Homology-mediated end-capping as a primary step of sister chromatid fusion in the breakage-fusion-bridge cycles

Michael Marotta¹, Xiongfong Chen², Takaaki Watanabe¹, Pieter W. Faber³, Scott J. Diede⁴, Stephen Tapscott⁴, Raymond Tubbs⁵, Anna Kondratova¹, Robert Stephens² and Hisashi Tanaka¹,6,*

¹Department of Molecular Genetics, Cleveland Clinic Lerner Research Institute, Cleveland, OH 44195, USA, ²Advanced Biomedical Computing Center, SAIC-Frederick, Inc., Frederick, MD 21702, USA, ³Genomics Facility, University of Chicago, Chicago, IL 60637, USA, ⁴Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, ⁵Department of Molecular Pathology, Cleveland Clinic, Cleveland, OH 44195, USA and ⁶Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH 44195, USA

Received May 13, 2013; Revised July 18, 2013; Accepted August 4, 2013

ABSTRACT

Breakage-fusion-bridge (BFB) cycle is a series of chromosome breaks and duplications that could lead to the increased copy number of a genomic segment (gene amplification). A critical step of BFB cycles leading to gene amplification is a palindromic fusion of sister chromatids following the rupture of a dicentric chromosome during mitosis. It is currently unknown how sister chromatid fusion is produced from a mitotic break. To delineate the process, we took an integrated genomic, cytogenetic and molecular approach for the recurrent MCL1 amplicon at chromosome 1 in human tumor cells. A newly developed next-generation sequencing-based approach identified a cluster of palindromic fusions within the amplicon at ~50-kb intervals, indicating a series of breaks and fusions by BFB cycles. The physical location of the amplicon (at the end of a broken chromosome) further indicated BFB cycles as underlying processes. Three palindromic fusions were mediated by the homologies between two nearby inverted Alu repeats, whereas the other two fusions exhibited microhomology-mediated events. Such breakpoint sequences indicate that homology-mediated fold-back capping of broken ends followed by DNA replication is an underlying mechanism of sister chromatid fusion. Our results elucidate nucleotide-level events during BFB cycles and end processing for naturally occurring mitotic breaks.

INTRODUCTION

Gene amplification, a selective copy-number increase of genomic segments through DNA rearrangements, is a clinically important form of genome instability in cancer, as gene amplification causes advanced tumors and acquired therapy resistance (1–4). Thus, a better understanding of the underlying mechanisms of gene amplification could improve prognosis of cancer patients. Cytogenetically, amplified genomic segments reside either in small chromosomes (double minute chromosomes, DM) or in intrachromosomal, homogenously staining regions (HSR) (5–9). Continuous DNA breaks and rearrangements through breakage-fusion-bridge (BFB) cycles have been implicated as an underlying mechanism for intrachromosomal gene amplification (10–12).

The BFB cycle was originally described by Barbara McClintock in 1939 as a fate of a dicentric chromosome during meiotic mitosis and endosperm development in maize (13). She observed the following: (i) ‘breakage’ of a dicentric chromosome in meiotic anaphase when the two centromeres pass to opposite poles, (ii) ‘fusion’ at the breakage site between two sister halves of the broken chromatid resulting in a duplicated chromatid with two centromeres, and (iii) the formation of a chromatid ‘bridge’ in the following mitotic anaphase. The bridge eventually ruptures, and broken chromatids enter into each daughter nuclei. Because the rupture can occur at any site between the two centromeres, the broken chromatids can inherit unequal amounts of genetic material: a partial inverted (palindromic) duplication in one chromatid and a partial deletion in the other (Figure 1a). The resulting broken chromatids repeat the cycle in following mitotic divisions, and, as a result, a segment between two

*To whom correspondence should be addressed. Tel: +1 216 444 9107; Fax: +1 216 444 0512; Email: tanakah@ccf.org

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Three events that can create a dicentric chromosome are listed on the top (see in the ‘Introduction’ section of the main text). Following a chromosome rupture and DNA replication, a broken chromatid undergoes inverted duplication (dotted arrows) with a fusion at the end (sister chromatid fusion), which results in the formation of a dicentric chromosome. Repeated occurrence of the cycle leads to the unequal distribution of chromosomal regions (rectangle). (b) Two models for the nucleotide-level mechanisms of sister chromatid fusion: end processing-capping-replication model (left) and NHEJ-dependent model (right).
web-based platform GALAXY and FASTQ files (Supplementary Figure S1) (32). Briefly, Bowtie (33) was used to map the reads in hg19. We used the commonly used settings except for allowing for a 3bp mismatch penalty for the 75 bp reads (34). The resulting SAM files were converted to BAM files using SAM Tools (Sequence Read Archive Accession No. SRP028202). Using BED Tools (35), we intersected the BAM alignments with a BED file containing coordinates for Segmental Duplications, Repeat Masked Data, Alpha Satellite DNA or Alu elements. These files were exported from the UCSC Genome Browser. For the analyses of the subsets of Repeat-Masked data, we custom sorted the data to filter out both Alpha Satellite DNA (Alpha/ALR) and Alu elements.

For copy number analyses, 1 µg of genomic DNA was processed by the Genome-Wide Human SNP Nsp/Sty 5.0/6.0 Assay kit (Affymetrix) and hybridized onto Affymetrix Human SNP6.0 arrays. Comparative data analysis and visualization was done using the Integrated Genome Browser (Affymetrix).

**Fluorescence in situ hybridization**

Semi-confluent cells in 15 cm dish were collected at 6 h after treatment with 100 ng/ml Colcemid (Sigma). The samples were fixed, denatured and spread as described previously (36). For probes, BAC clones were prepared using BACMAX DNA Purification Kit (Epicentre). The DNA was labeled with Biotin- or DIG-Nick Translation Mix (Roche) and hybridized onto metaphase samples at 37°C with human CotI DNA (Invitrogen) and salmon sperm DNA. After hybridization, the biotin- or digoxigenin-labeled DNA probe was visualized by the Alexa488-conjugated streptavidin (Invitrogen) and biotinylated anti-streptavidin antibody (Vector) or by Anti-Digoxigenin-Rhodamine (Roche), respectively. The samples were counterstained with ProLong Gold Antifade Reagent with DAPI (Invitrogen).

**Breakpoint sequences**

The breakpoint sequence for each palindromic junction was obtained using long-PCR and Sanger sequencing (Primer list in Supplementary Table S2). For long-PCR, genomic DNA was amplified using Phusion Hot Start Pol II (Thermo Scientific) and cloned into pSC-B (Stratagene) for sequencing. To confirm the breakpoint, genomic DNA was amplified using independent primer sets designed for each breakpoint.

**RESULTS**

**A genomic approach for identifying the locations of sister chromatid fusions**

To gain insight into the nucleotide-level mechanism of sister chromatid fusion, we developed an NGS-based approach that pinpoints the location of de novo palindromic junctions in the human genome. The NGS libraries were made through the sequential denaturation and rapid renaturation of genomic DNA (GAPF) (31,37) (Figure 2a). Briefly, intra-strand self complementarities (black and dotted lines) at palindromic junctions facilitate the in vitro folding-back of denatured, ssDNA molecules and produce double-stranded DNA (dsDNA) molecules. In contrast, non-palindromic DNA does not have such self-complementarities and remains as ssDNA molecules. The digestion of resulting DNA with ssDNA-specific nuclease S1 (left) can result in the steady increase in read depth (per 1 kb bin) toward a palindromic junction (right). The numbers indicate 1 kb bins on the reference genome. (d) The numbers of GAPF-reads/kb in Colo320DM (closed circles) and IMR90 (open circles) are shown for the 50 kb region of chromosome 1 where the CTSK gene is located. Two genomic features (GC content and SINE elements) within the genomic region are also shown.

![Figure 2.](https://academic.oup.com/nar/article-abstract/41/21/9732/1277663)
linker-specific primer (GAPF-seq). NGS libraries were obtained for a colorectal cancer cell line Colo320DM (Colo) and normal fibroblast IMR90. We previously showed that a palindromic junction within the CTSK gene sets the telomeric boundary of the 1q21 amplicon in Colo (30). Because this region contains the anti-apoptosis gene MCL1 and is amplified in 10% of all primary tumors (38), understanding the mechanisms of amplification has significance. To confirm that palindromic junctions were enriched in the libraries before NGS, PCR-amplified fragments were hybridized with a probe within the palindromic junction (CTSK cen) and outside of the junction (CTSK tel) (Figure 2b). Fragments with high signal intensity were only seen by the probe within the palindromic junction, thus confirming the enrichment.

Sequencing of the libraries (75-base single end sequencing) yielded 10 579 579 GAPF reads from Colo and 7 448 354 GAPF reads from IMR90 DNA that were mapped to the reference human genome (hg19) (Table 1). Consistent with the fact that our DNA processing steps eliminates non-duplicated DNA, we found a relative depletion of uniquely mapped sequences in the libraries. The intersections between GAPF reads and unique (non-repeat masked) sequences showed that only 26.52% of GAPF reads from Colo and 23.87% of GAPF reads from IMR90 are mapped uniquely to hg19. Considering that nearly 50% of the human genome is unique (49.37% in hg19), the efficiency of depletion is >50%. Instead, repeated sequences are enriched in GAPF reads. For example, we found an enrichment in both alpha satellite DNA (7.33-fold in Colo and 4.39-fold in IMR90) and SINE elements (Alu) (4.08-fold in Colo and 4.74-fold in IMR90). Segmental duplications (low-copy repeats), the recently duplicated genomic segments (39) that account for ~ 4.9% of hg19, were moderately enriched (1.4-fold in Colo and 1.5-fold in IMR90).

We then confirmed the enrichment of reads at the junction at CTSK in Colo and identified a unique pattern of read depth (the number of reads for each 1 kb bin) at and near palindromic junctions. Denatured DNA molecules should fold back more efficiently in the regions close to the middle of palindromic junctions than regions distant from the junctions. This should lead to a steady increase in the number of GAPF reads toward the palindromic junctions and a sharp drop beyond the junction (Figure 2c). To test this idea, read depth analysis was conducted for the entire 1.1 megabase amplicon at 1q21 (Chr 1:149 700 001–150 800 000) (Supplementary Table S1) (Figure 3a). On average, there were 7.98 reads per kb in the library from Colo DNA and 1.64 reads per kb from IMR90 DNA. Indeed, 1 kb bins with markedly increased reads were located within CTSK in Colo, but not in IMR90 (Figure 2d). At a palindromic junction that colocalizes with the boundary of the amplicon (30), Colo had > 180 reads per bin for three consecutive bins (150 771 001–150 774 000), whereas IMR90 had only 0–11 reads in the corresponding bins. Furthermore, we found a steady increase of reads toward these three bins from the centromeric side, which was followed by a sharp drop. These results demonstrate the enrichment and unique pattern of read depth at palindromic junctions.

### Table 1. Enrichment of repeated (duplicated) sequences in the GAPF-seq libraries

<table>
<thead>
<tr>
<th></th>
<th>hg19</th>
<th>GAPF-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GAPF-seq Reads</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total reads</td>
<td>17 498 356</td>
<td>15 351 029</td>
</tr>
<tr>
<td>Mapped reads</td>
<td>10 579 579</td>
<td>7 448 354</td>
</tr>
<tr>
<td>Intersected with segmental dup (# of Reads)</td>
<td>737 758</td>
<td>545 540</td>
</tr>
<tr>
<td>% of Mapped</td>
<td>4.90</td>
<td>7.32</td>
</tr>
<tr>
<td>Enrichment</td>
<td>1.42</td>
<td>1.50</td>
</tr>
<tr>
<td>Intersected with repeat masker (# of reads)</td>
<td>7 773 451</td>
<td>5 670 447</td>
</tr>
<tr>
<td>% of Mapped</td>
<td>50.63</td>
<td>76.13</td>
</tr>
<tr>
<td>Enrichment</td>
<td>1.45</td>
<td>1.50</td>
</tr>
<tr>
<td>Intersected with Alu (# of reads)</td>
<td>4 196 090</td>
<td>3 429 714</td>
</tr>
<tr>
<td>% of Mapped</td>
<td>9.72</td>
<td>46.05</td>
</tr>
<tr>
<td>Enrichment</td>
<td>4.08</td>
<td>4.74</td>
</tr>
<tr>
<td>Intersected with alpha satellite (# of reads)</td>
<td>184 018</td>
<td>77 681</td>
</tr>
<tr>
<td>% of Mapped</td>
<td>0.24</td>
<td>1.04</td>
</tr>
<tr>
<td>Enrichment</td>
<td>7.33</td>
<td>4.39</td>
</tr>
</tbody>
</table>

The palindromic junction at the telomeric boundary of the 1q21 amplicon would indicate the inverted duplication of the centromere-proximal part of the long arm of chromosome 1. Such duplication produces a dicentric chromosome that could initiate BFB cycles and gene amplification. Copy number analysis using high-density oligonucleotide microarrays defined the amplicon in Colo (Figure 3a, top). The moderately amplified 700 kb region extended up to a histone H4 gene cluster (HIST2H4) where naturally occurring inverted segmental duplications set the centromeric boundary (30). This moderately amplified region did not have a high peak of GAPF reads (Figure 3a, bottom). In contrast, within the highly amplified 400 kb region on the telomeric side, there was a series of high peaks (Figure 3b). Three peaks (RDRP2, ECM1 and ADAMS4L) had gradual increases of reads toward the centromeric sides (Figure 3c), indicating the palindromic duplication for the genomic region including MCL1. Importantly, the peak at RDRP2 was located at the boundary between the moderately amplified region and the highly amplified region, indicating the role of the palindromic junction in setting the amplicon boundary. The direction of duplication was somewhat ambiguous at the MCL1 locus. There was an increase toward the telomeric end of MCL1 that would indicate the palindromic duplication excluding MCL1. However, there were several bins with increased reads that do not show clear direction, suggesting more than one rearrangement within this region (see later in the text).

**Physical locations of amplicons**

We sought further evidence for the history of BFB cycles by defining the physical locations of amplicons using fluorescence in situ hybridization (FISH). Publically
available Spectral Karyotyping (SKY) data [SKY/M-FISH and comparative genomic hybridization (CGH) database] show a chromosome with complex fusions; a 5p chromosome fused to 1q and the 1q fused to 13q (Figure 4a). The fusion breakpoint was at 1q21 in a subset of cells where the CTSK gene is located. Such a fusion suggests a chromosome healing event by a translocation for a broken chromosome undergoing BFB cycles. To further define the physical locations of ampiclons, a BAC clone in the highly amplified region (RP11-152I9, red, also in Figure 3a) and a clone from a moderately amplified region (RP11-142H21, green) were labeled and hybridized to metaphase spreads (Figure 4b). Consistent with the SKY data, amplified red signals were seen in the middle of the long chromosome arm in the majority of metaphases. However, in 4 of 22 metaphases, multiple red and green signals are seen at the end of a small chromosome. In metaphases of Colo320HSR, a clone that originated from Colo320DM (40), an ampiclon at the end of a small chromosome was common (23/28) (Figure 5c). It is possible that the broken chromosome with the ampiclon at the end originally existed in Colo320DM. It could then fuse to chromosome 13 in most of the cells in Colo320DM, but not in Colo320HSR.

Together with the cluster of palindromic junctions, the physical location of ampiclons strongly suggests the history of BFB cycles during the establishment of the 1q21 ampiclon (Figure 4c). First, a break at the CTSK gene led to the duplication of the centromere-bearing chromosome and the loss of the distal 1q. The resulting dicentric chromosome broke at HIST2H4 and initiated another duplication, which set the boundary of the entire ampiclon. Subsequent breaks and fusions at RDRP2, ADAMSLT4 and MCL1 established the highly amplified region at the end of chromosome 1 (Figure 4c, circled). Such a BFB-based mechanism is also consistent with the copy number and GAPF profiles we observed (Figure 4c, bottom); (i) the palindromic junctions at CTSK and at RDRP2 set the boundaries for the highly amplified region and (ii) palindromic junctions are expected to be the most abundant at CTSK, followed by at RDRP2.

Central asymmetry at the junctions and homology at the breakpoints

Given the evidence of BFB cycles, breakpoint sequences at the palindromic junctions could elucidate the nucleotide-level mechanisms of sister chromatid fusion: how the breaks in mitosis are processed to produce sister chromatid fusion. To sequence the breakpoints, we designed two PCR primers for the minus-strand in the reference genome near the peaks of sequence reads (Figure 5a). The inverted duplication would bring the minus-strand to the plus-strand in the opposite orientation and thus could enable PCR to amplify breakpoints. Sequencing of the PCR products revealed striking similarities between fusion junctions (Figure 5b). First, there were few kb central asymmetries at the palindromic junctions: 2.1 kb in the RDRP2 junction, 1.7 kb in the ECM1 junction and 3.7 kb in the ADAMSLT4. We recovered two junctions at MCL1, each of which has 2.0 and 3.5 kb asymmetry, respectively. Second, there were sequence homologies at the breakpoints. For RDRP2 and MCL1 junctions, the breakpoints were located within Alu repeats that were a few kb apart in inverted orientations and exhibited high sequence similarities (>80%). In all three cases, there were more than 30 identical nucleotides surrounding the breakpoints (shown in red in Figure 5b). MCL1 junctions were located within a dense cluster of Alu repeats (Supplementary Figure S2), which could promote frequent breakages and fusions and explain the complex pattern of GAPF-read depth (Figure 3c). The ECM1 and ADAMSLT4 junctions were mediated by microhomologies of 3 and 15 bp, respectively. There were no non-template sequence insertions at these breakpoints. The homologies at the breakpoints favor the end processing—capping—replication model in which a broken end undergoes single-strand degradation and the resulting 3′ tail finds the homology in the same strand at few kb internal from the end.
followed by DNA synthesis and the formation of a capped end (Figure 1b).

**DISCUSSION**

A number of studies support a view that the BFB cycle is an important underlying process for genome instability and gene amplification in tumor cells (5,7,14,15). A critical step for BFB cycles leading to gene amplification is the fusion of sister chromatids; however, ambiguity still exists for the nucleotide-level mechanism underlying sister chromatid fusion. In this study, we address this issue using integrated cytogenetic, genomic and nucleotide-level analysis. The results from FISH were consistent with the integrated cytogenetic, genomic and nucleotide-level analysis. The results from FISH were consistent with the formation of the 1q21 amplicon through BFB cycles: the duplication of genomic segments at 1q21 at the end of a broken chromosome. GAPF-seq showed repeated breakage and sister chromatid fusions within the highly amplified region, as palindromic junctions are seen at 50 kb intervals. Finally, a few kb of central asymmetries and homologies at the breakpoints indicated that the end-capping followed by DNA replication is a primary mechanism of sister chromatid fusion.

Our study also provided a unique opportunity to understand the fate of naturally occurring broken ends in mitosis. The process preceding end-capping is end

**Figure 4.** 1q21 amplicon at the end of a broken chromosome. (a) The diagram of a chromosome with complex fusions in Colo320DM, obtained from the SKY/M-FISH and CGH database (http://www.ncbi.nlm.nih.gov/sky/skyquery.cgi). Note the fusion breakpoint between 1q21 and 13q12. (b) Pictures from dual color metaphase FISH using the probes for the regions of low (green) and high (red) level amplification are shown for Colo320DM (left) and Colo320HSR (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (yellow arrowheads) are magnified and superimposed. The genomic locations of BAC clones (right). The pictures for the broken chromosomes (y
resection. Whether broken ends undergo end resection depends on the stages of the cell cycle (41,42). End resection initiates homology-dependent repair of DSBs during S and G2 when the sister chromatid is available as repair templates. End resection is not a primary choice for DSB processing in G1, as DSBs are protected and repaired by NHEJ. An important aspect of broken ends in BFB cycles is that broken ends arise in mitosis and are one-ended. Although the fate of broken ends in mitosis has not been studied as extensively as DSBs in other stages of the cell cycle, recent studies indicate that mitotic breaks can undergo end resection. M phase breaks introduced by ionizing radiation (IR) trigger a DNA damage response in human cells. Those breaks are marked by IR-induced foci with γ-H2AX, the MRE11-RAD50-NBS1 (MRN) complex and MDC1 (43). However, 53BP1, a protein that preserves DSB ends for NHEJ (44), is excluded from the M phase IR-induced foci. The lack of 53BP1 hints that the ends are unprotected and are available for end resection. Indeed, Peterson et al. (45) has shown recently that broken DNA treated with the Xenopus M phase extract undergoes CtIP-dependent end resection. CtIP binds to the MRN complex and enhances the endonuclease activity (46). In S phase, the resection by MRN-CtIP removes DNA at short distance from the end and is followed by more extensive resection by EXO1 and DNA2 (47,48). It is currently unknown how extensive the end resection could be in M phase. Given the highly condensed chromatin in M phase, end resection could be limited to a few kb. A few kb of ssDNA tails produced in BFB cycles may have limited access to inter-homologue homology-dependent repair due to the lack of RAD51 loading (45), and instead fold back to find homologies.

We defined a characteristic pattern of breakpoints for sister chromatid fusion: homologies at the breakpoints and a few kb of central asymmetries between duplicated genomic regions. NGS-based paired end sequencing provides a large number of breakpoint sequences in an unbiased fashion. By searching breakpoints with characteristic pattern, we could infer the impact of BFB cycles on genome instability in tumors. A class of breakpoints called fold-back inversion exactly satisfies our criteria. Fold-back inversions are common in pancreatic tumors (16% of 381 breakpoints from 13 pancreatic tumors) and are often associated with amplified genomic regions (25). More than 70% of such fold back inversions have <3 kb of central asymmetries. Microhomologies (>1 bp) are seen in 60% of the breakpoints. Thus, the majority of breakpoints in fold-back inversions could represent sister chromatid fusions. Such inversions are also reported in breast, ovarian and gastric tumors (49,50). The fact that fold-back inversions are common in tumors demonstrates the impact of BFB cycles in genome instability. It is also important to mention that, although less common, there are fold-back inversions with the insertion of non-templated sequences. The insertion of non-templated sequences indicates the NHEJ of two ends as an underlying mechanism; however, it is not known whether this minor class of fold-back inversions derives from M phase DSBs during BFB cycles. It is possible that a chromosome with a spontaneous G1 break replicates in S phase and is repaired by NHEJ with the insertion of non-templated sequences. Such an event could also be important for BFB cycles because it could initiate BFB cycles when a centromere-bearing chromosome is joined and a dicentric chromosome is produced. Furthermore, non-DSB-initiated mechanisms, including a replication fork collapse followed by the strand invasion at an inverted ectopic homology and the subsequent initiation of break-induced replication (27,28), could produce characteristic junctions with central asymmetries and homologies at the breakpoints. Such a mechanism produces complex duplications and triplications of germline origin. However, asymmetries between duplicated segments are much longer, several hundred kb in some cases, suggesting that strand invasion typically occurs at regions of homology that are distant.

We show that BFB cycles use either homologies from divergent Alu repeats or microhomologies (3 and 15 bp) for the capping of broken ends. This is in contrast to the long homology seen at the HIST2H4 locus where a 29 kb segmental inverted duplication (96% sequence identity) could cap the end (30). Long stretches of homology certainly promote fold-back annealing (19,30). Whether divergent Alu repeats and microhomologies are preferred substrates for sister chromatid fusion in cancer cells needs to be addressed in future studies. In this regard, it is noteworthy that, in maize, a chromosome structure called a knob is a preferred site of breaks (13). A knob is a darkly stained, heterochromatic structure with highly repetitive sequences. Thus, breaks could occur non-randomly within or near repeated segments, and such segments could facilitate end-capping. Our targeted genomic approach for sister chromatid fusions will help to elucidate this issue.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
The authors thank Drs M-C Yao, V Boener and A Rattray for comments on the manuscript. The array-CGH experiments were conducted and supported by the Gene Expression and Genotyping Core Facility of the Case Comprehensive Cancer Center (Grant Number P30 CA43703).

FUNDING
National Cancer Institute [R01CA149385]; American Cancer Society and Cleveland Clinic (to H.T.). Funding for open access charge: National Cancer Institute [R01CA149385 to H.T.].

Conflict of interest statement. None declared.
REFERENCES


