Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

GENOME RESEARCH

A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome

Oliver A. Hampton, Petra Den Hollander, Christopher A. Miller, et al.

Genome Res. 2009 19: 167-177 originally published online December 3, 2008
Access the most recent version at doi:10.1101/gr.080259.108

References This article cites 41 articles, 15 of which can be accessed free at:
http://genome.cshlp.org/content/19/2/167.full.html#ref-list-1

Open Access Freely available online through the *Genome Research* Open Access option.

License Freely available online through the Genome Research open access option.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Copyright © 2009 by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

Letter

A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome

Oliver A. Hampton,[1,3,5] Petra Den Hollander,[4,5] Christopher A. Miller,[1,3] David A. Delgado,[4,5] Jian Li,[1,3] Cristian Coarfa,[1,2] Ronald A. Harris,[1,2] Stephen Richards,[2] Steven E. Scherer,[2] Donna M. Muzny,[2] Richard A. Gibbs,[2,3] Adrian V. Lee,[4,5,6] and Aleksandar Milosavljevic[1,2,3,5,7]

[1]Bioinformatics Research Laboratory, Baylor College of Medicine, Houston, Texas 77030, USA; [2]Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; [3]Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA; [4]Breast Center, Baylor College of Medicine, Houston, Texas 77030, USA; [5]Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA; [6]Department of Medicine, Baylor College of Medicine, Houston, Texas 77030, USA

By applying a method that combines end-sequence profiling and massively parallel sequencing, we obtained a sequence-level map of chromosomal aberrations in the genome of the MCF-7 breast cancer cell line. A total of 157 distinct somatic breakpoints of two distinct types, dispersed and clustered, were identified. A total of 89 breakpoints are evenly dispersed across the genome. A majority of dispersed breakpoints are in regions of low copy repeats (LCRs), indicating a possible role for LCRs in chromosome breakage. The remaining 68 breakpoints form four distinct clusters of closely spaced breakpoints that coincide with the four highly amplified regions in MCF-7 detected by array CGH located in the 1p13.1-p21.1, 3p14.1-p14.2, 17q22-q24.3, and 20q12-q13.33 chromosomal cytobands. The clustered breakpoints are not significantly associated with LCRs. Sequences flanking most (95%) breakpoint junctions are consistent with double-stranded DNA break repair by nonhomologous end-joining or template switching. A total of 79 known or predicted genes are involved in rearrangement events, including 10 fusions of coding exons from different genes and 77 other rearrangements. Four fusions result in novel expressed chimeric mRNA transcripts. One of the four expressed fusion products (*RAD51C–ATXN7*) and one gene truncation (*BRIP1* or *BACH1*) involve genes coding for members of protein complexes responsible for homology-driven repair of double-stranded DNA breaks. Another one of the four expressed fusion products (*ARFGEF2–SULF2*) involves SULF2, a regulator of cell growth and angiogenesis. We show that knock-down of *SULF2* in cell lines causes tumorigenic phenotypes, including increased proliferation, enhanced survival, and increased anchorage-independent growth.

[Supplemental material is available online at www.genome.org and through the Breast Cancer project page at www.genboree.org. All MCF-7 BAC clones are available from Amplicon Express under name HTA and plate/row/column names as indicated. The sequence data from this study have been submitted to the NCBI Trace and Short Read Archives (http://www.ncbi.nlm.nih.gov) under accession nos. 2172834909–2172901416 and 2172904832–2172911164, and SRR006762–SRR006767, respectively].

Many cancer genomes are characterized by mutability, including microsatellite instability (MIN) and chromosomal instability (CIN) (Lengauer et al. 1998). It is now generally anticipated that sequencing of cancer genomes using massively parallel sequencing technologies (Korbel et al. 2007; Campbell et al. 2008) will provide insights into structural mutability. Recent sequencing of four cancer amplicons (Bignell et al. 2007) derived from the HCC1954 breast cancer cell line and two lung cancer cell lines provided evidence for homologous and nonhomologous repair of double-strand DNA breaks induced by the breakage-fusion-bridge (BFB) mechanism.

[7]Corresponding author.
E-mail amilosav@bcm.edu; fax (713) 798-4373.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.080259.108. Freely available online through the *Genome Research* Open Access option.

Gene fusions and truncations that result from chromosomal rearrangements provide insight into the molecular mechanisms of cancer progression. Recurrent rearrangements of specific genes indicate increased mutability or positive selection (or a combination of both) in the evolution of tumor genomes. Recurrent fusions, translocations, and other aberrant joins are used as highly informative diagnostic and prognostic markers and drug targets in leukemias, lymphomas, and sarcomas. A total of 337 genes involved in fusions in cancer genomes have been recently surveyed (Mitelman et al. 2007). Four gene fusions have previously been reported in breast carcinomas (*ETV6–NTRK3*, *ODZ4–NRG1*, *TBL1XR1–RGS17*, *BCAS3-BCAS4*) (Mitelman et al. 2007, Ruan et al. 2007).

Breast cancer and carcinomas in general have proven less tractable to fusion discovery due to the typically higher degree of rearrangement. However, a prognostically significant rearrangement was recently discovered in the majority of prostate cancers (Tomlins et al. 2005). Of note, the initial discovery was not iden-

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

tified by analyzing DNA sequence or structure, but via the analysis of outlier gene expression, followed by a targeted locus-specific search for a fusion in genomic DNA. Here we demonstrate a method to detect gene fusions directly by the analysis of genomic DNA, even in highly rearranged breast cancer.

MCF-7 is the most widely used cell line model for estrogen-positive breast cancer. The cell line has been derived from a pleural effusion taken from a patient with metastatic breast carcinoma (Soule et al. 1973). Evidence of CIN in MCF-7 comes from apparent aneuploidy and significant genomic divergence in several sublines (Jones et al. 2000; Nugoli et al. 2003). Chromosomal aberrations in MCF-7 have previously been studied by spectral karyotyping (Kytola et al. 2000; Rummukainen et al. 2001), comparative genomic hybridization (CGH) (Kytola et al. 2000; Rummukainen et al. 2001), array CGH (Neve et al. 2006; Shadeo and Lam 2006; Jonsson et al. 2007), single nucleotide polymorphism arrays (Huang et al. 2004), and gene expression arrays (Neve et al. 2006).

More recently, bacterial artificial chromosome (BAC)-based end sequence profiling (ESP) (Volik et al. 2003, 2006; Raphael et al. 2008) has been applied to study genomic rearrangements in cancer genomes. Volik and colleagues sequenced a total of 19,831 BAC ends from the Amplicon Express MCF-7 BAC library, ~1× clone coverage of the human genome, to identify 582 BACs containing rearrangements.

As a starting point for our analysis, we constructed BAC pools from a nonredundant subset (n = 552) of rearranged BACs identified by Volik et al. (2003, 2006). To map chromosomal aberrations in the genome of the MCF-7 breast cancer cell line at sequence level resolution, we developed a method that combines end-sequence profiling and massively parallel sequencing. By analyzing sequences of the chromosomal breakpoints in the BAC pools, we gained insights into the mechanisms of chromosomal instability and repair. Specific gene fusions and truncations that have emerged during the pathological evolution of this cancer genome point to the molecular mechanisms of the disease. Additional products of our research are benchmarking reagents for the development of a new generation of methods for detecting structural genome variation, including well-characterized BAC pools and validated breakpoints in the MCF-7 genome.

Results

At least 157 breakpoints were induced by somatic rearrangements in MCF-7

Aberrant breakpoint-induced joins were identified by combining "bridging" and "outlining" steps, as illustrated in Figure 1A. The bridging step utilizes end-sequence information from fosmid-sized clone inserts to connect chromosomal loci brought together at aberrant rearrangement-induced joins in the cancer genome. End-sequences of breakpoint-spanning fosmids were recognized as those that do not map onto the reference genome in a manner consistent with the clone insert size or end-sequence orientation. The outlining step involves a precise localization of breakpoint sites by mapping short tags generated by the 454 Life Sciences (Roche) pyrosequencing machine onto the reference genome.

As illustrated in Figure 1B, three pools, each containing 192 BACs containing putative rearrangements, were constructed for the purpose of massively parallel sequencing using the 454 GS sequencing machine. Approximately 300,000 short (~100-bp) reads were sequenced from each pool, providing ~1× sequence coverage for the purpose of outlining. Six 96-BAC pools were formed from the same set of BACs for the purpose of fosmid library preparation, end-sequencing and bridging. Approximately 8000 to 10,000 fosmid inserts from each of the six pools were end-sequenced, providing 24× clone coverage and ~1× sequence coverage for the purpose of bridging.

Upon sequencing, the fosmid end-reads and the 454 reads together with the BAC end-sequences produced by Volik et al. (2003, 2006) were mapped onto the reference human genome. Independent aberrant mapping of two fosmids across a specific putative breakpoint was considered to constitute sufficient evidence to declare the breakpoint. BAC or fosmid ends that map onto different chromosomes are interpreted as interchromosomal breakpoints. The outlined regions were bridged using end-sequences from BACs and fosmids. The combination of outlining and bridging enabled identification of breakpoint locations down to a PCR-able distance. As indicated in Figure 1C, out of the total of 410 detected breakpoints, 157 could be confirmed by PCR across breakpoint joins as likely distinct somatic mutations. As indicated by the bars in the middle of Figure 1C, the remaining breakpoints failed the confirmation process for a number of different reasons, as we explain next.

A total of 47 breakpoints could not be unambiguously resolved down to a PCR-able distance using the outlining method. PCR primers were designed for the remaining breakpoints using a semi-automated primer design pipeline. When applied to pooled BACs, PCR primers failed to generate amplicons in expected size range for 23 predicted breakpoint joins. Further confirmation included amplification of a pool of genomic DNA from six MCF-7 cell lines (B, BK, C, D, L, and Neo). DNA isolated from MCF-10A and normal human female DNA (Novagen) were used as negative controls. A total of 123 PCR primer pairs that produced amplicons from the BAC pool did not produce amplicons from the genomic DNA derived from cell pools. A majority of these breakpoint sites contained HindIII restriction sites. Since the BAC library was prepared using HindIII partial-digestion of genomic DNA, those breakpoints were most likely created by fusion of digestion products in the course of BAC library preparation. Other sources of this discrepancy may include a number of cell line–specific aberrations generated over a number of passages that preceded preparation of the BAC library.

To identify structural polymorphic variants present in the germline of the MCF-7 donor, PCR amplification of breakpoint joins was performed on a pool of 90 Caucasian HapMap genomes (International HapMap Consortium 2005). Additionally, search for occurrences of the apparently somatic joins was performed in publicly available genomic sequences using the Pash program (Kalafus et al. 2004). A total of 40 apparently aberrant joins were present in the HapMap samples, as indicated by the presence of a PCR product, and thus correspond to structural alleles different from the structural alleles represented in the reference genome assembly. Finally, some breakpoints were identified to occur in more than one BAC, and the count was reduced by 20 to eliminate multiple counting, resulting in a total of 157 unique confirmed somatic breakpoint joins in the MCF-7 genome. Of the 157 MCF-7 somatic breast cancer breakpoints, 74 (47%) formed interchromosomal and 83 (53%) intrachromosomal joins, as illustrated in Figure 2, A and B.

A majority of the somatic breakpoints could be assigned to specific BACs

If a chromosomal segment outlined by 454 reads connected a BAC end-sequence and a breakpoint-spanning fosmid end-sequence,

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

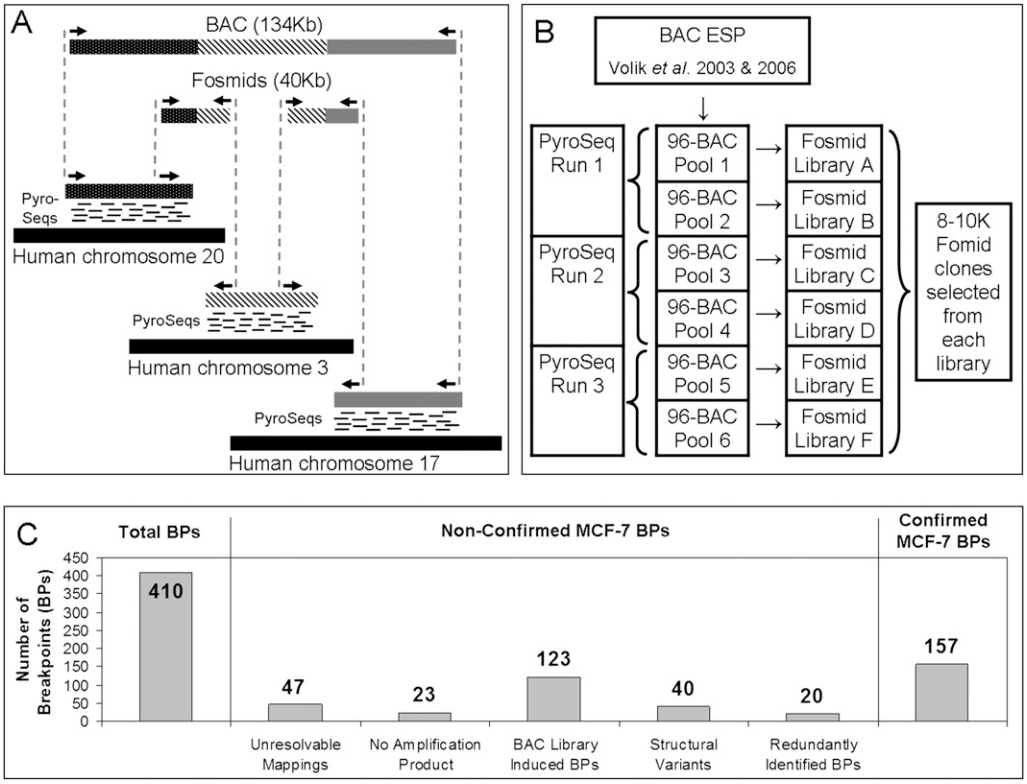

Figure 1. (*A*) An illustration of the principle of the method. Breakpoints within a BAC containing segments from chromosomes 20, 3, and 17 are detected using a combination of "bridging" and "outlining" steps. The bridging step maps fosmid end-sequences onto the reference genome. The outlining step maps short tags (labeled "PyroSeqs") using 454 technology from the BAC (in practice a pool of BACs) onto the reference genome. The results of bridging and outlining jointly allow precise mapping of breakpoints and reconstruction of rearranged BACs. (*B*) Organization of the mapping experiment. The nonredundant collection of 552 rearrangement containing BACs, 17 normal BAC negative controls, and seven positive controls was arrayed in six 96-well plates and pooled as indicated. Three 454 sequencing reactions (involving BACs pooled from plate pairs) produced tags for the purpose of outlining. Six fosmid libraries (one from each 96-well plate pool of BACs) were constructed for Sanger-based sequencing of fosmid ends and bridging. (*C*) Bar charts detailing the classification of detected MCF-7 breakpoints.

the breakpoint could be associated with the BAC. Out of 552 pooled BACs, at least one breakpoint could be assigned to 316 (57%) of them. The remaining BACs fall into the following two groups: First, in 129 (23%) cases, breakpoint assignment was inconclusive due to ambiguous mapping of reads onto the reference genome, mostly due to repetitive DNA regions, apparent overlaps between BACs, and other causes; second, in 107 (20%) cases, a single outlining block connected BAC ends, thus indicating lack of any rearrangement, contrary to previous reports (Volik et al. 2003, 2006).

To examine the source of the disagreement with the previous reports, the 107 disagreements were examined in detail. Most of the disagreements could be explained either by the differences between reference genome assemblies used in the previous and current studies or by mismapping of BAC-end sequence reads or by a combination of the two factors. Assemblies used in the previous studies were NCBI Build 30 of June 2002 (Volik et al. 2003) and NCBI Build 34 of July 2003 (Volik et al. 2006), while our study employed NCBI Build 36 of March 2006. The newer assembly is more likely to be more correct and complete, but some of the disagreements may also be explained by the presence of different structural alleles at sites of structural polymorphisms. The disagreements tended to occur in regions containing low copy repeats (LCRs). For example, Volik et al. (2003) identified MCF-7 BAC 9I10 as bridging apparent translocation t(11;11)(p11.12;q14.3) and apparently confirmed the rearrangement by fluorescent in situ hybridization (FISH). Examination of Build 36 reveals copies of an LCR at both 11p11.12 and 11q14.3. The LCR was absent from Builds 30 and 34, thus explaining the aberrant BAC-end sequence mapping and even the erroneous "confirmation" by FISH.

Examination of breakpoint sequences reveals signatures of DSB repair

To examine breakpoints at the sequence level, all the 157 breakpoint-spanning amplicons were used as substrates for sequencing from both ends. Most amplicons were of small enough size (less than 1 kb on average), allowing the Sanger read from at least one of the ends to reach the breakpoint. Difficultly of sequencing across breakpoints has been documented (Lee et al. 2007; Liu and Carson 2007), especially in repeat-rich regions. To ameliorate the problem, we sequenced DNA from specific BAC pools and employed nested sequencing primers in cases of first-pass sequencing failures. Breakpoint-straddling sequence could be obtained from 86 (55%) amplicons and could not be obtained for the remaining 71 (45%). Many of the failures were due to inability to design unique primers for sequencing across breakpoints that fall within repeat-rich regions.

Examination of 86 breakpoints that could be resolved to the base pair level (summarized in the chart in the middle of Fig. 2B)

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

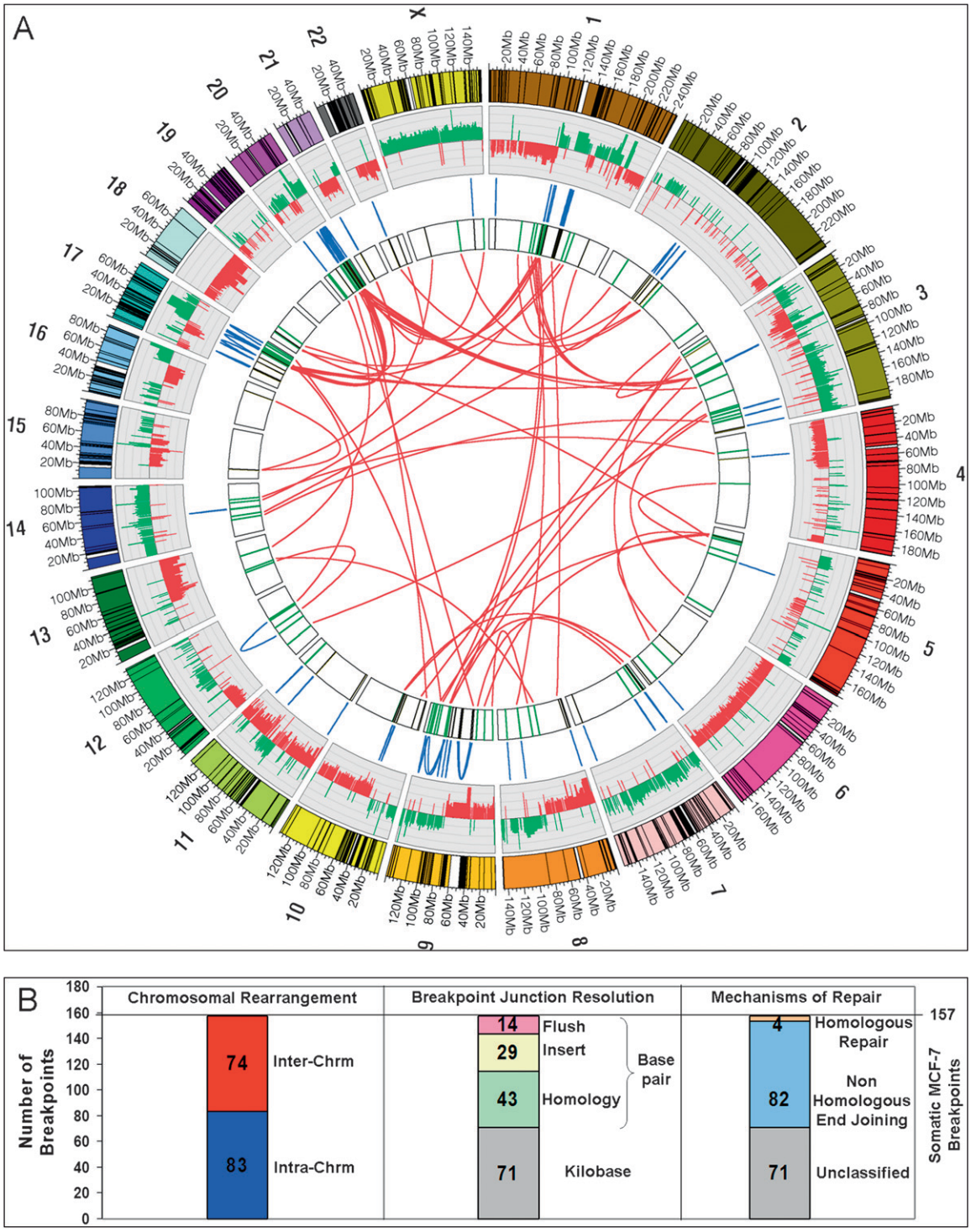

Figure 2. (*A*) Circular visualization of the MCF-7 genome obtained using Circos software. Chromosomes are individually colored with centromeres in white and LCR regions in black. MCF-7 BAC array comparative genome hybridization data (Jonsson et al. 2007) are plotted with gains in green and losses in red using $\log_2$ratio. The inner chromosome annotations depict 157 somatic MCF-7 breast tumor chromosomal rearrangements associated with LCRs (black) and breakpoints not associated with LCRs (green). Chromosomal rearrangements are depicted on each side of the MCF-7 breakpoints; intra-chromosomal rearrangements (blue) are located outside and interchromosomal rearrangements (red) are located in the center of the circle. (*B*) Bar charts indicating classification of somatic breakpoints in MCF-7.

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

revealed 14 flush joins without evidence of microhomology or intervening sequence, 29 joins with intervening inserts of unknown genomic origin averaging over 100 bp in length, and 43 joins where the joined segments exhibit homology. The extent of homology was in most (88%) cases restricted to ≤7 bp, consistent with microhomology observed in double-stranded breaks repaired by nonhomologous end-joining (NHEJ) or template switching (Sonoda et al. 2006). Due to the absence of straddling sequence, the remaining 71 breakpoints could only be analyzed at the ~1-kbp level of resolution.

Out of the 86 somatic breakpoints isolated to base pair resolution, only four (5%) exhibited sequence patterns—sequence identity and equal crossover between two homologous loci—-consistent with nonallelic homologous recombination (NAHR) (chart on the right of Fig. 2B). The dominant mechanism responsible for the repair of double-strand breaks in MCF-7 therefore appears to be NHEJ or template switching.

Two distinct types of breakpoints exist in MCF-7–clustered and LCR-associated

As evident from Figure 2, the breakpoints in MCF-7 are not evenly distributed across the genome. A number of clusters of closely spaced breakpoints are evident. To formally delineate the clustered breakpoints from the remainder, clusters of eight or more breakpoints that are less than 1.1 Mbp apart were identified. Four such clusters emerged in the following locations: 1p13.1-21.1, 3p14.1-p14.2, 17q22-q24.3, and 20q12-q13.33. These four rearrangement clusters, illustrated in Figure 3A, contain 43% of all MCF-7 somatic breakpoints, while representing only 1.5% of the normal reference genome.

The remaining nonclustered or dispersed breakpoints are highly associated with LCRs, showing a 5.2-fold enrichment for the presence of LCRs at the breakpoint site (P-value = 2.9×10^{-22}; see Fig. 3B). This is in contrast to the clustered breakpoints that do not exhibit enrichment for LCRs, with only five out of 68 clustered breakpoints being LCR-associated, well within the number expected by chance. Moreover, as illustrated in Figure 3C, the four clustered breakpoint locations exactly coincide with high copy number gain regions ("firestorms," the term proposed by Hicks et al. [2006]) in the MCF-7 genome described by Jonsson et al. (2007) and contain prognostic gene markers for breast cancer.

To further examine possible differences between the clustered breakpoints and the dispersed ones, we identified regions that show recurrent copy number amplification in cancer in previous studies involving 145 breast tumors and 56 breast cancer cell lines (Chin et al. 2006; Neve et al. 2006; Shadeo and Lam 2006; Jonsson et al. 2007). As illustrated in Supplemental Figure 5, almost three-fourths of breakpoints occurring in the four clusters are highly recurrently amplified (high recurrence is declared if at least 20% of the surveyed samples show amplification), a greater than twofold enrichment over other (dispersed) breakpoints. Additionally, the mean number of amplifications at each breakpoint location is significantly higher among clustered vs. dispersed breakpoints. These data suggest that genomic instability in these cluster regions is not specific to MCF-7.

Novel chimeiric transcripts could be predicted based on fusions of genomic DNA

Among the breakpoint fusions that involved genes, we first focused on those that occurred within introns and are predicted to lead to chimeric transcripts. We discovered 10 gene fusions (Table 1) where fusion breakpoints reside in intronic regions of the genes involved, implying in-frame translation of the original amino acid sequences.

To determine if the predicted chimeric mRNA transcript was created by these genomic fusions, we performed gene-specific reverse transcriptase reactions and a fusion-specific PCR on RNA extracted from MCF-7, MCF-10A, and normal breast tissue (the latter two serving as negative controls). Since the primers were designed to amplify the fusion product specifically, a band was only generated if a fusion product was present (for primers sequence see Supplemental Table 4). Out of 10 fusions, four showed a fusion mRNA transcript by RT-PCR, see Figure 4.

To identify if other sources reported the same fusion transcripts in MCF-7, other cell lines or primary tumors, we queried 70 MCF-7 and HCT116 (colon cancer) paired-end ditag fusion transcript sets reported by Ruan et al. (2007) and 237 fusion transcripts from the Cancer Genome Anatomy Project Recurrent Chromosome Aberrations in Cancer database reported by Hahn et al. (2004). Of the 10 MCF-7 gene fusions identified by our bridging and outlining method, the *BCAS3-BCAS4* fusion was found to be previously characterized Ruan et al. (2007) Interestingly, the *BCAS3-BCAS4* fusion is recurrently present in both the MCF-7 breast cancer and HCT116 colon cancer cell lines.

Some of the fusions and truncations may suppress function of normal gene product

Most fusions involve highly amplified clustered breakpoints, indicating possible positive selection and therefore functional significance. This is consistent with the fact that firestorm patterns indicate poor prognosis (Hicks et al. 2006) and that these highly amplified regions contain specific prognostic markers (Jonsson et al. 2007). However, not all the amplified loci contain oncogenes. Analysis and results below indicate that the oncogenic effects of some of the fusions may in fact be due to a suppression of normal function of a tumor suppressor gene. Observed amplification of gene fusions involving tumor suppressors is consistent with a dominant-negative effect of such gene fusions.

For example, the first two exons of *PTPRG*, comprising the carbonic anhydrase-like domain, are replaced by the first 10 exons of the unannotated inter-species *ASTN2* gene. Promoter hypermethylation in *PTPRG* in T-cell lymphoma leads to loss of gene expression and correlates with poor prognosis (van Doorn et al. 2005). Interestingly, Murine L cells producing *PTPRG* transcripts with a homozygous deletion of the carbonic anhydrase-like domain causes sarcomas in syngeneic mice (Wary et al. 1993).

To examine the effects of a possible suppression of *SULF2* function by the *ARFGEF2-SULF2* fusion, *SULF2* mRNA was knocked down using siRNA specifically targeting *SULF2* in MCF-7B, MDA MB231, and MCF-10A cells (Supplemental Fig. 6). Proliferation assays were performed on the three cell lines treated with knocked down *SULF2*, and all exhibited an advantage over the cells treated with control siRNA (Fig. 5A–C). To determine the effect on survival capabilities under stress conditions, *SULF2* siRNA and control siRNA treated cells were plated in serum-free conditions. Results indicate (Fig. 5D–F) that cells with knocked down *SULF2* survive better, and recover faster (seen by the steeper slope) in serum-free conditions then the control cells. This implies that knock-down of *SULF2* enhances survival compared to the control cells. Finally, knock-down of *SULF2* mRNA caused a twofold increase in anchorage-independent growth in MCF-7B and a threefold increase

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

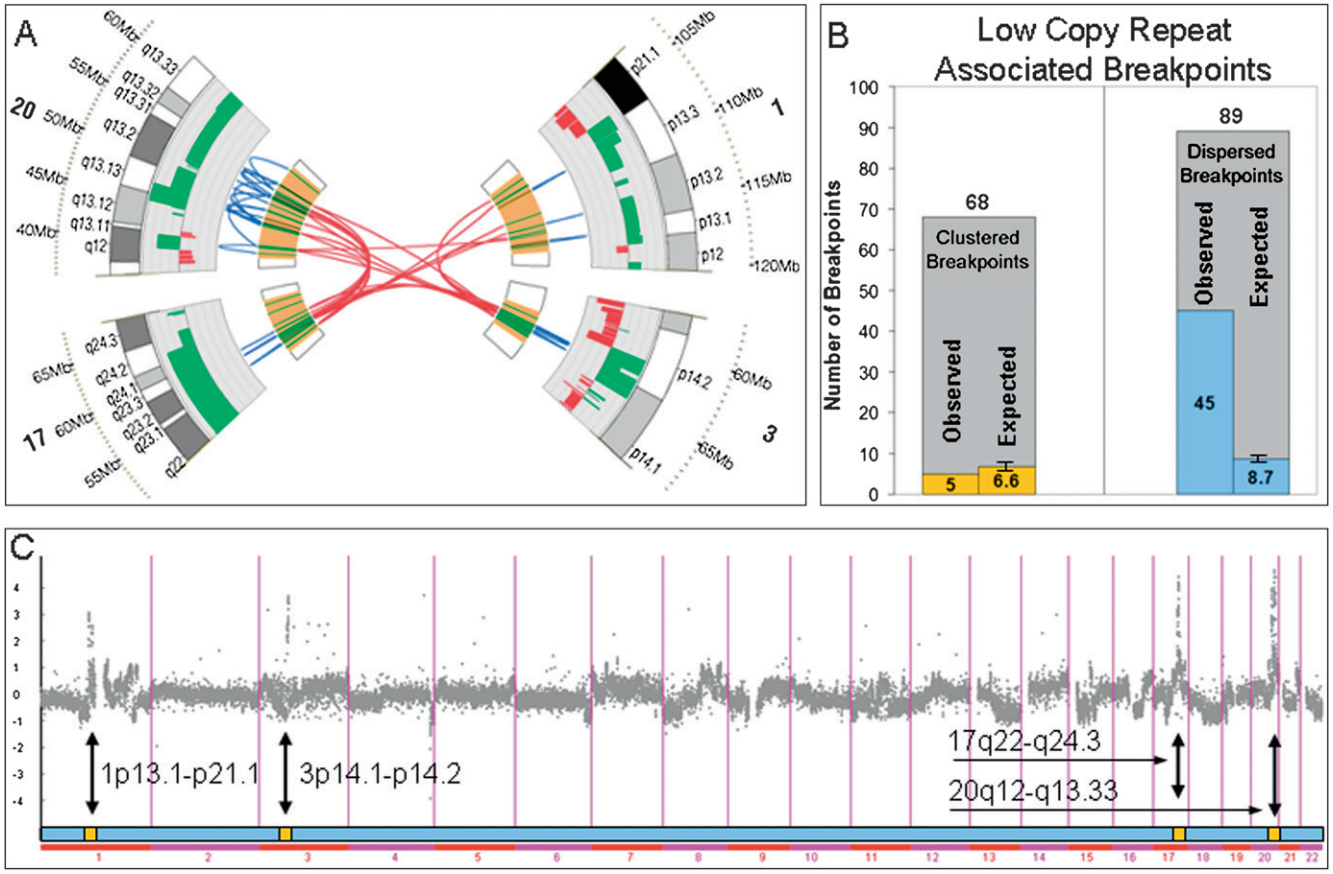

Figure 3. (*A*) Four clusters of breakpoints at 1p13.1-21.1, 3p14.1-p14.2, 17q22-q24.3, and 20q12-q13.33. (*B*) Low copy repeat (LCR) association with clustered and dispersed breakpoints. (*C*) The four clusters of breakpoints correspond exactly to the four highly amplified regions in MCF-7, as determined by array CGH.

in MDA MB231, as measured by the amount of colonies compared with controls (Fig. 5H). In summary, the data indicate that knockdown of *SULF2* causes tumorigenic phenotypes, including increased proliferation, enhanced survival, and increased anchorage-independent growth. *SULF2* may therefore act as a breast cancer suppressor.

Some genes are involved in numerous rearrangements

In addition to the 10 gene–gene fusions, a total of 77 genes were otherwise affected by the 157 breakpoints. We jointly refer to those events as "truncations" even though some, in fact, involve fusion of an upstream promoter with a protein coding gene. *PTPRG* and other genes were affected by multiple breakpoints, including both

Table 1. Gene fusions in MCF-7 that involve splicing of intact coding exons

Associated genes	Rearrangement type	Cytoband translocation	Comment
ARFGEF2-SULF2	Intrachromosomal inversion	20q13.13-20q13.13	Fusion of *ARFGEF2* exon 1 to *SULF2* exons 3–21; 1.2-Mb inversion
DEPDC1B-ELOVL7	Intrachromosomal translocation	5q12.1-5q12.1	Fusion of *DEPDC1B* N terminus exons 1–7 (out of 11) with *ELOVL7* exons 8–9
RAD51C-ATXN7	Interchromosomal rearrangement	3p14.1-17q22	Fusion of *RAD51C* exons 1–7 (out of nine) with *ATXN7* exons 6–13
SULF2-PRICKLE2	Interchromosomal rearrangement	3p14.1-20q13.13	Fusion of *SULF2* exon 1 with last exon of *PRICKLE2*
NPEPPS-USP32	Intrachromosomal inversion	17q21.32-17q23.2	Fusion of *NPEPPS* exons 1–12 (out of 23) with *USP32* exons 2–4; 13-Mb inversion
ASTN2-PTPRG	Interchromosomal rearrangement	3p14.2-9q33.1	Fusion of *ASTN2* exons 1–10 (out of 22) with *PTPRG* exons 3–30
BCAS3-BCAS4	Interchromosomal rearrangement	17q23.2-20q13.13	*BCAS4* exon 1 fused to *BCAS3* exons 23–24; also found by Ruan et al. (2007)
BCAS3-RSBN1	Interchromosomal rearrangement	1p13.2-17q23.2	Fusion of *RSBN1* first exon with *BCAS3* exons 6–24
ASTN2-TBC1D16	Interchromosomal rearrangement	9q33.1-17q25.3	Fusion of *ASTN2* exons 1–15 with *TBC1D16* exons 2–12
BCAS4-PRKCBP1	Intrachromosomal inversion	20q13.12-20q13.13	Fusion of *BCAS4* exon 1 with *PRKCBP1* exons 5–22; 3.5-Mb inversion

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

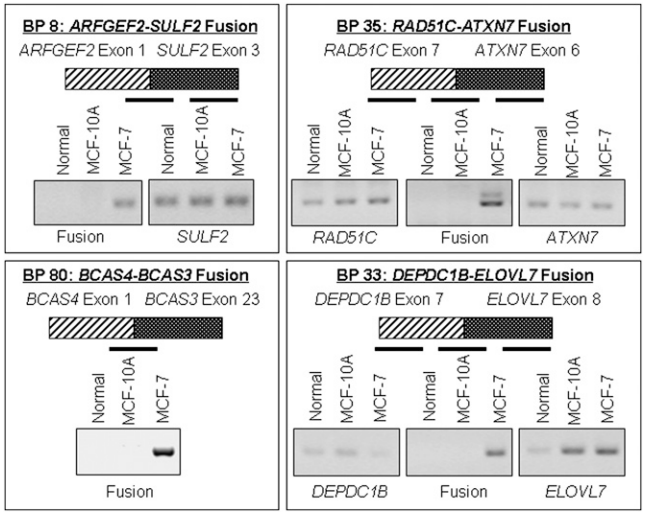

Figure 4. Confirmation of the presence of predicted processed chimeric mRNA transcripts in MCF-7 using RT-PCR.

fusion breakpoints and truncation breakpoints. The *PTPRG* breakpoints occur within the chromosome 3 breakpoint cluster and coincide within a known fragile site. Another example is the fusion of the *BMP7* promoter upstream of *ZNF217* breast cancer oncogene overexpressed in breast cancer (Collins et al. 2001) that we rediscovered but was also previously described Volik et al. (2003, 2006). The chromosome 20 rearrangement hotspot contains 37 breakpoints surrounding the *ZNF217* oncogene. Another extreme example of multiple rearrangements is the breast cancer amplified sequence 3 (*BCAS3*), occurring within the chromosome 17 rearrangement hotspot. There are seven breakpoints located within the intron–exon boundaries and an additional 19 nonfusion breakpoints surrounding the *BCAS3* gene region.

Rearrangements affect genes involved in homologous double-stranded break repair

We identified rearrangements in genes that code for members of protein complexes involved in double-stranded break repair (DSBR), raising the possibility that defects in DSBR genes may have contributed to genomic instability at certain stages of the evolution of the MCF-7 genome. One of the four MCF-7 gene fusions that produced a detectable predicted chimeric transcript is an interchromosomal fusion of *RAD51C* exons 1–7 to the neuronal-specific gene *ATXN7* exons 6–13. *RAD51C* is a paralog of *RAD51*, a gene central to DNA DSBR. RAD51C is an essential component of a complex reported to be involved in resolving holiday junctions (HJs) formed during DSBR (Liu et al. 2007) and as such is integral to the maintenance of genomic stability. The translocation we have identified eliminates the domain of *RAD51*C that binds other family member homologs such as RAD51D and Xrcc3 (Miller et al. 2004), possibly disrupting formation of the complex responsible for resolving HJs.

RAD51C is located at 17q23, a region of amplification that has been extensively studied in MCF-7 cells and breast cancer. One of the most studied oncogenes in breast cancer, *ErbB2,* is in close proximity to the 17q21.2 locus, which is amplified in a number of breast cancers (but not in MCF-7) but often independently of the 17q23 amplification. We examined *RAD51C* expression level in the microarray expression data set involving 50 breast cancer cell lines reported by Neve et al. (2006) and found that *RAD51C* levels are elevated in MCF-7, but much lower or absent in the majority of the other breast cancer cell lines.

We identified a translocation in another gene involved in DSBR, BRCA1-interacting protein-1 (*BRIP1,* also termed *BACH1*). BRIP1 was originally identified as a helicase-like protein that interacts directly with BRCA1 and contributes to its DNA repair function. BRIP1 binds to the BCRT repeat in BRCA1. The C terminus of BRIP1 is critical for its interaction with BRCA1, and a truncation mutant has been shown to block DSBR (Cantor et al. 2001; Yu et al. 2003; Lewis et al. 2005). Importantly, germline truncation mutations of *BRIP1* have been identified in familial breast cancer without mutations of *BRCA1/2*, and *BRIP1* truncations confer a twofold increased risk of developing breast cancer. We identified a translocation that results in the loss of the last three exons (exons 18–20); however, the fused DNA (3p14) downstream of *BRIP1* does not contain any exons or introns. The truncation at exon 17 of *BRIP1* would eliminate the C-terminal third of BRIP1 and eliminate binding to BRCA1. However, it is unclear at present whether the truncated mRNA would be stable as there is no transcription stop site or polyA tail.

Discussion

We have completed a sequence-level survey of rearrangements in a cancer genome. One major insight gained from this analysis is the presence of two types of breakpoints—clustered and dispersed, the latter being associated with LCRs. While we have not encountered previous reports of genome-wide association of LCRs with DSB breaks and chromosomal instability in tumors, the role of LCRs in promoting double-strand breaks through the replication fork stalling mechanism has recently been proposed in the context of genomic disorders (Lee et al. 2007).

A second major insight is that the two diverse types of breakpoints may have arisen during different stages of the evolution of the MCF-7 genome. Volik et al. (2006) hypothesized that 20q telomere loss initiated BFB cycles and a cascade of amplification resulting in small highly rearranged hotspots that colocalize DNA from different genomic regions. Our results show the same chromosomal rearrangement architecture, albeit at higher resolution and are consistent with the hypothesis that BFB cycles, possibly including extrachromosomal amplisomes, played an initial role in MCF-7 genome evolution. The chromosome 3 rearrangement hotspot encompasses the common fragile site *FRA3B*, prone to chromosomal instability, and a mediator of recurrent BFB amplification found in a variety of human tumors (Hellman et al. 2002). Recurrent breaks within common fragile sites propagated via BFB cycles amplify oncogenes and promote tumorgenesis (Huebner and Croce 2001; Hellman et al. 2002). Since both *RAD51C-ATXN7* fusion and *BRIP1* truncation belong to clusters possibly generated by the BFB mechanism, a possible effect is failure of the HR mechanism of DSBR and a consequent switch to NHEJ repair at stalled replication forks. A similar previously observed precedent is the switch from HR to NEHJ in *RAD54* homolog mutants (Sonoda et al. 2006). The switch to NHEJ at some point in the evolution of MCF-7 would have resulted in a mutator phenotype (Loeb 2001) and a pattern of extensive chromosomal rearrangements observed in MCF-7.

The switch to the rearrangement-creating NHEJ would have exposed the most breakage-prone sites—those containing LCRs—by converting simple replication-associated breaks into detectable

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

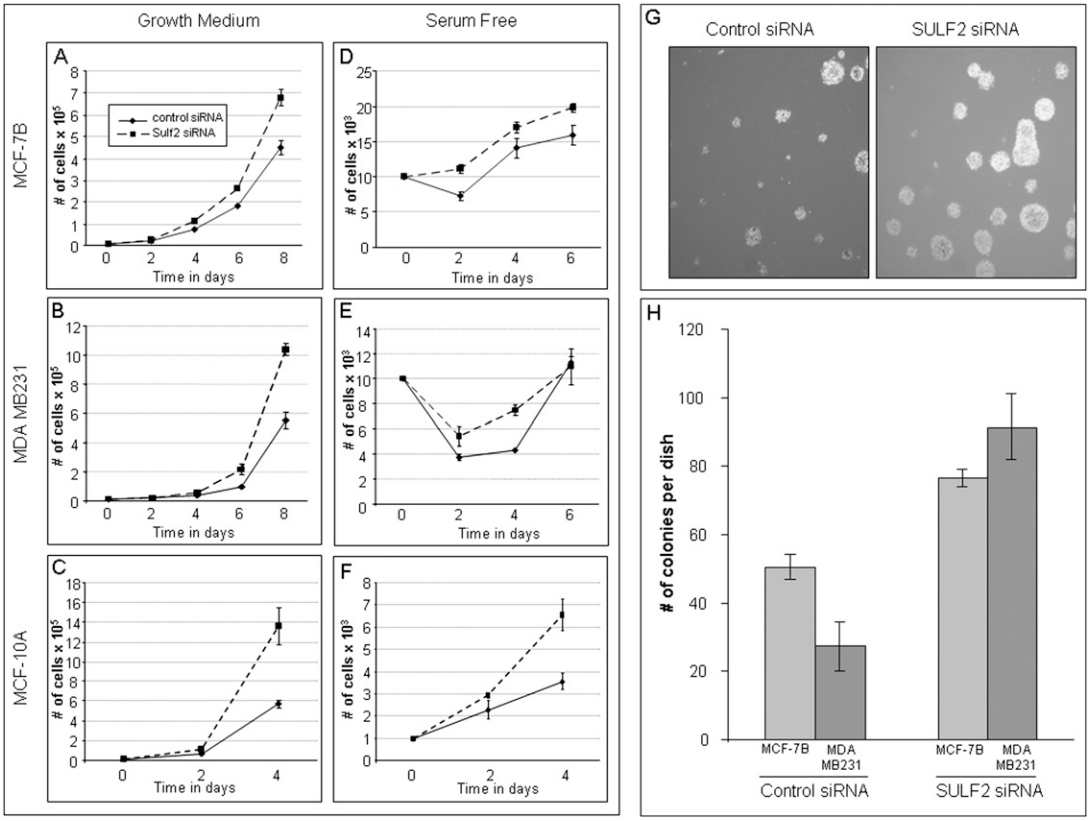

Figure 5. (*A*–*C*) Cells treated with *SULF2* siRNA have an enhanced proliferation compared with cells treated with control siRNA. MCF-7B (*A*; Mao et al. 2005), MDA MD231 (*B*), and MCF-10A (*C*) cells were transfected with 50 nM *SULF2* or control siRNA; 10^4 cells were plated in medium containing 10% FBS 48 h after transfection of the siRNA. Cells were counted on day 2, 4, 6, and 8. Experiments performed in triplicate; error bars show standard deviation. (*D*–*F*) Cells treated with *SULF2* siRNA have an enhanced survival compared with cells treated with control siRNA. MCF-7B (*D*; Mao et al. 2005), MDA MD231 (*E*), and MCF-10A (*F*) cells were transfected with 50 nM *SULF2* or control siRNA; 10^4 cells were plated in serum-free medium 48 h after transfection of the siRNA. Cells were counted on day 2, 4, and 6. Experiments performed in triplicate. Error bars, SD. (*G*,*H*) Treatment of MCF-7B and MDA MB231 cells with siRNA for *SULF2* increases the anchorage-independent growth capabilities. After treatment with siRNA, 10^4 cells were plated in 0.3% agar in growth medium, MCF-7B colony formation is shown in *G*. Plates were incubated for 21 d, and colonies were counted; bar chart results shown in *H*. Experiments performed in triplicate. Error bars, SD.

rearrangements. An analogy here exists between LCRs and DSB repair on one hand and microsatellites and mismatch repair on the other (Lengauer et al. 1998): By presenting challenges to DNA replication, LCRs and microsatellites, expose weaknesses in DSB repair and mismatch repair mechanisms, respectively. We should note that our extensive sequencing did not indicate increased mutability of MCF-7 at the base pair level, indicating highly functional mismatch repair.

The two-stage model also accounts for the typical curve indicating increase in genome complexity during the typical evolution of a breast cancer genome (Chin et al. 2004). While the BFB may account for the steep slope of rise in genomic complexity in MCF-7 during the stage of in situ carcinoma and telomere crisis, the subsequent instability mediated by the failure of the homology-based DSB repair mechanism resulting in breaks at LCR loci may account for the subsequent less steep slope that typically follows completion of the telomere crisis stage and accompanies metastasis. The two-stage model is also consistent with ongoing plasticity of the MCF-7 genome as evidenced by polyclonality and divergence of MCF-7 sublines (Jones et al. 2000; Nugoli et al. 2003).

The third insight is abundance of genes affected by rearrangements, and particularly of gene fusions, which exceeds current estimates of the abundance of gene fusions in breast cancer (Mitelman et al. 2007). Our unbiased screen of MCF-7 cell lines identified seventy nine genes involved in rearrangement events. Ten gene fusions were identified, nine novel and one previously reported by Ruan et al. (2007), and 77 other fusions involving genes and gene truncations.

The fourth insight is that at least a fraction of genes affected by fusions and truncations may in fact be tumor suppressors (e.g., *PTPRG*, *SULF2*) or may be responsible for genome stability (e.g., *RAD51C*, *BRIP1*). Both *BRIP1* and *RAD51C* fall within the cluster of breakpoints at 17q23 and are amplified in MCF-7 cells, indicating possible positive selection for the amplification. Such positive selection would be consistent with previously reported dominant-negative effects observed in genes responsible for genome stability (Milne and Weaver 1993).

The fifth insight is that chimeric transcripts can in fact be discovered by directly mapping rearrangements at the level of genomic DNA and then predicting specific chimeric transcripts. This opens the possibility of discovering recurrent, mechanistically and prognostically significant rearrangements by simply mapping a sufficient number of genomes and directly observing recurrent events.

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

In conclusion, this study validates the utility of mapping rearrangements in cancer genomes by providing mechanistically significant insights into cancer evolution and identifying genes likely involved in cancer progression. Building on the benchmarks developed in this study, next steps include technological and methodological improvements that will allow scale-up to whole genomes and to multiple cell lines and tumor samples at a more affordable cost, thus broadening applications in the research context and eventually in clinical settings.

Methods

Fosmid library preparation and end-sequencing of clone inserts

Fosmid libraries were prepared from each of the six 96-BAC pools indicated in Figure 1B using the Epicentre EpiFOS Fosmid Library Production Kit.

DNA sequencing

The ends of fosmid inserts were obtained using Sanger-based sequencing on an ABI 3730XL. Approximately 300,000 short (100-bp) reads were obtained from each of the three 192-BAC pools indicated in Figure 1B using the 454 Life Sciences (Roche) GS machine. Detailed sequencing statistics are included in the Supplemental Table 1. The sequencing reads are available for download from the public project pages at http://www.genboree.org.

Mapping reads onto the reference genome

Fosmid-end reads, 454 Life Sciences (Roche) shotgun reads, and BAC-end reads were mapped onto the reference human genome (March 2006 assembly, Build 36) using the BLAT program. BLAT parameters used for mapping are described in Supplementary Materials and coordinates are available through the Genboree site on the Breast Cancer project page at http://www.genboree.org.

PCR primer design pipeline

PCR primers were designed for amplifying breakpoint regions using repeat-masked human genome assembly (March 2006 assembly, Build 36) using a semi-automated primer design pipeline. Primer 3 primer design program was run to obtain a set of nested primers using two categories or parameters, "stringent" and "relaxed." Primer pairs in each category were scored, and the highest-scored primer pair was selected for initial round of PCR amplification. Priority was also given to the stringent category. In case of failure, additional lower-scoring primer pairs were employed. More details, including Primer 3 parameters, can be found in Supplemental materials.

PCR amplification of genomic DNA from cell lines

Breakpoint confirmation included PCR amplification of a pool of genomic DNA from six different sublines of MCF-7 cells (B, BK, C, D, L, and Neo). DNA isolated from immortalized but nontransformed mammary epithelial cells (MCF-10A) and normal human female DNA (Novagen) were used as negative controls. Genomic cell line DNA was isolated with the DNeasy kit (Qiagen). PCR bands were visualized on a 2% agarose gel.

Breakpoint clustering algorithm

Consecutive breakpoints that are closer than 1.1 Mbp in the reference genome assembly were connected. Runs of consecutive connected breakpoints with eight or more members are declared to constitute a cluster. Four clusters on chromosomes 1, 3, 17, and 20 indicated in Figure 3 were obtained in this fashion.

Identification of LCR regions

Each of the 157 MCF-7 breakpoints was examined for the presence of LCR. Intrachromosomal and interchromosomal LCRs were detected by applying a novel algorithmic method to the human genome assembly (March 2006 assembly, Build 36). The method involved self-comparison of the human genome using the Pash program (Kalafus et al. 2004) and an automated pipeline for segmentation, clustering, and parsing of LCRs based on sequence feature analysis. The LCRs detected by this method cover 6.15% of the whole genome in length, of which 18.7% are gene-containing regions. A detailed description of the algorithm is available in Supplemental materials.

Analysis of recurrent copy number changes in 157 somatic breakpoint loci

Copy number variation in the 157 somatic breakpoint loci identified in this study was examined. In order to identify recurrent copy number changes in breakpoint loci, array CGH data from 201 breast cancer cell lines and tumors (Chin et al. 2006; Neve et al. 2006; Shadeo and Lam 2006; Jonsson et al. 2007) were integrated. A locus was declared recurrently amplified if amplification was reported in more than 20% cases for the specific locus. Detailed results are compiled in a table where breakpoints are sorted by their level of recurrent copy number amplification (for details, see Supplemental materials and Supplemental Table 3).

Analysis of recurrent expression and copy number changes in 79 breakpoint-associated genes

Patterns of recurrent copy number and expression level variation were examined for 79 genes associated with the 157 somatic breakpoints identified in this study. Expression data from 50 breast cancer cell lines (Neve et al. 2006) were combined with copy number data from 201 breast cancer cell lines and tumors (Chin et al. 2006; Neve et al. 2006; Shadeo and Lam 2006; Jonsson et al. 2007). Detailed results are compiled in a table where genes are sorted by their level of recurrent alteration. (for details, see Supplemental Materials and Supplemental Table 2). Additionally, copy number data from an Affymetrix 100k SNP chip were used to identify breakpoint genes that also associate with regions of copy number alteration (see Supplemental Table 3).

Detection of predicted fusion transcripts by RT-PCR

mRNA from exponentially growing MCF-7 and MCF-10A cells were isolated with the RNeasy kit (Qiagen). To determine the presence of a fusion transcript, primers were designed across the fusion point on cDNA using Primer3. Control primers were designed on either side of the fusion. cDNA was generated by using gene specific primers. PCR amplification of the mRNA was restricted to 35 cycles. PCR bands were visualized on a 2% agarose gel, and verified by sequencing to confirm that the product contained mRNA from both genes involved.

Cell growth and soft-agar experiments

For the cell growth experiments, 10,000 cells were plated in triplicate in 24-well plates. The cells were grown in growth medium, containing 10% FBS, or in serum-free medium. Growth rate was

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

measured on days 0, 2, 4, and 6 with a Coulter Counter (Beckman Coulter).

Colony growth assays were performed as followed: 1 mL of solution of 0.5% noble agar in growth or serum-free medium was layered onto 30 × 10-mm tissue culture plates. A total of 1×10^4 cells was mixed with 1 mL of 0.3% agar solution prepared in a similar manner and layered on top of the 0.5% agar layer. Plates were incubated at 37°C in 5% CO_2 for 21 d. The experiment was performed in triplicate.

Knock-down of SULF2 using short interfering RNA (siRNA)

Transfections with *SULF2* and control nonspecific siRNA (Dharmacon) were carried out using 50 nM pooled siRNA duplexes and 4 μL of Dharmafect (Dharmacon) in six-well plates according to the manufacturer's protocol. After 48 h, the cells were prepared the respective assays.

Acknowledgments

We thank Andrew R. Jackson and Dr. Manuel Gonzalez-Garay for their computational support in providing the Genboree Discovery System, and Dr. Martin Krzywinski for providing the Circos circular genome visualization software. This project was funded by the NIH-NHGRI grant 1 R01 HG02583 and NIH-NCI grants R33 CA114151 and R21 CA128496 to A.M.

References

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.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., et al. 2008. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40:** 722–729.

Cantor, S.B., Bell, D.W., Ganesan, S., Kass, E.M., Drapkin, R., Grossman, S., Wahrer, D.C., Sgroi, D.C., Lane, W.S., Haber, D.A., et al. 2001. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105:** 149–160.

Chin, K., de Solorzano, C.O., Knowles, D., Jones, A., Chou, W., Rodriguez, E.G., Kuo, W.L., Ljung, B.M., Chew, K., Myambo, K., et al. 2004. In situ analyses of genome instability in breast cancer. *Nat. Genet.* **36:** 984–988.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., et al. 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10:** 529–541.

Collins, C., Volik, S., Kowbel, D., Ginzinger, D., Ylstra, B., Cloutier, T., Hawkins, T., Predki, P., Martin, C., Wernick, M., et al. 2001. Comprehensive genome sequence analysis of a breast cancer amplicon. *Genome Res*. **11:** 1034–1042.

Hahn, Y., Bera, T.K., Gehlhaus, K., Kirsch, I.R., Pastan, I.H., and Lee, B. 2004. Finding fusion genes resulting from chromosome rearrangement by analyzing the expressed sequence databases. *Proc. Natl. Acad. Sci*. **101:** 13257–13261.

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.

Hicks, J., Krasnitz, A., Lakshmi, B., Navin, N.E., Riggs, M., Leibu, E., Esposito, D., Alexander, J., Troge, J., Grubor, V., et al. 2006. Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res*. **16:** 1465–1479.

Huang, J., Wei, W., Zhang, J., Liu, G., Bignell, G.R., Stratton, M.R., Futreal, P.A., Wooster, R., Jones, K.W., and Shapero, M.H. 2004. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics* **1:** 287–299.

Huebner, K. and Croce, C.M. 2001. FRA3B and other common fragile sites: The weakest links. *Nat. Rev. Cancer* **1:** 214–221.

International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* **437:** 1299–1320.

Jones, C., Payne, J., Wells, D., Delhanty, J.D., Lakhani, S.R., and Kortenkamp, A. 2000. Comparative genomic hybridization reveals extensive variation among different MCF-7 cell stocks. *Cancer Genet. Cytogenet*. **117:** 153–158.

Jonsson, G., Staaf, J., Olsson, E., Heidenblad, M., Vallon-Christersson, J., Osoegawa, K., de Jong, P., Oredsson, S., Ringner, M., Hoglund, M., et al. 2007. High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. *Genes Chromosomes Cancer* **46:** 543–558.

Kalafus, K.J., Jackson, A.R., and Milosavljevic, A. 2004. Pash: Efficient genome-scale sequence anchoring by positional hashing. *Genome Res*. **14:** 672–678.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., et al. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318:** 420–426.

Kytola, S., Rummukainen, J., Nordgren, A., Karhu, R., Farnebo, F., Isola, J., and Larsson, C. 2000. Chromosomal alterations in 15 breast cancer cell lines by comparative genomic hybridization and spectral karyotyping. *Genes Chromosomes Cancer* **28:** 308–317.

Lee, J.A., Carvalho, C.M., and Lupski, J.R. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131:** 1235–1247.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. 1998. Genetic instabilities in human cancers. *Nature* **396:** 643–649.

Lewis, A.G., Flanagan, J., Marsh, A., Pupo, G.M., Mann, G., Spurdle, A.B., Lindeman, G.J., Visvader, J.E., Brown, M.A., and Chenevix-Trench, G. 2005. Mutation analysis of FANCD2, BRIP1/BACH1, LMO4 and SFN in familial breast cancer. *Breast Cancer Res*. **7:** R1005–R1016.

Liu, Y.T. and Carson, D.A. 2007. A novel approach for determining cancer genomic breakpoints in the presence of normal DNA. *PLoS One* **2:** e380. doi: 10.1371/journal.pone.0000380.

Liu, Y., Tarsounas, M., O'Regan, P., and West, S.C. 2007. Role of RAD51C and XRCC3 in genetic recombination and DNA repair. *J. Biol. Chem*. **282:** 1973–1979.

Loeb, L.A. 2001. A mutator phenotype in cancer. *Cancer Res*. **61:** 3230–3239.

Mao, J.H., Li, J., Jiang, T., Li, Q., Wu, D., Perez-Losada, J., DelRosario, R., Peterson, L., Balmain, A., and Cai, W.W. 2005. Genomic instability in radiation-induced mouse lymphoma from p53 heterozygous mice. *Oncogene* **24:** 7924–7934.

Miller, K.A., Sawicka, D., Barsky, D., and Albala, J.S. 2004. Domain mapping of the Rad51 paralog protein complexes. *Nucleic Acids Res*. **32:** 169–178.

Milne, G.T. and Weaver, D.T. 1993. Dominant negative alleles of RAD52 reveal a DNA repair/recombination complex including Rad51 and Rad52. *Genes & Dev*. **7:** 1755–1765.

Mitelman, F., Johansson, B., and Mertens, F. 2007. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7:** 233–245.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F., et al. 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10:** 515–527.

Nugoli, M., Chuchana, P., Vendrell, J., Orsetti, B., Ursule, L., Nguyen, C., Birnbaum, D., Douzery, E.J., Cohen, P., and Theillet, C. 2003. Genetic variability in MCF-7 sublines: Evidence of rapid genomic and RNA expression profile modifications. *BMC Cancer* **3:** 13. doi: 10.1186/1471-2407-3-13.

Raphael, B.J., Volik, S., Yu, P., Wu, C., Huang, G., Linardopoulou, E.V., Trask, B.J., Waldman, F., Costello, J., Pienta, K.J., et al. 2008. A sequence-based survey of the complex structural organization of tumor genomes. *Genome Biol*. **9:** R59. doi: 10.1186/gb-2008-9-3-r59.

Ruan, Y., Ooi, H.S., Choo, S.W., Chiu, K.P., Zhao, X.D., Srinivasan, K.G., Yao, F., Choo, C.Y., Liu, J., Ariyaratne, P., et al. 2007. Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs). *Genome Res*. **17:** 828–838.

Rummukainen, J., Kytola, S., Karhu, R., Farnebo, F., Larsson, C., and Isola, J.J. 2001. Aberrations of chromosome 8 in 16 breast cancer cell lines by comparative genomic hybridization, fluorescence in situ hybridization, and spectral karyotyping. *Cancer Genet. Cytogenet*. **126:** 1–7.

Shadeo, A. and Lam, W.L. 2006. Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res*. **8:** R9.

Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y., and Takeda, S. 2006. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair (Amst.)* **5:** 1021–1029.

Soule, H.D., Vazguez, J., Long, A., Albert, S., and Brennan, M. 1973. A human cell line from a pleural effusion derived from a breast carcinoma. *J. Natl. Cancer Inst.* **51:** 1409–1416.

Downloaded from genome.cshlp.org on May 2, 2026 . Published by Cold Spring Harbor Laboratory Press

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310:** 644–648.

van Doorn, R., Zoutman, W.H., Dijkman, R., de Menezes, R.X., Commandeur, S., Mulder, A.A., van der Velden, P.A., Vermeer, M.H., Willemze, R., Yan, P.S., et al. 2005. Epigenetic profiling of cutaneous T-cell lymphoma: Promoter hypermethylation of multiple tumor suppressor genes including BCL7a, PTPRG, and p73. *J. Clin. Oncol.* **23:** 3886–3896.

Volik, S., Zhao, S., Chin, K., Brebner, J.H., Herndon, D.R., Tao, Q., Kowbel, D., Huang, G., Lapuk, A., Kuo, W.L., et al. 2003. End-sequence profiling: Sequence-based analysis of aberrant genomes. *Proc. Natl. Acad. Sci.* **100:** 7696–7701.

Volik, S., Raphael, B.J., Huang, G., Stratton, M.R., Bignel, G., Murnane, J., Brebner, J.H., Bajsarowicz, K., Paris, P.L., Tao, Q., et al. 2006. Decoding the fine-scale structure of a breast cancer genome and transcriptome. *Genome Res.* **16:** 396–404.

Wary, K.K., Lou, Z., Buchberg, A.M., Siracusa, L.D., Druck, T., LaForgia, S., and Huebner, K. 1993. A homozygous deletion within the carbonic anhydrase-like domain of the Ptprg gene in murine L-cells. *Cancer Res.* **53:** 1498–1502.

Yu, X., Chini, C.C., He, M., Mer, G., and Chen, J. 2003. The BRCT domain is a phospho-protein binding domain. *Science* **302:** 639–642.

Received April 29, 2008; accepted in revised form November 19, 2008.