Published in final edited form as:

Nature. 2015 June 11; 522(7555): 179–184. doi:10.1038/nature14493.

CHROMOTHRIPSIS FROM DNA DAMAGE IN MICRONUCLEI

Cheng-Zhong Zhang[1,2,3,4,*], **Alexander Spektor**[2,4,5,*], **Hauke Cornils**[2,4,*], **Joshua M. Francis**[1,3,*], **Emily K. Jackson**[2,4,8], **Shiwei Liu**[2,4], **Matthew Meyerson**[1,3,6,7], and **David Pellman**[2,3,4,8,9]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts.

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute

[3]Broad Institute of Harvard and MIT, Cambridge, Massachusetts

[4]Department of Cell Biology, Harvard Medical School

[5]Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[6]Department of Pathology, Harvard Medical School, Boston, Massachusetts

[7]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts

[8]Howard Hughes Medical Institute

Abstract

Genome sequencing has uncovered a new mutational phenomenon in cancer and congenital disorders called chromothripsis. Chromothripsis is characterized by extensive genomic rearrangements and an oscillating pattern of DNA copy number levels, all curiously restricted to one or a few chromosomes. The mechanism for chromothripsis is unknown, but we previously proposed that it could occur through the physical isolation of chromosomes in aberrant nuclear structures called micronuclei. Here, using a combination of live-cell imaging and single-cell genome sequencing, we demonstrate that micronucleus formation can indeed generate a spectrum of genomic rearrangements, some of which recapitulate all known features of chromothripsis. These events are restricted to the missegregated chromosome and occur within one cell division. We demonstrate that the mechanism for chromothripsis can involve the fragmentation and subsequent reassembly of a single chromatid from a micronucleus. Collectively, these experiments establish a new mutational process of which chromothripsis is one extreme outcome.

[9]Correspondence: david_pellman@dfci.havard.edu.
[*]Co-first authors

AUTHOR CONTRIBUTIONS

D.P and M.M. initiated the project. D.P. conceived the idea and supervised all aspects of the work. H.C, J.F., D.P, A.S., and C.-Z. Z. designed the experiments. D.P., A.S., C.-Z. Z., and M.M., wrote the manuscript, with edits from all authors. C.-Z. Z. developed and performed the computational analysis; H.C. and A.S. developed and performed the Look-Seq protocol; E.J., S. L. and A.S. performed the cell biological analysis experiments in Fig. 1 and Extended Data Fig. 1. J. F. and H. C. generated sequencing libraries; J. F. and C.-Z. Z. designed and performed the PCR validation of rearrangements.

AUTHOR INFORMATION

The sequencing data in this study have been deposited into the Sequence Read Archive (SRP052954). M.M. is a founder and equity holder of Foundation Medicine, a for-profit company that provides next-generation sequencing diagnostic services.

Most cancer genomes are extensively altered by point mutations and chromosome rearrangements. Although mutations are generally thought to accumulate gradually, over many cell division cycles[1,2], recent cancer genome sequencing provides evidence for mutational processes that generate multiple mutations “all-at-once”, during a single cell cycle[3]. The most striking example of such an event is “chromothripsis”, where a unique pattern of clustered rearrangements occurs, typically involving only a single chromosome or a few chromosomes[4-7].

Several models have been proposed to explain the rearrangements in chromothripsis. One proposal is that the affected chromosome is somehow fragmented, with random joining of some segments and loss of others[4]. This model explains the characteristic pattern of DNA copy number in chromothripsis—oscillation between two copy number states, with islands of DNA retention and heterozygosity interspersed with regions of DNA loss. An alternative hypothesis is that chromothripsis is generated by DNA replication errors: Collapsed replication forks trigger cycles of microhomology-mediated break-induced replication (MMBIR), where distal sequences are copied to the sites of replication fork collapse by template-switching[8]. Evidence for the latter model comes from templated insertions detected at translocation junctions and sequence triplications[8,9]. Both models have only indirect support from genomic sequencing and have not been tested experimentally[10].

We recently proposed that the physical isolation of chromosomes in aberrant nuclear structures called micronuclei might explain the localization of DNA lesions in chromothripsis[11]. Micronuclei are a common outcome of many cell division defects, including mitotic errors that missegregate intact chromosomes, and errors in DNA replication or repair that generate acentric chromosome fragments[12,13]. We previously found that the partitioning of intact chromosomes into newly formed micronuclei leads to cytological evidence of DNA damage, specifically on the missegregated chromosome[11]. After mitosis, chromosomes from micronuclei can be reincorporated into daughter nuclei[11], potentially integrating mutations from the micronucleus into the genome.

Here, using an approach combining live-cell imaging with single-cell genomic analysis that we call “Look-Seq”, we demonstrate that micronucleus formation can generate a spectrum of complex chromosomal rearrangements, providing the first direct experimental evidence for a mechanism leading to chromothripsis.

DAMAGE TO MICRONUCLEI AFTER S PHASE ENTRY

To determine if micronucleus formation leads to chromosome rearrangements, we first sought to clarify the cell population where rearrangements would most likely occur. Previously, we found that newly-formed micronuclei do not have significant levels of DNA damage in G1, but damaged micronuclei accumulate as cells progress into the S and G2 phases of the cell cycle[11], suggesting a link between DNA damage and DNA replication. Additionally or alternatively, the nuclear envelopes of micronuclei are prone to irreversible “rupture” as defined by the abrupt loss of soluble nuclear proteins[14]. Nuclear envelope rupture in micronuclei is strongly associated with DNA damage, but occurs at random, not specifically during S phase[14].

To reexamine the timing of DNA damage, micronuclei were generated in synchronized cells by a nocodazole release procedure[11]. As expected[11,14], no significant DNA damage was detected in ruptured micronuclei during G_1, but damage was common during S and G_2 phases as indicated by fluorescence labeling for γ-H2AX, or Gam, a bacteriophage protein that marks double strand breaks[15] (Extended Data Fig. 1a,b). Moreover, micronuclei from serum-starved G_0 cells showed little detectable DNA damage, despite rupture of the micronuclear envelope during G_0[14] (Extended Data Fig. 1c). Therefore, DNA damage is not triggered by nuclear envelope rupture alone, but also requires entry into S phase.

Consistent with this conclusion, EdU-labeling demonstrated that most damaged micronuclei had initiated DNA replication (Extended Data Fig. 1d). However, overall EdU incorporation was markedly lower in micronuclei as compared to the cell's primary nucleus, irrespective of whether the micronuclei were ruptured or intact[11,14] (Fig. 1a). Thus, chromosomes in micronuclei are underreplicated in general, even though the majority of the damaged micronuclei have initiated DNA replication. These results focused the experiments below on micronuclei that rupture after S phase entry.

LOOK-SEQ STRATEGY

We designed a procedure to determine the genomic consequences of DNA damage in ruptured micronuclei (Fig. 1b). Non-transformed human RPE-1 cells were synchronized by nocodazole release, sorted into 384 well plates, and wells containing a single micronucleated cell were identified. By live-cell imaging, we identified cells where the micronuclear envelope ruptured after the beginning of S phase (Methods). These experiments were performed after siRNA-mediated knockdown of p53 because chromosome missegregation after nocodazole release induces a p53-dependent G1 cell cycle arrest[16,17]. After one division of the micronucleated cell, we selected daughters with no detectable micronuclei, indicating the micronuclear chromosome was reincorporated into daughters' primary nuclei. Cells with reincorporation were selected because nuclear envelope rupture inactivates nuclear processes such as DNA replication and transcription[14], and we presume that the damaged chromosome from the micronucleus may require exposure to a normal nucleoplasm to undergo repair and generate rearrangements. Daughter cells were then separated, amplified (multi-strand displacement amplification, MDA), sequenced, and analyzed independently[18] (Methods, Supplementary Table 1).

IDENTIFYING THE MICRONUCLEAR CHROMOSOME

Micronucleated cells will be of two kinds: disomic for the chromosome in the micronucleus (Fig. 2a, left scheme), if the lagging chromosome was segregated into the correct daughter cell (albeit partitioned into a micronucleus); or trisomic for the affected chromosome (Fig. 2a, right scheme), if the lagging chromosome was missegregated. The division of a disomic micronucleated cell will produce one near disomic daughter with the underreplicated chromosome from the micronucleus, and one monosomic daughter, hereafter referred to as a 2:1 asymmetric copy number pattern. Similarly, the division of a trisomic micronucleated cell will produce one near trisomic daughter with the micronuclear chromosome, and one disomic daughter, or a 3:2 pattern. In either pattern, we refer to the daughter cell with the

higher DNA copy number as the “plus” cell and the other daughter as the “minus” cell. Thus, reduced DNA replication in the ruptured micronucleus creates an odd chromosome number and a copy number asymmetry between the daughters. This asymmetry identifies the missegregated chromosome and demonstrates that missegregation is a *de novo* event that occurred during the last cell division (Fig. 2a). Because the plus daughter has the chromosome from the micronucleus, rearrangements, if observed, should be concentrated in the plus daughter and associated with the “extra” chromatid from the micronucleus; we refer to the haplotype of this chromatid as the “gained” or “missegregated” haplotype.

We sequenced 10 control daughter cell pairs from non-micronucleated mother cells and 9 experimental pairs derived from micronucleated cells (Supplemental Videos 1-9). All controls underwent RNAi-mediated knockdown of p53 (C1-C4 and N1-N6), and six of these also were carried through the nocodazole release protocol (N1-N6). DNA copy-number analysis (Methods, Supplementary Table 2) identified the known clonal gains of Chr. 10q and subclonal gains of Chr. 12 in the RPE-1 line (Ref. 18 and Fig. 2b), validating our arm-level copy number measurements. As expected, sporadic chromosome gains or losses—shared between the daughters—were also detected. These shared aneuploidies were presumably present in the mother-cell primary nucleus and underwent normal replication and segregation.

In contrast to the non-micronucleated controls (Fig. 2b, top and middle panels), the 9 daughter pairs derived from cells with ruptured micronuclei all contained at least one chromosome with copy number asymmetry, with either a 2:1 (MN1-6) or 3:2 (MN7-9) segregation pattern (Fig. 2b, bottom panel). In the pair of MN8 daughter cells, the 3:2 ratio for both Chr. 4 and Chr. 11 (Fig. 2b) suggests that these chromosomes were both in the micronucleus observed by imaging. In the MN7 daughters, we observed a 3:2 ratio only for the q arm of Chr. 1 (Fig. 2b; also see Extended Data Fig. 3). Most likely, an acentric Chr. 1q fragment was generated by the cleavage of a chromosome bridge from the previous mitosis[19], and partitioned into the micronucleus in the MN7 mother.

We developed a method to determine loss-of-heterozygosity (LOH) in single-cell genomes (Methods) that is insensitive to the amplification bias inherent to MDA[20]. This analysis confirmed genuine monosomy of chromosomes in the minus daughters of 2:1 missegregations (Extended Data Fig. 2a-c). From the sequencing of these hemizygous chromosomes we determined the haplotype phase (genotypes at polymorphic sites for each homologue) and devised a method to measure the copy number for each homologue (Extended Data Fig. 3a, Methods). This haplotype copy number information enabled us to identify amplification bias affecting both homologues equally (Extended Data Fig. 3b) and distinguish it from true copy number alterations affecting one homologue (Extended Data Fig. 3c); this analysis further confirmed the predicted 3:2 missegregation pattern (Fig. 2a, right scheme).

LOCALIZED CHROMOSOME REARRANGEMENTS

We next tested the prediction that there would be a concentration of rearrangements on the missegregated chromosome. *De novo* rearrangements were detected by clustering of

discordant read pairs[18] (Methods, Supplementary Table 3). Rearrangements on the normally segregated control chromosomes were uniformly distributed (Extended Data Fig. 4a), as would be expected for background errors from MDA-based whole genome amplification. By contrast, there was a significant enrichment of rearrangements (median: 12.5 fold) on the missegregated chromosomes identified by copy number asymmetry relative to the normally segregated control chromosomes (Extended Data Fig. 4b, top panel). Remarkably, the concentration of rearrangements on the missegregated chromosomes was observed in 8 of 9 post-micronuclear cell pairs ($p < 10^{-4}$, Bonferroni corrected one-sided Poisson test).

As short-range inversions are amplification errors reported to occur frequently during MDA, we analyzed inverted-type and non-inverted type rearrangements separately (Extended Data Fig. 5a-c, Methods). Power-law scaling analysis revealed that inversions were enriched at breakpoint distances <150 kb (Extended Data Fig. 5a,b) and were randomly distributed (Extended Data Fig. 4b), as expected for MDA errors (Methods). Long-range rearrangements (>150 kb or interchromosomal) on the normally segregated chromosome were also randomly distributed (Supplementary Tables 4 and 5). By contrast, the enrichment of long-range rearrangements specifically on the missegregated chromosomes was even more significant after elimination of short-range rearrangements (Fig. 3a, Extended Data Fig. 4b).

As predicted, the rearrangements in the missegregated chromosome occurred predominantly in the plus daughter cell (Fig. 3b, Extended Data Fig. 4c), with a few informative exceptions to be discussed below. PCR amplification across rearrangement junctions from whole genome-amplified DNA (Extended Data Fig. 6a) confirmed the rearrangements, but could not exclude amplification errors. However, by sequencing rearrangement junctions with nearby heterozygous sites (Extended Data Fig. 6b-d, Methods), we found that every rearrangement tested was associated with the gained haplotype (Extended Data Table 1, Supplementary Table 5), indicating that the rearrangements occurred on the missegregated chromatid. Interestingly, we sometimes detected an unaltered product in addition to the rearranged product with the missegregated haplotype. We hypothesize that these two products may be generated by breakage of a partially replicated sequence near a replication fork with only one side of the break participating in the rearrangement (Extended Data Fig. 6e). Thus, there is a marked concentration of long-range rearrangements associated with the gained haplotype in the plus daughter cell, indicating that these rearrangements originate from the breakage of the chromatid in the ruptured micronucleus.

The translocation junctions for long-range rearrangements had several notable features. Unlike control chromosomes, microhomology was observed at >50% of the junctions from the missegregated chromosomes, a higher frequency than expected by chance[21] (Extended Data Fig. 7a). Microhomology could originate from alternative non-homologous end joining or MMBIR[8]. In the MN6 plus daughter, 14 out of 16 breakpoints formed an uninterrupted chain (Extended Data Fig. 7b). These chained translocations resemble examples of germline chromothripsis[22] that are presumed to preserve chromosome copy number because of the selection for viability[3]. At some translocation junctions, we also identified short insertions (50-500 bp) originating from other widely dispersed sites on the missegregated chromosome (Extended Data Fig. 7c). In one example (MN9), 8 short segments from all over Chr. 8 were

inserted into a single junction, also on Chr. 8. Such short “templated” insertions are a well-described characteristic of MMBIR[9].

In the MN8 daughters, both Chrs. 4 and 11 were missegregated with a 3:2 ratio (Fig. 2b). Interestingly, in the plus cell, we not only detected intrachromosomal rearrangements in Chrs. 4 and 11, but also 8 translocations between these chromosomes (Fig. 4a, left panel). Thus, the MN8 mother probably had copies of Chrs. 4 and 11 within the single micronucleus detected by imaging, generating translocations both within and between chromosomes.

EVIDENCE FOR CHROMOSOME FRAGMENTATION

One striking feature of chromothripsis is the oscillation of DNA copy number between only two states[7]. This feature of chromothripsis was identified on the missegregated chromosome in three cell pairs, MN2, MN4, and MN8 (Fig. 4a, Extended Data Fig. 8a-c), most strikingly in the MN4 daughters. For this pair, one intact Chr. 3 haplotype was distributed evenly to both daughters (Extended Data Fig. 8b) whereas the other haplotype displayed a pattern of alternating retention and loss (Fig. 4b). Combining the copy number for both haplotypes yields an oscillating pattern: hemizygous regions with one copy alternate with heterozygous regions with two copies. Strikingly, segments of the fragmented haplotype that were gained in one daughter were almost always lost from the other and vice versa. Of all fragments of the missegregated Chr. 3 haplotype detected in either daughter, 97% were mutually exclusively distributed between the two daughters, with only 3% shared by both. Moreover, long-range rearrangements in both daughter cells were almost completely (> 90%) restricted to regions where the fragmented haplotype was retained ($p \sim 10^{-4}$ for the plus daughter and 10^{-12} for the minus daughter, binomial test) and directly associated with the gained haplotype when adjacent polymorphisms were present (Fig. 4b). Therefore, the division of the MN4 mother cell generated the canonical features of chromothripsis in both daughters. A single chromatid from a micronucleus was fragmented and randomly distributed between daughter cells, followed by the joining of fragments in random order and orientation. Thus, the majority of DNA segments that are “lost” in one cell, are, in fact, distributed to the other.

POTENTIAL NASCENT DOUBLE MINUTES

One way that chromothripsis may promote tumor development is by generating double minute chromosomes[3,5], small circular acentric chromosomes that can be present at very high copy number and carry oncogenes[23]. Intriguingly, in the MN4 daughters, we detected four 1-2 Mb circular chromosomes, which may represent examples of the initial step in generating double minutes [24] (Fig. 5). Evidence that these are true circular chromosome fragments not only comes from sequencing reads spanning the junctions, but also because the junctions fall at boundaries where the missegregated haplotype is deleted. These deletions confirm that the junctions occur at genuine break sites and also exclude the possibility that the junctions result from tandem duplications.

DISCUSSION

The experiments described here define a new mutational process that provides one mechanistic explanation for chromothripsis, a unique pattern of localized chromosome

rearrangements observed in cancer and congenital diseases. By recapitulating chromothripsis in the laboratory, we establish that it can occur after an intact chromosome is partitioned into a micronucleus. These findings highlight the critical importance of nuclear architecture and nuclear envelope integrity for the maintenance of genome stability in eukaryotic cells[25].

After the division of micronucleated cells, we observed extensive localized chromosome rearrangements, some of which bear all the hallmarks of chromothripsis. The following evidence indicates that these rearrangements occurred on the chromosome from the micronucleus. First, chromosomes in micronuclei are underreplicated and accumulate marked evidence of DNA damage[11,14]. Underreplication of the chromosome in the micronucleus creates a copy number asymmetry between the daughters, identifying the missegregated chromosome and establishing missegregation as a *de novo* event. Second, there is a striking enrichment of chromosome rearrangements only on the missegregated chromosome after the division of micronucleated cells; it is never observed on normally segregated chromosomes after division of either micronucleated or control cells. Third, the rearrangements are associated with the chromatid in the micronucleus identified from the gained haplotype. Thus, the partitioning of chromosomes into micronuclei is an “all-at-once” catastrophe that can trigger extensive mutagenesis at a surprisingly high frequency.

Our results show that partitioning chromosomes into micronuclei can have diverse consequences for genome structure. In addition to intrachromosomal rearrangements, we find that the partitioning of more than one chromosome into one micronucleus can generate translocations between chromosomes. We speculate that the reassembly of fragments from one or more chromosomes in a micronucleus could generate ring chromosomes, which are observed in a number of human cancers, and whose formation has recently been suggested to involve chromothripsis[26]. We also observed circularized fragments originating from the missegregated chromosome. Circularization of shattered chromosome fragments provides an appealingly simple mechanism for the first step in generating double minute chromosomes[24], which are frequent conduits for oncogene amplification in tumors[4,5,23] and were previously linked to chromothripsis[3].

Our experiments provide insight into the mechanism of chromothripsis. Elegant statistical analysis led to the proposal that chromothripsis could involve the “shattering” and reassembly of a chromosome, with “loss” of some segments[4]. However, the molecular mechanism or even the feasibility of such an event was not clear. Our analysis directly establishes that a chromatid can indeed be fragmented, with fragments distributed between daughter cells. In addition to validating the shattering and reassembly mechanism, our findings also explain the segmental DNA loss that characterizes many examples of chromothripsis: Loss may simply occur by partitioning of chromosome segments into a daughter cell that does not expand and does not contribute to the final population. Chromothripsis has also been suggested to originate from DNA replication errors that generate MMBIR[9]. MMBIR could be an independent mechanism causing chromothripsis, or an additional contributing factor. In agreement with this possibility, we detect short, potentially “templated” insertions at some translocation junctions that are consistent with co-occurring MMBIR[27].

The data here establish that the rupture of micronuclei during S phase is one significant source of mutagenesis. However, as we previously proposed[11], there may be more than one defect in micronuclei that generates DNA damage. Intact micronuclei exhibit reduced or delayed DNA replication and also fail to normally accumulate several DNA replication and repair proteins (Ref. 18; A. Spektor, E. Jackson, and D. Pellman, unpublished data). If the short insertions we observe at translocation junctions result from replication defects and MMBIR[8], given that nuclear envelope rupture appears to terminate most nuclear activity[14], MMBIR likely occurred in the intact micronucleus prior to rupture, or after reincorporation of the micronucleus into a daughter cell nucleus. Because micronuclei replicate asynchronously with the primary nucleus, many cells enter mitosis with micronuclei that are still undergoing DNA replication[11,28]. This results in premature chromosome compaction, which is thought to cause DNA breakage, best documented at chromosome fragile sites[29].

Our findings define a new mutagenesis pathway that generates a spectrum of localized chromosomal rearrangements, some of which have all the features of chromothripsis. Consistent with other recent work[16,30,31], the results here show that mitotic chromosome segregation errors can be heavily mutagenic, which has important implications for how mitotic errors and the accompanying aneuploidy might contribute to cancer or other human diseases. Mitotic chromosome segregation errors occur frequently, resulting in 1-5% aneuploidy in normal tissues in mice[32]. Chromothripsis is reported in a few percent of human cancers[33,34] and in rare human congenital disorders[6]. However, the actual rate of chromothripsis is likely to be much higher because most events are expected to compromise cellular fitness, and these events would only be detected by single cell analysis[32]. Furthermore, we find that DNA damage from micronuclei can lead to a moderate degree of rearrangement that might not, *ex post facto*, be recognized as related to chromothripsis. Micronuclei may therefore be a significant, but previously unappreciated, source of genetic variation.

METHODS

Cell culture and treatment

U2OS, telomerase-immortalized RPE-1 cells (ATCC), and all derivative cell lines generated for this study were tested for Mycoplasma, and were grown in phenol red-free DMEM:F12 media containing 10% FBS, 100 IU/ml penicillin, and 100 μg/ml streptomycin. All cells were maintained at 37°C with 5% CO_2 atmosphere. For cell cycle synchronization and induction of MN, cells were treated with 100 ng ml^{-1} nocodazole (Sigma) for 6 h. Mitotic cells were collected and washed three times with fresh medium containing 10% FBS before plating. To arrest cells in G0, nocodazole-treated cells were washed 3 times with media containing 0.01% FBS following mitotic shake-off and then re-plated in media containing 0.01% FBS. After 3 hours, cells were placed in a serum-free DMEM:F12 media. For EdU incorporation experiments, cells were incubated in the presence of 10 mM EdU from the time of mitotic shakeoff.

Generation of cell lines

Lentivirus or retrovirus carrying genes of interest were generated by transfection of 293FT cells with the appropriate packaging plasmids (Lentivirus: pMD2.G and psPAX2; Retrovirus: pUMVC and pVSV-G) using Lipofectamine 2000 (Life Technologies), according to the manufacturer's instructions. RPE-1 cells were infected for 16-24 hours with virus in the presence of 10 μg/ml polybrene, washed, and allowed to recover for 24 h before selection with an antibiotic or by fluorescence cell sorting.

RNA interference

Sequence information of the small interference RNA (siRNA) pools used from Dharmacon are as follows: Human TP53 ON-TARGETplus SMARTpool siRNA L-003329-00-0005, (J-003329-14) GAAAUUUGCGUGUGGAGUA, (J-003329-15) GUGCAGCUGUGGGUUGAUU, (J-003329-16) GCAGUCAGAUCCUAGCGUC, (J-003329-17) GGAGAAUAUUUCACCCUUC. Cells were transfected with 40 nM siRNA using Lipofectamine 3000 transfection reagent (Life Technologies) per manufacturer's instructions.

DNA constructs

Plasmid encoding cDNA for H2B-GFP was obtained from Addgene (Plasmid 11680). Constructs encoding TDRFP-NLS and NLS-eGFP were a kind gift from Adrian Salic. GamEmGFP (Emerald GFP-Gam)[15] was a kind gift of Susan Rosenberg.

Reagents and antibodies

DNA damage was detected using Phospho-Histone H2A.X (Ser139) antibody for γH2AX (1:300-500, Cell Signaling Catalog Number 2577S). Nocodazole was purchased from Sigma-Aldrich. Secondary antibodies used were Alexa Fluor 488 (green), 594 (red) and 647 (far red) (1:1000, Life Technologies).

Detection of EdU incorporation

Detection of EdU incorporation was performed using *Click*-iT® *Plus* EdU Alexa Fluor® *imaging kits* 594 and 647 (Life Technologies) per manufacturer's instructions.

Live cell imaging, single cell isolation and daughter cell separation

RPE-1 cells expressing H2B-GFP and GFP-NLS were treated as described above to induce micronuclei after depletion of p53 by siRNA. After mitotic shake-off, cells were re-plated and allowed to progress into G1 phase for 4 hours. Afterwards cells were trypsinized and single-cell sorted into 384-well μClear plates (Greiner) using FACS. Following single cell sorting, cells were incubated for 2h to allow for cell attachment and spreading. Plates were mounted on a Nikon TE2000-E2 inverted microscope equipped with the Nikon Perfect Focus system. The microscope was enclosed within a temperature- and CO_2-controlled environment that maintained an atmosphere of 37°C and 3-5% humidified CO_2. Wells containing single cells of interest were identified manually and fluorescence and DIC images were captured every 30 minutes with a 20X 0.5 NA Plan Fluor objective for up to 48

hours or until the majority of cells had progressed through mitosis. All captured images were analyzed using NIS-Elements software.

Wells containing cells of interest, having completed mitosis, were washed with PBS and cells were subsequently trypsinized. After addition of an excess of fresh medium, daughter cells were separated by limited dilution into new wells in a fresh 384-well μClear plate. Successful separation and transfer into new wells was monitored using a fluorescence microscope. In cases where both daughters ended up in the same well, separation by limited dilution was repeated. After separation, the cells were left to attach for up to 4 hours prior to cell lysis.

Indirect immunofluorescence

Cells were washed in PBS and fixed in 4% paraformaldehyde for 20 min; cells were then extracted in PBS-0.5% Triton X-100 for 5 min, washed 3 times with PBS, blocked for 30 min in PBS containing 3% BSA (PBS-BSA) and incubated with primary antibodies diluted in PBS-BSA for 60 min. Samples were washed 3 times for 5 min with PBS-0.05% Triton X-100 and primary antibodies were detected using species-specific fluorescent secondary antibodies (Life Technologies). Samples were washed 3 more times for 5 min with PBS-0.05% Triton X-100 prior to DNA detection (2.5 μg/ml Hoechst). For pre-extraction, cells were washed once with PBS, and then incubated in CSK buffer (100 nM NaCl, 300 mM sucrose, 3 mM $MgCl_2$ and 10 mM PIPES pH 6.8) containing 0.5% Triton X-100 for 5 minutes on ice. Cells were then washed 3 times with PBS, fixed and processed as above.

Image acquisition and analysis

Immunofluorescence images were collected with a Yokogawa CSU-22 spinning disk confocal system with Borealis modification, which was attached to a Nikon Ti-E inverted microscope (Nikon Instruments, Melville, NY). Laser excitation of the fluorophores was performed sequentially using the 405nm, 488nm, and 561nm and 642 nm lasers. Images were acquired using a 60x Plan Apo NA 1.4 oil objective with a CoolSnapHQ2 CCD camera (Photometrics). Acquisition parameters, shutters, filter positions and focus were controlled by Metamorph software (Molecular Devices), which was also used for image analysis. Regions of interest were defined for micronuclei and corresponding primary nuclei, and average fluorescence intensities were determined. Background values were subtracted from a region the same size and shape as the micronucleus, set equidistant from the primary nucleus.

Thresholds were set to exclude cells that were not synchronized after mitotic shakeoff. In G1 phase samples, cells were excluded if the primary nucleus contained visible EdU. In S or G2 phase samples, cells were excluded if there were low levels of replication in the primary nucleus, operationally defined by an average fluorescence intensity threshold. Additional average fluorescence intensity thresholds for γ-H2AX were used to exclude rare cells from all samples (<5%) where the primary nucleus had significant DNA damage, and to define γ-H2AX positive micronuclei. For EdU detection, micronuclei were scored as EdU positive if their EdU signal was >3 standard deviations above the mean background. Thresholds were set by eye and used consistently for samples within each experiment.

Each immunofluorescence experiment included two biological replicates. Number of cells counted (N) for each experiment were as follows (Biological Replicate 1, Biological Replicate 2, Total): Fig. 1A: Primary Nuclei (83, 83, 166), Intact Micronuclei (53, 50, 103); Ruptured Micronuclei (51, 51, 102); Extended Data Fig. 1A: G1 (63, 69, 132), S (55, 50, 105), G2 (62, 63, 125); Extended Data Fig. 1B: γ-H2AX (−) micronuclei (87, 63, 150), γ-H2AX (+) micronuclei (109, 73, 182); Extended Data Fig. 1C: G0 (63, 63, 126), S (53, 50, 103); Extended Data Figure 1D: γ-H2AX (+) micronuclei (62, 49, 111).

Multi-strand displacement amplification and library construction of single-cell genomic DNA

We chose to amplify single-cell genomes by multi-strand displacement amplification (MDA) with the Phi-29 polymerase[36] for four main reasons: First, MDA gives better overall genome coverage than PCR-based methods[37] and also gives comparable uniformity[38-40] to other methods such as MALBAC[41]; this is required for the detection of chromosomal rearrangements. Second, Phi-29 polymerase has the highest processivity and the lowest error rate among existing polymerases[37,42]. Third, amplification bias due to MDA has been characterized as largely random, even between the two homologues of the same chromosome[20]; this enables us to estimate the coverage for each homologue, the detection sensitivity for *de novo* variants, and to accurately calculate the copy number for each homologue from the coverage at heterozygous sites. Finally, the high processivity of Phi-29 polymerase consistently generates large amplicons above 10 kb[43,44]; this enables us to perform Sanger sequencing on the MDA product after PCR to generate phasing information of rearrangements and validate their association with the missegregated chromosome, which is crucial in establishing the relationship between chromosomal rearrangements and DNA damage in the micronuclei.

DNA from isolated cells was subject to MDA following lysis using the REPLI-g Single Cell Kit (Qiagen) with minor modification. (Note that we achieved the best overall coverage uniformity with this latest version of REPLI-g from Qiagen as compared to earlier versions of REPLI-g or the RepliPhi enzyme from Epicentre. Comparison of the coverage and uniformity of the single-cell libraries in the current study with previous studies[41,44] is summarized in Supplementary Table 1 and in Ref. 20.) Samples were washed once with PBS and cells were lysed using 10 μl of a 1:1 mixture of the provided lysis buffer and PBS by a brief vortex and spin down, followed by a 10 min incubation at 50°C and then an additional 10 min at room temperature. Lysis was stopped by adding 5 μl of stop solution with vortex and spin down, followed by incubating at room temperature for further 10 min. Whole genome amplification by MDA was carried out in a total of 50 μl for 80 min at 30°C. Purified genomic DNA (for bulk RPE-1) or amplified DNA (from single cells) was sheared to 300-500 base pair size and utilized for multiplex genome sequencing libraries as previously described[18].

Library quality was assessed by low-pass sequencing ~ 0.1x on a MiSeq instrument (Illumina). DNA libraries that passed MiSeq quality control[18] were then sequenced to ~ 5x per cell on the HiSeq platform (Illumina). Two samples, one from the control group (N3) and one from the MN group (MN4), were subject to additional sequencing to a total depth of

~ 9x per cell for validation of the enrichment of chromosomal rearrangements on the missegregated chromosomes. For the MN9 daughters the plus cell did not divide but the minus cell divided twice. The four progeny cells from the MN9 minus daughter were sequenced to ~ 6x combined coverage and the plus cell that did not divide was sequenced to ~ 4x depth. (The main goal of this experiment had been to obtain biological replicates of cells with potentially chromothriptic damage; however, in this example the plus daughter with rearranged Chr. 8 did not divide). For the three N6 progeny, one cell divided and the other did not prior to library preparation, all three progeny were sequenced to a ~ 4x depth.

Processing of single-cell sequencing data

Sequencing reads were aligned to the human genome reference (hg19/GRCh37) using **bwa** (http://bio-bwa.sourceforge.net/) in the paired-end mode by "bwa mem". For both primary and supplementary alignments, duplicated sequencing fragments were removed by **MarkDuplicates** from the PICARD software suite (http://picard.sourceforge.net/). We also generated genotype information by running the UnifiedGenotyper module from GATK (https://www.broadinstitute.org/gatk/) at HapMap SNP sites (ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/current/b37/hapmap_3.3.b37.vcf) across all sequenced samples (bulk, controls, MN samples). We identified 884,944 heterozygous sites in the bulk sample sequencing with quality scores ≥ 100. Genotypes at these heterozygous sites were later utilized to calculate haplotype coverage and copy number.

DNA copy number analysis from sequencing read depth

The average read depth was calculated by dividing the number of properly aligned read pairs (insert size < 1 kb and having the "forward-reverse" pair orientation) by the total ungapped length of each bin (for binned coverage) or by the ungapped length of the relevant chromosomal arm (for arm-level coverage). The read depth was then normalized by the median value of all bins (or arms) for each sample to calculate the average DNA copy number of each bin.

We corrected for GC-content amplification bias in each library by modifying the strategy described in Ref. 45 The overall strategy was to statistically infer read depth variation due to GC content and normalize regional coverage by a GC-content-dependent correction factor. Importantly, we expected the dominant GC-dependent variation to have been generated during MDA prior to the library preparation; therefore, the average GC-content was calculated for 25 kb bins because the average amplicon size of MDA was previously estimated to be in the range of 10-50 kb[20,44] .

To generate a reference for the relationship between read depth and GC content, we analyzed the sequence coverage of a chromosome (Chr. 6) that was present in all samples at two copies for each library. For each sequencing fragment (read pair) we calculated the average GC-content in the 25 kb bin centered at the leftmost alignment position of the read pair. We then grouped all read pairs into GC strata that differed by 1% GC content and calculated the average read depth for each stratum. A normalizing weight was then derived for each GC stratum to bring the median coverage within each GC stratum to the same value. Each sequencing fragment was then assigned a weight based on the GC composition

in its 25 kb neighborhood. The GC corrected sequencing coverage was generated by random sampling based on the "GC-correcting" weight for each read.

Although GC correction at the amplicon level does reduce the amplification noise at the sub-megabase scale, significant variation at longer scales (> 1 Mb) is still evident after GC correction. To additionally correct for long-range systematic amplification biases, especially in the small chromosomes (most evident in the 19p arm, Fig. 3b and Supplementary Table 2), we normalize the average copy-number CN_a^i for arm *a* in sample *i* as

$$x = \left(log_2 CN_a^i - log_2 \overline{CN_a}\right) \times \frac{\text{median}\left[std\left(log_2 CN_a\right)\right]}{std\left(log_2 CN_a\right)}.$$

The first term on the right side of the equation is the standard log2 copy number; $\overline{CN}_a$ is the average log2 arm-level copy number across all samples (but excluding samples with significant gains or losses defined as 10% difference from the median). Subtracting the average eliminates systematic (recurrent) amplification bias at the arm level (most noticeably on Chr. 19). The second term corrects for variation in amplification bias across different chromosomes: the denominator, std(•), is the standard deviation that reflects variability in amplification for a given chromosome arm occurring between the different, independently generated, samples; the numerator is the median of the standard deviation across all chromosomes. The standard deviation is small for large chromosomes but becomes bigger for small chromosomes. The normalized log2-scale copy number was plotted in Fig. 2.

The genome-wide coverage of each single-cell library is presented in 5 Mb bins in the CIRCOS plots in Fig. 3b and Fig. 4a. The standard deviation of the normalized bin-level coverage was estimated to be ~ 0.1 from the disomic Chr. 6 for all daughters, except for sample MN7 (std. dev. ~ 0.3). The thresholds for significant gains (mean coverage ratio >1.35) and losses (mean coverage < 0.6) were chosen because they are ~ 3.5x standard deviations from the mean value (0.95) for MN1-6, 8-9, and for coverage >1.4 and coverage < 0.5 for MN7, corrected for the additional variance. Where there were arm-level copy number alterations present in single cell library samples, but absent from the bulk sequencing, haplotype copy-number analysis (below) was used to discriminate systematic amplification bias (affecting both haplotypes) from true copy number gains or losses (Extended Data Fig. 3).

Detection of loss-of-heterozygosity

The model that the chromosome in a ruptured micronucleus undergoes *de novo* damage (i.e. alterations that are not shared between daughters) has the clear prediction that the damage should occur on one of the homologous chromosomes but not the other. Our sequence-read-depth based copy number analysis indicated that some missegregated chromosomes are distributed between the daughters in ~ 2:1 ratio (Fig. 2a, left panel). This implies that the cell with two copies of the affected chromosome contains the potentially damaged chromosome from the micronucleus and that the other cell should only contain the intact homologue of the missegregated chromosome (i.e., monosomic). A further prediction, which

would independently validate these copy number results, is that the daughter with two copies of the chromosome should be heterozygous at polymorphic sites whereas the other daughter should be hemizygous. To test this prediction, we developed a method to determine the presence or absence of loss-of-heterozygosity (LOH) from the genotypes observed at heterozygous sites.

Because the genotype data are derived from single-cell libraries that are subject to variable amplification, it is non-trivial to distinguish true LOH from incomplete coverage due to uneven amplification by the MDA procedure. To address this problem, we derived an expected relationship between the observed number of sites showing heterozygosity and the observed number of sites showing the reference base (or equivalently, the number of sites showing the alternate base), given the sequence coverage. This relationship was derived with the assumption that the cell has one copy of each homologous chromosome (i.e., "1:1 heterozygosity"). Knowing the expected heterozygosity at the desired level of coverage, we can then infer LOH, if the sequencing data deviate significantly from what is predicted for 1:1 heterozygosity. Because reference or alternate bases are counted as "present" or "not", and the number of reads when present is ignored, this relationship is relatively insensitive to variable amplification and holds as long as the two homologues are amplified independently[20].

For a cell with a single copy of each homologous chromosome, assume that there are a total of M heterozygous sites and the fractional coverage of each homologous chromosome above a certain sequencing depth is p. (For example, when we required ≥ 3 reads to call a structural variant, p would correspond to the fraction of each homologue covered with ≥ 3 reads, which defined the detection sensitivity discussed below under **Estimation of SV detection sensitivity**.) If the two homologues have identical copy number and their amplification bias is independent, then the expected percentage of sites where we should observe both the reference and the alternate bases (i.e., both homologues are covered at or above the specified depth) is given by p^2. If the observed percentage of sites showing heterozygous coverage deviates from this prediction, then it implies that the two homologues have different copy numbers. In particular, it is straightforward to see that when there is complete LOH for a given chromosome, the expected percentage of sites where we should observe both the reference and the alternate bases should be zero, if there are no genotyping, sequencing or systematic amplification errors. However, to determine if there are partial deletions in either homologue, we need to estimate the value of p.

Estimating the fraction of a homologue that should be covered at any given depth by the sequencing data, p, is straightforward if the haplotype phases (i.e., the order of reference and alternate bases at heterozygous sites in a single haplotype) of both homologues are known. However, this can be achieved even without knowledge of the haplotype phase as long as each homologue has a complete copy and their average coverage is equivalent, which is true for MDA when the coverage is evaluated over regions that are significantly larger than the amplicon size[20]. We can account for p as follows. For a total of M heterozygous sites, assume there are fM reference and $(1-f)M$ alternate bases in haplotype 1. (It will become obvious below that the haplotype composition f will not affect the estimate for p as long as the amplification of different homologues is independent.) For this haplotype we expect to

observe the reference base at $p \times fM$ heterozygous sites that are covered by more than the threshold number of sequence reads. For the other haplotype, there are $(1\text{-}f)M$ reference and fM alternate bases, which by the same reasoning predicts that we should observe the reference base at $p \times (1\text{-}f)M$ heterozygous sites above the threshold of coverage. Thus, without separating the two haplotypes, we expect to observe pfM reference bases from the coverage of haplotype 1 and $p(1\text{-}f)M$ from the coverage of haplotype 2, which gives a total number of pM reference bases at all heterozygous sites. The same number is expected for the total number of alternate bases being covered by more than the threshold number of sequence reads. Note that this estimation for p comes directly from the coverage of the reference or alternate bases at the heterozygous sites but does not rely on knowledge of the haplotype phase or its composition f.

In summary, for a chromosome that consists of two single-copy homologues ("1:1" heterozygosity) and having a total number of M heterozygous sites where the two homologues differ, we expect the following coverage at heterozygous sites:

The number of sites showing reference base coverage: pM

The number of sites showing alternate base coverage: pM

The number of sites showing heterozygous coverage: p^2M

With these expected values we can derive an expected relationship between the fraction of heterozygous coverage relative to the average fractional coverage for the reference and for the alternate bases:

$$\frac{[\text{fraction of sites w. heterozygous coverage}]}{([\text{fraction of sites w. reference coverage}]/2+[\text{fraction of sites w. alternate coverage}]/2)^2}=\frac{p^2}{p^2}=1,$$

which holds for the "1:1 heterozygosity" scenario. Deviation from this ratio in the observation reflects changes in the copy number of either or both homologues. We therefore define the ratio on the left hand side of the above equation evaluated from the experimental data as a "heterozygosity coefficient" to gauge the deviation of the homologue copy number from 1:1 heterozygosity.

Importantly, we do not require direct knowledge of the absolute fractional coverage (p) to test whether the ratio on the left side of the equation gives the expected value close to 1. [We will however use this relationship below (section on detection sensitivity) to estimate p as a measure of the detection sensitivity.] It is based only on the relative value for the fraction of heterozygosity divided by the fraction of sites with reference (or of alternate) coverage. This relationship is therefore more robust than a direct test of the observed coverage on a given chromosome against p estimated from a different disomic chromosome, because the coverage p may vary slightly between different chromosomes due to sequence-specific amplification bias.

Results shown in Extended Data Fig. 2 confirmed that for heterozygous chromosomes the heterozygosity coefficient is overall slightly above but close to unity, reflecting only a small degree of correlation for the amplification of the homologues[20]. For chromosomes that are

completely or partially hemizygous (i.e., one homologue has complete or partial loss) the heterozygosity coefficients are significantly smaller. Of note, the heterozygosity coefficient is a sensitive method for detecting even a relatively small segment of heterozygosity in a chromosome that is otherwise near completely hemizygous. This is illustrated by the analysis of the MN2 minus daughter, where we could detect a 15 Mb segment of heterozygosity from Chr. 2q (~7% of Chr. 2) that is lost from the plus daughter (Extended Data Fig. 8a).

The method is not, however, sensitive for detecting gains above 1:1 heterozygosity, as occurs when cells undergo 3:2 segregation patterns for the missegregated chromosome. For this circumstance, we directly calculated the haplotype copy number based on the haplotype phase information (see below). Note that the determination of the haplotype phase required the LOH analysis: The haplotype phase was obtained from the sequencing data of chromosomes that are inferred to be completely hemizygous by their heterozygosity coefficients, an inference that is independent from and more robust than the read depth-based arm-level copy number data shown in Fig. 2b and Supplementary Table 2.

Haplotype copy number analysis

If the haplotype phase is known, then the copy number ratio of the two homologues, based on their haplotypes, can be directly determined in a given cell. Because the LOH analysis provided confidence of genuine hemizygosity (Extended Data Fig. 2), we were able to use the sequencing data to directly extract the haplotype phase for one homologue and then infer the phase of the other, and deconvolute the sequence coverage for each homologue (Extended Data Fig. 3a).

This analysis was first applied to Chr. X where we observed recurrent gains in the sequence coverage of Xq that is nonetheless absent in the bulk sequencing data (Supplementary Table 2). As can be seen from the example of the haplotype copy number for Chr. X in the MN3 daughters (Extended Data Fig. 3b, top panel), the haplotypes were present in equal copy number in both daughters, enabling us to attribute the read depth variation (between the Xp and the Xq arms) to systematic amplification bias. Variable penetrance of this amplification bias presumably caused the apparent copy-number asymmetry in the C1 and MN7 daughters (Fig. 2b), which were the samples with the most read depth noise. Indeed, haplotype copy-number analysis confirmed that there was no difference in the relative copy number of the two Chr. X homologues in both C1 daughters (Extended Data Fig. 3b, middle panel). It also demonstrated a true gain of one Chr. X haplotype was shared in both MN7 daughters (Extended Data Fig. 3b, bottom panel).

By contrast with the amplification bias, haplotype copy-number analysis verified that there was a gain of a single homologue of the missegregated chromosome in the plus daughters from the 3:2 segregations (MN7 and MN8, Extended Data Fig. 3c), excluding the possibility that the gain is due to mixed segments from both homologues or uniparental trisomy. Moreover, this analysis confirmed that the minus daughters from these cells had a 1:1 ratio of the two homologues. Therefore, the haplotype copy number established the segregation model depicted in Fig. 2a (right panel) for the 3:2 segregation samples.

Knowledge of the haplotype phase was also critical for obtaining the fine structure of copy number alterations. For the daughters with a 2:1 distribution of the missegregated chromosome, knowing the haplotype phase enabled a digital readout for whether a segment of that haplotype was deleted or retained (Fig. 4b, Extended Data Fig. 8b). Like the LOH analysis, this readout is minimally affected by amplification noise: heterozygosity is scored as present or absent and differences in the numbers of reads at different heterozygous sites are ignored. (To capture copy number gains, e.g., in the 3:2 segregations, the read depth signal is needed and amplification bias will have some effect on the detection of focal gains. We are currently working on methods to address this problem.)

We utilized the haplotype phase information for Chr. 3 (inferred from the MN3 minus daughter) to generate segmented copy number profiles for the missegregated haplotype in both MN4 daughters that contained interspersed deletions (Fig. 4b). Each phaseable SNP served as a probe for the altered haplotype; the boundaries of deletions were identified by a transition from “covered” (i.e. present) probes to “uncovered” probes or vice versa. For each SNP, we provisionally designated it as being “covered” if there were reads supporting the haplotype, and “uncovered” if there were zero supporting reads. This raw signal at each SNP was then smoothed using the coverage of other SNPs within the local bin to eliminate false negative assignments of SNPs not being covered because of amplification bias. We set this bin size to 100 kb because it is much larger than the typical MDA amplicon and therefore unlikely to have been completely lost because of variation in amplification[20]. In the rare regions with sparsely distributed SNPs, we extended the bin to contain at least 8 flanking phaseable SNPs, which was chosen because with 8 phaseable SNPs, the probability that half of them (4 out of 8) were absent due to variable coverage (false LOH) was less than 0.1% based on the current level of variant detection sensitivity and heterozygous coverage (Supplementary Table 1). (The probability of having sequencing or amplification errors generating false heterozygosity in 4 out of 8 sites is even smaller.) For each SNP, more than 50% of the nearby SNPs were covered, the SNP of interest was designated as covered; if that value was less than 50%, that SNP was designated as being not covered; if the local coverage was 50%, the SNP was designated as being at the boundary of a deletion. This process was iterated until convergence; the final designation for each SNP can only be “covered (1)”, “uncovered (0)”, or “boundary (0.5)”. We then connected all SNPs with the same copy number states and identified copy-number changes directly. This procedure was used to generate the segmented copy-number profiles for MN4 daughters in Fig. 4b, with the average coverage for each bin shown as gray dots.

Detection of chromosomal rearrangements

Chromosomal rearrangements were detected from clusters of discordant read pairs. Read pairs are designated to be discordant if both mates mapped to the reference genome and the inferred insert size exceeded 20 kb, irrespective of the orientation of the mapped pair mates. The 20 kb threshold is significantly longer than the average length of the sequencing fragment (~ 1 kb) but is comparable to the typical size of the amplicon of multi-strand displacement amplification. The choice of a 20 kb threshold thus excludes a large number of short-range artificial chimeras (mostly of inverted orientation) that result from amplicon annealing or phi-29 polymerase template switches[40,46]. As explained below,

intrachromosomal rearrangements within the 20-100 kb range were also populated by artifacts and we used the frequency of these events on control chromosomes to estimate the background frequency of MDA-generated chimeras. The control chromosomes included all the chromosomes from the daughters derived from non-micronucleated mothers (C1-4 and N1-6) as well as all chromosomes from both daughters from the micronucleated mothers, with the exception of the missegregated chromosomes.

We initially applied a low threshold of two discordant read pairs to search for candidate rearrangements. Discordant pairs include both primary discordant pairs, where the two pair mates are aligned to discordant loci, and split reads, where a single read is split into two parts and aligned to non-contiguous segments. A primary discordant pair can contain one mate that itself constitutes a split read; in this scenario, the discordant pair and the split read were counted as a single discordant fragment. Although split read alignment potentially provides base-level breakpoint resolution, the smaller size of each split alignment means that they are more likely to be misaligned to the reference genome due to short interspersed repeats. We therefore required each discordant cluster to consist of at least one primary discordant pair with each mate aligned to discordant loci with a mapping quality of 30 or greater (from BWA MEM, ~ 0.1% alignment uncertainty). From this first-pass analysis there were a total of 8,179 clusters of discordant pairs in all samples.

Next, we excluded all discordant clusters that consisted of reads from daughters with different mothers, or reads from the bulk RPE-1 line. Such clusters could reflect recurrent artifacts due to inaccuracy in the alignment or in the reference assembly. Although expected to be infrequent, we also cannot exclude the possibility that some of these clusters reflect clonal or subclonal chromosomal rearrangements that accumulated during cell culture. The remaining 2,665 non-recurrent clusters are expected to include both putative *de novo* chromosomal rearrangements and random artifacts due to MDA. There were 2,403 intra-chromosomal and 262 inter-chromosomal clusters that consisted of two or more discordant read pairs. There were 1,088 intra-chromosomal and 76 inter-chromosomal clusters with three or more discordant read pairs. Discordant clusters supported by only two discordant read pairs occurred at a higher frequency than predicted by the haplotype coverage and were especially enriched in short-range events (data not shown). Because of the expectation that amplification artifacts (excepting those that occur very early in the reaction) would be supported by fewer discordant read pairs, we expected more frequent artifacts among the SV clusters supported by only two discordant read pairs and required putative rearrangements be supported by at least three discordant pairs and also by at least one split read. We relaxed these criteria only in a few exceptions where there was additional supporting evidence, such as for the chained rearrangements depicted in Extended Data Fig. 7b, but subsequently performed PCR validation of all the detected events.

Enrichment analysis of chromosomal rearrangements in the missegregated chromosome

Random MDA-derived artificial chimeras and true *de novo* chromosomal rearrangements cannot be distinguished by the variant allele frequency or by the number of discordant pairs. We therefore developed statistical criteria to determine if clusters of rearrangements on the

missegregated chromosome in the plus daughter cells were significantly enriched relative to the background observed for other chromosomes.

Template switch events in multi-strand displacement amplification can result in frequent short-range chimeras of the inverted orientation[46]; by contrast, long-range artificial chimeras or *de novo* chromosomal translocations are not expected to have a preferred orientation. The observed SV events in the single-cell DNA libraries indeed confirmed this prediction (Extended Data Fig. 5a). A substantial enrichment of short-range SVs (20-50 kb) of the inverted type relative to the non-inverted type was observed in all control chromosomes: both daughters from the non-micronucleated mothers (C1-C4, N1-N6) and all chromosomes, excluding the missegregated chromosomes, in the daughters from the micronucleated mothers. We therefore excluded short-range SVs from the list of candidate *de novo* rearrangements, even though this likely underestimates the numbers of true *de novo* events.

To rigorously establish a threshold to distinguish short-range from long-range SVs, we looked at the cumulative distribution of inverted and non-inverted type SVs relative to the breakpoint distance (Extended Data Fig. 5b). By simple visual inspection it is clear that at short breakpoint distances, inverted-type SVs are more frequent than non-inverted type SVs. However, the frequencies of inverted and non-inverted type SVs became equivalent at longer breakpoint distances. Power law fitting revealed that events of the intermediate length range (85 non-inverted type SVs in the range of 200 kb ~ 5 Mb) followed a power law scaling, $P(\ \#\ \text{SVs} \geq l\ ) \sim l^{-0.32}$. By contrast, short-range inverted-type SVs (632 inverted type SVs < 100 kb) decayed as $\sim l^{-1.18}$, reflecting a much higher frequency of inverted-type SVs at shorter breakpoint separations. The intersection of these two decay curves indicated that at or above 150 kb breakpoint distances, inverted and non-inverted type SVs approximately followed the same distribution. This analysis led us to adopt 150 kb as an operational cut-off for “long-range” rearrangements that excludes most systematic MDA artifacts. We note that this cut-off does not exclude signal from the missegregated chromosomes, because these chromosomes had a highly significant enrichment of rearrangements separated by > 500 kb breakpoint distances (Extended Data Fig. 5c).

We then estimated the frequency of short- and long-range (including interchromosomal) rearrangements on all control chromosomes to establish the background rates of these events (Extended Data Fig. 4a). The frequency of breakpoints within each chromosome arm was calculated as the ratio of the total number of observed breakpoints divided by the length of the relevant arm, corrected for the copy number of that arm and for the detection sensitivity for each sample (i.e., the percentage of each homologue that is covered at or above the threshold depth for variant detection). The arm-level copy number was determined as described above and shown in Supplementary Table 2). The detection sensitivity estimated from sequence coverage (to be discussed further below) is shown in Supplementary Table 1. Short-range rearrangements occurred at an average frequency of 1.5 events per 100 Mb; long-range rearrangements occurred at an average frequency of 0.5 events per 100 Mb. By a χ^2 test, the distribution of either short-range or long-range events on the control chromosomes was indistinguishable from that expected for a uniform frequency with normally distributed error (minimum $p = 0.18$ for short non-inverted type rearrangements, 3 degrees of freedom).

To determine if the observed structural variants follow a uniform distribution, we performed one-sided Poisson tests of the number of SV's observed in the test chromosome against the background frequency estimated from all chromosomes including the test chromosome on a per-sample basis. In 8 out of 9 daughter cell pairs from micronucleated mothers, there was a significant enrichment of SVs (combining short- and long-range events) on the missegregated chromosome (Extended Data Fig. 4a). By contrast, there was no enrichment for short-range rearrangements (Bonferroni corrected $p > 0.3$) on the missegregated chromosomes (Extended Data Fig. 4b, middle panel). Moreover, the enrichment of long-range SVs on the missegregated chromosome was even more dramatic, the most dramatic case being the MN3 pair, with an estimated p-value $< 10^{-100}$. Meanwhile, the only exception, MN5, served as a negative control for the statistical test.

We also performed a Fisher's exact test of the observed number of long-range and short-range events on the missegregated as compared with the remaining chromosomes, assuming the short-range events represent an empirical sampling of background events (summarized in Extended Data Table 1). The Fisher's exact test also confirmed the enrichment of long-range events in the missegregated chromosome relative to the rest of the genome with the background given by short-range events. MN5 was the only example where the partitioning of a chromosome (Chr. 7) into a micronucleus appeared not to have led to detectable rearrangements.

Finally, we also tested the enrichment of long-range rearrangement breakpoints in each of the normally segregated chromosomes in every sample. The background frequency was estimated by the average among all control chromosomes in each sample, i.e., for all chromosomes in each control daughter pair, and for all but the missegregated chromosome(s) in each micronucleated daughter pair. No normally segregated chromosome or chromosomal arm reached the significance level of $p = 0.05$ after Bonferroni correction (results for the statistical tests are not shown but raw data are in the Supplementary Tables).

Estimation of SV detection sensitivity

Whole-genome amplification bias results in varied coverage across each chromosome. Even at the same locus, this coverage can be different between the homologous chromosomes. With a requirement of ≥ 3 reads to support a structural variant, we will only detect SV events when the sequence coverage at the sites of rearrangements exceeds this threshold. Importantly, this threshold is required for reads from a single homologue. (For example, even at loci that are covered by more than three reads, it is possible that by random chance all the reads were derived from amplification of the intact homologue, and the variant homologue is missing from the sequencing data.) Therefore, the SV detection sensitivity is equivalent to the expected fraction, p, of each chromosome homologue that is covered, in this case by three or more reads, which can be estimated from coverage at heterozygous sites as discussed above in the section on LOH detection.

For chromosomes that have a single copy of each homologue, the fraction of one homologue that is covered by three or more reads is equal to the fraction of heterozygous sites at which we observe the reference base (or equivalently, the alternate base) by three or more reads. Because there can be subtle variation in the amplification of different chromosomes, we

generated a per-cell reference for detection sensitivity, using the coverage at heterozygous sites in Chrs. 5 and 6 consistently. These chromosomes were chosen because the LOH analysis (Extended Data Fig. 2b) and the read-depth analysis (Fig. 2b) indicated that both homologues were present at a 1:1 ratio in every cell that we analyzed (1:1 heterozygosity).

Three metrics of haplotype coverage are reported in Supplementary Table 1: the fraction of genome coverage (% of heterozygous sites covered with ≥ 1 read), the fraction of heterozygous coverage (% of heterozygous sites covered with ≥ 1 read corresponding to each genotype), and the detection sensitivity for variants [(% of sites covered by ≥ 3 reads of the reference base + % of sites covered by ≥ 3 reads of the alternate base)/2].

Based on the latter metric, we estimate that at the current sequencing depth ~ 5x, we should detect 30-40% of all *de novo* structural rearrangements that occur on a single homologue. For two samples, N3 and MN4, we doubled the sequencing depth to ~ 9x per cell and the estimated detection sensitivity correspondingly reached 60-70%. For example, in the N3 pair, the total number of rearrangements detected from the entire data set was 17, whereas at half the depth, 10 events were detected. Similarly, in the MN4 pair, the total number of rearrangements detected from the entire data set was 38 for the missegregated chromosome and 26 for the remainder of the genome; at half the depth, we observed 21 events in the missegregated chromosome and 15 events in the rest of the genome. These observations support the validity of our estimation of the detection sensitivity. Importantly, at higher sequencing depths, the statistical significance associated with the enrichment of rearrangement events in the missegregated chromosome only became more pronounced. Thus, although we are almost certainly underestimating the absolute numbers of rearrangements at the current level of sequencing coverage, we can nevertheless support with confidence the conclusion that there is enrichment of rearrangements on the missegregated chromosome.

Knowing the detection sensitivity is also important for the statistical inference that the observed unique SVs in each daughter cell are genuine, *de novo* events rather than a random sampling of shared events that are incompletely detected. If we assume that there are a total of N SVs that are shared between the two cells and the detection sensitivity is p_A in the plus daughter, and p_B in the minus daughter, then the number of SV events that are expected to be detected in both cells is $N\, p_A\, p_B$. Moreover, the number of SVs that are unique to cell A is $N\, p_A\, (1- p_B)$, and the number of SVs that are unique to cell B is $N\, p_B\, (1- p_A)$. Even if we do not know the total number of events, N, the relative fraction of shared versus unique events can be derived from the detection sensitivity. We therefore performed a multinomial test of the number of events observed in each of these categories (shared, unique to cell A, unique to cell B) against these ratios. Based on the results of the multinomial test, the hypothesis that the observations are due to incomplete detection can be rejected except in MN2 (Extended Data Fig. 4c, MN5 does not show enrichment of rearrangements in the missegregated chromosome.).

Rearrangement validation by PCR

The junction sequence for each rearrangement was constructed from the reference sequence and the breakpoint coordinates of the partner loci. For each rearrangement two primer pairs

were designed spanning the rearrangement junction: one pair for the rearranged sequence, the other for the reference sequence (Extended Data Fig. 6, primer data available upon request). When adjacent SNPs were present, PCR primers were designed to incorporate these sites to generate genotype information on the rearranged and the wild-type products. PCR was performed in both daughters: 10 ng of whole-genome amplified DNA was used per reaction and subject to 35 cycles of PCR. The product was gel purified and Sanger sequenced in both directions to validate the rearrangement and to infer the haplotype associated with each product.

Association of rearrangements with the gained haplotype

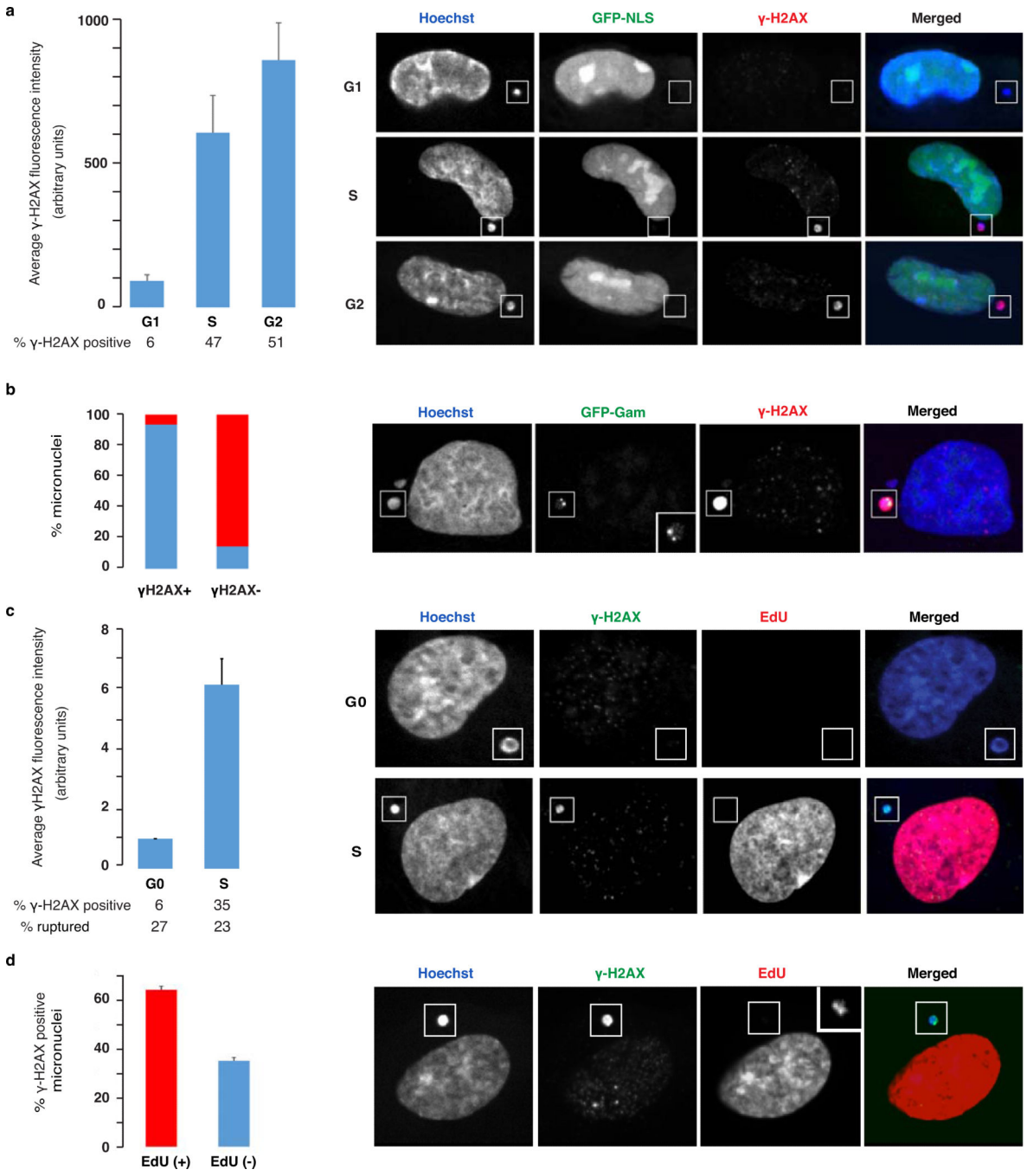
If rearrangements result from alteration of a single chromatid, then they should all be associated to the same haplotype. Furthermore, a strong prediction of the model that rearrangements occur from damage in micronuclei is that all rearrangements detected in the missegregated chromosome should be associated with the gained haplotype.

Haplotype phasing (associating rearrangements with a specific haplotype by the genotypes at polymorphic sites) was done in two ways. If a polymorphic site was close enough to a rearrangement junction (i.e. within the size range of an average DNA sequencing fragment), we looked for sequencing reads that both covered the polymorphic site and supported the rearrangement (either the read pair was a discordant pair or one pair mate was a split read supporting the rearrangement). We also performed long-range PCR (~ 1 kb) on the MDA amplified DNA to generate a product that would include the rearrangement junction and incorporate one or more polymorphic sites that were further away and could not be phased by sequencing reads directly. Sanger sequencing of this product was then used to test for association of the rearrangement with either haplotype. Importantly, Sanger sequencing can demonstrate that the rearranged product was derived only from the gained haplotype. PCR of the reference sequence was also performed on DNA from both daughters to determine the haplotype associated with the wild-type product.

The PCR validation strategy and representative examples are illustrated in Extended Data Fig. 6 and results of the PCR validation and haplotype association are summarized in Extended Data Table 1 and Supplementary Table 5. In each case where we PCR-validated a rearrangement with an informative nearby polymorphism, the rearranged product showed that the rearrangement was linked to the gained haplotype at one or more polymorphic sites. In addition, in all cases, the PCR validation of the wild-type products in the minus daughters always showed the genotype from the intact haplotype. For 2:1 segregations, this polymorphic site was hemizygous, consistent with the segregation model.

Interestingly, the reference sequence PCR product in the plus daughters was hemizygous in some events and heterozygous in others (Supplementary Table 5, Extended Data Fig. 6d). In each case where this product was hemizygous, the genotype corresponded, as expected, to the intact haplotype, i.e., the normally segregated chromosome. In other cases where the reference sequence product was heterozygous, the presence of both rearranged and reference sequence products with the gained haplotype is most easily explained by a DNA break at one side of a DNA replication fork (Extended Data Fig. 6e). This result is consistent with the cytological evidence suggesting partial DNA replication (Fig. 1).

Extended Data



Extended Data Figure 1. DNA damage and double-strand breaks occur in micronuclei when replication is coincident with nuclear envelope rupture

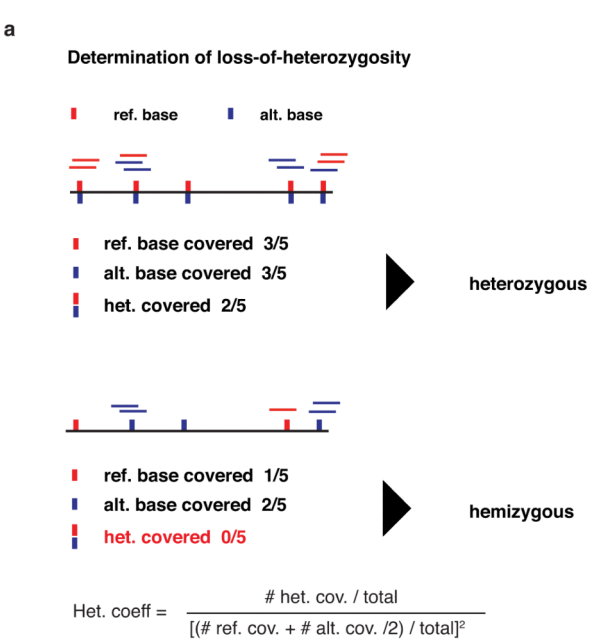
a. Nuclear envelope rupture in G1 is not sufficient to induce DNA damage in micronuclei. Left: Graph shows the percentage of ruptured micronuclei, determined by the loss of GFP-NLS, that were positive for γ-H2AX (>3 standard deviations above the mean of S-phase primary nuclei), and the fluorescence intensities for γ-H2AX labeling in the indicated

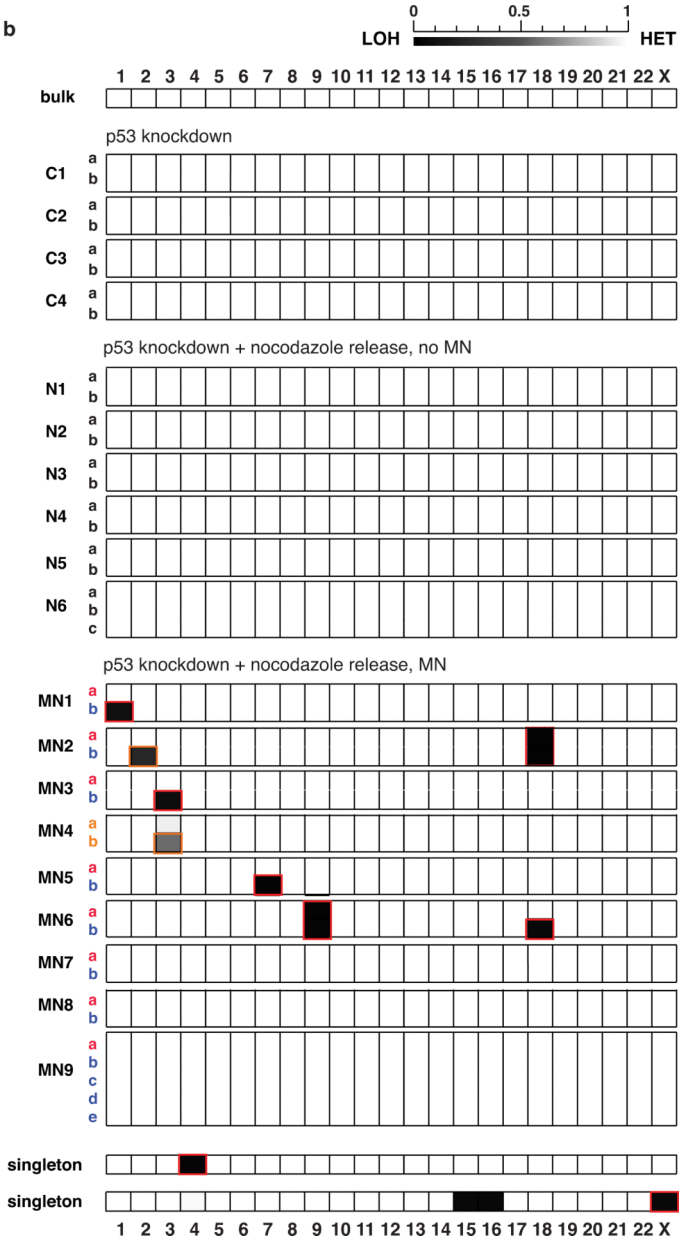
samples. N >100 from two experiments for each time point (see Methods). Error bars: standard error of the mean. Right: images of representative cells.

b. Ruptured micronuclei have double-strand breaks detected by GFP-Gam. Left: Graph shows the percentage of micronuclei with one or more GFP-Gam positive foci in γ-H2AX-positive and negative micronuclei in S-phase. N > 100 from two experiments for each category (see Methods). Right: images of representative cells. Inset: magnification of GFP-Gam signal.

c. Nuclear envelope rupture of micronuclei in G0 phase cells does not result in significant DNA damage. Micronucleation was induced in RPE-1 cells by a nocodazole block-and-release protocol, and cells were released into serum-depleted (G_0) or serum replete medium (S) for 17 hr. Left: graph shows the percentage of micronuclei that were positive for γ-H2AX (> 3 standard deviation above the average level in S-phase primary nuclei) as well as the distribution of the fluorescence intensities for γ-H2AX labeling in the indicated samples. Percentage of ruptured micronuclei was from a parallel sample with a GFP-NLS-expressing RPE-1 line. N > 100 from two experiments for each time point (see Methods). Right: images of representative cells. Error bars show standard error of the mean.

d. The majority of damaged micronuclei have initiated DNA replication. Replication was detected by continuous EdU labeling following nocodazole release and integrated EdU signal normalized over nuclear or micronuclear area. Left: percentage of γ-H2AX positive micronuclei that were positive (> 3 standard deviation above the background) or negative for EdU. N > 100 from two experiments (see Methods). Right: images of representative cells. Error bars show standard error of the mean.





c

cell id	Chr.	heterozygous sites in bulk	genotype supported by ≥ 1 read			heterozygous coefficient
			Ref.	Alt.	both (Het.)	
MN1(b)	1	64,528	23,555	23,641	919	0.10
MN3(b)	3	60,531	23,438	22,969	777	0.09
MN5(b)	7	49,915	19,677	18,888	216	0.03
MN6(b)	18	24,523	9,792	9,785	225	0.06
MN6(a/b)	9	42,251	16,034	15,880	191	0.03
MN2(a/b)	18	24,523	9,811	9,702	87	0.02
singleton	4	54,813	19,654	17,768	229	0.04
singleton	X	21,843	9,536	9,773	216	0.05
MN2(b)	2	74,605	28,794	28,250	2,656	0.24
MN4(a)	3	60,531	43,562	42,430	28,685	0.94
MN4(b)	3	60,531	32,730	33,401	10,354	0.57
	other					1.08 ± 0.06

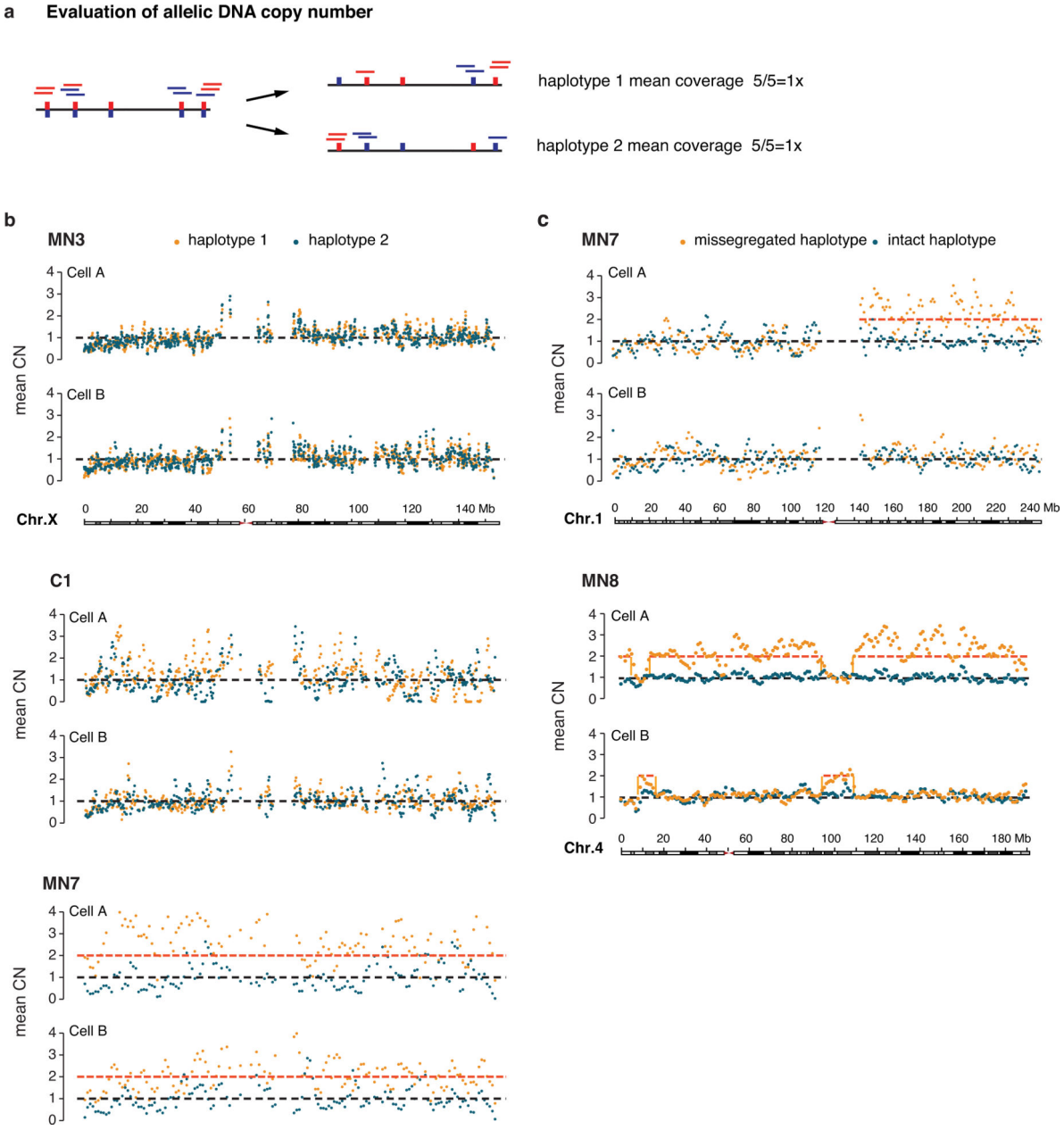
Extended Data Figure 2. Determination of loss-of-heterozygosity

a. Cartoon comparing the expected single-cell sequencing coverage of heterozygous and hemizygous chromosomes at polymorphic sites. Loss-of-heterozygosity (LOH) can be inferred from the scarcity of heterozygous genotypes without knowing the haplotype phase (i.e. the genotypes at polymorphic sites for each homologue). The presence or absence of the reference or the alternate base provides a digital read-out of heterozygosity or LOH that is insensitive to read-depth noise common in single-cell sequencing data. This can be quantified as a heterozygosity coefficient: the ratio of the observed fraction of heterozygosity relative to the expectation for a heterozygous chromosome consisting of a single copy of each homologue (“1:1” heterozygosity). For a diploid cell with 1:1 heterozygosity, if the fraction of sites that are covered (≥ 1 read per site) for each homologue

is denoted as p, then the expected fraction of sites with heterozygous coverage is approximately p^2. If the chromosomes are equally amplified then $p \approx 1/2$(observed reference base + observed alternate base)/total sites (Methods).

b. Heatmap of the heterozygosity coefficients for all chromosomes from all single cell samples included in Fig. 2b plus two additional single cells ("singletons") with monosomies that were sequenced to identify the haplotype phase of the monosomic chromosomes. Near complete LOH in the MN1, MN3, MN5, and MN6 minus daughter cells independently confirmed the monosomy of the missegregated chromosome, as determined from DNA copy number analysis (Fig. 2b). Note that the MN2 and MN6 daughters had monosomies (Chr. 18 in MN2 and Chr. 9 in MN6) shared in both daughters, indicating that the monosomy was preexisting in the mother cell. Two "singletons" were identified as having monosomy in Chr. 4 and in Chr. X based on low-pass MiSeq sequencing. They were each sequenced to ~ 4.5x to generate the haplotype phase of Chr. 4 and Chr. X. Note that the second singleton also had monosomies in Chrs. 15 and 16; the haplotype phases for these two chromosomes were not used in the current study, so they were omitted from the table in **c**, but the data are available upon request.

c. Table summarizing results from the loss-of-heterozygosity analysis. Heterozygosity coefficients are shown for the boxed chromosomes in **b**. Red heterozygosity coefficients indicate complete LOH; Orange are partial LOH. For the cell ID, we indicate the individual daughters as "MN#(a)" or "MN#(b)". The cases denoted as "MN#(a/b)" are monosomies shared in both daughters. The third column lists the number of heterozygous sites detected for the indicated chromosome from the bulk sequencing data (Methods). Columns 4-6 summarize the number of sites at which sequencing coverage from each cell supports the presence of reference bases ("Ref."), alternate bases ("Alt."), or both ("Het."). The last column lists the heterozygosity coefficients calculated for the indicated chromosome in the specified cell or cells. The average heterozygosity coefficient is 1.08 for all chromosomes that did not show LOH from all samples (last row). For chromosomes with near complete LOH (rows 1-8) the small number of heterozygous sites is likely due to genotyping errors (e.g., duplicated/homeologous sequences on the same chromosome) or amplification and sequencing errors. The incomplete LOH in the MN2 and MN4 daughters results from the reciprocal distribution of a fragmented chromosome between the two daughters (Extended Data Fig. 8). Note that our calculated heterozygosity coefficient can sensitively detect even a small region of heterozygosity in the MN2 minus cell (Extended Data Fig. 8a).



Extended Data Figure 3. Calculation of haplotype copy number from phased haplotypes

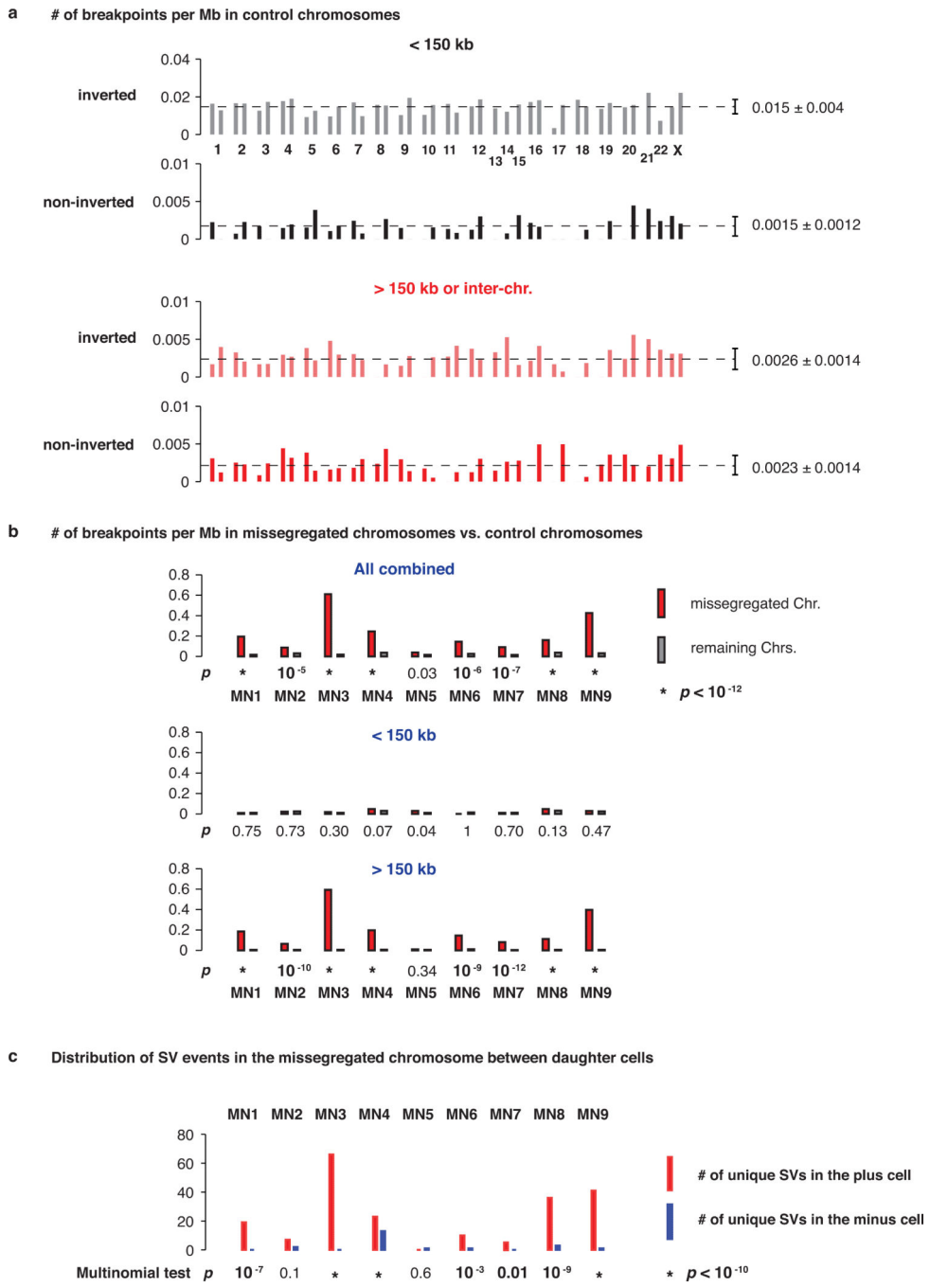
By sequencing monosomic cells we were able to determine the haplotype phases for chromosomes 1, 3, 4, 18, and X (Extended Data Fig. 2).

a. Cartoon illustrating the strategy to calculate the copy number for each haplotype (homologue) from the coverage at individual polymorphic sites. The example is shown for a cell having one copy of each homologue. The left shows the aggregate sequence data; the right shows the deconvolution of the sequence based on the determined haplotype phase. The haplotype coverage is calculated by dividing the number of sequencing reads from the indicated haplotype by the total number of heterozygous sites. Copy-number alterations affecting only one homologue can be directly identified by calculating the ratio between the

two homologues. Note that this approach is robust to any recurrent amplification bias that equally affects both homologues.

b. Validation of copy-number alterations in Chr. X. The haplotype phase was inferred from the sequence of a singleton cell with monosomic Chr. X. The DNA copy number analysis alone suggested frequent Chr. Xq gains shared in many daughter pairs (including all controls, Supplementary Table 2). We considered it unlikely that these inferred copy number alterations were genuine because they were not present in the bulk sample (Supplementary Table 2). We calculated the haplotype copy number to distinguish true copy-number alterations from a potential systematic amplification bias for Chr. X. Each dot represents the haplotype copy number calculated as the average coverage at all sites within each bin for which phase information could be obtained (i.e. where there was coverage in the reference cell that we sequenced with Chr. X monosomy). Haplotype copy number of Chr. X in 0.1 Mb bins in MN3 confirmed that the small yet significant gain in Xq relative to Xp affects both haplotypes equally, thus excluding the Xq gain as a genuine copy number alteration. Initial read-depth-based copy number analysis (Supplementary Table 2, Methods) also implied that there could be a copy number asymmetry for Chr. Xq for the two daughters from sample C1 as well as for Chr. X between the MN7 daughters (Fig. 2b). Calculation of the copy number of Chr. X for 0.25 Mb bins in C1 confirmed that there is no difference between the two haplotypes in Xq; thus, any variation in the combined sequence coverage likely also reflects amplification bias. By contrast, the copy number of Chr. X in MN7 (1 Mb bins) indicated that there is a true gain of a single haplotype that generates trisomy for Chr. X that is shared (preexisting) in both daughters.

c. Use of haplotype copy number to validate the 3:2 segregation patterns in the MN7 and MN8 daughters inferred from DNA copy number. The haplotype phase for Chr. 1 was determined from the sequence of the MN1 minus cell. Coverage of the intact haplotype (blue dots) is evenly distributed between the daughters and throughout the chromosome; the mean coverage of this intact haplotype was used to calculate the normalized copy number of the missegregated/gained haplotype. For the MN7 daughters, there is a single copy gain of Chr. 1q in the plus daughter. By contrast, there is no gain of Chr. 1p in either cell, providing an internal control. In MN8, nearly an entire copy of Chr. 4 was gained in the plus daughter, with the exception of two segments partitioned into the minus daughter. The gains and losses both occurred only to the missegregated haplotype (orange). The reciprocal gain and loss of these segments in the two daughter cells illustrates the sensitivity of the method.



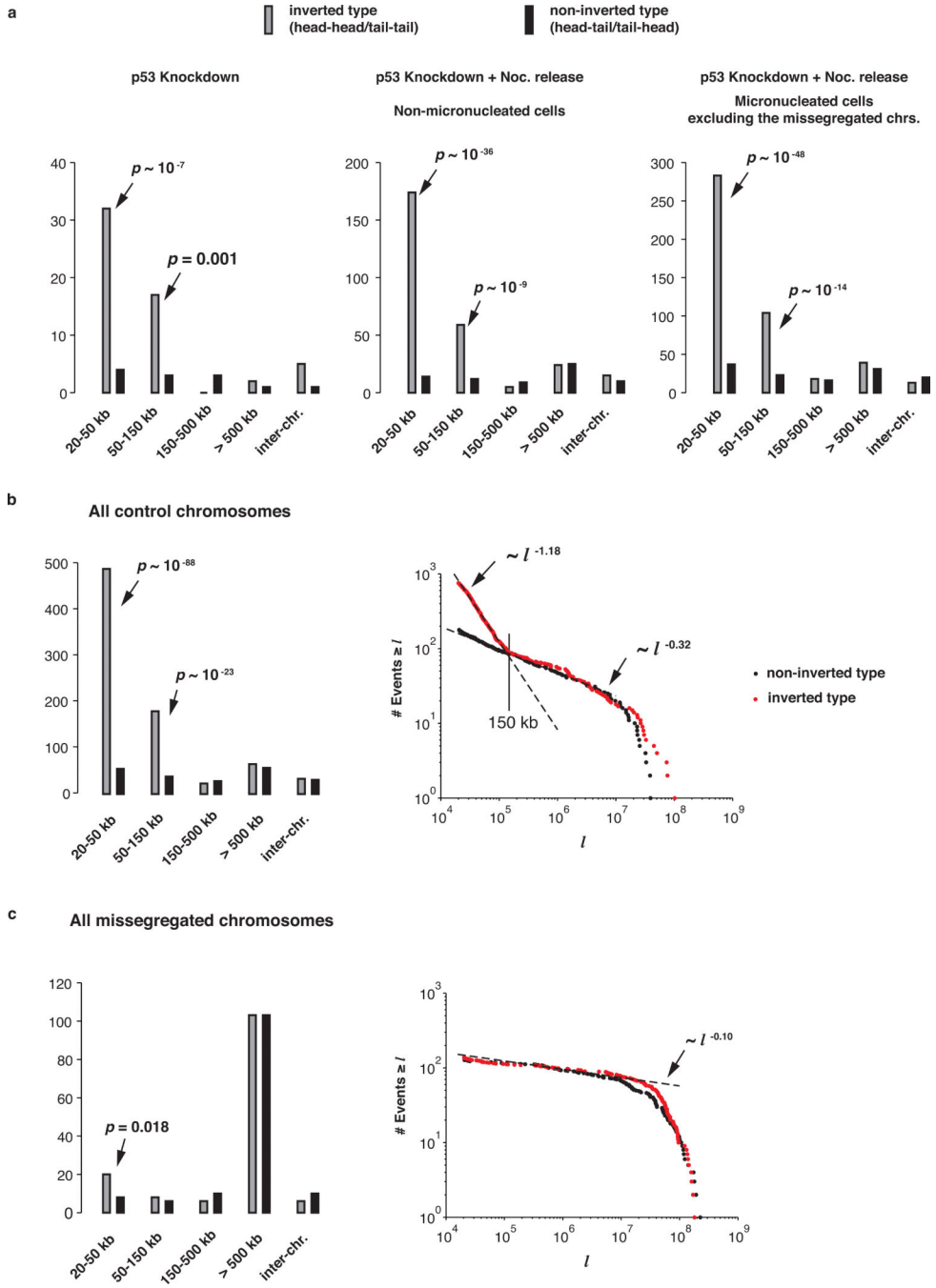
Extended Data Figure 4. Statistical analysis of the enrichment of structural variants on missegregated chromosomes

a. No enrichment of structural variants was observed in control chromosomes, including all chromosomes after the division of non-micronucleated mothers plus all normally segregated chromosomes after the division of micronucleated mothers. Shown are the frequencies of short- and long-range chromosomal rearrangements, at inverted (“head-to-head/tail-to-tail”) or non-inverted (“head-to-tail/tail-to-head”) orientations, detected in all control chromosomes, plotted for each chromosome arm. The frequencies were calculated by dividing the total number of rearrangement breakpoints of each type for each chromosome

by the total length of the chromosome and after correcting for copy-number alterations for each chromosome and the detection sensitivity in each sample. Fluctuation around the mean value across the genome is insignificant for all groups ($p > 0.05$, χ^2 test for a normal distribution based on the observation).

b. Enrichment of structural variants specifically on the missegregated chromosomes identified by asymmetric copy number. Top: frequency of all structural variants (breakpoints per Mb, normalized for DNA copy number and detection sensitivity) detected in the missegregated chromosomes (both plus and minus cell combined) as compared with all the remaining, normally segregated chromosomes; middle: frequency of intrachromosomal SVs with breakpoint distance < 150 kb, showing no enrichment; bottom: frequency of long-range SVs (intrachromosomal SVs with breakpoint distance > 150 kb or joining different chromosomes). *P*-values are derived from a one-sided Poisson test (Methods).

c. Mutually exclusive distribution of SV breakpoints between the two daughters. There are three categories of events: those unique to cell “a”, those unique to cell “b”, and those shared between both cells. The frequencies in each category can be estimated given the detection sensitivity in each library, which can be inferred from the sequence coverage at heterozygous SNPs (Methods). By a multinomial test, the large number of SVs detected in the missegregated chromosome that are unique to each daughter cell cannot be explained by incomplete detection of preexisting SVs shared between the cells. This contrasts with the few shared events (one each in MN1 and in MN3; none in the others). This conclusion holds for all daughter cell pairs except those from the MN2 and MN5 mothers. For the MN2 daughters the small number of evaluable events does not reach statistical significance. The MN5 daughters are the one negative example where there does not appear to be chromosomes that underwent any significant rearrangement.



Extended Data Figure 5. Length distribution of structural variants detected in single cells

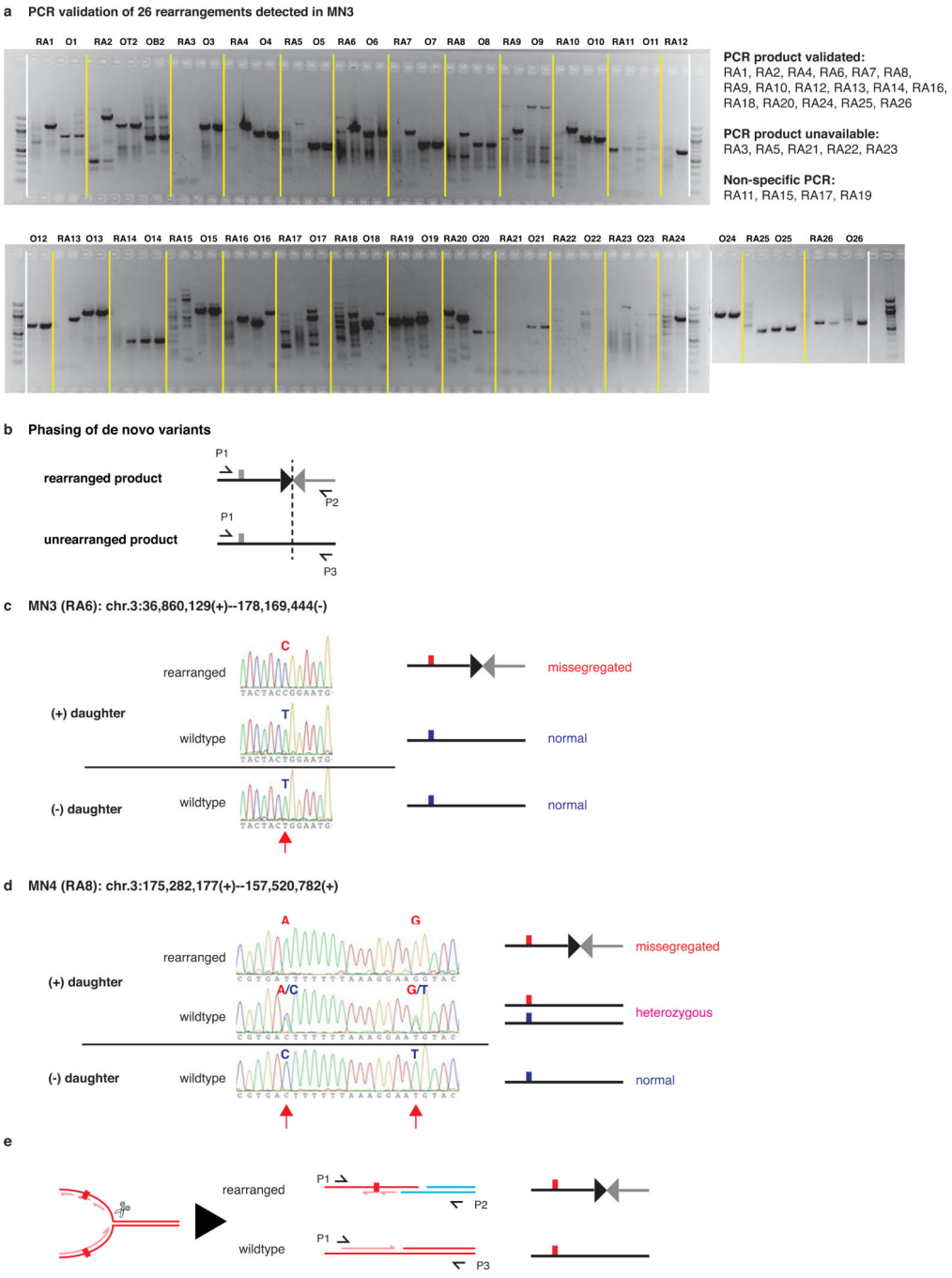
a. Structural variants in control samples: Short-range rearrangements are enriched for the inverted orientation. The number of structural variants (SV) detected in three groups of controls broken down by the distance between the rearranged sequences. Left: p53 knockdown cells (C1-C4); Middle: p53 knockdown + nocodazole release (N1-N6); Right: micronucleated cells (MN1-MN9), but excluding the missegregated chromosome. Note a significant enrichment of events in the inverted orientation is observed in the 20-150 kb range in all three groups (binomial test, $p < 0.05$), but there was no such enrichment for SV

events with breakpoint distances exceeding 150 kb or joining different chromosomes. The enrichment of short-range inverted type rearrangements can be explained by the fact that Phi-29-based multi-strand displacement (MDA) reaction used to amplify DNA generates frequent short-range inverted chimeras[46].

b. Control chromosomes: Scaling relationship between the frequency of SVs and the distance between breakpoints. Left: combined frequency of SVs from all control chromosomes from **a** above. Right: The cumulative distribution of inverted (red dots) or non-inverted (black dots) SV events as a function of breakpoint distance. Note: the graph shows the accumulation from infrequent long-range rearrangements (starting on the right) to more frequent short-range rearrangements (finishing on the left). We expect that genuine rearrangements should equally favor inverted or non-inverted orientations and attribute the bias towards the inverted-type to MDA artifacts. Thus, as a filter for potential MDA artifacts, we used power-law scaling to identify the breakpoint distance above which inverted and non-inverted type rearrangements occur with approximately equal frequency. In the 20-100 kb range, the data for inverted-type rearrangements (632 events total) were best fit by a power law decay of −1.176 (+/− 0.006, 95% confidence interval) and for the non-inverted type rearrangements (85 events total) by a power law decay of −0.395 (+/− 0.005). This difference is lost in the range of 200 kb-5 Mb, where the data for inverted-type rearrangements (53 events total) were best fit by a power law decay of −0.337 (+/− 0.02) and for the non-inverted type events by a power law decay of −0.31 (+/− 0.005). The power law fitting for the inverted-type events in the 20-100 kb range and the power law fitting for events in both groups in the 200 kb-5 Mb range (−0.322 +/− 0.007) intersected at ~150 kb. This established an operational cutoff of 150 kb to define “long-range” rearrangements, above which there should be no enrichment of inverted-type MDA-generated artificial chimeras.

c. Missegregated chromosomes: Scaling relationship between the frequency of SVs and the distance between breakpoints. Left: frequency of SVs from all missegregated chromosomes. By contrast with the controls in **b**, there is a marked enrichment of long-range SVs in this group with breakpoints > 500 kb apart. Indeed, the frequency of long-range rearrangements in these samples is substantially higher than the frequency of short-range rearrangements. The fact that these SVs are concentrated on the missegregated chromosome in the plus daughter cell (Fig. 3b, Extended Data Table 1) suggests that these are genuine rearrangements of the underreplicated chromosome from the micronucleus. Considering only the missegregated chromosome, we find a less pronounced difference in the ratio between inverted and non-inverted type rearrangements, even for short-range events. We speculate that this occurs because the missegregated chromosome also has a high frequency of genuine short-range rearrangements that are not biased to be in an inverted orientation. These numbers could also reflect a smaller sample size of rearrangements or possibly be due to rearrangement of the damaged chromosome prior to amplification, altering the relationship of the starting sequence relative to the reference genome. By establishing the 150 kb cutoff to filter for artifacts, we thus likely exclude some genuine short-range events. Finally, we note that our power law scaling analysis for SVs detected in the missegregated chromosomes is consistent with other independent estimates for the likelihood of intrachromosomal contacts. The power law dependence of $\sim l^{-0.10}$ (power decay −0.105 +/− 0.003 from all events in the range of 150 kb ~ 5 Mb) is equivalent to a density distribution

of $p(l) \sim l^{-1.10}$. This dependence is close to the length distribution of somatic copy number alterations ($\sim l^{-1}$) observed in cancers[47], and also to the distribution of breakpoint distances for somatic chromosomal rearrangements (J. Wala and R. Beroukhim, personal communication). Moreover, it is consistent with the probability of intra-chromosomal contacts ($\sim l^{-1.03}$) inferred from Hi-C experiments[48].The distribution of the long-range breakpoint distances from the missegregated chromosomes shown here (150 kb ~ 10 Mb) is significantly different from the distribution of rearrangements from all control chromosomes, shown above in **b** ($p = 0.0043$, Kolmogorov-Smirnov test). By contrast, pairwise comparisons of the distribution of the control samples (**a**, above) showed no significant differences ($p = 0.6$ for all events, $p = 0.9$ for long range events, K-S test).



Extended Data Figure 6. Validation of rearrangements by PCR and haplotype phasing

a. PCR validation of 26 (out of 66) intra-chromosomal rearrangements detected in Chr. 3 of the MN3 plus daughter. The complete results for all samples are summarized in Supplementary Table 5 with examples of the Sanger sequencing results shown below (**c**,**d**). For each rearrangement, two PCR reactions ("RA" across the putative rearrangement junction, "O" for the reference sequence) were performed on the MDA amplified DNA from both daughter cells. For each sample indicated above the gel, the leftmost lane is the PCR product for the rearrangement-specific primers in the minus daughter; the left middle lane

the rearrangement-specific primers in the plus daughter; the right middle lane, the reference-specific primers in the minus daughter; and the right lane, the reference-specific primers in the plus daughter. In “RA2” where there are heterozygous SNPs on both ends of the rearrangement junction, PCR was performed to generate (and have validated) the genotypes at both sites.17 out of 26 PCR products confirmed the rearrangement junction after Sanger sequencing. The remaining 9 PCR reactions generated no product or products that did not correspond to the rearrangement sequence; several of these were due to low sequence complexity near the rearrangement junction that resulted in non-specific primer pairs.

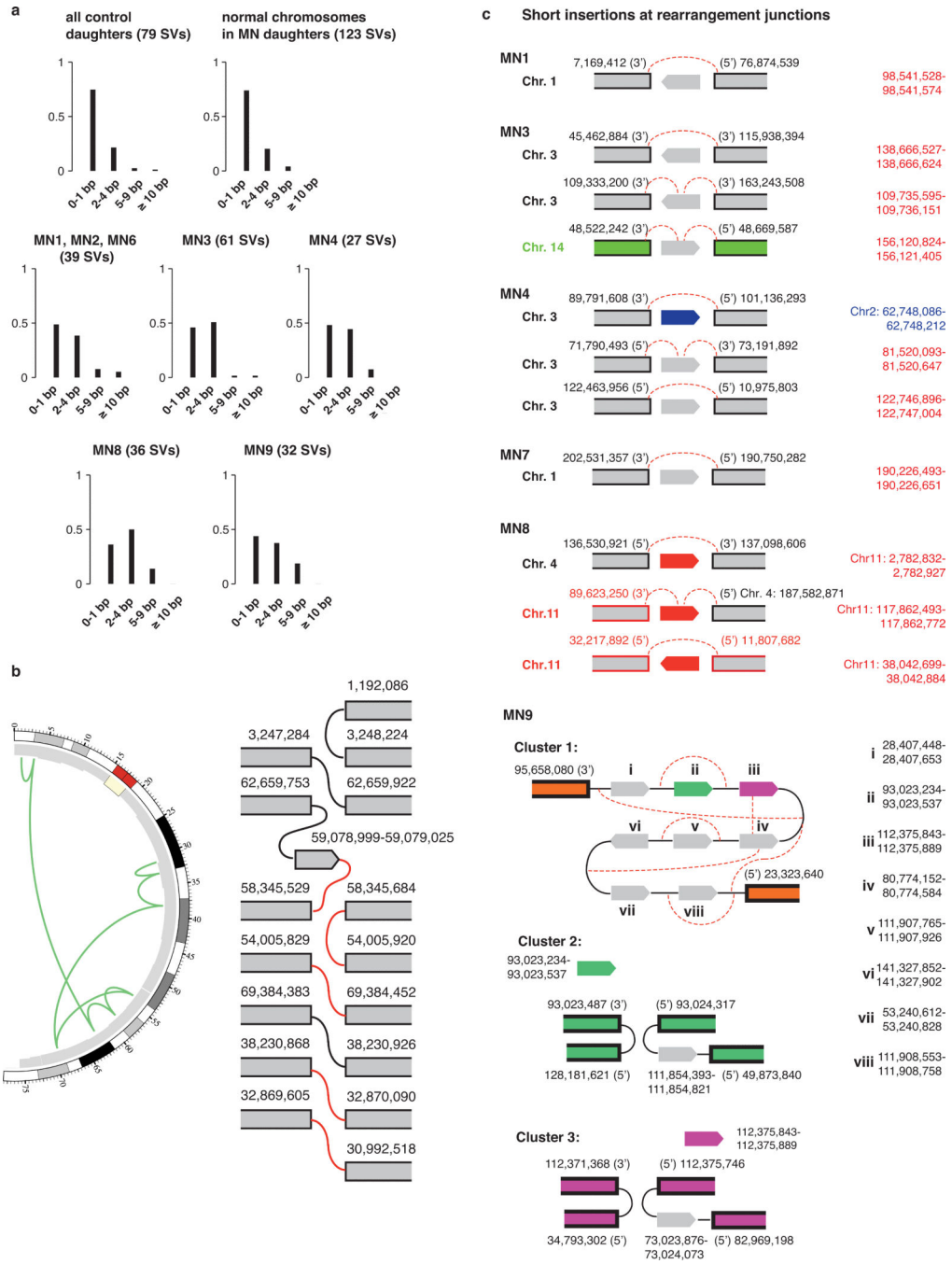
b. Cartoon showing the strategy for validating rearrangements and associating these rearrangements with the missegregated haplotype. Forward PCR primers were chosen 5’ from an informative heterozygous site and reverse primers were chosen 3’ of the rearrangement breakpoint junction, either in the rearranged DNA sequence or in the reference DNA sequence. PCR was performed to amplify both the rearranged product and the control reference genome product, followed by Sanger sequencing. Here the undetermined base at the polymorphic site is colored grey.

c. Example of haplotype validation for a rearrangement in MN3 based on an adjacent “C/T” SNP. From the minus daughter, we were only able to amplify the reference product, which showed a “T” at the polymorphic site. Because MN3 underwent a 2:1 segregation, the missegregated haplotype was inferred to have a “C” at this polymorphic site. From the plus daughter, we amplified both the rearranged and reference products. As expected, the rearranged showed a “C” at the polymorphic site, indicating that the rearrangement occurred on the missegregated chromosome. Also as expected, there was a “T” at the polymorphic site on the reference product. The base associated with the missegregated haplotype is in red; the base for the normal haplotype is in blue.

d. Example of haplotype validation for a rearrangement in MN4 yielding three products. In this case, there are two informative polymorphic sites near the rearrangement: the “A+G” pair is associated with the missegregated haplotype and the “C+T” pair is associated with the intact reference haplotype. As expected, the minus daughter had the reference product showing the “C+T” haplotype. Also as expected, the plus daughter had the rearranged product showing the “A+G” haplotype as well as the reference product showing the “C+T” haplotype. Somewhat unexpectedly, the plus daughter also had a third product: a reference product associated with the “A+G” missegregated haplotype. We speculate that the presence of both rearranged and reference products with the missegregated haplotype results from partial replication of this region of the missegregated chromosome.

e. Proposed partial replication/replication fork breakage model to explain the presence of three products detected in **d** above. Shown on the left is a replication fork on the missegregated chromosome, in the middle are the products of replication or recombination, and on the right are the products with the red bar indicating that the products are associated with the missegregated haplotype. The original DNA strands are in dark red; the newly synthesized ones are in light red; DNA from a distal rearrangement partner locus is in blue. We hypothesize that breakage of the replication fork (scissors) generates a single-end break that recombines with the distal locus shown in blue. The other end of the replication fork generates a reference product. Importantly, both products should contain the base(s) associated with the missegregated haplotype. The presence of both rearranged and reference products containing the missegregated haplotype could also occur if the rearrangement is an

artificial chimera that arose during MDA amplification. However, such artifacts should not be restricted to a single homologous chromosome: the highly significant enrichment of rearrangements on the missegregated chromosome and their association with the missegregated/gained haplotype establish that most of these rearrangements are genuine. Notes: 1. Rearrangements were not only associated with the missegregated haplotype by PCR, but in some cases this association could be made directly by sequencing read-based phasing utilizing either discordant read pairs or split reads that covered heterozygous SNP sites close to either side of the breakpoint. This analysis enabled us to determine the haplotype association for 10 events in MN3 (3 in addition to Sanger sequencing), 5 events in MN4 (3 in addition to Sanger sequencing), and 6 events in MN8 (5 in addition to Sanger sequencing), all confirmed to be associated with the gained haplotype. 2. For the daughters with a 3:2 segregation pattern, even if the plus cell contained an intact copy of the missegregated chromosome, we expect a replicate of this homologue to be present in the minus cell because that copy of the homologue was normally segregated. Because we did not detect any rearrangements in the minus cell, the rearrangements detected in the plus daughter can only come from the missegregated copy of the homologue in that daughter.

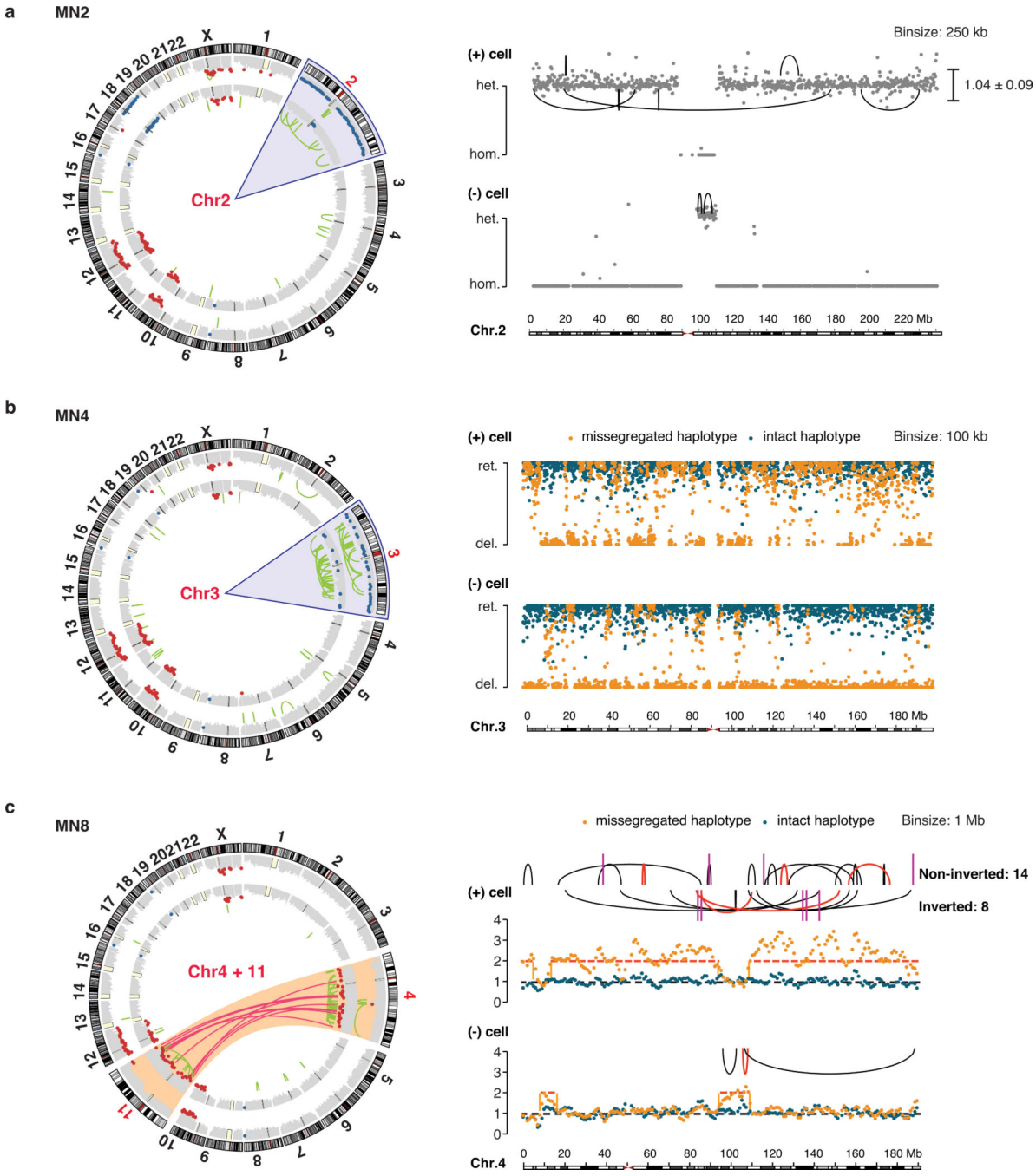


Extended Data Figure 7. Sequence features at the rearrangement junctions

a. Length distribution of microhomology at the junctions of rearrangements in control chromosomes (top), in the missegregated chromosomes of daughters with 2:1 segregations (middle), and in the missegregated chromosomes from daughters with 3:2 segregations (bottom). The distribution of microhomology at rearrangement junctions detected in all control daughters is indistinguishable to that detected in the control chromosomes in the micronucleated daughters, with ~ 75% of events showing 0-1 bp homology. By contrast,

rearrangements in the missegregated chromosomes contain a higher percentage of microhomology: more than 50% of all events exhibited > 1bp homology in every sample. **b**. Chained translocations between breakpoints on Chr. 18 in the MN6 plus daughter. Left: CIRCOS plot for the translocation chain in Chr.18; Right: translocation links between 16 breakpoints, 14 of which had paired break ends forming a chain. All events were validated by PCR, and red links reflect rearrangements that were associated to the gained haplotype through nearby SNPs (Supplementary Table 5).

c. Examples of short (50-500 bp) insertions at breakpoint junctions. Dashed red links represent read pairs supporting the given junction. In addition to insertions derived from the missegregated chromosome and inserted into rearrangements in the missegregated chromosome, we identified additional examples as follows: For the MN3 sample we identified one example of Chr. 3 insertion into a rearrangement between two loci in (normally segregated) Chr. 14. For the MN4 sample we also identified one example where a short segment from (normally segregated) Chr. 2 was inserted into a rearrangement between loci in the missegregated Chr. 3. For the MN8 sample, where both Chr. 4 and Chr. 11 were inferred to have been fragmented in the same micronucleus, we identified one example where a rearrangement between loci on Chr. 4 contained a 95 bp insertion from Chr. 11 in the rearrangement junction and another example of a 279 bp segment originated from Chr. 11 inserted into a Chr. 4-Chr. 11 rearrangement junction. In MN9, we have identified >20 insertions at sites of long-range rearrangements on Chr. 8 via local sequence assembly (Supplementary Table 6). Here we show one translocation junction containing 8 short segments from all over Chr. 8. The segments at the boundaries of the rearrangement are in bold outline. Between the boundaries are 8 short insertions (47-433 bps, grey, green or purple bars) from different parts of Chr. 8 (cluster 1). Green and purple bars indicate insertions originating at or near breakpoints of other rearrangements (clusters 2 and 3, correspondingly green or purple). As the average detection sensitivity was ~ 35% for each library (Supplementary Table 1), it is likely that some insertions could have been missed. Importantly, the insertions only come from distal sites on the missegregated chromosome(s) and such insertions are not detected in any control samples. Note: 1. We can only determine the presence of insertion sequences < 500 bp as most sequencing fragments are shorter than 600 bp (99% of fragments are shorter than 600 bp in the DNA library of the MN9 plus daughter, < 513 bp for the MN1 plus daughter, < 380 bp for the MN3 plus daughter, MN4 daughters, MN8 plus daughter, and < 350 bp for the MN7 plus daughter). 2. The MN9 sample has many more insertions than the other samples with the inserted segments frequently derived from sequences near other rearrangement breakpoints. The explanation for this is not clear and future work will require experiments of a larger sample size. However, unlike the MN1-MN8 daughters, which were isolated shortly after division of the micronucleated mother, the MN9 plus cell remained arrested for a ~ 2 day period of time while the minus daughter divided twice. We speculate that the missegregetaed chromosome in the plus daughter from the MN9 daughter pair have undergone MMBIR as part of the mechanism that combined these Chr. 8 fragments. It is also possible that breakpoint ends in the MN9 plus cell could have been fragmented into small segments.



Extended Data Figure 8. Evidence of chromosomal fragmentation detected from haplotype copy-number analyses in three examples, MN2, MN4, and MN8

a. Inference of chromothripsis of Chr. 2 in the MN2 daughters without knowledge of the Chr. 2 haplotype phase. Left: CIRCOS plot. Right: Plot of the heterozygosity coefficients in 250 kb bins, with rearrangement links indicated (above: non-inverted type; below: inverted type). Note that the MN2 daughters underwent a 2:1 distribution of the missegregated chromosome, implying that any chromosomal loss generates loss-of-heterozygosity. The heterozygosity plot demonstrates that a pericentric fragment of 2q is partitioned into the minus cell, whereas the remainder of Chr. 2 is in the plus daughter. Chromosomal

rearrangements are only observed in heterozygous regions, consistent with heterozygosity originating from the damaged/underreplicated homologue from the micronucleus. Each dot represents the heterozygosity coefficient in a 250 kb bin (~ 50 heterozygous sites per bin). Bins with fewer than 25 phaseable heterozygous sites or showing only 1~2 observed heterozygous sites are not shown.

b. Haplotype copy number of Chr. 3 in the MN4 daughters (100 kb resolution). Left: CIRCOS plot. Right: Chr.3 haplotype copy number in MN4 daughters calculated from the chromosomal haplotype phase derived from the sequence of the MN3 minus daughter. Each dot represents average haplotype copy number in a 100 kb bin: the normally segregated haplotype (blue dots) is equivalently detected in both daughters whereas the fragmented haplotype (orange dots) shows oscillating and reciprocal retention and loss between the two daughters.

c. Haplotype copy number of Chr.4 in the MN8 daughters with rearrangement links. The haplotype phase for Chr. 4 was inferred from the sequence of a single cell with Chr. 4 monosomy (Extended Data Fig. 2c). Gains of the missegregated haplotype (orange dots) are reciprocal in both daughters; except for one rearrangement in the plus daughter, all detected breakpoints, including both intrachromosomal events (links) and interchromosomal events with Chr. 11 (vertical magenta lines), are restricted to regions of gains in the missegregated haplotype. Red links indicate translocations that are associated with the missegregated haplotype by informative SNPs near the breakpoints; black links indicate rearrangements for which phasing validation was not performed (a subset of which have no adjacent SNPs).

Extended Data Table 1

Summary table of the rearrangements detected in the daughters of micronucleated mothers and their enrichment on the missegregated chromosome.

Sample	MS Chr.	cell	mean CN of MS Chr.	# of de novo SVs		# of short-range SVs		Poisson test	Fisher's test	PCR validation	haplotype validation
				MS. Chr.	other Chr.	MS. Chr.	other Chr.				
MN1	**Chr. 1**	**(+) cell**	**2**	**19**	4	3	10	**$<10^{-10}$**	**10^{-12}**	7/10	3/3
		(-) cell	**1**	none	5	none	14				
MN2	**Chr. 2**	**(+) cell**	**1.7**	**7**	8	0	32	**10^{-10}**	**10^{-8}**	3/5	2/2
		(-) cell	**1.1**	**2**	3	3	40				
MN3	**Chr. 3**	**(+) cell**	**2**	**67**	12	4	20	**$<10^{-10}$**	**10^{-24}**	17/26	15/15 + 3[†]
		(-) cell	**1**	none	1	none	9				
MN4	**Chr. 3**	**(+) cell**	**1.5**	**25**	12	3	77	**$<10^{-10}$**	**10^{-28}**	10/14	5/5 + 3
		(-) cell	**1.2**	**13**	13	5	62				
MN5	Chr. 7	**(+) cell**	2	1	8	3	17	0.34	0.28		
		(-) cell	1	none	6	none	11				
MN6	**Chr. 18**	**(+) cell**	**2**	**5**	4	0	21	**10^{-9}**	**10^{-4}**	8/8[*]	5/5
		(-) cell	**1**	1	20	0	13				
MN7	**Chr. 1q**	**(+) cell**	**2.9**	**8**	1	1	9	**$<10^{-10}$**	**10^{-7}**	3/6	1/1
		(-) cell	**2.1**	none	5	none	17				
MN8	**Chrs. 4 and 11**	**(+) cell**	**2.9**	**36**	8	15	51	**$<10^{-10}$**	**10^{-22}**	7/8	1/1 + 5
		(-) cell	**2**	**3**	6	4	23				
MN9	**Chr. 8**	**(+) cell**	**2.8**	**41**	7	3	29	**$<10^{-10}$**	**10^{-28}**		
		(-) cell	**2**	1	2	1	20				

Abbreviations: MS. Chr.: missegregated chromosome; CN: copy number. *P*-value for enrichment was derived from a one-sided Poisson test of the observed number of rearrangements in the missegregated chromosome against the average frequency of rearrangements in the whole genome for each daughter pair. Fisher's test compared the observed numbers of short and long-range rearrangements on the missegregated chromosome (both daughters) with the corresponding numbers from the remaining chromosomes in the same daughter cell pair. The enrichment for long-range rearrangements on the missegregated chromosomes was highly significant as compared to short-range events, which had no enrichment. PCR validation confirmed the presence of a PCR amplified product and the mapping of the sequence to the two partner loci of the rearrangement (but does not exclude artificial chimeras); haplotype validation further associated the rearrangement to the gained (missegregated) haplotype inferred from haplotype copy number analysis (Extended Data Fig. 3; Methods). Validation results are presented as the number of validated cases divided by the number of total attempts; in haplotype validation, the number of attempts only includes cases where the informative SNP base can be confidently determined by Sanger sequencing that span the rearranged junctions (Supplementary Table 5).

[/]We included additional events in the haplotype validation ("+") when the Sanger sequencing was unable to generate the haplotype information, but direct phasing of the rearrangement was possible through supporting short-reads with an adjacent SNP. In both MN3 and MN4 plus daughters, there was one translocation between the missegregated chromosome and a different chromosome (see Extended Data Fig. 7c): these events were counted as "*de novo* events in the missegregated chromosome" although excluding them does not change the statistical analysis.

[*]In the MN6 plus daughter, the standard rearrangement analysis detected five SV events in Chr. 18, three of which were inferred to be in a chain of eight translocations between multiple double-strand breaks. All these eight events were validated by PCR and sequencing, with five out of eight associated to the missegregated haplotype. In all statistical analysis of this chromosome (Extended Data Fig. 4b,c) we only counted the five events to ensure consistency with the estimated detection sensitivity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We would like to thank R. Beroukhim, J. Walter and N. Ganem for comments on the manuscript, V. Adalsteinsson and C. Love for help with preliminary experiments, and A. Salic and S. Rosenberg for reagents. A.S. was a resident in the American Board of Radiology Holman Research Pathway; H.C. was a fellow of the Leukemia and Lymphoma Society; C.Z.Z, J.F. and M.M. were supported by the Bridge Project of the Dana-Farber Cancer Institute and the Koch Institute of MIT; M.M. and D.P. were supported by the Claudia Adams Barr Program in Innovative Cancer Research; D.P. is a HHMI investigator and is supported by NIH grant GM083299-18.

REFERENCES

1. Yates LR, Campbell PJ. Evolution of the cancer genome. Nature reviews. Genetics. 2012; 13:795–806. doi:10.1038/nrg3317.
2. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nature reviews. Genetics. 2014; 15:585–598. doi:10.1038/nrg3729.
3. Zhang CZ, Leibowitz ML, Pellman D. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. Genes & development. 2013; 27:2513–2530. doi:10.1101/gad.229559.113. [PubMed: 24298051]
4. Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011; 144:27–40. doi:10.1016/j.cell.2010.11.055. [PubMed: 21215367]
5. Rausch T, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell. 2012; 148:59–71. doi:10.1016/j.cell.2011.12.013. [PubMed: 22265402]
6. Kloosterman WP, et al. Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. Cell Rep. 2012; 1:648–655. doi:10.1016/j.celrep.2012.05.009. [PubMed: 22813740]
7. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer genomes. Cell. 2013; 152:1226–1236. doi:10.1016/j.cell.2013.02.023. [PubMed: 23498933]
8. Liu P, Carvalho CM, Hastings PJ, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Genet Dev. 2012; 22:211–220. doi:10.1016/j.gde.2012.02.012. [PubMed: 22440479]
9. Liu P, et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. Cell. 2011; 146:889–903. doi:10.1016/j.cell.2011.07.042. [PubMed: 21925314]
10. Kinsella M, Patel A, Bafna V. The elusive evidence for chromothripsis. Nucleic Acids Res. 2014; 42:8231–8242. doi:10.1093/nar/gku525. [PubMed: 24939897]
11. Crasta K, et al. DNA breaks and chromosome pulverization from errors in mitosis. Nature. 2012; 482:53–58. doi:10.1038/nature10802. [PubMed: 22258507]
12. Terradas M, Martin M, Tusell L, Genesca A. Genetic activities in micronuclei: is the DNA entrapped in micronuclei lost for the cell? Mutat Res. 2010; 705:60–67. doi:10.1016/j.mrrev.2010.03.004. [PubMed: 20307686]
13. Hoffelder DR, et al. Resolution of anaphase bridges in cancer cells. Chromosoma. 2004; 112:389–397. doi:10.1007/s00412-004-0284-6. [PubMed: 15156327]
14. Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. Catastrophic nuclear envelope collapse in cancer cell micronuclei. Cell. 2013; 154:47–60. doi:10.1016/j.cell.2013.06.007. [PubMed: 23827674]
15. Shee C, et al. Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. Elife. 2013; 2:e01222. doi:10.7554/eLife.01222. NOW SUPPLEMENT. [PubMed: 24171103]
16. Ganem NJ, Pellman D. Linking abnormal mitosis to the acquisition of DNA damage. J Cell Biol. 2012; 199:871–881. doi:10.1083/jcb.201210040. [PubMed: 23229895]

17. Uetake Y, Sluder G. Prolonged prometaphase blocks daughter cell proliferation despite normal completion of mitosis. Curr Biol. 2010; 20:1666–1671. doi:10.1016/j.cub.2010.08.018. [PubMed: 20832310]

18. Francis JM, et al. EGFR variant heterogeneity in glioblastoma resolved through single- nucleus sequencing. Cancer Discov. 2014; 4:956–971. doi:10.1158/2159-8290.CD-13-0879. [PubMed: 24893890]

19. Hayashi MT, Cesare AJ, Fitzpatrick JA, Lazzerini-Denchi E, Karlseder J. A telomere-dependent DNA damage checkpoint induced by prolonged mitotic arrest. Nature structural & molecular biology. 2012; 19:387–394. doi:10.1038/nsmb.2245.

20. Zhang C-ZA, V.A. Francis J, Cornils H, Jung J, Maire C, Ligon KL, Meyerson M, Love JC. Calibrating genomic and allelic coverage bias in single-cell sequencing. Nature Communications. 2015; 6:6822. doi: 10.1038/ncomms7822.

21. Simsek D, et al. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet. 2011; 7:e1002080. doi:10.1371/journal.pgen.1002080. [PubMed: 21655080]

22. Chiang C, et al. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. Nat Genet. 2012; 44:390–397. S391. doi:10.1038/ng.2202. [PubMed: 22388000]

23. Storlazzi CT, et al. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. Genome Res. 2010; 20:1198–1206. doi:10.1101/gr.106252.110. [PubMed: 20631050]

24. Carroll SM, et al. Double minute chromosomes can be produced from precursors derived from a chromosomal deletion. Molecular and cellular biology. 1988; 8:1525–1533. [PubMed: 2898098]

25. Misteli T, Soutoglou E. The emerging role of nuclear architecture in DNA repair and genome maintenance. Nat Rev Mol Cell Biol. 2009; 10:243–254. doi:10.1038/nrm2651. [PubMed: 19277046]

26. Garsed, Dale W., et al. The Architecture and Evolution of Cancer Neochromosomes. Cancer Cell. 2014; 26:653–667. doi:10.1016/j.ccell.2014.09.010. [PubMed: 25517748]

27. Haber JE. Genome stability : DNA repair and recombination. Garland Science. 2014

28. Sen S, Hittelman WN, Teeter LD, Kuo MT. Model for the formation of double minutes from prematurely condensed chromosomes of replicating micronuclei in drug- treated Chinese hamster ovary cells undergoing DNA amplification. Cancer Res. 1989; 49:6731–6737. [PubMed: 2819718]

29. El Achkar E, Gerbault-Seureau M, Muleris M, Dutrillaux B, Debatisse M. Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. Proc Natl Acad Sci U S A. 2005; 102:18069–18074. doi:10.1073/pnas.0506497102. [PubMed: 16330769]

30. Sheltzer JM, et al. Aneuploidy drives genomic instability in yeast. Science. 2011; 333:1026–1030. doi:10.1126/science.1206412. [PubMed: 21852501]

31. Janssen A, van der Burg M, Szuhai K, Kops GJ, Medema RH. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. Science. 2011; 333:1895–1898. doi:10.1126/science.1210214. [PubMed: 21960636]

32. Knouse KA, Wu J, Whittaker CA, Amon A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. Proc Natl Acad Sci U S A. 2014; 111:13409–13414. doi:10.1073/pnas.1415287111. [PubMed: 25197050]

33. Zack TI, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013; 45:1134–1140. doi:10.1038/ng.2760. [PubMed: 24071852]

34. Malhotra A, et al. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. Genome Res. 2013; 23:762–776. doi:10.1101/gr.143677.112. [PubMed: 23410887]

36. Dean FB, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. (USA). 2002; 99:5261–5266. [PubMed: 11959976]

37. Voet T, et al. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. Nucleic Acids Research. 2013:1–20. [PubMed: 23143271]

38. De Bourcy, et al. A quantitative comparison of single-cell whole-genome amplification methods. PLoS One. 2014; 9:e105585. [PubMed: 25136831]

39. Wang Y, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature. 2015; 512:155–160. [PubMed: 25079324]

40. Evrony GD, et al. Cell lineage analysis in human brain using endogeneous retroelements. Neuron. 2015; 85:49–59. [PubMed: 25569347]

41. Zong C, et al. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2012; 338:1622–1626. [PubMed: 23258894]

42. Pinard R, et al. Assessment of whole genome amplification-induced bias through high- throughput, massively parallel whole genome sequencing. BMC Genomics. 2006; 7:216. [PubMed: 16928277]

43. Lage JM, et al. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. Genome Research. 2003; 13:294–307. [PubMed: 12566408]

44. Evrony GD, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell. 2012; 151:483–496. [PubMed: 23101622]

45. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high- throughput sequencing. Nucleic Acids Research. 2012; 40:e72. [PubMed: 22323520]

46. Lasken RS, Stockwell TB. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC biotechnology. 2007; 7:19. [PubMed: 17430586]

47. Beroukhim R, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899–905. doi:10.1038/nature08822. [PubMed: 20164920]

48. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. doi:10.1126/science.1181369. [PubMed: 19815776]

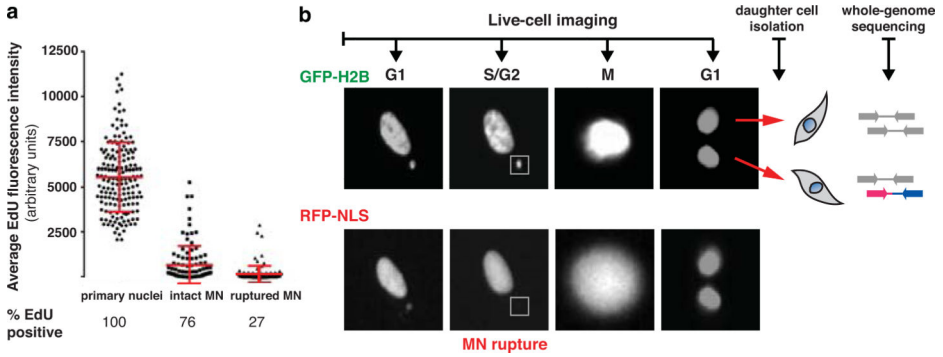


Figure 1. Look-Seq procedure to analyze DNA damage from the underreplicated chromosomes in micronuclei

a. Reduced DNA replication in both intact and ruptured micronuclei. Left: Fluorescence intensity measurements after continuous EdU labeling following release from a nocodazole block. EdU intensity is normalized to nuclear area (N >100 from two experiments for each category, see Methods). Red bars: mean and standard deviation. **b.** Look-Seq strategy. Disruption of the micronuclear envelope is visualized by the loss of a nuclear localized fluorescent protein (RFP-NLS); reincorporation of the micronucleus is inferred from the absence of micronuclei in either G1 daughter.

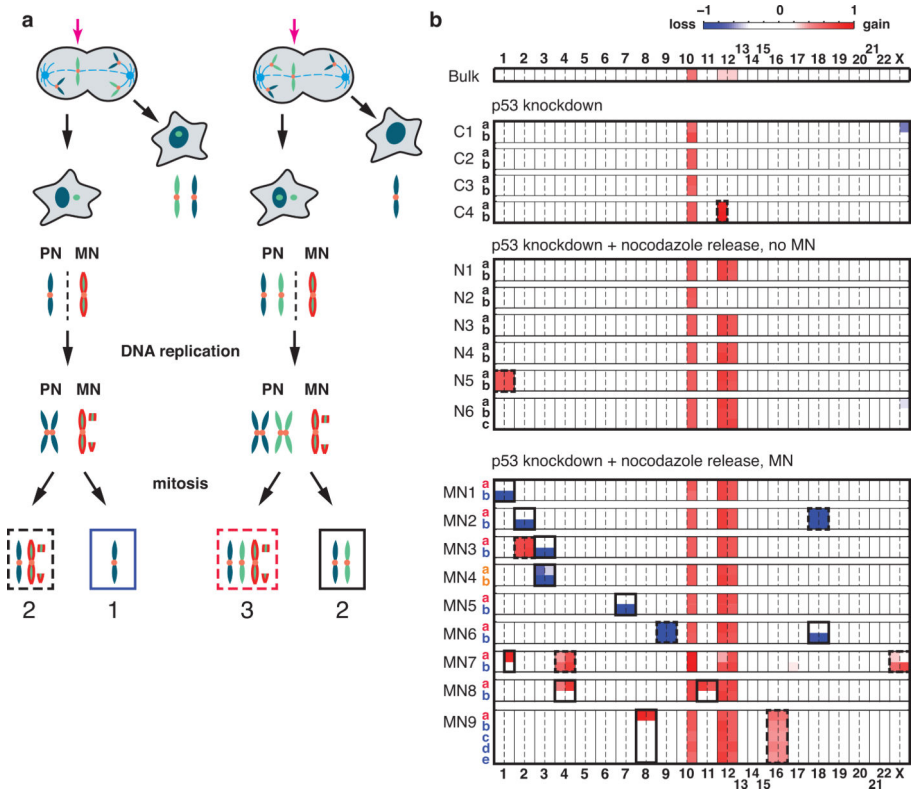


Figure 2. Identification of the missegregated chromosome by DNA copy number analysis

a. Predicted DNA copy number outcomes for daughter cells derived from a micronucleated mother cell. Left: Lagging chromosome (magenta arrow) is correctly segregated, but partitioned into a micronucleus (MN); Right: lagging chromosome is missegregated into a micronucleus. The chromosome in the micronucleus is underreplicated and asymmetrically segregated, resulting in either a 2:1 (left) or 3:2 (right) copy number ratio in daughter cells. Homologue in the micronucleus: light green; homologue in the primary nucleus (PN): blue; red outline indicates the underreplicated chromosome from the micronucleus. Hatched box: "plus" daughter; solid box: "minus" daughter. **b**. Heatmap of arm-level DNA copy number (log2 ratio) in daughter cells derived from micronucleated mothers (MN1-MN9) or non-micronucleated controls (C1-C4, N1-N6, see Supplementary Table 2). Mother cells are listed on the left with daughters, "a" and "b". Chromosomes are shown on the x-axis; dotted lines separate p and q arms. Preexisting aneuploidies in mother cell: dashed boxes; *de novo* missegregations with copy number asymmetry: solid boxes. Top panel: bulk RPE-1 line; top middle panel: daughters from non-micronucleated mothers; middle panel: non-micronucleated controls after nocodazole release; bottom panel: daughter cells from micronucleated mothers. For N6, one daughter cell divided once, producing three cells. For MN9, extra time was provided for additional cell divisions: The plus daughter did not divide whereas the minus daughter divided twice, generating five cells in total.

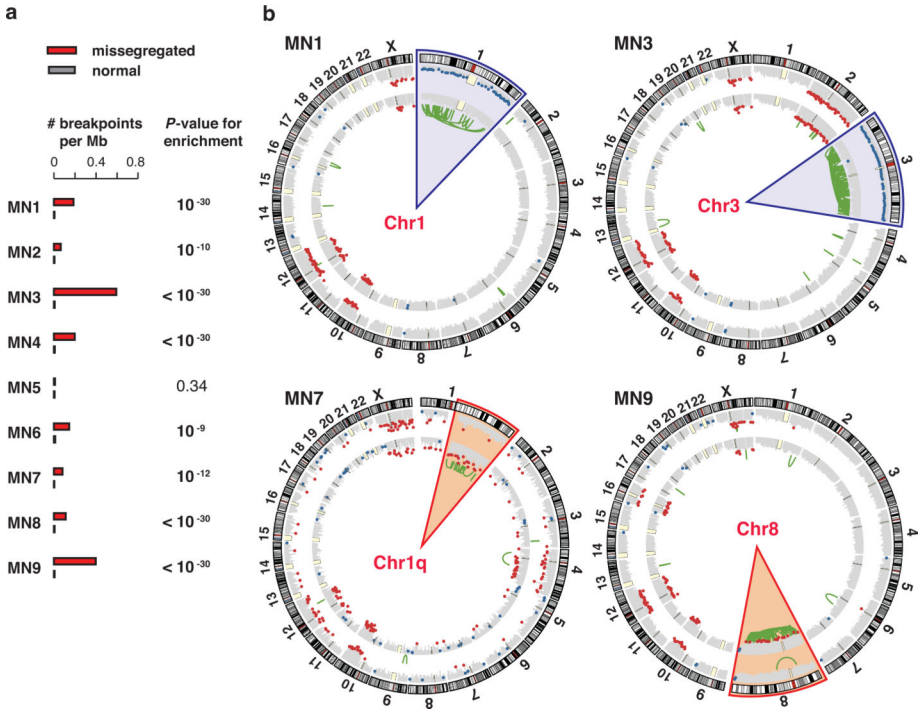


Figure 3. Enrichment of long-range rearrangements on the missegregated chromosome in the predicted daughter cell

a. Left: frequency of long-range rearrangements detected in the missegregated chromosomes (both plus and minus cell, red bars) as compared with the remaining chromosomes (grey bars): # breakpoints per Mb is normalized for DNA copy number and the detection sensitivity. Right: *P*-values for enrichment derived from a one-sided Poisson test (Methods). **b**. CIRCOS plots showing DNA copy number (grey histograms) and long-range intra-chromosomal rearrangements (green links) for four daughter pairs from micronucleated mothers. Outside ring: chromosome banding pattern. Grey histograms: DNA copy number (5 Mb bins) for the minus cell (outer ring) and the plus cell (inner). (For MN9, one "grandchild" of the minus daughter is shown, cf. Fig. 2). Red dots: bins with significant gains (>1.35x mean); blue dots: bins with significant loss (<0.6x mean). Loss of 19p and gain of Xq are seen in all single cell samples but not in the bulk and are attributed to systematic amplification bias (Extended Data Fig. 3b). Light blue cones: chromosomes with 2:1 missegregations; orange cones: 3:2 missegregations.

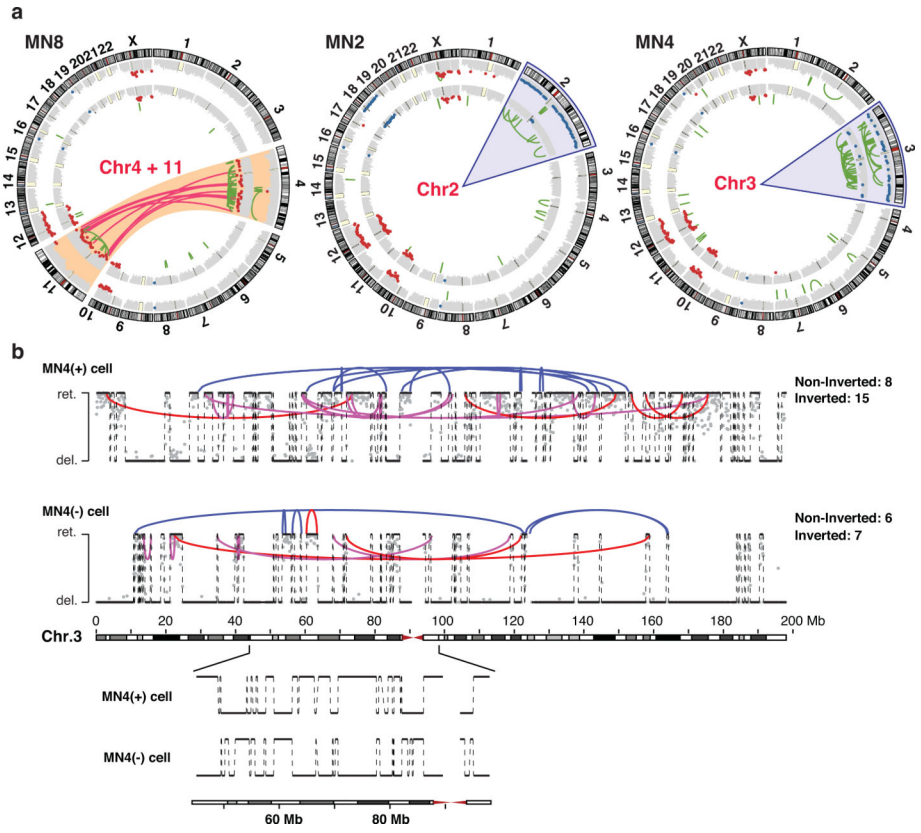


Figure 4. Two-state oscillating copy-number patterns characteristic of chromothripsis

a. CIRCOS plots for three daughter cell pairs where both daughters received fragments of the missegregated chromatid. The MN8 minus daughter contains two Chr. 4 fragments missing from the plus daughter; the MN2 minus daughter contains one segment of Chr. 2q missing from the plus daughter; for the MN4 pair, dozens of Chr. 3 fragments are reciprocally distributed between the daughters. **b**. Reciprocal distribution of fragments of the missegregated chromatid results in two-state copy number oscillations (retention or loss) of the missegregated haplotype in both daughters. Grey dots: average copy number of the missegregated haplotype (100 kb bins). The other haplotype is intact in both daughters (shown in Extended Data Fig. 8b). Blue links indicate non-inverted type rearrangements; magenta links indicate inverted type rearrangements. Red links indicate rearrangements directly associated with the missegregated haplotype by nearby SNPs (Supplementary Table 5). Reciprocal distribution of Chr. 3 fragments is evident from the complementarity of DNA copy-number in both daughters shown for the 50-100 Mb region (inset, bottom).

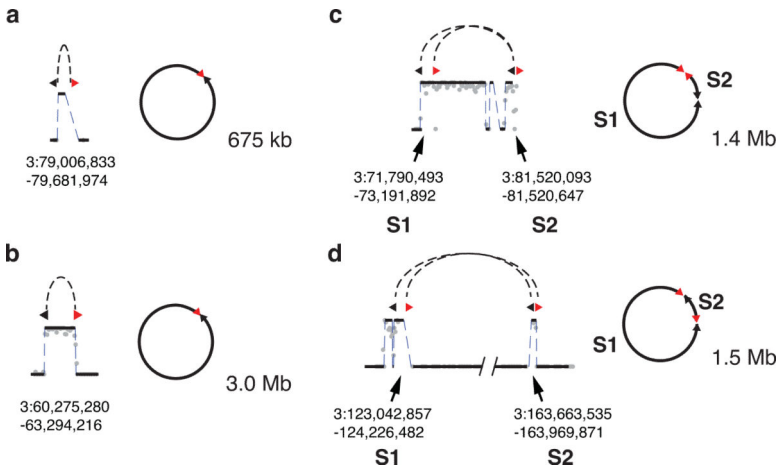


Figure 5. Circular chromosomal structures resulting from chromothripsis

Shown are four circular structures formed by single segments (**a** and **b**) or by multiple separate segments (**c** and **d**); **a** and **c** are from the MN4 plus daughter; **b** and **d** are from the MN4 minus daughter. In each panel, circularized fragments are shown with the reference coordinates on the left, with 5' breaks (black triangles) and 3' breaks (red triangles) linked by rearrangements indicated by dashed lines; resulting circular structures are illustrated on the right (not to scale).