9732–9740 *Nucleic Acids Research, 2013, Vol. 41, No. 21* *Published online 23 August 2013*
doi:10.1093/nar/gkt762

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

Michael Marotta[1], Xiongfong Chen[2], Takaaki Watanabe[1], Pieter W. Faber[3], Scott J. Diede[4], Stephen Tapscott[4], Raymond Tubbs[5], Anna Kondratova[1], Robert Stephens[2] and Hisashi Tanaka[1,6,*]

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

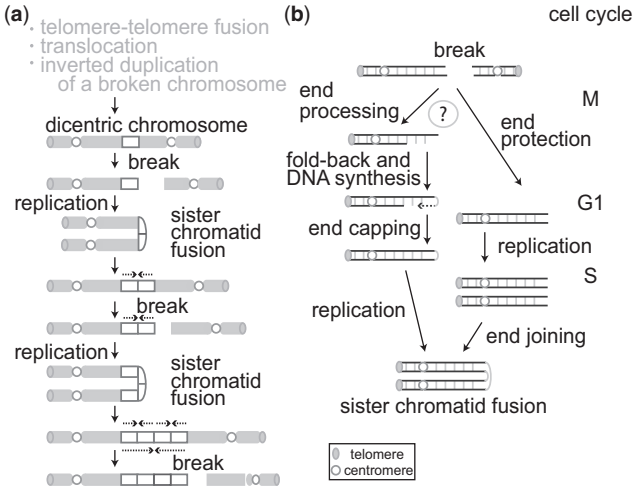

Figure 1. (**a**) Sister chromatid fusion during the BFB cycle promotes gene amplification. Only the fate of one broken chromatid is shown. Three events that can create a dicentric chromosome are listed on the top (see in the 'Introduction' section of the main text). Following a chromosome rupture and 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 chromsomal 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).

centromeres is amplified in some cell descendants. This was observed phenotypically in kernels with extremely dark color because the gene required for pigment production was located between two centromeres. These observations inform about key steps leading to gene amplification: DNA double-strand breaks (DSBs) in mitosis and the subsequent sister chromatid fusion.

In cancer cells, dicentric chromosomes can arise from a variety of events: the fusion of two chromosomes with extremely short telomeres (14,15), the fusion between two centromere-bearing broken non-homologous chromosomes (translocation) (16–18) and the inverted duplications of centromere-bearing broken chromosomes (19,20). Once formed, dicentric chromosomes can enter into BFB cycles and initiate gene amplification through sister chromatid fusion, as McClintock described for maize. If the resulting amplified segments harbor genes that promote cell proliferation, the cells can become dominant in a cancer cell population and confer aggressive tumor phenotypes. Despite the clinical importance, how sister chromatids fuse in tumor cells remains elusive. Cytogenetic studies provide ample evidence for the occurrence of BFB cycles in tumor cells (21–24) but do not provide base-pair resolution on palindromic junctions. Base-pair resolution can be obtained by recent next-generation sequencing (NGS)-based breakpoint analyses (25,26), but breakpoint sequences themselves are not direct evidence of sister chromatid fusion and can be produced by other mechanisms, such as replication fork stalling and template switching (27,28).

Two plausible mechanisms have been proposed for sister chromatid fusion (Figure 1b) (25,26,29,30). A broken end in mitosis would undergo end resection and leave a 3′ single-stranded DNA (ssDNA) tail. The ssDNA would fold back and anneal using homologies. DNA synthesis would fill the gap and complete the end capping. The entire chromosome would duplicate in the S-phase of next cell cycle to produce chromatids fused at the broken end. This intra-strand fold-back annealing of ssDNA between homologous sequences has been shown for endonuclease-induced DSBs in both mammalian cells and yeast (29,30). Alternatively, a non-homologous end joining (NHEJ)-based mechanism predicts that a broken end in mitosis would be protected in G1 phase and would enter into S phase. After DNA replication, the two ends of sister chromatids would fuse to each other by NHEJ.

Distinguishing these possibilities requires (i) identifying the amplicon that is created by BFB cycles and (ii) defining the DNA sequences at the breakpoints. A homology-dependent end-capping mechanism could exhibit homologies at breakpoints, whereas a NHEJ-dependent process could predict breakpoints either without homologies or with the insertion of non-templated sequences. In this manuscript, we address this issue for the recurrent chromosome 1q21 amplicon in human cancer cells using an integrated NGS-based genomic, cytogenetic and molecular approach.

MATERIALS AND METHODS

Genome-wide analysis of palindrome formation (GAPF-seq) and array-comparative genomic hybridization

The colorectal cancer cell line Colo320DM and primary fibroblast IMR90 were obtained from ATCC. Genomic DNA was processed for genome-wide analysis of palindrome formation (GAPF) as previously described with minor modifications (31). Two micrograms of genomic DNA was used as starting material, of which 1 μg was digested with KpnI and 1 μg was digested with SbfI for 16 h (20 μl of reaction with 10 U of each enzyme). After the heat inactivation of the enzyme (65°C for 20 min), digests were combined for denaturation (7 min in boiling water) and rapid renaturation in the presence of 100 mM NaCl and 50% formamide (90 μl of reaction mix consisting of 40 μl of DNA digests, 3 μl of 3 M NaCl, 45 μl of formamide and 2 μl of water). Following the digestion with nuclease S1(Invitrogen) (120 μl of reaction consisting of 90 μl of DNA, 8 μl of 3 M NaCl, 12 μl of S1 nuclease buffer, 2 μl of 100 U/μl S1 nuclease and 8 μl of water), DNA was purified using the ChargeSwitch PCR Clean-up Kit.

After fragmentation using the M220 focused ultrasonicator (Covaris), fragments between 100 and 250 bp were purified for NGS library construction using 5500 SOLiD™ Fragment Library Core Kit (Applied Biosystems). Each DNA sample was ligated to linkers with unique barcodes. Libraries were then pooled together and sequenced in a single flowcell run using a SOLiD 5500 × l sequencer. The barcodes were then used to trace the sequence data back to a specific sample. BAM files were generated using LifeTech LifeScope software. Read depth analyses were done using Genomics Suite (Partek). All repeat content analysis was performed using the open

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 3 bp 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 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 the 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 reconfirm the breakpoint, genomic DNA was amplified using independent primer sets designed for each breakpoint.

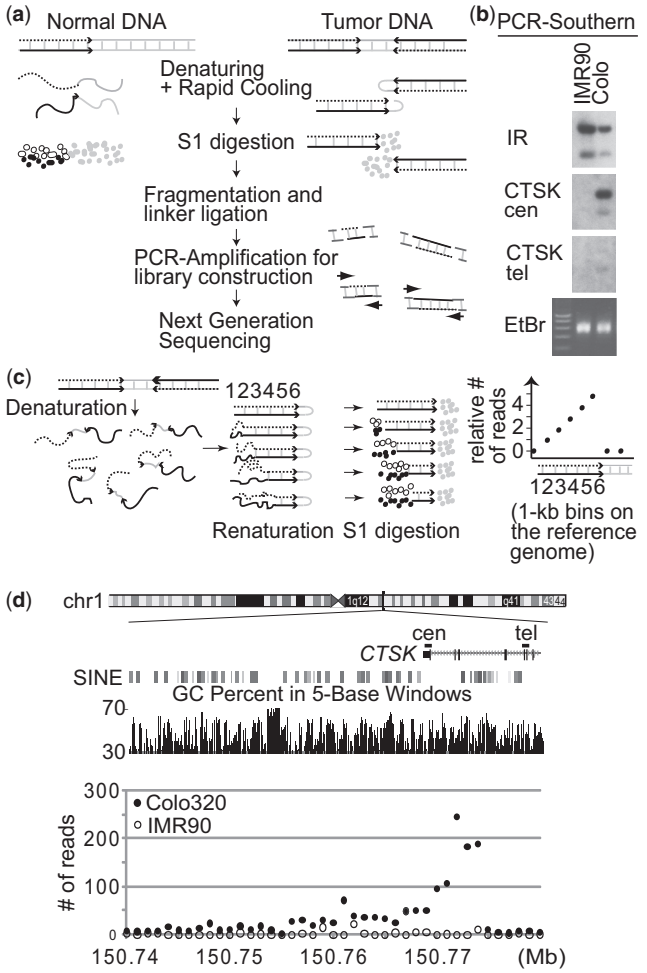

Figure 2. (**a**) A strategy for enriching DNA from palindromic junctions. DNA is shown by either a black, dotted, or gray line, with the black and dotted lines being complementary to each other. (**b**) Southern analyses of the PCR products using probes in the human genome. The locations of probes (CTSK cen and tel) are shown in (**d**). (i) enrichment is only seen in Colo320DM by probe CTSK cen that is within the palindromic junction, but not by probe CTSK tel, and (ii) probes from a naturally existing DNA inverted repeat at chromosome 19p13.2 (chr19:7049026–7058989 in hg19) showed strong enrichment in both IMR90 and Colo320DM. EtBr, ethidium bromide stained gel. (**c**) Schematic drawings show that *in vitro* fold-back of palindromic DNA after denaturation/renaturation and subsequent digestion by nuclease S1 (left) can results 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.

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 results in the enrichment of DNA from palindromic junctions. The enriched DNA was fragmented, ligated to linkers and PCR-amplified for NGS using a

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

	hg19	GAPF-seq	
GAPF-seq Reads		COLO320DM	IMR90
Total reads		17 498 356	15 351 029
Mapped reads		10 579 579	7 448 354
Intersected with segmental dup (# of Reads)		737 758	545 540
% of Mapped	4.90	6.97	7.32
Enrichment		1.42	1.50
Intersected with repeat masker (# of reads)		7 773 451	5 670 447
% of Mapped	50.63	73.48	76.13
Enrichment		1.45	1.50
Intersected with Alu (# of reads)		4 196 090	3 429 714
% of Mapped	9.72	39.66	46.05
Enrichment		4.08	4.74
Intersected with alpha satellite (# of reads)		184 018	77 681
% of Mapped	0.24	1.74	1.04
Enrichment		7.33	4.39

Palindromic junctions cluster within the highly amplified genomic regions

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 *ADAMSL4*) 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

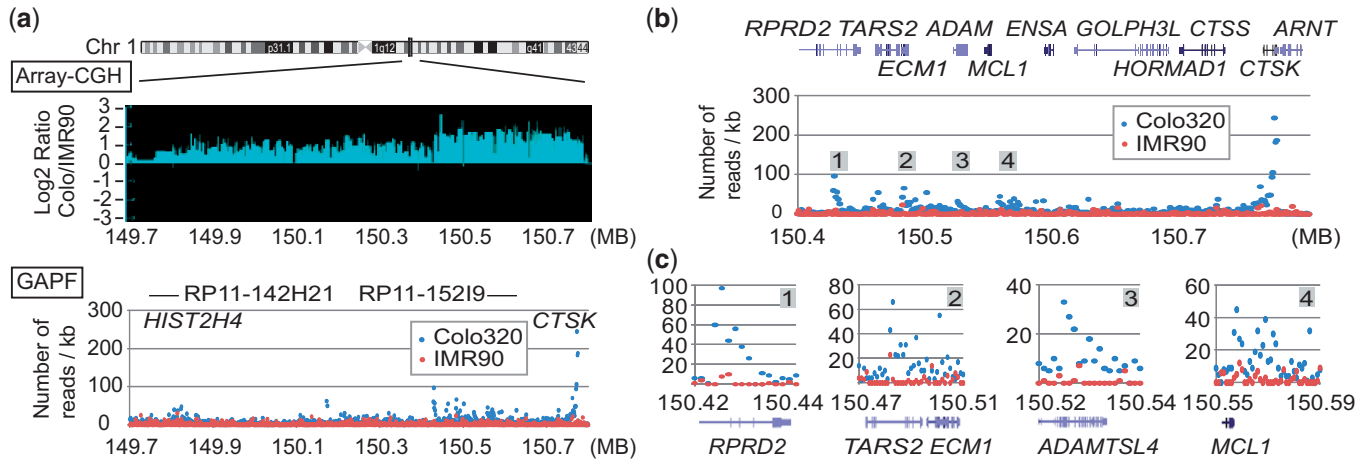

Figure 3. (**a**) Copy number ratio (Colo/IMR90, Array-CGH, *top*) and the numbers of GAPF-reads/kb in the 1.1 Mb region (*bottom*) of chromosome 1. The numbers of GAPF-reads for Colo (blue circles) and IMR90 (red circles) for each 1 kb bin are shown. The locations of BAC clones used for the probes for FISH (Figure 4a) are shown. (**b**) The number of GAPF-reads/kb over the 400 kb highly amplified region. Four peaks are marked as 1–4. Genes within the region are shown on the top. (**c**) The detailed views of read depths for GAPF peaks 1–4.

available Spectoral Karyotyping (SKY) data [SKY/M-FISH and comparative genomic hybridization (CGH) database] show a chromosome with complex fusions; a 5 p 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 amplicons, 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 amplicon at the end of a small chromosome was common (23/28) (Figure 5c). It is possible that the broken chromosome with the amplicon 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 amplicons strongly suggests the history of BFB cycles during the establishment of the 1q21 amplicon (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 amplicon. Subsequent breaks and fusions at *RDRP2*, *ADAMSLT4* and *MCL1* established the highly amplified region at the end of chromosome (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 sequencse 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 are 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,

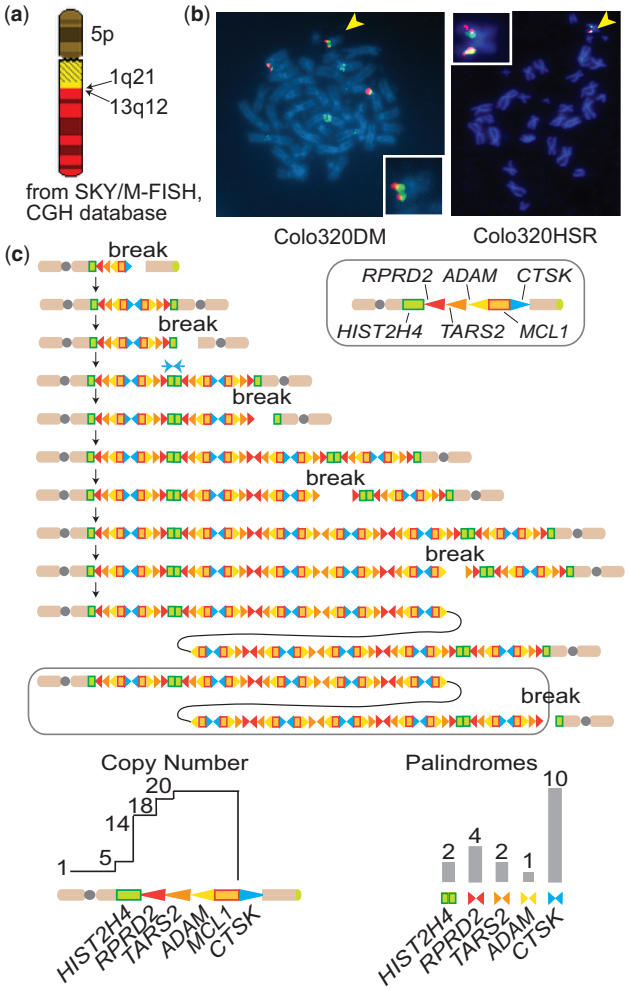

Figure 4. 1q21 amplicon at the end of a broken chromosome. (**a**) The diagram of a chromosome with complex fusions in Colo320DM, obtained from the SKY/M-FISH and CGH database (http://www.ncbi.nlm.nih.gov/sky/skyquery.cgi). Note the fusion breakpoint between 1q21 and 13q12. (**b**) Pictures from dual 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 are shown in Figure 3a. (**c**) The BFB cycle-based process for giving rise to the 1q21 amplicon in Colo320 cells. Genomic segments are shown either by rectangles or triangles. Triangles also represent the direction of read-depth increases, and palindromic junctions are shown by pairs of triangles. Both copy number profiles and GAPF read depth predicted by the process are shown at the bottom.

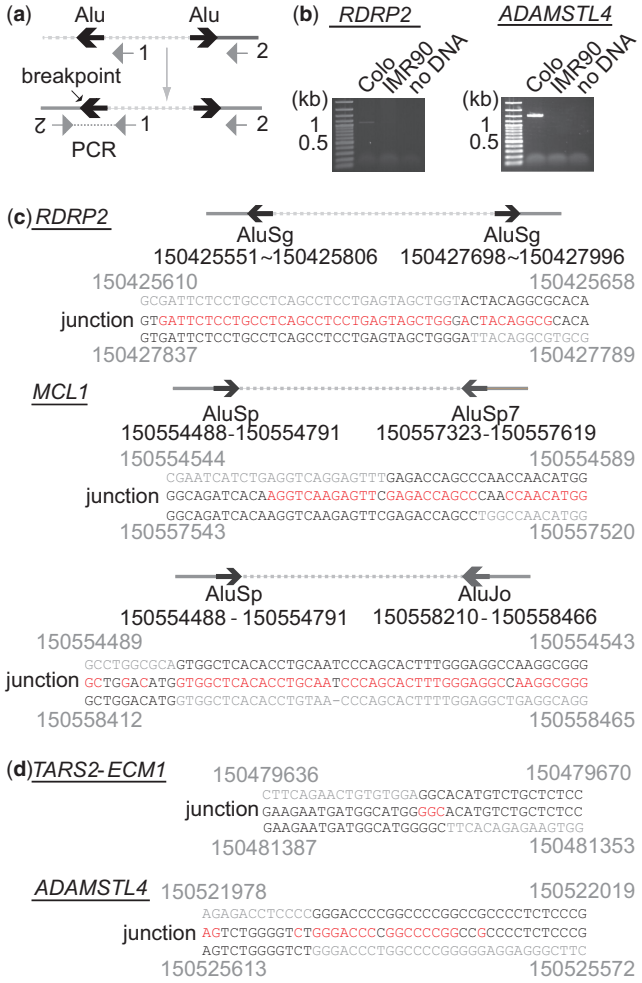

Figure 5. (**a**) PCR strategy for the amplification of palindromic fusion breakpoints. Lines represent genomic segments. Two PCR primers are indicated by gray arrows. Inverted Alu repeats (or microhomologies) are indicated by a pair of black arrows. (**b**) Ethidium bromide-stained agarose gels after PCR amplification for RDRP2 and ADAMSTL4 junctions. Normal fibroblast IMR90 DNA was used as a negative control. (**c**) DNA sequences of palindromic junctions with Alu inverted repeats. The locations (coordinates of chromosome 1) of inverted Alu repeats and the coordinates (top and bottom) of breakpoints and junction sequences (middle) are shown. Sequence homologies are shown in red. (**d**) DNA sequences of microhomology-mediated palindromic junctions.

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 formation of the 1q21 amplicon through BFB cycles: the duplication of genomic segments at 1q21 at the end of 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

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 cycles, 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

1. Gorre,M.E., Mohammed,M., Ellwood,K., Hsu,N., Paquette,R., Rao,P.N. and Sawyers,C.L. (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*, **293**, 876–880.
2. Engelman,J.A., Zejnullahu,K., Mitsudomi,T., Song,Y., Hyland,C., Park,J.O., Lindeman,N., Gale,C.M., Zhao,X., Christensen,J. *et al.* (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science*, **316**, 1039–1043.
3. Slamon,D.J., Clark,G.M., Wong,S.G., Levin,W.J., Ullrich,A. and McGuire,W.L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, **235**, 177–182.
4. Brodeur,G.M., Seeger,R.C., Schwab,M., Varmus,H.E. and Bishop,J.M. (1984) Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science*, **224**, 1121–1124.
5. Debatisse,M. and Malfoy,B. (2005) Gene amplification mechanisms. *Adv. Exp. Med. Biol.*, **570**, 343–361.
6. Mondello,C., Smirnova,A. and Giulotto,E. (2010) Gene amplification, radiation sensitivity and DNA double-strand breaks. *Mutat. Res.*, **704**, 29–37.
7. Tanaka,H. and Yao,M.C. (2009) Palindromic gene amplification—an evolutionarily conserved role for DNA inverted repeats in the genome. *Nat. Rev. Cancer*, **9**, 216–224.
8. Kaufman,R.J., Brown,P.C. and Schimke,R.T. (1979) Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes. *Proc. Natl Acad. Sci. USA*, **76**, 5669–5673.
9. Nunberg,J.H., Kaufman,R.J., Schimke,R.T., Urlaub,G. and Chasin,L.A. (1978) Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. *Proc. Natl Acad. Sci. USA*, **75**, 5553–5556.
10. Marotta,M., Chen,X., Inoshita,A., Stephens,R., Thomas Budd,G., Crowe,J.P., Lyons,J., Kondratova,A., Tubbs,R. and Tanaka,H. (2012) A common copy-number breakpoint of ERBB2 amplification in breast cancer colocalizes with a complex block of segmental duplications. *Breast Cancer Res.*, **14**, R150.
11. Smith,K.A., Gorman,P.A., Stark,M.B., Groves,R.P. and Stark,G.R. (1990) Distinctive chromosomal structures are formed very early in the amplification of CAD genes in Syrian hamster cells. *Cell*, **63**, 1219–1227.
12. Coquelle,A., Pipiras,E., Toledo,F., Buttin,G. and Debatisse,M. (1997) Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell*, **89**, 215–225.
13. McClintock,B. (1941) The stability of broken ends of chromosomes in Zea Mays. *Genetics*, **26**, 234–282.
14. Maser,R.S. and DePinho,R.A. (2002) Connecting chromosomes, crisis, and cancer. *Science*, **297**, 565–569.
15. Murnane,J.P. (2012) Telomere dysfunction and chromosome instability. *Mutat. Res.*, **730**, 28–36.
16. Tlsty,T.D. (2002) Functions of p53 suppress critical consequences of damage and repair in the initiation of cancer. *Cancer Cell*, **2**, 2–4.
17. Zhu,C., Mills,K.D., Ferguson,D.O., Lee,C., Manis,J., Fleming,J., Gao,Y., Morton,C.C. and Alt,F.W. (2002) Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell*, **109**, 811–821.
18. Difilippantonio,M.J., Petersen,S., Chen,H.T., Johnson,R., Jasin,M., Kanaar,R., Ried,T. and Nussenzweig,A. (2002) Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. *J. Exp. Med.*, **196**, 469–480.
19. Tanaka,H., Tapscott,S.J., Trask,B.J. and Yao,M.C. (2002) Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. *Proc. Natl Acad. Sci. USA*, **99**, 8772–8777.
20. Narayanan,V., Mieczkowski,P.A., Kim,H.M., Petes,T.D. and Lobachev,K.S. (2006) The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell*, **125**, 1283–1296.
21. Shuster,M.I., Han,L., Le Beau,M.M., Davis,E., Sawicki,M., Lese,C.M., Park,N.H., Colicelli,J. and Gollin,S.M. (2000) A consistent pattern of RIN1 rearrangements in oral squamous cell carcinoma cell lines supports a breakage-fusion-bridge cycle model for 11q13 amplification. *Genes Chromosomes Cancer*, **28**, 153–163.
22. Ciullo,M., Debily,M.A., Rozier,L., Autiero,M., Billault,A., Mayau,V., El Marhomy,S., Guardiola,J., Bernheim,A., Coullin,P. *et al.* (2002) Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of PIP gene duplication from a break at FRA7I. *Hum. Mol. Genet.*, **11**, 2887–2894.
23. Hellman,A., Zlotorynski,E., Scherer,S.W., Cheung,J., Vincent,J.B., Smith,D.I., Trakhtenbrot,L. and Kerem,B. (2002) A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell*, **1**, 89–97.
24. Kitada,K. and Yamasaki,T. (2008) The complicated copy number alterations in chromosome 7 of a lung cancer cell line is explained by a model based on repeated breakage-fusion-bridge cycles. *Cancer Genet. Cytogenet.*, **185**, 11–19.
25. Campbell,P.J., Yachida,S., Mudie,L.J., Stephens,P.J., Pleasance,E.D., Stebbings,L.A., Morsberger,L.A., Latimer,C., McLaren,S., Lin,M.L. *et al.* (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, **467**, 1109–1113.
26. Bignell,G.R., Santarius,T., Pole,J.C., Butler,A.P., Perry,J., Pleasance,E., Greenman,C., Menzies,A., Taylor,S., Edkins,S. *et al.* (2007) Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome Res.*, **17**, 1296–1303.
27. Hastings,P.J., Ira,G. and Lupski,J.R. (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.*, **5**, e1000327.
28. Carvalho,C.M., Ramocki,M.B., Pehlivan,D., Franco,L.M., Gonzaga-Jauregui,C., Fang,P., McCall,A., Pivnick,E.K., Hines-Dowell,S., Seaver,L.H. *et al.* (2011) Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat. Genet.*, **43**, 1074–1081.
29. Rattray,A.J., Shafer,B.K., Neelam,B. and Strathern,J.N. (2005) A mechanism of palindromic gene amplification in *Saccharomyces cerevisiae*. *Genes Dev.*, **19**, 1390–1399.
30. Tanaka,H., Cao,Y., Bergstrom,D.A., Kooperberg,C., Tapscott,S.J. and Yao,M.C. (2007) Intrastrand annealing leads to the formation of a large DNA palindrome and determines the boundaries of genomic amplification in human cancer. *Mol. Cell. Biol.*, **27**, 1993–2002.
31. Diede,S.J., Guenthoer,J., Geng,L.N., Mahoney,S.E., Marotta,M., Olson,J.M., Tanaka,H. and Tapscott,S.J. (2010) DNA methylation of developmental genes in pediatric medulloblastomas identified by denaturation analysis of methylation differences. *Proc. Natl Acad. Sci. USA*, **107**, 234–239.
32. Goecks,J., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
33. Langmead,B. (2010) Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinformatics*, **Chapter 11**, Unit 11.17.
34. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
35. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
36. Watanabe,T., Tanabe,H. and Horiuchi,T. (2011) Gene amplification system based on double rolling-circle replication as a model for oncogene-type amplification. *Nucleic Acids Res.*, **39**, e106.
37. Tanaka,H., Bergstrom,D.A., Yao,M.C. and Tapscott,S.J. (2005) Widespread and nonrandom distribution of DNA palindromes in cancer cells provides a structural platform for subsequent gene amplification. *Nat. Genet.*, **37**, 320–327.
38. Beroukhim,R., Mermel,C.H., Porter,D., Wei,G., Raychaudhuri,S., Donovan,J., Barretina,J., Boehm,J.S., Dobson,J., Urashima,M.

et al. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature*, **463**, 899–905.
39. Bailey,J.A., Gu,Z., Clark,R.A., Reinert,K., Samonte,R.V., Schwartz,S., Adams,M.D., Myers,E.W., Li,P.W. and Eichler,E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003–1007.
40. Quinn,L.A., Moore,G.E., Morgan,R.T. and Woods,L.K. (1979) Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions. *Cancer Res*., **39**, 4914–4924.
41. Chapman,J.R., Taylor,M.R. and Boulton,S.J. (2012) Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell*, **47**, 497–510.
42. Symington,L.S. and Gautier,J. (2011) Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet*., **45**, 247–271.
43. Giunta,S., Belotserkovskaya,R. and Jackson,S.P. (2010) DNA damage signaling in response to double-strand breaks during mitosis. *J. Cell Biol*., **190**, 197–207.
44. Difilippantonio,S., Gapud,E., Wong,N., Huang,C.Y., Mahowald,G., Chen,H.T., Kruhlak,M.J., Callen,E., Livak,F., Nussenzweig,M.C. *et al*. (2008) 53BP1 facilitates long-range DNA end-joining during V(D)J recombination. *Nature*, **456**, 529–533.
45. Peterson,S.E., Li,Y., Chait,B.T., Gottesman,M.E., Baer,R. and Gautier,J. (2011) Cdk1 uncouples CtIP-dependent resection and Rad51 filament formation during M-phase double-strand break repair. *J. Cell Biol*., **194**, 705–720.
46. Sartori,A.A., Lukas,C., Coates,J., Mistrik,M., Fu,S., Bartek,J., Baer,R., Lukas,J. and Jackson,S.P. (2007) Human CtIP promotes DNA end resection. *Nature*, **450**, 509–514.
47. Peng,G., Dai,H., Zhang,W., Hsieh,H.J., Pan,M.R., Park,Y.Y., Tsai,R.Y., Bedrosian,I., Lee,J.S., Ira,G. *et al*. (2012) Human nuclease/helicase DNA2 alleviates replication stress by promoting DNA end resection. *Cancer Res*., **72**, 2802–2813.
48. Gravel,S., Chapman,J.R., Magill,C. and Jackson,S.P. (2008) DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev*., **22**, 2767–2772.
49. Hillmer,A.M., Yao,F., Inaki,K., Lee,W.H., Ariyaratne,P.N., Teo,A.S., Woo,X.Y., Zhang,Z., Zhao,H., Ukil,L. *et al*. (2011) Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. *Genome Res*., **21**, 665–675.
50. McBride,D.J., Etemadmoghadam,D., Cooke,S.L., Alsop,K., George,J., Butler,A., Cho,J., Galappaththige,D., Greenman,C., Howarth,K.D. *et al*. (2012) Tandem duplication of chromosomal segments is common in ovarian and breast cancer genomes. *J. Pathol*., **227**, 446–455.