM (IX, Fig. 4B) present in ≏30% of the cells (X, Fig. 4B). All three highly amplified fragments were present as exogenous amplified regions (RP11-98I18, RP11-422A6 and RP11-355H10, IV–VI, Fig. 4B) and they co-existed on one DM structure (VII, Fig. 4B).

FISH of TR-14 revealed that three apparently normal copies of chromosome 2 and several different architectures of DM co-exist in one metaphase. The MYCN-amplified region (RP11-355H10) co-existed with three other amplified regions (RP11-352J11, RP11-1293J14 and RP11-98I18) on one DM (indicated by arrows, I–III, Fig. 4B), but also formed exogenous smaller DM (indicated by arrowheads, I, Fig. 4B). Both DM structures harboring MYCN were detected in 100% of all metaphases. The region spanning 10.22–10.78 Mb was less amplified (with a log2 ratio of 2.46) than the MYCN-containing DM, and represents a third DM present in 30% of the cells analyzed (IV, Fig. 4B), in which, 70% harbored a diploid status of this region (V, Fig. 4B). The same situation was detected for the region spanning 16.98–17.98 Mb, but with a DM incidence of 50% in all cells (RP11-41904, data not shown). These results indicate that the instability of FRA2Ctel and FRA2Ccen may lead to two different types of genomic rearrangements. DSB within or in close proximity to the cFS could cause the formation of additional copies of the distal portion of 2p and their translocation to other chromosomes. Independent amplicon formation also frequently originates at FRA2C and its neighboring sequences. This indicates that cFS are prone to form circular DNA structures, which could later exist as exogenous DM in the cell. Moreover, the lack of deletion of the amplified sequences at their original location on 2p indicates that fragile site sequences are not only particularly unstable, but are also predisposed to form DNA structures facilitating aberrant replication of chromosome segments.

**DISCUSSION**

We identify here the two cFS, FRA2Ctel and FRA2Ccen, with their corresponding genes that map to the boundaries of 56.5% of MYCN amplicons from NBs. We further provide evidence that FRA2Ctel and FRA2Ccen are involved in generating chromosomal rearrangements in other tumors. We suggest a dual role for cFS in generating rearrangements by DNA breakage and high-level amplicons. To date, most cFS have been predicted only by G-banding without providing fine mapping of the genomic regions they involve or the genes affected by rearrangements in them. Identification of the full content of genes affected within a fragile site is an important task in cancer and genome research, since DNA damage at these chromosomal regions occurs during the early stages of tumorigenesis (28,29,45,46). In this study, we determined the genomic location of FRA2C at a BAC clone resolution level of ≏150 kb, and identifying the cFS, FRA2Ctel and FRA2Ccen, at 2p24.3 and 2p24.2, respectively. Only eight other cFS to date have had their centromeric and telomeric borders defined to this level of detail (2–10). About 1% of metaphases showed breaks in FRA2Ctel and 3% in FRA2Ccen after aphidicolin treatment. In concordance with a systematic genome-wide screen for cFS in the human genome (47), FRA2Ccen belongs to the minor group (21/230) of cFS, which acquired breaks and gaps at a frequency of ≏1%, whereas FRA2Ctel (a.k.a. FRA2M) belongs to the intermediate group (91/230) acquiring breaks at a frequency between 0.1 and 1%. Based on the cFS inter-individual frequency of breakage variation, FRA2Ctel could be classified as a low-frequency and FRA2Ccen as high-frequency site of breakage.

Sequence motifs may contribute to the inherent fragility of cFS (11). FRA2Ctel and FRA2Ccen are AT-rich, as are all other cFS described to date. Both cFS show a similar distribution of their repeat composition compared with the average repeat composition reported for the human genome in areas with an AT content of 60–62% (42). We did detect significantly higher flexibility peaks in both FRA2Ctel and FRA2Ccen, compared with non-fragile regions. FRA2Ccen is 3.18-fold more enriched in DNA flexibility peaks compared with the average value for cFS. The enriched DNA flexibility of FRA2Ccen and FRA2Ctel is in line with reports for a number of cFS also showing enrichment in sequences predicted to have high DNA flexibility (12,13,48,49). Perfect polymorphic AT repeats within AT-rich areas are proposed to facilitate the formation of strong secondary DNA structures, thus, increasing chromosomal breakage (13). Stretches of (AT/TA)20 perfect repeats have been suggested to cause the fragility of FRA16D (14). In concordance with this hypothesis, we detected an (AT/TA)23 repeat in FRA2Ccen directly located in the maximum peak of flexibility, which is similar to the situation reported for FRA16D. We detected no
Figure 4. Different architectures of DM exist in NB cell lines. (A) Gains and amplicons are detected by array CGH and represented as increased log2 ratio (y-axis) in Vi-856 and TR-14 in the region from telomere to 20 Mb on 2p (x-axis). The genomic positions of FRA2Ctel and FRA2Ccen are shown by grey shading. The positions and the colors of the arrows correspond to the locations and fluorescence dye colors of BAC-probes used for FISH in (B). Black arrows indicate the probes used for FISH, but not depicted in (B). (B) The architecture of gains and multiple DM was revealed by FISH in Vi-856 and TR-14. RP11-342E23 (I) represents a 7-fold gain, RP11-314E10 an 8-fold (II) and RP11-120J4 a 9-fold (III) in Vi-856. RP11-98I18 (IV), RP11-422A6 (V) and RP11-355H10 (VI) exposed a high-level amplification of their corresponding genomic regions visible as DM in all metaphases of Vi-856. Panel VII represents the merged image of these BAC probes showing that all three amplified regions are co-localized on the same DM. The signal overlap is indicated by arrows in green, blue and pink. RP11-65N17 (VIII) visualizes gains of the 2p region from 16.31 to 16.88 Mb either integrated at other chromosomes (green arrows) or exogenous as DM (green arrowheads). DM with other amplified regions are indicated by blue arrowheads. The merged image of RP11-65N17 and RP11-355H10 signals (IX) showed two types of DM, at which one contained green and pink signals, and a second one containing only RP11-355H10 signals (pink). Panel X shows interphase nuclei, where RP11-65N17 is amplified only in some Vi-856 cells. FISH of TR-14 revealed the co-existence of three genomic regions originating from 2p at the same DM indicated by arrows for RP11-355H10 (I), RP11-98I18 (II) and RP11-1293J14 (III). Arrowheads in panel I indicate DM harboring only signals of the genomic MYCN region (RP11-355H10). The existence of a third type of DM is shown in the merged image in panel IV by RP11-320M2 (yellow) and RP11-355H10 (pink). Both signals are present on different DM. Panel V shows one metaphase and interphases of TR-14 hybridized with RP11-320M2, which demonstrates two cell populations: one with a diploid and a second one with an amplified status of the corresponding region. RP11-342E23, RP11-98I18 and RP11-65N17 were labeled by FITC-dUTP (green signals); RP11-120J4 and RP11-1293J14 by Cy3-dUTP (red signals); RP11-355H10 by Cy3.5-dUTP (pink signal); RP11-314E10 and RP11-320M2 by Cy5-dUTP (yellow signals); RP11-422A6 by Cy5.5 (blue signal). Metaphase chromosomes were counterstained with DAPI. All data in this figure are illustrated according to the NCBI build 36.1/March 2006.
(AT/TA)_{>20} perfect repeats in FRA2Ctel. Thus, the contribution of sequence to fragility may be more complex than only the presence of specific motifs. Our results emphasize the importance of combining sequence analysis with experimental assays to assess the roles of cFS in tumorigenesis.

Common fragile sites frequently exhibit rearrangements in tumors. Homozygous and heterozygous deletions spanning several kilobases are prevalent at these regions (1,16,50,51). These rearrangements are suggested to either result from DSB incorrectly repaired by HR or NHEJ (17), or the rearrangement process at cFS is mediated by microheterogeneities (50). As fragile site activation is thought to be an early event during tumorigenesis (51), several studies aim to correlate the localization of cFS with the borders of genomic rearrangements and breaks in cancer cells (52). Previously, the formation of the cancer-specific RET/PTC1 translocation has been linked to activated FRA10C and FRA10G in papillary thyroid carcinoma (53). Here we report that rearrangements are present at FRA2Ctel and FRA2Ccen in different tumors. We analyzed unbalanced genomic rearrangements in NBs without MYCN amplification, and from breast and colon carcinoma cell lines, and report that almost 40% of samples harbored gains and losses within or near FRA2Ctel and FRA2Ccen. Rearrangements ranged from a deletion/duplication of a single contiguous DNA segment to complex rearrangements with up to 14 breakpoints, and FISH analysis revealed inversions and balanced translocations at both cFS. Lesions at FRA2Ctel and FRA2Ccen indicate that cells were exposed to replication stress during tumorigenesis.

We show here that FRA2Ctel and FRA2Ccen span three genes, which can be affected during rearrangements. Approximately 20 large genes have been localized within cFS regions to date, ranging from 370 kb large CNTLN in FRA19G (8) to the 2 Mb DMD in FRA1C (54). Whether the co-localization of large genes in cFS is due to their AT-rich intrinsic sequences or whether it reflects common functional aspects, such as signaling in the cellular stress response (55), remains to be investigated. In line with the observation that most cFS include large genes, FRA2Ctel contains NBAS, a novel member of large cFS gene group. It has been suggested that NBAS links p31 with Zw10-RINT-1, and is involved in protein transport from the Golgi apparatus to the endoplasmic reticulum (56). NBAS dysfunction or haploinsufficiency caused by DNA damage at FRA2Ctel could contribute to tumorigenesis, although this remains to be elucidated. NBAS overexpression via MYCN co-amplification may perturb the quantitative balance of complexes involved in cell cycle-related events and membrane trafficking (56). This would be a rare event, since we detected only a complete co-amplification of NBAS in 2/24 samples. We show that RDH14 and NT5C1B localize to FRA2Ccen. These genes encode proteins belonging to large gene families, none of which as yet implicated in human cancers. RDH14 is a p53 transcriptional target (57), which could implicate it as a surrogate way to de-regulate p53 tumor suppressor function via chromosomal alterations accompanying MYCN amplification. Rearrangements at FRA2C resulting in copy number changes of NBAS, RDH14 and NT5C1B may create a favorable situation for cancer progression.

High-level gene amplification is one of the most frequent genomic rearrangements in tumors, enhancing protein expression that can lead to the deregulation of cell proliferation and promoting tumorigenesis (31,58,59). MYCN amplification in NBs results in a 50- to 100-fold copy number change (60–62). The amplicons are predominantly organized as extrachromosomal DNA visualized as DM. DM may also integrate at other chromosomal sites and undergo several cycles of in situ amplification, resulting in tandem-arranged unit-sized HSR amplicons. These MYCN-amplicon features demonstrate that amplification of MYCN is BFB-independent. Hence, high copy number amplification of MYCN may be explained, for instance, by the replication excision model (63) or chromosome fragmentation (41). Neither of these models have been examined in vitro or in vivo or connected to an activated fragile site to date. Here, we provide evidence that MYCN amplicons may originate from genomic rearrangements triggered by recombination events at FRA2Ctel or FRA2Ccen. MYCN amplicons and gains mapped to FRA2Ctel, FRA2Ccen or both in ~80% of NB cell lines and 33% of NB primary tumors. It is possible that initiation of MYCN amplification is a consequence of perturbed replication at FRA2Ctel or FRA2Ccen, since especially FRA2Ctel maps to the boundaries of MYCN amplicons in the majority of all NB samples. Our report is the first, to our knowledge, to link sites of high-level amplification to cFS in cancer. Though chromosomal fragmentation is able to connect amplification to cFS activation, the replication–excision model is better suited to explain MYCN amplification, since amplification occurs without a deletion at the resident chromosomal loci in the vast majority of NB tumor cells (44). Genes including RDH14 and NT5C1B appear not to provide strong selective advantages, possibly explaining why amplicons derived from this region occur less often in NB cells. Some NB cell lines and tumors exhibited discontinuously amplified sequences on 2p, most probably caused by independent amplification events occurring in or near FRA2Ctel and FRA2Ccen, and indicating that DM structures form recurrently and continuously at FRA2Ctel and FRA2Ccen, reflecting that its sequences are especially prone to form circular DNA. These large DM structures harboring non-syntenic genomic regions were also detected using FISH, which may be fusions of structures produced by the initial amplification of single regions. Thus, cFS sequences may trigger high-level amplification via the extensive looping out of the DNA and formation of self-replicating circular structures even without DNA damage. This new role for cFS differs from their formerly known role in amplification via BFB cycles after DSB occurrence. The cytogenetic position of other cFS relative to specific amplicons linked to certain human cancers further support the hypothesis that cFS play a role in extrachromosomal amplification of at least some genes during tumor progression. For instance, the amplification of MYC in variant forms of small-cell lung cancer and ERBB2 in 25–30% of mammary carcinomas may be amplified via activation of FRA8C and FRA17B, respectively. Nevertheless, final support of this hypothesis requires detailed genomic localization of FRA8C and FRA17B, including their borders.

The organization and architecture of FRA2Ctel and FRA2Ccen rearrangements resulting in gains and amplicons
in NB cells were further elucidated using FISH in this report. Most gains reflected an accumulation of additional copies of chromosomal regions several kilobases large and encompassing MYCN, which were often translocated to other chromosomes. Duplications and gains are thought to originate from DSB, and the breakpoints were often localized to FRA2Ctel or FRA2Ccen in our samples. These results indicate that activation of a fragile site driving gains and duplications may be due to the occurrence of DSB at a stalled or collapsed replication fork. FISH of MYCN amplicons in tumor cells indicated that gains may serve as precursors for high-level amplification. We detected different copy number levels in Vi-856 cells that probably represent the progression from low- to high-level rearrangements.

Taken together, the data we present in this report suggest that MYCN amplification, duplications and deletions are attributed to the highly recombinogenic nature of FRA2Ctel and FRA2Ccen in the human cancer genome. On the one hand, FRA2Ctel and FRA2Ccen may drive gains and deletions by DNA breakage, and on the other hand, they probably induce amplification via extra replication rounds of unbroken DNA, which are facilitated by the extensive looping out of the secondary DNA structures. We suggest the latter serves as a common mechanism at all cFS, as it does in generating gains and deletions.

**MATERIALS AND METHODS**

**Fragile site induction**

Lymphocytes were isolated from three healthy individuals, and transformed with Epstein-Barr virus to establish lymphoblastoid cell lines. The cells were grown in the RPMI-1640 medium supplemented with 10% fetal bovine serum and antibiotics. cFS were induced by 0.4 μM aphidicolin for 24–28 h before harvesting. Routine cytogenetic techniques were applied to harvest cells and prepare chromosomes.

**Cell lines, tumor samples and DNA isolation**

The 22 NB cell lines, GI-ME-N, IMR-32, HD-N-16, Kelly, NPA, LA-N-1, LA-N-2, LA-N-5, LA-N-6, Vi-856, LS, TR14, NBS-124, SIMA, SMS-KCNR, CHP-126, SK-N-BE(2)c, NBL-S, CHLA-90, SK-N-FI and SK-N-AS were grown in the RPMI-1640 medium supplemented with 10% fetal bovine serum. CHP-126 and SIMA were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ), and the remaining cell lines were obtained from different institutes where the cell lines were established. Nine localized primary tumors, collected by the national multicenter trial, the German Neuroblastoma Study of the German Children’s Hospital, Department of Pediatric Oncology and Hematology, University of Cologne. Tumor samples were chosen with amplified MYCN and a minimum of 90% tumor content. None of the patients had undergone treatment prior to tumor biopsy. According to the criteria of the International Neuroblastoma Staging System (INSS), three tumors were classified as stage 3 and three as stage 4. Three samples are still unclassified. Nine breast cancer cell lines (MX-1, MFM-223, T47D, BT-20, BT-474, Colo-824, EFM-192B, MDA-MB-436 and SK-BR-3) and nine colon cell lines (HDC-9, HDC-54, HDC-73, HDC-75, HDC-82, HDC-87, HDC-101, HDC-142 and HDC-143) were cultured in the RPMI-1640 medium supplemented with 10% fetal bovine serum. All breast cancer cell lines were from the tumor bank of the German Cancer Research Center (DKFZ), Heidelberg. The colon cell lines were established in the Division of Tumor Genetics, DKFZ, Heidelberg (64).

**Fluorescence in situ hybridization**

Extracted BAC and fosmid DNA were used as probes for FISH experiments. The BAC clones were chosen from the Ensembl database according to their genomic position and the fosmid clone (G248P85637C4) was chosen from the UCSC genome browser database. The 50 BAC clones from the RPCI-11 library and the fosmid clone (Table 1 and Supplementary Material, Table S2) were obtained from the Children’s Hospital Oakland Research Institute (CHORI). Direct sequencing of BAC ends was carried out by GATC Biotech (Konstanz, Germany) using standard T7 and SP6 primers. Sequences were aligned with the human genome using the BLAT search at the UCSC hg18 assembly. Six-color FISH was performed using modified nucleotides. DEAC-, FITC (Molecular Probes), Cy3, Cy3.5, Cy5 and Cy5.5 (Amersham, GE Healthcare, Germany) fluorescent dyes were coupled to allylamide-dUTPs (Sigma, Deisenhofen, Germany) as described previously (65). BAC DNA was labeled with DEAC-, FITC-, Cy3-, Cy5.5-, Cy5- and Cy5.5-dUTPs by nick translation, hybridized to metaphase chromosomes, counterstained with 4,6-diamidino-2-phenylindol (DAPI) (Sigma, Munich, Germany), and analyzed on a Leica DMRA 2 microscope using Leica CW 4000 FISH software. A BAC was considered to bridge a fragile site when a hybridization signal was observed on both sides of the breakage, or if a signal appeared proximal (centromeric) in one metaphase and distal (telomeric) in a separate metaphase.

**Array CGH**

Two 385 K array CGH designs were used, representing either chromosome 2 or chromosome 2p (NimbleGen Systems, Inc.). Approximately 385 000 60-mer oligonucleotide probes were spotted on each array, yielding average resolutions of 200 kb (chr2p array) and 600 kb (chr2 array). Pooled female and male DNA isolated from human lymphocytes were used as reference samples. Sample labeling and array CGH processing was performed according to the manufacturer’s protocol (NimbleGen Arrays User’s Guide, CGH Analysis, version 3.1). The labeling efficiency of test and reference products was tested by measuring Cy3 and Cy5 quality on the ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The minimum was set to 8.0 pmol/μl for each dye. Slides were scanned at a 5 μm resolution using an Axon GenePix 4000B scanner (Axon Instruments, Molecular Devices Corp., Sunnyvale, CA, USA). Array CGH images were
processed using Roche NimbleGen NimbleScan software (version2.4). Data were analyzed using the R statistical software packages and the Ringo package for array CGH analysis provided via the Bioconductor project. Raw data files from the 2 and 2p NimbleGen arrays were imported into the Ringo library (66). The arrays were loess normalized, and segmented using GLAD (67). A threshold of a log2 ratio >0.15 in NB cell lines and log2 ratio >0.16 in NB primary tumors was chosen to be a gain. Gains in NB-S230 and NB-S498 were defined as a log2 ratio >0.35, due to the oscillating signal-to-noise ratio. Amplification was defined as a log2 ratio >1 for all array CGH analyses. All raw and normalized array CGH data are available from the iCHIP platform (Integration Center of HIfh throughPut experiments).

WEB RESOURCES

SUPPLEMENTARY MATERIAL
Supplementary material is available at HMG online.

ACKNOWLEDGEMENTS
We thank Elisa Maria Hess for technical assistance, Kathy Astrahantseff for her advice during manuscript preparation, the German Neuroblastoma Tumor Bank for providing tumor samples, the German Neuroblastoma Study Group (study chair Frank Berthold) for providing clinical data and the reference laboratories for providing molecular data.

Conflict of Interest statement. None declared.

FUNDING
This work was supported by the Helmholtz-Russia Joint Research Groups (HRJRG-006) and the Bundesministerium für Bildung und Forschung Neuroblastom Genom Forschungs Netzwerk (BMBF NGFN Plus Program 01GS0896).

REFERENCES


