Article

Chromothripsis drives the evolution of gene amplification in cancer

https://doi.org/10.1038/s41586-020-03064-z

Received: 16 December 2018

Accepted: 26 November 2020

Published online: 23 December 2020

Check for updates

Ofer Shoshani[1,2], Simon F. Brunner[3], Rona Yaeger[4], Peter Ly[1,2,10,11], Yael Nechemia-Arbely[1,2,12,13], Dong Hyun Kim[1,2], Rongxin Fang[1,2], Guillaume A. Castillon[5], Miao Yu[1,2], Julia S. Z. Li[1,2], Ying Sun[6], Mark H. Ellisman[5,7,8], Bing Ren[1,2], Peter J. Campbell[3,9]✉ & Don W. Cleveland[1,2]✉

Focal chromosomal amplification contributes to the initiation of cancer by mediating overexpression of oncogenes[1–3], and to the development of cancer therapy resistance by increasing the expression of genes whose action diminishes the efficacy of anti-cancer drugs. Here we used whole-genome sequencing of clonal cell isolates that developed chemotherapeutic resistance to show that chromothripsis is a major driver of circular extrachromosomal DNA (ecDNA) amplification (also known as double minutes) through mechanisms that depend on poly(ADP-ribose) polymerases (PARP) and the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs). Longitudinal analyses revealed that a further increase in drug tolerance is achieved by structural evolution of ecDNAs through additional rounds of chromothripsis. In situ Hi-C sequencing showed that ecDNAs preferentially tether near chromosome ends, where they re-integrate when DNA damage is present. Intrachromosomal amplifications that formed initially under low-level drug selection underwent continuing breakage–fusion–bridge cycles, generating amplicons more than 100 megabases in length that became trapped within interphase bridges and then shattered, thereby producing micronuclei whose encapsulated ecDNAs are substrates for chromothripsis. We identified similar genome rearrangement profiles linked to localized gene amplification in human cancers with acquired drug resistance or oncogene amplifications. We propose that chromothripsis is a primary mechanism that accelerates genomic DNA rearrangement and amplification into ecDNA and enables rapid acquisition of tolerance to altered growth conditions.

Gene amplification was first identified[4] in cells that developed resistance to methotrexate, an inhibitor of dihydrofolate reductase (DHFR) that has been used for more than 50 years as a treatment for cancer. Amplified DNA can be found in small, circularized DNAs referred to as double minutes (DMs)[5], a sub-type of extrachromosomal DNA (ecDNA)[2], or in intrachromosomal homogenously staining regions (HSRs)[6]. Previous studies have suggested that DMs can integrate into host chromosomes[7,8] and that HSRs could be a potential source of DMs[2,9,10]. The use of whole-genome sequencing (WGS) has provided high-resolution views of genome rearrangements in human cancers[11–13], including gene amplification and chromothripsis[14–16]—the catastrophic shattering of a chromosome followed by religation of fragments in a random order. Different mechanisms have been proposed to drive the formation of DMs, including chromosome breakage leading to simple circularization of an excised chromosome segment or co-ligation of multiple fragments induced by chromothripsis[14,17–21]. The formation of HSRs can also result from multiple mechanisms: breakage–fusion–bridge (BFB) events (as identified by McClintock in the 1940s)[22], intrachromosomal tandem duplications[23], or neochromosomes, which are found in sarcomas[24]. Here we examine the interplay between intra- and extrachromosomal amplifications, and the evolution of these structures in response to increased selection pressure.

Chromothripsis drives amplification of ecDNA

To identify the mechanisms that underlie the early acquisition of drug resistance and their subsequent dynamics, we applied tunable selection pressure using different concentrations of methotrexate (Fig. 1a). Although the initial karyotype of HeLa cells is abnormal, it was found to be stable (Extended Data Fig. 1a), with chromosome 5 content in all primary clones (and the parental HeLa population) consistent with previous reports[20,25]. We subjected clones to a one-step methotrexate

[1]Ludwig Cancer Research, University of California at San Diego, La Jolla, CA, USA. [2]Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA, USA. [3]Wellcome Sanger Institute, Hinxton, UK. [4]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [5]National Center for Microscopy and Imaging Research (NCMIR), University of California at San Diego, La Jolla, CA, USA. [6]Department of Pediatrics, University of California at San Diego, La Jolla, CA, USA. [7]Department of Neurosciences, University of California at San Diego, La Jolla, CA, USA. [8]Department of Bioengineering, University of California at San Diego, La Jolla, CA, USA. [9]Department of Haematology, University of Cambridge, Cambridge, UK. [10]Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA. [11]Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA. [12]Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. [13]Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, PA, USA. ✉e-mail: pc8@sanger.ac.uk; dcleveland@health.ucsd.edu

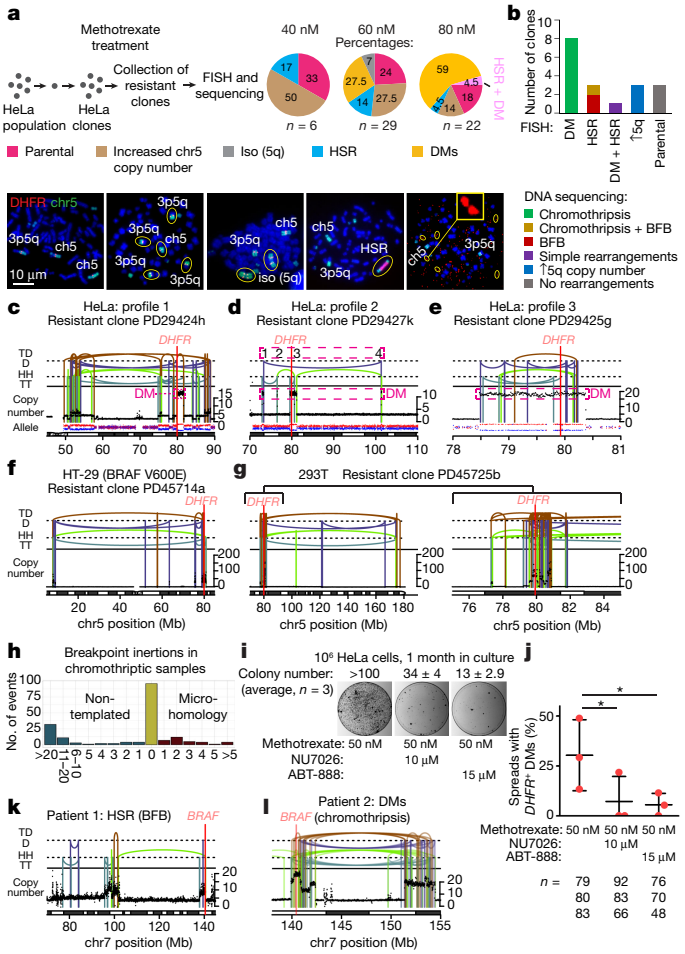


Fig. 1 | Chromothripsis drives ecDNA formation and amplification as DMs early during drug resistance development in a PARP- and NHEJ-dependent manner. a, Strategy used to generate unique and independent methotrexate-resistant HeLa cell clones. Pie charts show the distribution of chromosome 5 abnormalities. Representative DNA-FISH images (from the 57 clones) of each abnormality are shown below. **b**, Summary of the genomic mechanisms (determined by whole-genome paired-end sequencing) that lead to each type of karyotypic outcome (determined by FISH). **c**–**g**, **k**, **l**, Copy number and structural variation profiles of DHFR- and DM-positive methotrexate-resistant clones (**c**–**g**) and vemurafenib-resistant tumours (**k**, **l**). Allelic ratios are shown for **c**–**e**. TD, tandem duplication; D, deletion; HH, head-to-head inversion; TT, tail-to-tail inversion. **g**, Right, expanded view of *DHFR* region. **h**, Sequence analysis of structural variation breakpoints in chromothriptic DM HeLa clones. **i**, Colony assay of methotrexate-treated naive HeLa cells with or without inhibition of DNA repair. Images are representative of three independent experiments (mean ± s.d.). **j**, Quantification of cells with *DHFR*⁺ DMs as determined by FISH under the indicated conditions. Average of three independent experiments. **P* < 0.05 (repeated measures one-way analysis of variance (ANOVA) with post hoc Tukey test, mean ± s.e.m.).

treatment using one of three concentrations (based on testing of effective dose responses; Extended Data Fig. 1b) and collected 57 resistant colonies, of which 28 were shown by DNA fluorescence in situ hybridization (FISH) to have amplified *DHFR* within DMs or HSRs (Fig. 1a). Application of weak, continuous selection pressure typically generated simple, low-level copy number gains, whereas continuous, stronger selection pressure promoted DM formation (Fig. 1a). Both types of amplification occurred early, as in initial resistant colonies (100–200 cells) all cells contained HSRs or DMs (Extended Data Fig. 1c).

We used paired-end WGS to determine the genomic rearrangements of 5 control and 18 resistant clones (Fig. 1b). WGS data perfectly coincided with karyotyping by DNA-FISH, and RNA expression measured with RNA sequencing (Extended Data Fig. 1d). The DNA copy number and RNA level of *DHFR* were highly correlated ($R^2 = 0.85$; Extended Data Fig. 1e). Resistant clones without *DHFR* amplification showed no increase in *DHFR* expression but had a distinct gene expression profile (1,438 differentially expressed genes; Extended Data Fig. 1f).

As expected, control clones had acquired no new genome rearrangements, but in resistant clones we found three different profiles consistent with chromothripsis being the trigger of ecDNA formation and amplification as DMs (Fig. 1c–e). In the first, a chromothriptic rearrangement of one chromosome 5 excised a single fragment containing the *DHFR* gene, which then circularized into a DM (Fig. 1c, Extended Data Fig. 2a–c). In the second profile, multiple (4–17) chromothriptically produced, non-contiguous fragments from chromosome 5 (as far as 80 Mb apart on the initial chromosome and including *DHFR)* were circularized and amplified (Fig. 1d, Extended Data Fig. 2d–k). In the third profile, an entire copy of chromosome 5 was lost except for a DM circularized from a single fragment containing *DHFR* (Fig. 1e, Extended Data Fig. 2l–n), which appeared subsequently to have undergone additional rearrangements[26].

These profiles are representative of rearrangements arising from selection, as two additional cell lines (HT-29 colon cancer cells and 293T human embryonic kidney cells) treated with methotrexate generated similar *DHFR* amplification through chromothripsis (Fig. 1f, g, Extended Data Fig. 2o–p). Breakpoint junction analysis of the eight chromothriptic DM-positive HeLa clones (summarized in Extended Data Fig. 2q) revealed non-templated insertions with minimal microhomology sequences (Fig. 1h). Non-homologous end joining (NHEJ) was needed for efficient DM formation, as inhibition of NHEJ (with DNA-PKcs or PARP inhibitors added at concentrations with minimal effect on viability; Extended Data Fig. 2r) significantly decreased the frequency of both resistant colony formation (Fig. 1i) and DM production (Fig. 1j).

Finally, to test the generality of chromothripsis in drug-mediated gene amplifications, we analysed pre- and post-treatment biopsies from colorectal cancers or their metastases, from two individuals whose cancers were caused by the kinase-activating mutation V600E in BRAF and who developed resistance to therapy with vemurafenib, a selective inhibitor of this BRAF mutation[27]. Vemurafenib resistance in one patient was accompanied by copy number jumps of $BRAF^{V600E}$ and inversions typical of BFB events (Fig. 1k, Extended Data Fig. 3a, b, Supplementary Table 1). The second patient developed amplification of the $BRAF^{V600E}$ gene through chromothripsis, with a *BRAF* gene-containing fragment being ligated together with a second fragment that was initially 10 Mb upstream into a DM containing multiple additional rearrangements, consistent with one or more subsequent rounds of chromothripsis (Fig. 1l, Extended Data Fig. 3c, d, Supplementary Table 1).

Chromothripsis drives ecDNA evolution

Next, we examined the evolution of DMs during adaptation to an increased methotrexate concentration (Extended Data Fig. 4a). Stepwise increases in dose led to increases in DM copy number (Fig. 2a, Extended Data Fig. 4b) and major structural changes that resulted in up to a tenfold increase in *DHFR* gene copies per DM, as determined by the intensity of the FISH signal (Fig. 2b, Extended Data Fig. 4c). WGS revealed that two clones contained DMs with a common rearrangement profile at the basal drug concentration (Extended Data Fig. 4d). An initial 'simple' 2.1-Mb DM in clone PD29429h (Fig. 2c) underwent chromothriptic rearrangement during adaptation to increased methotrexate, forming one or more DMs (see variation in read coverage, Extended Data Fig. 4e) with more complex rearrangements (Fig. 2d) that were undetectable before the increase in methotrexate (Extended Data Fig. 4f), consistent with multiple co-evolving DMs produced by subsequent round(s) of chromothripsis acting on one or more copies of the initial DM.

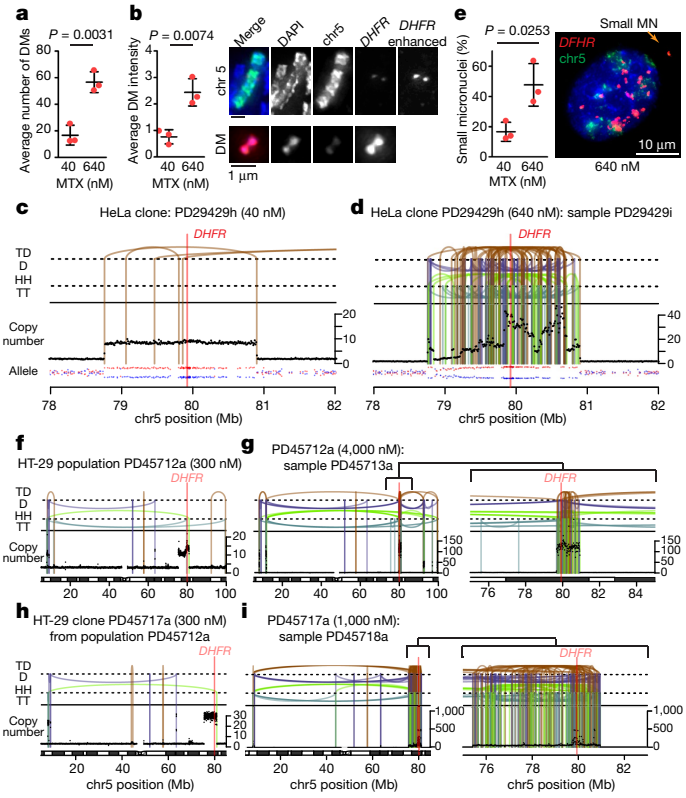


Fig. 2 | DM structure is optimized by chromothripsis during adaptation to increased selection pressure. a, Quantification of DM numbers using DNA-FISH. Mean ± s.d. of three independent clones (Extended Data Fig. 4b); two-tailed *t*-test. MTX, methotrexate. **b**, Left, average intensity of *DHFR* signal in DMs relative to the intensity of the endogenous *DHFR* signal on chromosome 5 from the same spread, as determined by DNA-FISH. Mean ± s.d. of three independent DM clones (Extended Data Fig. 4c); two-tailed *t*-test. Right, DNA-FISH image from a chromosome spread of clone PD29429i (Extended Data Fig. 4c) showing a DM with increased signal intensity. **c**, **d**, **f**–**i**, Copy number and structural variation profiles of indicated samples. **c**, **d**, Allelic ratios for HeLa clone PD29429h and a derivative population following adaptation to 640 nM methotrexate. **f**, **g**, A population of HT-29 cells with *DHFR*$^{+}$ DMs that was resistant to initial 300 nM methotrexate (**f**, PD45712a), and after increasing methotrexate concentration to 4,000 nM (**g**, PD45713a). **h**, Results of a clone (PD45717a) derived from population PD45712a. **i**, As in **h**, after increasing methotrexate concentration to 1,000 nM (sample PD45718a). Note that the massive rearrangements are different from PD45713a (**g**), highlighting the random nature of the rearrangements. **g**, **i**, Right, expanded view of *DHFR* region. **e**, Quantification of *DHFR*$^{+}$ DM-sized (small) micronuclei (MN) using DNA-FISH (representative image presented). Mean ± s.d. of three independent clones; two-tailed *t*-test; 40 nM, $n = 147, 130, 98$; 640 nM, $n = 74, 216, 133$.

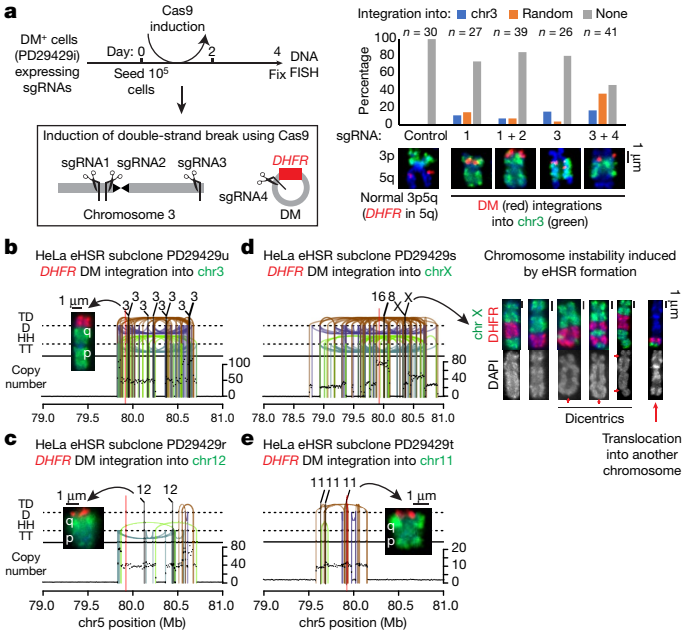


Fig. 3 | DMs preferentially integrate into broken DNA sites near chromosome ends and form ectopic HSRs. a, Quantification of DM integration into ectopic sites on chromosome 3 following induction of double-strand breaks using Cas9. DNA-FISH images of chromosomes with DM integration from each experiment are shown. **b**–**e**, Copy number, structural variation, and translocation profiles of four independent subclones derived from clone PD29429i treated with 1,500 nM methotrexate and 15 μM ABT-888 (PARP inhibitor). DNA-FISH images of chromosomes with DM integration in each clone are shown (integrations observed in all spreads from each clone, $n = 15$ per clone).

Indeed, *DHFR*$^{+}$ micronuclei were present in almost two-thirds (65%) of cells that were resistant to the higher concentration of methotrexate (Fig. 2e, Extended Data Fig. 4g), which indicates ongoing production of substrates for additional chromothriptic events[26,28]. We observed similar structural instability in a second HeLa DM clone (Extended Data Fig. 4d) and in HT-29 cells (Fig. 2f–i) subjected to increased methotrexate concentrations. We performed in silico simulations of DM selection with the assumption that DMs assorted randomly to daughter cells (as determined in Extended Data Fig. 4h) from a cell with an initially defined set of DMs, only one of which carried multiple copies of the resistance gene. These analyses confirmed the advantage conferred by DMs with more than one copy of the resistance gene independent of initial DM number (Extended Data Fig. 4i–l). Thus, escalating selection pressure drives the structural evolution of an initial DM through repeated rounds of chromothripsis, yielding DMs with an increased number of copies of the target gene per DM.

DNA damage drives ecDNA integration

Previous studies have shown that ecDNA elements can integrate into chromosomes[2,7,8]. We initially tested the role of DNA damage in this process by inducing random double-strand breaks in clonal DM-positive cells using ionizing radiation or doxorubicin (Extended Data Fig. 5a, b) or by inducing specific double-strand breaks using nucleases targeting a site 3.5 kb from *DHFR* that was found on chromosome 5 as well as on the DMs (Extended Data Fig. 5c). We identified DM integrations forming ectopic HSRs (eHSRs) in up to 30% of scored spreads following the induction of either random or specific double-strand breaks (Extended Data Fig. 5a–c). To directly test whether DMs integrate into sites of DNA damage, we used Cas9 to induce double-strand breaks at three locations on chromosome 3; up to 17% of cells had DM integrations at these specific ectopic DNA sites (Fig. 3a).

In most (70%) cases following random DNA damage, DM integration occurred at an ectopic site near a chromosome end (Extended Data Fig. 5a–c). Treatment with a PARP inhibitor, but not with an inhibitor of DNA-PKcs, resulted in a threefold increase in the frequency of such eHSRs, most of which (78%) inserted within 10 Mb of a chromosome end (Extended Data Fig. 5d). PARP inhibition also increased the frequency of DM conversion into eHSRs in Colo320DM-GFP cells[29] (Extended Data Fig. 5e), and of enlarged DMs (Extended Data Fig. 5f, g). Consistent with previous suggestions that integrations occur near telomeres[7,8], unbiased, genome-wide in-situ Hi-C sequencing showed that there was significant linkage of DMs with chromosome ends in both mitotically arrested cells and cycling cells (Extended Data Fig. 5h–j).

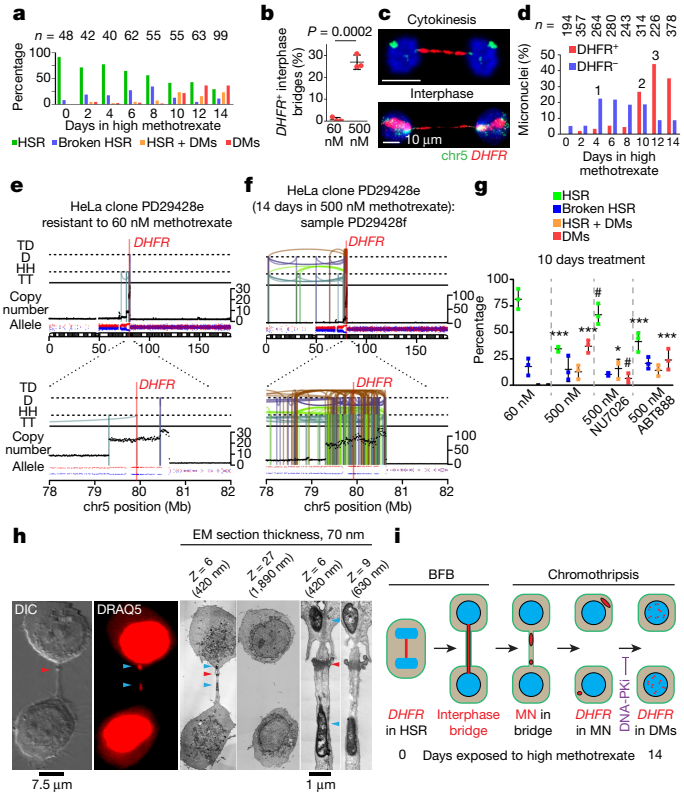


Fig. 4 | HSRs fragment in interphase bridges and form micronuclei and DMs through chromothripsis and NHEJ repair. a, Location of *DHFR* amplification, determined using DNA-FISH in clone PD29428e over the course of two weeks when exposed to increased methotrexate concentration (500 nM). **b**, Quantification of *DHFR*$^+$ interphase bridges in clone PD29428e using DNA-FISH. Mean ± s.d. of three independent experiments (number of cells scored: 60 nM, 144, 221, 204; 12 days in 500 nM, 163, 218, 214); two-tailed *t*-test. **c**, Representative DNA-FISH images of *DHFR*$^+$ interphase bridges (from data shown in **b**). **d**, Quantification of *DHFR*$^+$ micronuclei over the course of two weeks in clone PD29428e exposed to increased methotrexate concentration as determined by DNA-FISH (representative image from the data is shown in Extended Data Fig. 8k). Day 4 (time point 1) shows increased *DHFR*$^-$ micronuclei (owing to the random DNA damage from increased methotrexate). Day 12 (time point 2) shows increased *DHFR*$^+$ micronuclei (products of HSR fragmentation), followed by a reduction in *DHFR*$^-$ micronuclei two days later (time point 3). **e**, **f**, Copy number and structural variation profiles of indicated samples. **g**, Quantification of amplification patterns in clone PD29428e exposed to increased methotrexate concentration for 10 days (from 60 nM to 500 nM), with or without DNA repair inhibitors. Bars represent mean ± s.d. of three independent experiments (spreads counted: 60 nM, 42, 32, 33; 500 nM, 36, 50, 38; 500 nM + 10 μM NU7026, 37, 35, 45; 500 nM + 15 μM ABT-888, 35, 34, 32). $^*P < 0.05$, $^{***}P < 0.001$ (compared to 60-nM condition); $^{\#}P < 0.001$ (compared to 500-nM condition); two-way ANOVA. **h**, Correlative light and electron microscopy (CLEM) showing an interphase bridge (of 17 bridges imaged) with micronuclei on both sides (blue arrowheads) of the midbody (red arrowhead). DIC, differential interference contrast; DRAQ5, deep red anthraquinone 5, DNA dye. **i**, Summary of the events underlying BFB chromothripsis. DNA-PKi, inhibitor of DNA-PKcs.

To investigate the structures of eHSRs, we isolated 25 subclones that were resistant to increased methotrexate in the presence of a PARP inhibitor from an initial DM-containing clone. Four (16%) contained clonal events of DM integration that produced an eHSR in all cells, three of which were located near a chromosome end (Fig. 3b–e). The eHSRs of all four clones showed distinct rearrangement profiles derived from the parental DM population (see also breakpoint PCR, Extended Data Fig. 4e). One clone carried a large (40-Mb) eHSR that initially integrated 60 Mb from a telomere of chromosome X (Fig. 3d), producing a region susceptible to breakage and the formation of dicentric chromosomes that drive BFB cycles (Fig. 3d). Finally, two eHSRs shared the same amplified region (Fig. 3b, d) with one presenting massive rearrangements with multiple proximal (kilobase-scale) integration sites within 1.1 Mb of the tip of chromosome 3, which suggests either multiple DM integrations or a single integration of a massively rearranged DM containing multiple *DHFR* copies (Fig. 3b).

BFB with subsequent chromothripsis

We examined the evolution of HSRs in response to a stepwise elevation in selection pressure. We identified a small HSR, initially generated by two cycles of BFB following methotrexate treatment, that contained four copies of *DHFR* (Extended Data Fig. 6a–h). Adaptation to elevated drug concentration resulted in the acquisition of DMs coupled with loss of the HSR (Extended Data Fig. 6i). WGS of one clone revealed a complex, 1.4-Mb DM with 14 non-adjacent regions of the HSR derived by chromothriptic shattering and re-stitching in random order, and with variable copy numbers of individual pieces (Extended Data Fig. 6j). Other resistant clones showed evidence of BFB followed by chromothripsis. Clone PD29427p presented an HSR that underwent complex rearrangements, consistent with localized chromothripsis of chromatin trapped in an interphase DNA bridge (Extended Data Fig. 7a–c). Two DM-containing clones (Fig. 1c, Extended Data Fig. 2b) that maintained the chromothriptic chromosome of origin showed features of BFB-chromothripsis (including complete loss of the 5q region telomeric to *DHFR* and an increased copy number of some fragments). Kataegis, a hypermutation signature that is associated with chromothripsis of dicentric chromosomes[30], was present in nearly half (three of eight) of DM clones, consistent with evolution from an initial HSR (Extended Data Fig. 7d, e).

We followed the evolution of an HSR into DMs during adaptation to a higher dose of methotrexate (Fig. 4a, Extended Data Fig. 8a–c). Over two weeks, HSR size tripled (Extended Data Fig. 8d, e), with a sharp (55%) increase by day 10 in the frequency of dicentric HSRs (Extended Data Fig. 8f, g). Live-cell imaging revealed that at day 10, 40% of mitoses had anaphase bridges, which frequently persisted into interphase (Extended Data Fig. 8h, Supplementary Video 1). In the next division, either daughter cells formed another bridge (40%), consistent with ongoing BFB (Extended Data Fig. 8i, Supplementary Video 1), or the dicentric state was resolved, possibly through chromothripsis or telomere capture (60%). Notably, by day 12 almost 30% of cells had *DHFR*$^+$ HSRs trapped within interphase bridges (Fig. 4b, c). Rupture of interphase bridges observed by live-cell imaging resulted in the formation of *DHFR* gene-containing nuclear bulges (Extended Data Fig. 8j) and production of *DHFR*$^+$ micronuclei (Fig. 4d, Extended Data Fig. 8k).

WGS of one HSR-containing clone and a derivative DM-positive population resulting from a two-week exposure to increased methotrexate revealed that during selection the initial HSR (produced through five BFB cycles) was shattered (along with adjacent regions of chromosome 5) to produce highly rearranged DMs (Fig. 4e, f). We also found BFB with chromothripsis in methotrexate-resistant 293T cells, which initially contained a *DHFR*$^+$ HSR that converted into DMs with increased drug concentration (Extended Data Fig. 8l, m). In HeLa cells that were resistant to the mitotic kinesin inhibitor *S*-trityl-L-cysteine (STLC), chromothripsis promoted the resolution of BFB-induced amplification of *KIF15* by capturing a telomere from chromosome 1 (Extended Data Fig. 9). Finally, the formation of DMs through BFB-chromothripsis was highly dependent on NHEJ, as DM formation was inhibited by the addition of a DNA-PKcs inhibitor (Fig. 4g).

Examination of interphase DNA bridges in HSR-containing cells adapting to increased methotrexate or in cells adapting to the DNA-damaging agent zeocin identified many instances of micronuclei forming within the bridge (Fig. 4h, Extended Data Fig. 10a–c, Supplementary Video 2), consistent with a previous report that micronuclei

can form following anaphase bridge resolution[31]. Ingression of the cleavage furrow disrupted the nuclear envelope structure of a nucleus or micronucleus caught between nascent daughter cells (Extended Data Fig. 10d). DNA in bridges juxtaposed to the midbody had aberrant nuclear envelopes and appeared compartmentalized (Extended Data Fig. 10e). DNA and midbody signals were mutually exclusive, consistent with the midbody being a site of DNA cleavage (Extended Data Fig. 10f). Micronuclei that formed after bridge resolution had highly abnormal nuclear envelopes (Extended Data Fig. 10g), serving as a chromothriptic source[26] for DM formation (Fig. 4i).

Chromothriptic oncogene amplifications

We next examined whether the chromothriptic origin of gene amplification seen in response to DHFR, *BRAF*^{V600E}, and kinesin inhibitors was representative of additional human cancers. Comparison of the genomic profiles of HeLa cells with *DHFR* amplification and glioblastoma multiforme (GBM), a cancer with a high frequency of DMs[2], revealed that both shared a similar random distribution of rearrangements (Extended Data Fig. 11a, b). Furthermore, when we compared copy number profiles of the amplified *DHFR* locus to a panel of 15 different cancers containing oncogene amplifications (Extended Data Fig. 11c, d) we found similar localized amplifications in randomly rearranged chromosomes (Extended Data Fig. 11d, e). Finally, breakpoint junction analysis revealed that non-homologous end-joining was the dominant form of repair (Extended Data Fig. 11f) leading to amplifications.

Chromothripsis and cancer therapy

We have identified mechanisms accessed by cancer cells in which chromothripsis drives the evolution of amplified DNA (Extended Data Fig. 11g). Our findings bridge two biological phenomena discovered more than 70 years apart: cycles of break–fusion–bridge[22] (with subsequent HSR formation) and chromothripsis[14] of either micronucleated DNAs[26,28] or interphase chromosome bridges[30] (Fig. 4). Our work identifies NHEJ and PARP-dependent repair after chromothriptic shattering both as a mechanism of initial oncogene amplification or anti-cancer drug resistance and as a driver of subsequent rearrangements. Given that chemotherapeutic drugs (for example, methotrexate or vemurafenib) can lead to chromothriptically initiated rearrangements, the combination of a targeted therapy with DNA repair inhibitors[32] may represent an effective approach to prevent an initial cancer from becoming more aggressive or drug resistant.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-03064-z.

1. Benner, S. E., Wahl, G. M. & Von Hoff, D. D. Double minute chromosomes and homogeneously staining regions in tumors taken directly from patients versus in human tumor cell lines. *Anticancer Drugs* **2**, 11–25 (1991).
2. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
3. Albertson, D. G. Gene amplification in cancer. *Trends Genet*. **22**, 447–455 (2006).
4. Alt, F. W., Kellems, R. E., Bertino, J. R. & Schimke, R. T. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. *J. Biol. Chem*. **253**, 1357–1370 (1978).
5. Kaufman, R. J., Brown, P. C. & Schimke, R. T. Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes. *Proc. Natl Acad. Sci. USA* **76**, 5669–5673 (1979).
6. Nunberg, J. H., Kaufman, R. J., Schimke, R. T., Urlaub, G. & Chasin, L. A. Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. *Proc. Natl Acad. Sci. USA* **75**, 5553–5556 (1978).
7. Carroll, S. M. et al. Double minute chromosomes can be produced from precursors derived from a chromosomal deletion. *Mol. Cell. Biol*. **8**, 1525–1533 (1988).
8. Ruiz, J. C. & Wahl, G. M. Chromosomal destabilization during gene amplification. *Mol. Cell. Biol*. **10**, 3056–3066 (1990).
9. Coquelle, A., Rozier, L., Dutrillaux, B. & Debatisse, M. Induction of multiple double-strand breaks within an hsr by meganucleasel-SceI expression or fragile site activation leads to formation of double minutes and other chromosomal rearrangements. *Oncogene* **21**, 7671–7679 (2002).
10. Nathanson, D. A. et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).
11. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
12. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
13. Cortes-Ciriano, I. et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet*. **52**, 331–341 (2020).
14. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
15. deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet*. **50**, 708–717 (2018).
16. Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat. Rev. Cancer* **19**, 283–288 (2019).
17. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
18. Nones, K. et al. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat. Commun*. **5**, 5224 (2014).
19. Ly, P. et al. Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat. Genet*. **51**, 705–715 (2019).
20. Singer, M. J., Mesner, L. D., Friedman, C. L., Trask, B. J. & Hamlin, J. L. Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks. *Proc. Natl Acad. Sci. USA* **97**, 7921–7926 (2000).
21. Windle, B., Draper, B. W., Yin, Y. X., O'Gorman, S. & Wahl, G. M. A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration. *Genes Dev*. **5**, 160–174 (1991).
22. McClintock, B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).
23. Glodzik, D. et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat. Genet*. **49**, 341–348 (2017).
24. Garsed, D. W. et al. The architecture and evolution of cancer neochromosomes. *Cancer Cell* **26**, 653–667 (2014).
25. Landry, J. J. et al. The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda)* **3**, 1213–1224 (2013).
26. Zhang, C. Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
27. Yaeger, R. et al. Mechanisms of acquired resistance to BRAF V600E inhibition in colon cancers converge on RAF dimerization and are sensitive to its inhibition. *Cancer Res*. **77**, 6513–6523 (2017).
28. Ly, P. et al. Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat. Cell Biol*. **19**, 68–75 (2017).
29. Shimizu, N., Hashizume, T., Shingaki, K. & Kawamoto, J. K. Amplification of plasmids containing a mammalian replication initiation region is mediated by controllable conflict between replication and transcription. *Cancer Res*. **63**, 5281–5290 (2003).
30. Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163**, 1641–1654 (2015).
31. Hoffelder, D. R. et al. Resolution of anaphase bridges in cancer cells. *Chromosoma* **112**, 389–397 (2004).
32. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **8**, 193–204 (2008).

Methods

Cell culture

HeLa S3 (authenticated using WGS and FISH) and 293T (authenticated by frequent use for transfection and lentiviral particle preparation, and by morphology and growth in culture) cells, both from Cleveland laboratory cell stock, and DLD-1 cells (fresh from ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM, Gibco) containing 10% fetal bovine serum (FBS, Omega), supplemented with 100 U/ml penicillin, 100 U/ml streptomycin, and 2 mM L-glutamine, at 37 °C under 5% CO_2. When treating with methotrexate, dialysed FBS was used instead (Gibco, 26400044). HT-29 cells were obtained fresh from ATCC and grown in McCoy's 5A (Gibco) containing 10% FBS (Omega), supplemented with 100 U/ml penicillin, 100 U/ml streptomycin, and 2 mM L-glutamine, at 37 °C under 5% CO_2. T-47D cells were obtained fresh from ATCC and grown in RPMI-1640 with supplements as above. Colo320DM-GFP cells (authenticated by having GFP-labelled DMs, and by morphology in culture), a kind gift from Noriaki Shimizu (Hiroshima University, Japan), were grown in RPMI (Gibco) containing 10% FBS (Omega), supplemented with 100 U/ml penicillin, 100 U/ml streptomycin, and 2 mM L-glutamine, at 37 °C under 5% CO_2. Cells were periodically tested for mycoplasma and confirmed free of contamination. Cell lines used in this study are not found on the list of commonly misidentified cell lines maintained by the International Cell Line Authentication Committee.

Isolation of methotrexate-resistant clones

HeLa S3 cells were subjected to two rounds of single-cell cloning. In the first round, five primary clones were isolated from a 96-well plate in which 0.5 cells per well were seeded. Subsequently, each of the five primary clones was seeded at 0.5 cells per well in 96-well plates to isolate six secondary clones from each of the primary clones (a total of 30 secondary clones were collected). All clones collected were verified microscopically to be derived from a single cell. HeLa clones were then subjected to treatment with 40 nM, 60 nM, or 80 nM methotrexate for up to one month, with medium replaced twice per week. Individual surviving colonies were picked and transferred to a 24-well plate and later to a 6-well plate, from which cells were collected for FISH, RNA and DNA sequencing, and storage.

Colony-forming assay

Cells were seeded at low dilution (number of cells seeded and time of culture are indicated in each experiment) and colonies formed were methanol fixed and stained with a 0.5% crystal violet, 20% ethanol solution. Plates were photographed using Gel Doc XR+ (Bio-Rad) and colonies were manually quantified.

Reagents

DNA damage repair inhibitors were dissolved in DMSO and used at the following concentrations: 15 μM ABT-888 (PARP inhibitor, Enzo), 10 μM olaparib (PARP inhibitor, ApexBio), and 10 μM NU7026 (DNA-PKcs inhibitor, Abcam). Methotrexate (454126, Calbiochem) was dissolved in DMSO and used at the indicated concentrations. Doxorubicin (Sigma) was dissolved in water and cells were treated at a final concentration of 0.5 μM for 1 h. Hydroxyurea (Sigma) was dissolved in PBS and used at a final concentration of 100 μM. *S*-Trityl-L-cysteine (STLC, Sigma) was used at a final concentration of 10 μM, and zeocin (Gibco) was used at a final concentration of 100 μg/ml.

Site-specific nucleases

For Fig. 3a (dox-inducible lentiviral expression of SpCas9 construct), pCW-Cas9 was obtained from Addgene (50661). Lentiviral sgRNA-expressing constructs were made by inserting guide sequences into two BbsI sites of pLH-sgRNA1 (Addgene, 75388). To generate lentiviruses, 293T cells were transfected with the sgRNA construct, psPAX2 and pMD2.G plasmids. Guide sequences were designed using the Broad Institute portal (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). gRNA#1: GAAGCCACCATGTCTCACCA; gRNA #2: CTGAAAGTCCTCGGTCTGAC; gRNA #3: TGATATCACAG; gRNA #4: GCTGAAGAGATTTACAGTCC.

For Extended Data Fig. 5c, transcription activator-like effector nuclease (TALEN) targeting a location approximately 3,500 bp centromeric to the *DHFR* locus (80622611–80622657 Mb on chromosome 5) and containing the sequence: t**ctgctgaagagattt**acagtcctggtcccc**attttctctt attct**a (TALEN recognition sites in bold) was assembled using the golden gate reaction as previously described[33]. CRISPR–Cas9 targeting the site 5′ GCTGAAGAGATTTACAGTCC 3′, near the same location as TALENs (80622635–80622654 Mb) was constructed by inserting the sgRNA sequence into a wild-type Cas9 vector (Addgene 42230) as described previously[34]. HeLa cells were transfected with either the TALEN or CRISPR–Cas9 expression vector by nucleofection (Lonza) using solution R and program I-013, and efficient cutting was validated using specific primers and the surveyor assay kit (Transgenomics) as previously described[34].

DNA-FISH

Chromosome spreads were obtained by dropping mitotic cells onto glass slides. In brief, cells were arrested for 3 h with 100 ng/ml colcemid (KaryoMAX, Thermo Fisher), and then incubated in 75mM KCl for 15 min at 37 °C. Cells were fixed by adding methanol/acetic acid (3:1), washed with fixative three times, and kept in fixative at −20 °C until use. For interphase FISH, cells growing on poly-L-lysine-coated coverslips (Corning) were washed with PBS and fixed using methanol/acetic acid (3:1) and kept in fixative at −20 °C until use. To make custom BAC probes, the following BACs were ordered from bacpac (https://bacpacresources.org/): RP11-90A9 (*DHFR*, used throughout the study), RP11-538B23 (*DHFR*, used to validate first BAC), RP11-440N18 (*MYC*), for chromosome 5 rearrangements analysis: RP11-958F12, RP11-314L7, RP11-33C19, and for *KIF15* amplification analysis: RP11-659N22 and RP11-372G3. BACs were isolated from 50-ml bacterial cultures using the BACMAX Bac extraction kit (Epicentre). Isolated BACs were sonicated (×10 cycles of 15 s 'on'/45 s 'off' at constant intensity with power set to 3, Branson sonifier 450) and labelled with either TM-rhodamine or cy5 using Label-IT (Mirus). Labelled BAC probes were suspended in commercial chromosome paint probes (Metasystems). DNA-FISH was performed by applying probes onto samples and covering with a glass coverslip. Genomic DNA and probes were co-denatured at 75 °C for 2 min by placing slides on a pre-heated metal plate. Samples were hybridized overnight at 37 °C in a dark humidified chamber. Slides were subsequently washed with 0.4 × SSC at 72 °C for 2 min and rinsed in 2 × SSC, 0.05% Tween-20 at room temperature for 30 s. Slides were then rinsed in PBS, counterstained with DAPI, and mounted using pro-long gold (Invitrogen). FISH images were acquired on a DeltaVision elite system (Applied Precision) at either 60× or 100× magnification (10 × 0.5 μm *z*-sections). Deconvolution and maximum intensity projections were generated using the softWoRx program.

Measurements of gene amplification

DMs and micronuclei were manually counted. DM intensities relative to the endogenous *DHFR* on chromosome 5 were determined using the ROI feature in Fiji (ImageJ)[35] by averaging the intensities of the brightest four single DM dots in each spread and subtraction of similarly sized adjacent background regions, divided by the average intensity of the endogenous *DHFR* signals within the same spread (with background subtracted). HSR lengths were measured (Fiji) and normalized to the normal chromosome 5 length within the same spread for comparison.

Live-cell imaging

To study the events that occur in HSR-positive cells after increasing methotrexate concentration, one million HSR-positive clone 28e cells (resistant to 60 nM methotrexate) were treated with 500 nM

methotrexate for 10 days in a 10-cm culture dish from which 5,000 cells per well were seeded into a polystyrene glass 96-well plate. One day later, cells were stained with DNA SiR (1:2,000 from a stock of 50 nmol in 50 µl DMSO, Cytoskeleton, CY-SC007) and verapamil (1:2,000 from a stock of 1 µmol in 100 µl DMSO, Cytoskeleton, CY-SCV01), and 2 h later were imaged using a CQ1 spinning disk confocal systems (Yokogawa Electric Corporation) with a 40× magnification at 37 °C under 5% CO_2. Image acquisition and data analysis were performed using CQ1 software and ImageJ, respectively. Images of 20 fields per well at 8 × 3-µm *z*-sections in cy5 were acquired at 10-min intervals for 48 h. The mitotic fates of parental and daughter cells (two cell cycles) were analysed and scored for formation of micronuclei and/or DNA bridges. Proliferation rate was automatically determined using CQ1 software by automated counting of the number of cells per field at each time point, and data were plotted using Prism GraphPad.

Immunofluorescence

Cells were plated onto poly-L-lysine-coated coverslips (Corning) and fixed in 4% formaldehyde for 10 min at room temperature. Cells were incubated in blocking solution (0.2 M glycine, 2.5% FBS, 0.1% Triton X-100, PBS). The following primary antibodies were used at 1:200 dilution in blocking solution and washed with 0.05% Tween in PBS: Aurora B (Abcam, ab2254), Lamin B (Proteintech, 12987-1-AP). Actin was labelled using Alexa Fluor 647 phalloidin (Thermo Fisher Scientific). Images were acquired on a DeltaVision elite system at 40–60× magnification (8 × 0.5 µm *z*-sections) and deconvolved maximum intensity projections were generated using softWoRx program.

Cell cycle analysis

Mitotic synchronization was achieved using 100ng/ml nocodazole (Sigma) treatment for 6 h, and mitotic cells were collected by plate shake-off. Ethanol-fixed cells were stained with 10 µg/ml propidium iodide and 50 µg/ml RNase A and analysed for DNA content by flow cytometry on a BD LSR II instrument (BD Biosciences) using FACS Diva version 8.0.1.

Correlative light and electron microscopy

For DNA-labelled electron microscopy (ChromEM) used in Fig. 4h and in Extended Data Fig. 10d, e, g, the samples were treated as previously described[36]. Cells were briefly washed in PBS without calcium and magnesium, then fixed with 2.5% EM grade glutaraldehyde (Electron Microscopy Sciences) in 5 mM $CaCl_2$, 0.1 M sodium cacodylate buffer, pH 7.4, at room temperature for 2 min and for an additional hour on ice. From this step on, the cells were always treated either on ice or on a cold stage set at 4 °C. All solutions were at 4 °C before they were applied to cells. Fixed cells were washed three times for two minutes in 0.1 M sodium cacodylate buffer. Cells were blocked for 15 min using blocking buffer (10 mM glycine, 10 mM potassium cyanide in 0.1 M sodium cacodylate buffer). DNA was stained with DRAQ5 (5 µM) in 0.1 M cacodylate buffer for 30 min followed by 3 × 5-min washes with blocking buffer. After excess dye was washed away, cells were bathed in 2.5 mM diaminobenzidine tetrahydrochloride (Sigma) in 0.1 M sodium cacodylate buffer. Cells were placed on a Leica SPE-II confocal microscope with a custom cold stage set at 4 °C and imaged using a 63× oil immersion objective lens (NA = 1.3) and Cy5 filter set (620 ± 30 nm). After 3D imaging in the DIC channel and DRAQ5 fluorescence, cells were photo-oxidized by continuous epi-fluorescence illumination (150 W Xenon Lamp) for 2 min. DIC was taken one more time to verify the reaction of photo-oxidation. Cells were rinsed 5 × 2 min with 0.1 M sodium cacodylate buffer. Cells were stained for 30 min with 2% osmium tetroxide, 2 mM $CaCl_2$, 1.5% potassium ferrocyanide in 0.15 M sodium cacodylate buffer. Cells were washed with double distilled water 5 × 2 min. Cells were additionally incubated for 2 h in 2% uranyl acetate in double distilled water and washed with double distilled water 5 × 2 min before dehydration.

EM sample preparation and TEM

Cells were dehydrated in ethanol and embedded in Durcupan epoxy resin as described previously[37]. Epoxy-embedded cells were cut into 70-nm sections and mounted on 200 hexagonal mesh copper grids. Sections were post-stained for 3 min in 2% uranyl acetate in double distilled water followed by four washes in double distilled water, then incubated for 2 min in Sato Lead, and finally washed four times in double distilled water. Sections were imaged at 120 keV with a FEI/Thermofisher Tecnai G2 Spirit. Images and montages were acquired with serial EM and stitched with Etomo.

Breakpoint PCR

Primers specific for breakpoints (junctions of non-contiguous fragments) found were designed using WGS data. Primers for breakpoints found in clone 1 640 nM and eHSR clone 29s: 5′ ACCACAATTCAG CTAACTCCACT 3′ and 5′ TTTGAAGAGTCCTTGCCATTGTC 3′. Primers for breakpoints found only in clone 1 640 nM: 5′ CCTCGTGTT ATTCTTTGGCCTG 3′ and 5′ TTGAAAGCGAACGTCACTGAG 3′. Primers for breakpoints found in clone 1160 nM and 640 nM and eHSR clone 29s: 5′ CCCCACATAACCCTGAGCTG 3′ and 5′ TGGTAGGAGCAACCTGCTTT 3′. Full scans of gels are provided in Supplementary Information.

In situ Hi-C sequencing

In-situ Hi-C experiments were performed as previously described[38]. Cells were fixed with 1% of formaldehyde at room temperature for 10 min. After inactivation with glycine, crosslinked cell pellets were washed with cold PBS once before snap freezing with liquid nitrogen. One to three million crosslinked cells were used per Hi-C experiment as follows: in brief, the crosslinked cell pellets were thawed on ice, lysed on ice for 15 min (lysis buffer: 10 mM Tris, pH 8.0, 10 mM NaCl, 0.2% IGEPAL CA-630 with proteinase inhibitor) followed by a wash with lysis buffer. The pellets were then resuspended in 50 µl 0.5% of SDS and incubated at 62 °C for 10 min. Permeabilization was quenched by addition of 25 µl 10% Triton X-100 and 145 µl water, and incubation at 37 °C for 15 min. Digestion was performed overnight at 37 °C, shaking at 1,000 rpm with 1× NEBuffer 2 and 100 units of MboI, followed by inactivation of MboI at 62 °C for 20 min. Biotin fill-in reaction was performed for 1.5 h at 37 °C with 15 nmol each of dCTP, dGTP, dTTP and biotin-14-dATP and 40 U Klenow. In situ ligation was then performed at room temperature with slow rotation in a total volume of 1.2 ml containing 1× T4 ligase buffer, 0.1 mg/ml BSA, 1% Triton X-100 and 4,000 U T4 ligase. After ligation, the DNA was reverse crosslinked at 55 °C for 30 min followed by 68 °C overnight in the presence of SDS, NaCl and proteinase K. Afterwards, DNA was extracted by ethanol precipitation and then sheared to an average of 400 bp with Covaris M220. Ampure beads were used to further purify the sheared DNA and select fragments between 300 and 700 bp. Biotin pull-down enrichment and library construction were performed as previously described[38] and sequencing used HiSeq 4000 (Illumina). Hi-C data processing was performed as follows: paired-end HiC sequencing reads were mapped using BWA-MEM to the reference genome (hg19) in single-end mode with default parameter setting for each of the two ends separately. The independently mapped ends were then paired up and the read pairs were kept if both ends uniquely mapped to the genome (MQAL > 10). Next, read pairs were further removed if either end was mapped more than 500 bp apart away from the closest MboI cutter site. Read pairs were next sorted based on genomic coordinates followed by PCR duplicate removal using samtools rmdup. Processed Hi-C data were converted to .hic files and visualized using juicebox without normalization. The following analysis was performed to quantify the interaction frequency between DM sequences and the rest of the genome. (1) Each Hi-C read pair contains two reads that are mapped to two different loci on the genome, indicating one captured interaction between these two loci. We first extracted all Hi-C read pairs (called DM pairs) that had one read mapped within the DM region and the other end

aligned outside the DM region (excluding chr5). (2) We then segmented the genome into a set of bins of equal length (1 Mb). (3) Next, using the DM pairs obtained from step 1, for each 1-Mb bin obtained from step 2, we counted the number of reads falling into each bin. (4) Finally, for every sample, the interacting frequency was normalized to RPM (reads per million reads) by the total number of DM pairs. (5) To compare two samples, we normalized the interacting frequency of one sample with the interacting frequency of another sample (control sample).

RNA sequencing

Global gene expression profiling was done on total RNA extracted using the Nucleospin RNA kit (Macherey-Nagel) and processed using the Illumina TruSeq Stranded mRNA Sample Preparation Kit. Then, cDNA libraries were sequenced on an Illumina HiSeq 4000 using single read, 50-cycle runs. The quality of sequencing reads was assessed using FastQC (Babraham Bioinformatics) and aligned to a reference genome (hg19, UCSC Genome Browser) using TopHat. Sequencing yielded on average 46.3 million reads per sample with a ~71% mapping rate to generate ~33 million reads per sample uniquely mapped to the reference genome. Cufflinks was used to generate transcript abundance as fragments per kilobase of transcript per million mapped reads (FPKM), and statistical analysis of FPKM values was calculated using R (Bioconductor). RNA expression (relative to control) was plotted on a linear map of chromosome 5. Genes which expressed [FPKM ≥ 1] were kept for downstream principal component analysis (PCA). Semi-unsupervised (gene unsupervised and sample supervised) hierarchical clustering with complete method was applied to the heatmap.

Colorectal cancer analysis

Patient data have been previously published[27]. Collection and analysis of patient samples was conducted under appropriate Institutional Review Board/Privacy Board protocols and waivers (MSKCC protocols 19-323, 06-107). Participating patients provided signed written informed consent. The study was conducted in accordance with the ethical guidelines in the Declaration of Helsinki. Patient DNA was sequenced at the Sanger Institute using whole-genome paired-end sequencing at a coverage of 60–130×. Pre-treatment biopsies showed no evidence of *BRAF* amplification. Progression samples were assigned a sample number with no patient identifying information before submission for DNA extraction and genomic analysis. No one performing genomic analysis had access to the patients' identifying information.

X10 sequencing

DNA from the pre- and post-methotrexate exposure samples was extracted using the DNeasy kit (Qiagen, 69504) according to the manufacturer's guidelines and was sequenced on the Illumina X10 platform to a target coverage of 30× whole human genome. Sequencing libraries were synthesized from genomic DNA according to the manufacturer's protocols. Cluster generation and sequencing were performed using the manufacturer's pipelines. Average sequence coverage achieved across the samples was 33.3× (range, 27.4–35.9×).

Mapping and analysis

Mapping to the human genome was performed using the BWA algorithm[39], using the BWA mem version 0.7.8. The exact genome build used was hs37d5.

Variant calling

Copy numbers were called using ascatNgs[40]. The variant-calling pipeline of the Cancer Genome Project, Wellcome Trust Sanger Institute was used to call somatic mutations[41]. We used the following algorithms with standard settings and no additional post-processing: CaVEMan for substitutions and the BRASS algorithm for rearrangements. For all algorithms, the parental, un-exposed samples were used as the reference samples.

Chromothripsis calling

Published chromothripsis criteria[42] were used to assess regions with complex clusters of chromothripsis. 1. Clustering of breakpoints: applicable. Five out of nine regions with high density of rearrangements show strong clustering of rearrangement breakpoints. 2. Regularity of oscillating copy-number stages: applicable. As opposed to 'conventional chromothripsis', copy numbers oscillate over multiple copy number states because chromothripsis takes place on chromosomes amplified through BFBs. 3. Interspersed loss and retention of heterozygosity: applicable. Interspersed loss and retention of heterozygosity were present in multiple chromosomes with chromothripsis. 4. Prevalence of rearrangements affecting a specific haplotype: applicable. Karyotyping analysis confirmed the prevalence of rearrangements on a specific haplotype. 5. Randomness of DNA segment order and fragment joins: applicable. The orientations of rearrangement joins in chromothripsis regions are consistent with random draws from a uniform multinomial distribution. However, the chromothripsis events involve fairly low numbers of intrachromosomal rearrangements, which would decrease power in finding statistical departures from a uniform multinomial distribution. 6. Ability to walk the derivative chromosome: not applicable, as chromothripsis takes place on chromosomes with preceding duplication through BFBs. In this scenario, it would be impossible to walk the derivative chromosome[43].

Mutation calling and kataegis

Kataegis mutation clusters were detected where we observed six or more consecutive mutations each with an inter-mutation distance of <1 kb. This method of detection corresponds to a previously used approach[44].

Haplotype phasing of SNPs

Inspection of the variant allele frequencies (VAFs) of germline single-nucleotide polymorphisms (SNPs) in the parental, unexposed samples revealed peaks at 1/3, 2/3 and 1, corresponding to heterozygous SNPs present on one or two copies of the triploid chromosome, as well as homozygous SNPs. Thus, we assigned SNPs to alleles based on their VAFs. We defined two groups: those with VAFs between 0.23 and 0.43 and those with VAFs between 0.56 and 0.76. In methotrexate-exposed samples, we used the allelic assignments of the germline SNPs to phase copy number segments to each parental allele.

Simulation of DM evolution

We wondered whether the dominance of multi-copy DMs is plausible when each DM is given an equal chance to randomly segregate to a daughter cell, or if an alternative mechanism, such as the frequent fusion of DMs, is responsible. To address this question, we performed in silico simulations of DM selection. Essentially, we allowed each cell in a population to divide depending on the number of resistance genes that it contained. A pseudocode describing the simulation procedure can be found in Supplementary Fig. 3. The simulation was implemented in Python 2.7. In each simulation cycle and for each cell, we recorded the maximum number of resistance genes per DM. We then grouped the cells depending on their maximum DM gene copy number and evaluated the size of each group. We could then observe the growth of each cell group over the simulation cycles, which we used as a proxy for time.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Paired-end WGS data are available at the European Nucleotide Archive (ENA), accession number ERP107458. In-situ Hi-C sequencing data are

available at the Gene Expression Omnibus (GEO), accession number GSE119825. RNA sequencing data are available at GEO, accession number GSE119979. TCGA database can be accessed at https://portal.gdc.cancer.gov/. Source data are provided with this paper.

Code availability

A pseudocode of the simulation performed in Extended Data Fig. 4i–l is provided in Supplementary Fig. 3. Code can be found at https://github.com/sfbrunner/chromothripsis-gene-amp-cancer.

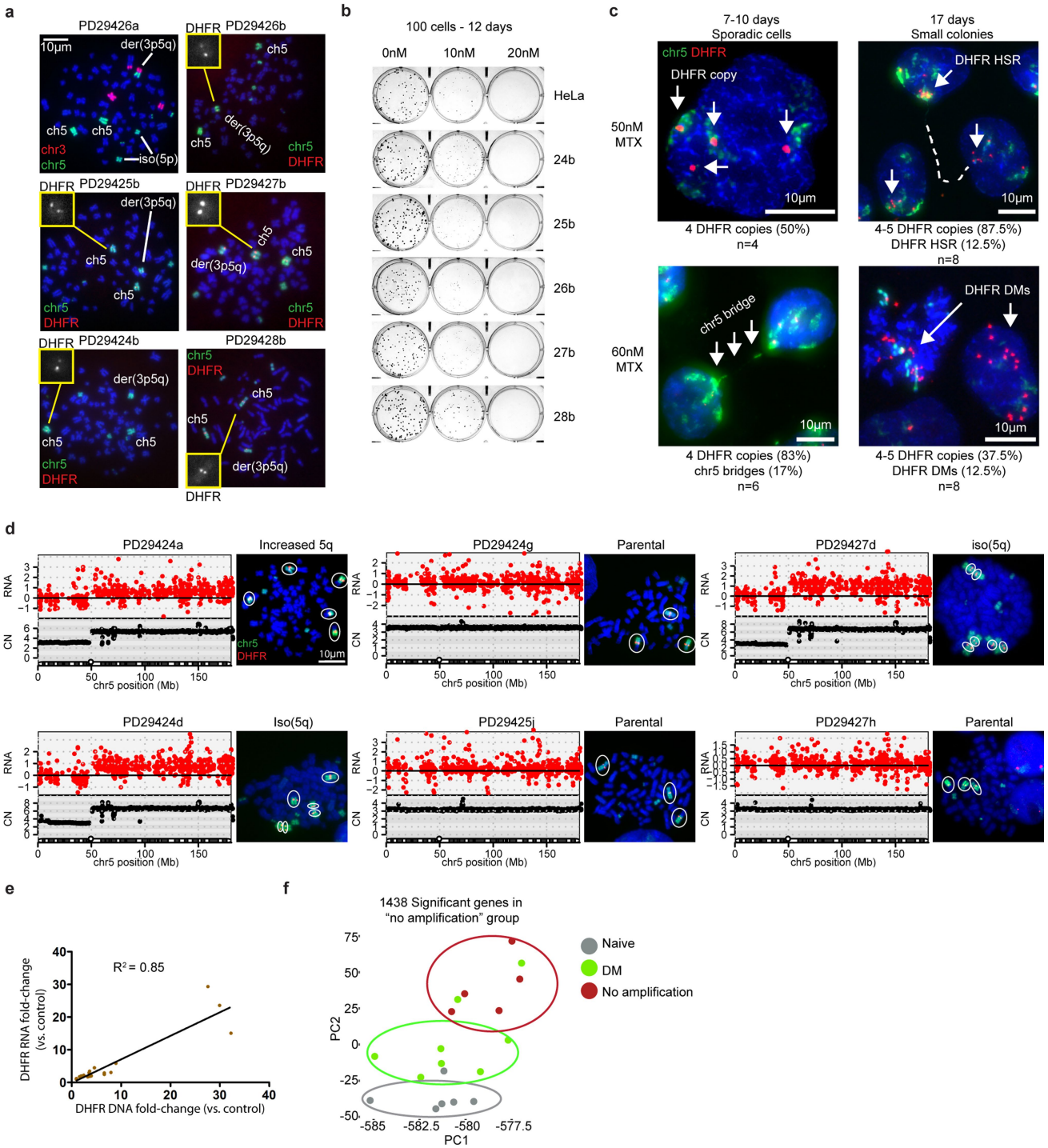
33. Cermak, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**, e82 (2011).
34. Fachinetti, D. et al. DNA sequence-specific binding of CENP-B enhances the fidelity of human centromere function. *Dev. Cell* **33**, 314–327 (2015).
35. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
36. Ou, H. D. et al. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).
37. Ou, H. D., Deerinck, T. J., Bushong, E., Ellisman, M. H. & O'Shea, C. C. Visualizing viral protein structures in cells using genetic probes for correlated light and electron microscopy. *Methods* **90**, 39–48 (2015).
38. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
39. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
40. Raine, K. M. et al. ascatNgs: identifying somatically acquired copy-number alterations from whole-genome sequencing data. *Curr. Protoc. Bioinformatics* **56**, 15.9.1–15.9.17 (2016).
41. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
42. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).
43. Li, Y. et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).
44. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).

Acknowledgements This work was funded by grants from the US National Institutes of Health (R35 GM122476 to D.W.C.), the Wellcome Trust (WT088340MA to P.J.C.), the US National Institutes of Health (K99 CA218871 to P.L.), the Swiss National Science Foundation (P2SKP3-171753 to S.F.B.), the Ludwig Institute for Cancer Research (D.W.C. and B.R.), an MSK Cancer Center Core Grant from the NIH (P30 CA 008748 to R.Y.), the National Institute of General Medical Sciences (P41GM103412, R24GM137200 to M.H.E.), and the High End Instrumentation Award (S10OD021784 to M.H.E.). D.W.C. and B.R. receive salary support from the Ludwig Institute for Cancer Research. We thank A. Shiau for providing access to the CQ1 spinning disk confocal system and N. Shimizu for providing the Colo320-DM-GFP cell line.

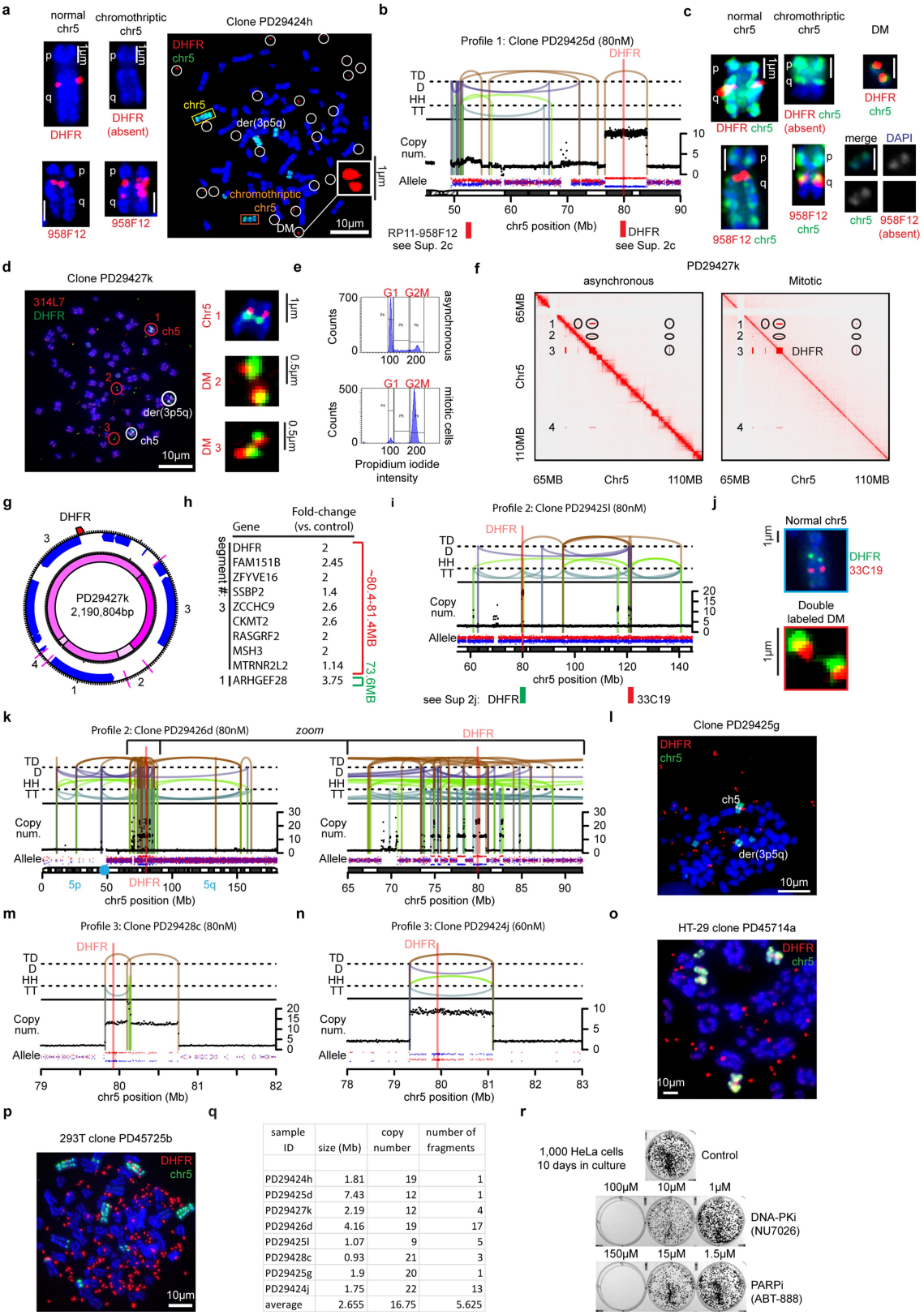
Author contributions O.S. and D.W.C. conceived the project and wrote the manuscript. O.S. designed, performed, and analysed all experiments unless otherwise specified here. O.S. and P.L. performed HSR evolution experiments. O.S. and Y.N.-A. tested drug treatment concentrations, O.S. and D.H.K. performed live-cell imaging and performed DM integration experiments following Cas9 expression. R.Y. obtained human DNA samples and performed FISH of human biopsies. O.S. and J.S.Z.L. designed sgRNA for Cas9 experiments. S.F.B., P.J.C., and O.S. performed and analysed the DNA sequencing experiments. O.S., S.F.B., and Y.S. analysed the RNA sequencing data. O.S. and M.Y. performed the HiC experiments. O.S., R.F., and B.R. analysed the HiC sequencing data. G.A.C. and O.S. performed the CLEM experiments. O.S., G.A.C., and M.H.E. analysed the CLEM experiments. All authors provided input on the manuscript. D.W.C., P.J.C., and O.S. supervised all aspects of the work.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-03064-z.
Correspondence and requests for materials should be addressed to P.J.C. or D.W.C.
Peer review information *Nature* thanks Jan Korbel, Roel Verhaak and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



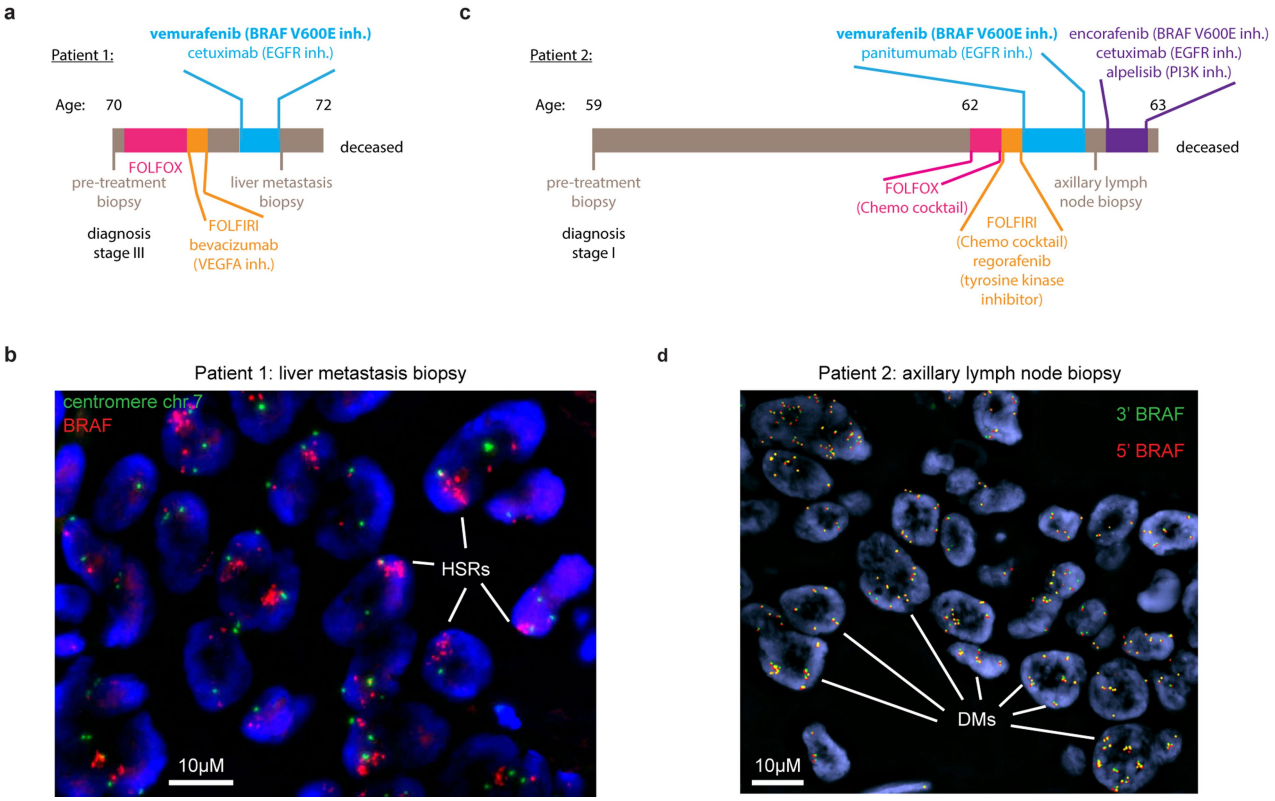
Extended Data Fig. 1 | Genomics and transcriptomics before and after methotrexate resistance. a, Representative DNA-FISH images showing parental HeLa karyotype of chromosome 5 found in the five primary clones used in the study. The parental der(3p5q) is shown in the top left image. **b**, Colony assay showing methotrexate sensitivity of the parental HeLa and five derivative clones. **c**, Representative DNA-FISH images (of the indicated independent experiments n displayed below each image) of surviving cells and colonies of naive HeLa cells treated with methotrexate at indicated concentrations for the indicated times. Increased *DHFR* signals over wild-type (>3 signals) were frequently observed, and *DHFR* aggregates indicative of HSR formation (with chromosome 5 interphase bridge detected in some cases – outlined with white dashed line in top right image) or *DHFR*+ DMs (dispersed signal) were found in surviving colonies of 100-200 cells at day 17. **d**, RNA expression and DNA copy number levels plotted on the linear maps of chromosome 5 of six methotrexate-resistant clones with no *DHFR* amplification. Representative DNA-FISH images are displayed (of at least 10 different chromosome spreads from each clone). **e**, Linear regression comparing DNA copy number and RNA expression levels of *DHFR* in methotrexate-resistant HeLa clones. **f**, Principal component analysis (PCA) of naive and resistant HeLa clones (with DMs or without *DHFR* amplification).



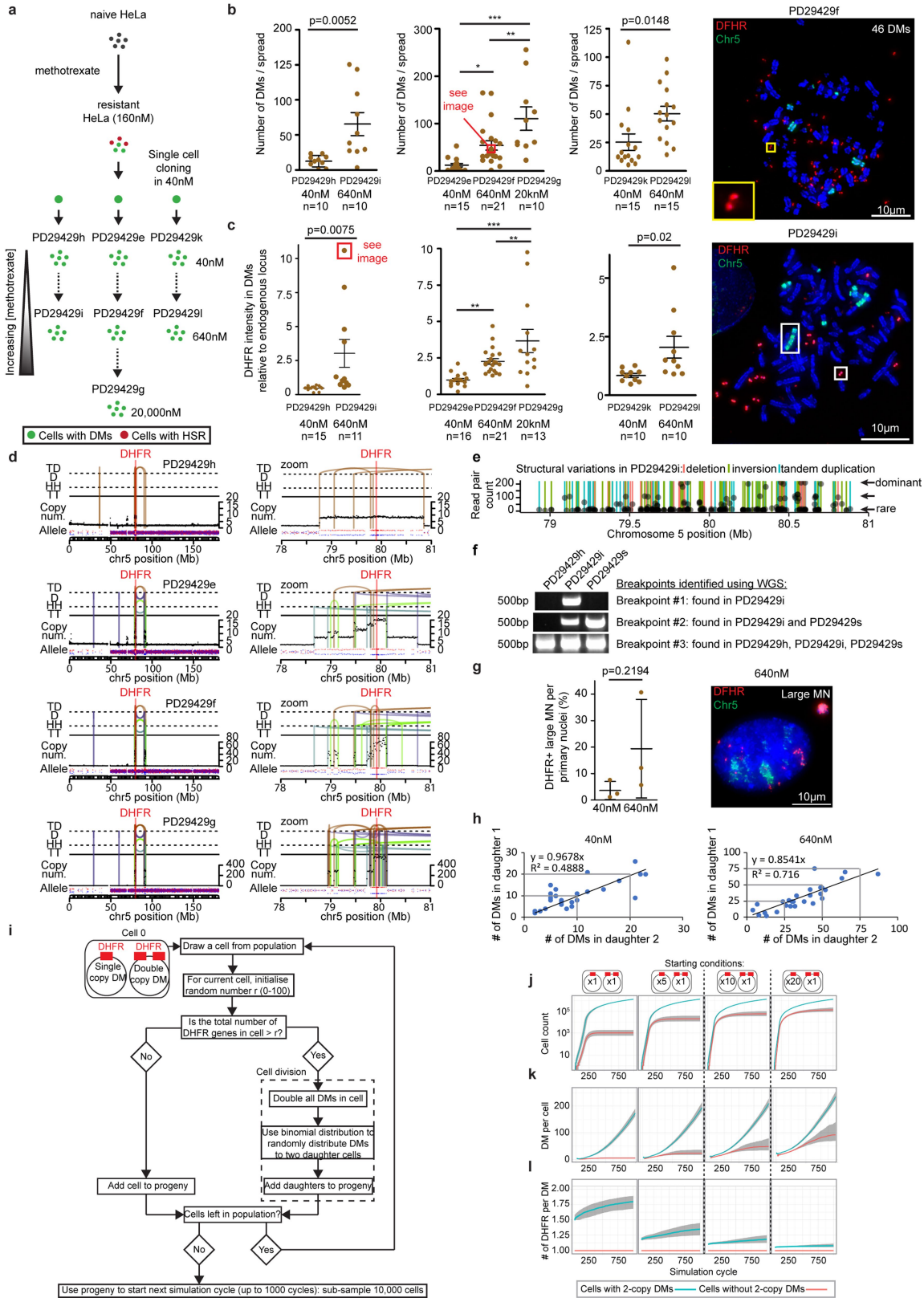
sample ID	size (Mb)	copy number	number of fragments
PD29424h	1.81	19	1
PD29425d	7.43	12	1
PD29427k	2.19	12	4
PD29426d	4.16	19	17
PD29425l	1.07	9	5
PD29428c	0.93	21	3
PD29425g	1.9	20	1
PD29424j	1.75	22	13
average	2.655	16.75	5.625

Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Formation of DMs through chromothripsis in methotrexate-resistant cells. a, Representative DNA-FISH of metaphase spread prepared from clone PD29424h (see Fig. 1c) showing *DHFR* amplified in DMs (marked with white circles). Alignment of the normal chromosome 5 (*DHFR*+) and shorter chromothriptic chromosome 5 (*DHFR*-) is presented. Representative DNA-FISH from a metaphase spreads hybridized with chromosome 5 paint probe (green) and BAC probe (RP11-958F12, red, ~30 Mb away from DHFR) found on both the normal chromosome 5 (single location) and shorter chromothriptic chromosome 5 (multiple dispersed locations) is also presented. **b**, **i**, **k**, **m**, **n**, Copy number, allelic ratio, and structural variation profiles of indicated samples. **c**, Representative FISH images from clone PD29425d showing a normal chromosome 5 (*DHFR*+, RP11-958F12 positive), a shorter chromothriptic chromosome 5 (*DHFR*-, RP11-958F12 positive), and a DM (*DHFR*+, RP11-958F12 negative). See panel **b** for BAC probe target. **d**, Representative DNA-FISH image of metaphase spreads prepared from clone PD29427k and hybridized with BAC RP11-314L7 (labels fragment #1 of the DM as shown in Fig. 1d) and *DHFR* locus probes. Insets show the co-localization of the non-contiguous genomic locations within the DMs. **e**, Representative FACS analysis using propidium iodide staining showing cell cycle distribution in asynchronized and mitotic clones prepared for Hi-C experiments. For gating strategy see Supplementary Fig. 2. **f**, In situ Hi-C map of chromosome 5 (65-110 Mb) of asynchronized and mitotic cells from methotrexate-resistant clone PD29427k. Interactions between distant chromosome locations (preserved also in the mitotic sample where topologically associated domains are erased) are circled in black (above the diagonal red line) and presented natively (below the line). **g**, Reconstruction of the circular map of the DM present in clone PD29427k. Numbers represent the four DM segments appearing in panel **f**. **j**, Representative DNA-FISH from clone PD29425l showing co-localization of *DHFR* (green) and BAC 33C19 probes signals. See panel **i** for BAC probe target. **l**, Representative DNA-FISH image of metaphase spreads prepared from clone PD29425g (see Fig. 1e) hybridized with chromosome 5 paint (green) and *DHFR* locus (red) probes. Inset shows the *DHFR* positive DM. Notice only one chromosome 5 and one der(3p5q) are present. **o**, **p**, Representative DNA-FISH image of metaphase spreads prepared from clones PD45714a (**o**, see Fig. 1f) and PD45725b (**p**, see Fig. 1g) hybridized with chromosome 5 paint (green) and *DHFR* locus (red) probes showing DHFR amplification in DMs. **q**, List of DM+ HeLa clones showing size, copy number, and number of non-contiguous fragments. **r**, Colony assay of naive HeLa cells treated with DNA repair inhibitors (control – untreated). Images are representative of two independent experiments.

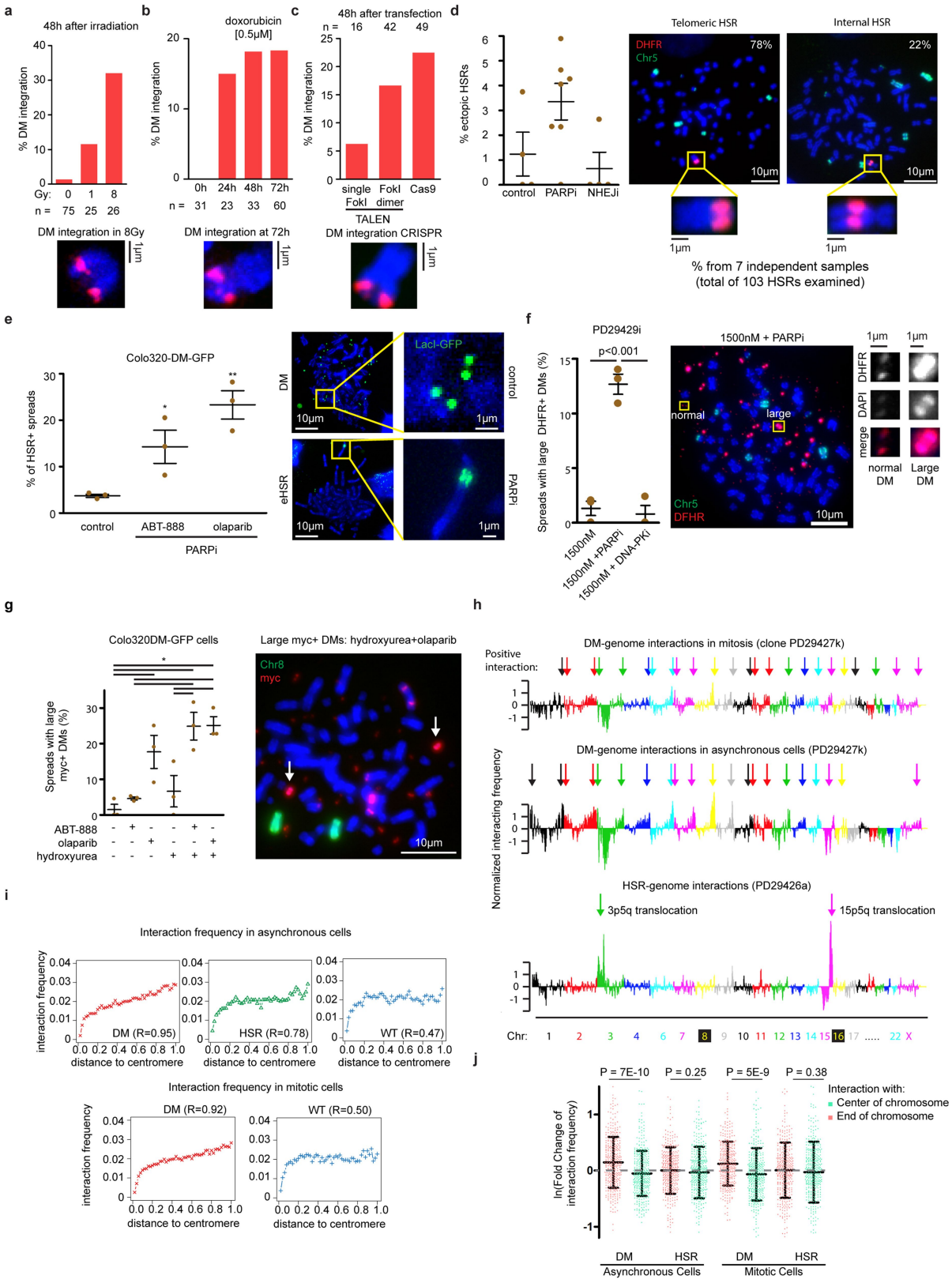


Extended Data Fig. 3 | *BRAF*^{V600E} amplification in patients with drug-resistant colorectal cancer. a, **c**, Treatment and biopsies collection timeline of patients 1 and 2. **b**, **d**, Representative FISH images (of entire section stained) showing BRAF amplification in biopsies from post-treatment biopsies from patients 1 and 2. Probes used are listed in the methods section.



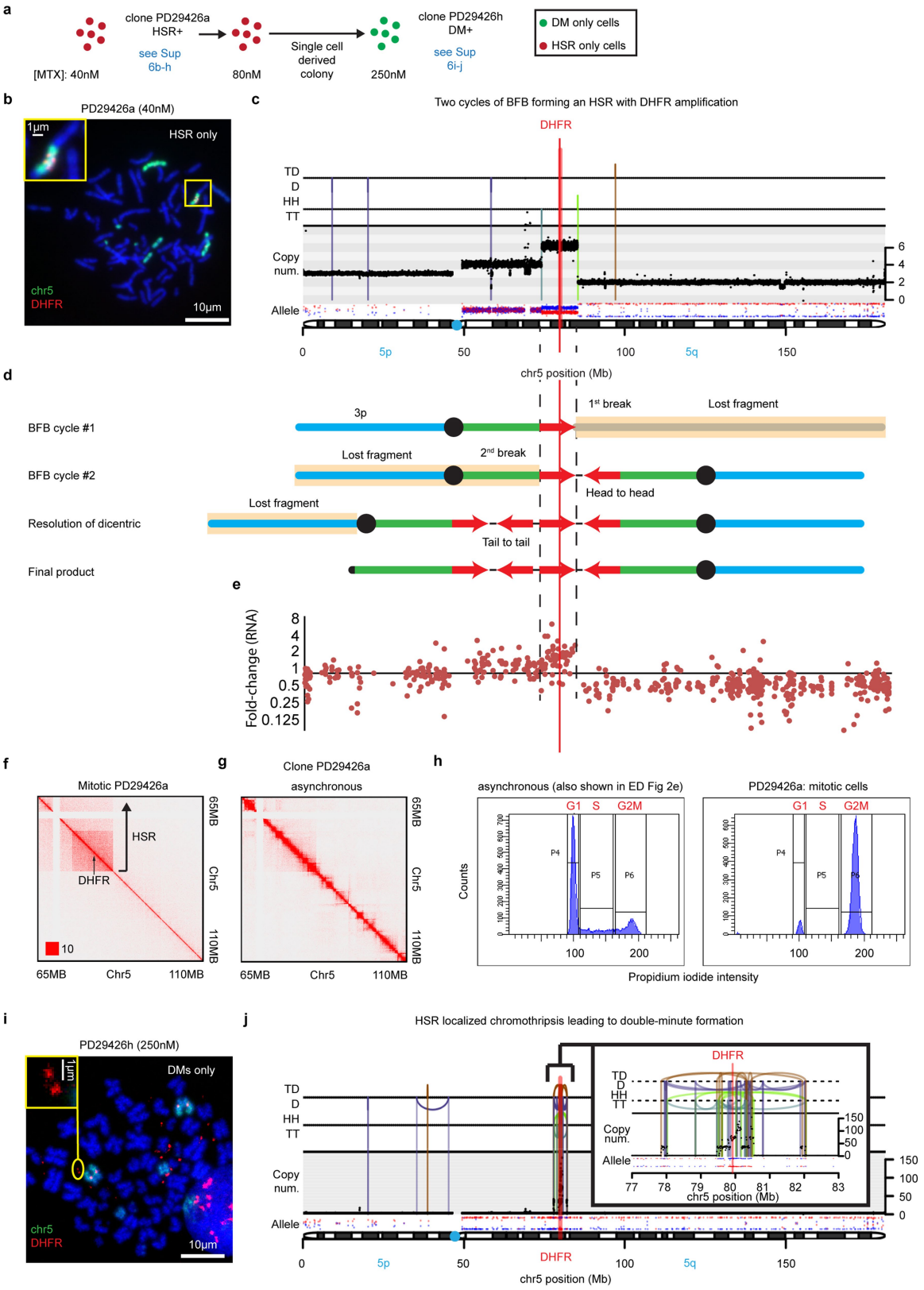
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | DMs are numerically and structurally unstable. **a**, Strategy used to isolate DM or HSR positive clones from a heterogeneous population of methotrexate-resistant HeLa cells. **b**, Left: Average number of *DHFR*+ DMs in three clones derived from a methotrexate-resistant HeLa population was determined using DNA-FISH. Analyses with indicated p-values above plots were performed using Student *t*-test. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ are p-values calculated using one-way ANOVA. Error bars represent mean ± s.e.m., number of spreads examined for each condition is written below each graph. Right: Representative DNA-FISH image of metaphase spreads from clone PD29429f (640 nM) containing 46 *DHFR*+ DMs as determined by DNA-FISH. Inset shows a representative DHFR positive DM. **c**, Average intensity of *DHFR* signal in DMs relative to the intensity of the endogenous *DHFR* signal on chromosome 5 from the same spread, as determined by DNA-FISH using a probe (RP11-90A9) specific for the *DHFR* locus. Analyses with indicated p-values above plots were performed using Student *t*-test. $^{**}P < 0.01$ and $^{***}P < 0.001$ are p-values calculated using one-way ANOVA. Error bars represent mean ± s.e.m. Representative DNA-FISH image of chromosome spread from which insets are shown in Fig. 2b is presented. **d**, Copy number, allelic ratio, and structural variation profiles of indicated samples. **e**, Read counts in structural variation breakpoints in sample PD29429i showing potential different DM subspecies forming through chromothripsis. **f**, Breakpoint PCR using primers specific for three different rearrangements detected using WGS. For gel source data, see Supplementary Information Fig. 1. **g**, Percentage of DM-positive cells exposed to low or high methotrexate concentrations with large micronuclei, as determined by DNA-FISH. Results represent an average of 3 clones per methotrexate concentration (as seen in Extended Data Fig. 4a), error bars represent mean ± s.d. Representative DNA-FISH image of a cell with a large micronucleus is provided. 40 nM: $n = 147, 130, 98$; 640 nM: $n = 74, 216, 133$. **h**, Analysis of DM inheritance to two daughter cells after one cell division. Cells were seeded at low dilution and 24 h later DNA-FISH on interphase daughter cells from PD29429h (40 nM) or PD29429i (640 nM) was performed using a DHFR probe. DM numbers in each daughter cell were counted and plotted on an x-y plot ($n = 54$ daughters per condition). **i**, Schematic explaining the steps of a simulation testing the effect of harbouring DMs with more than one DHFR gene. **j**–**l**, In-silico simulation showing that cells containing a 2-copy *DHFR* DM have a selection advantage (**i**), with increased DM content (**j**), and that 2-copy *DHFR* DMs are selected over DMs with a single copy of *DHFR* (**k**). Plotted is the median and first and third quartile (grey ribbon). For simulation pseudocode see Supplementary Information Fig. 3.



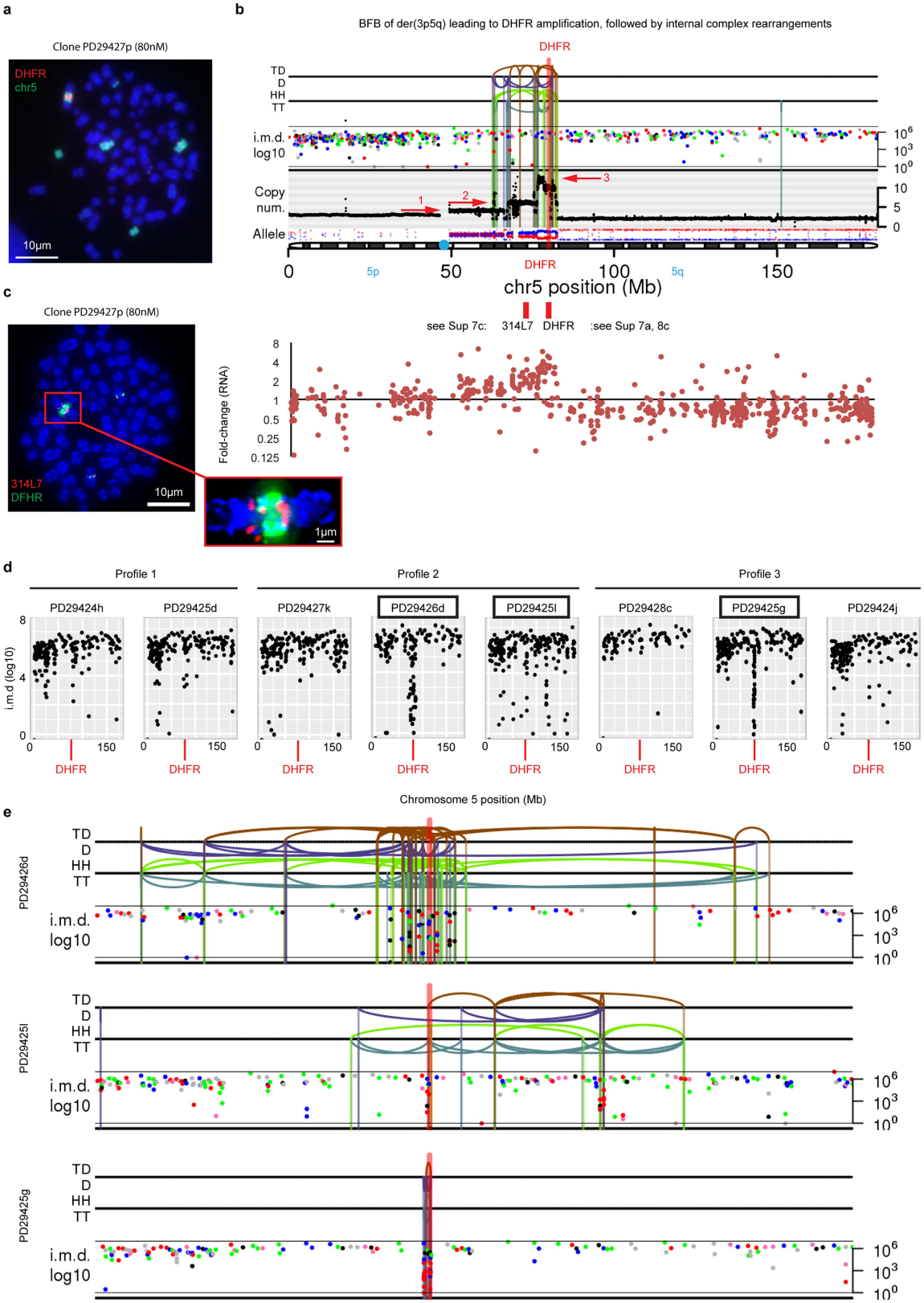
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | DMs integrate into ectopic chromosomes following DNA damage. a–**c**, Percentage of DM integration as detected using DNA-FISH with probes against chromosome 5 (green) and *DHFR* (red) in clone PD29429i after ionizing irradiation (**a**), doxorubicin (**b**), or transfection with nucleases specific for a region near the *DHFR* locus (**c**). Representative images of DNA-FISH of DMs integrated into ectopic chromosomes are presented below each graph. **d**, Percentage of ectopic HSRs detected in DM-positive clones treated with increased methotrexate concentration (x2.5 fold higher) with or without (control, 4 clones) the addition of ABT-888 (PARPi, 7 clones) and NU7026 (NHEJi, 4 clones) for 3 weeks. Means ± s.e.m. are presented. control: $n = 107, 96, 46, 164$; PARPi: $n = 67, 247, 94, 128, 17, 130$; NHEJi: $n = 51, 45, 28, 38$. Representative FISH images of ectopic HSRs at the end of chromosome (78% of the cases) or in the middle of chromosomes (22% of the cases) are presented. **e**, Percentage of GFP+ HSRs detected in colo320DM-GFP cell line treated with DMSO (control) or with ABT-888 (PARPi) or olaparib (PARPi) for 2 weeks. Means ± s.e.m. of three independent experiments are shown. p-values $^{*}P < 0.05$ and $^{**}P < 0.001$ calculated using one-way ANOVA are presented. control: $n = 26, 47, 33$; ABT-888: $n = 42, 44, 37$; Olaparib: $n = 29, 32, 34$. Representative fluorescent images of GFP+ DMs (from control) and GFP+ HSRs (from PARPi treated cells) are presented. **f**, Percentage of chromosome spreads in which large *DHFR*+ DMs were detected using DNA-FISH when PD29429i was exposed to 1,500 nM methotrexate only (control) or with addition of 15μM ABT-888 (PARPi) or 10 μM NU7026 (DNA-PKi) for 3 weeks. Means ± s.e.m. of three independent experiments are presented. * p-value <0.001 calculated using one-way ANOVA and Tukey's Multiple Comparison Test. Number of spreads scored per condition/experiment: 1,500 nM MTX (52, 53, 49), 1,500 nM MTX + ABT-888 (50, 76, 55), 1,500 nM MTX + NU7026 (57, 16, 42). Representative image (of three independent experiments as shown in the plot to the left) of a chromosome spread containing multiple normal sized and large DMs is presented. **g**, Percentage of chromosome spreads in which large myc+ DMs were detected using DNA-FISH in colo320DM-GFP cells with indicated types of treatment for 2 weeks (DMSO served as control, PARP inhibitors used were 15μM ABT-888 and 10 μM Olaparib, hydroxyurea used at 100 μM). Representative image of a chromosome spread containing multiple normal sized and large DMs (white arrows) is presented. Means ± s.e.m. of three independent experiments with $^{*}P < 0.05$ calculated using one-way ANOVA are presented. control: $n = 22, 22, 21$; ABT-888: $n = 26, 19, 21$; Olaparib: $n = 20, 22, 21$; hydroxyurea: $n = 20, 22, 20$; hydroxyurea+ABT-888: $n = 19, 22, 24$; hydroxyurea+Olaparib: $n = 20, 22, 22$. **h**, Quantification of the interaction frequency between DM or HSR sequences and the rest of the genome in mitotic or asynchronous cells. Linear chromosome maps (chromosomes 1-X, excluding chromosome 5) are presented from beginning to end (p arm to q arm direction), and the normalized interaction frequency is presented. **i**, Interaction frequency as a function of distance from the centromere (R - pearson correlation). **j**, Fold-change (relative to cells with no amplification) of *DHFR* sequences interaction in DM+ and HSR+ with centre of chromosomes (50% of sequences in chromosome centre) and with chromosome ends (25% of sequences in each chromosome end). Means ± s.d. and *P* values calculated using two-sided paired *t*-test are presented.



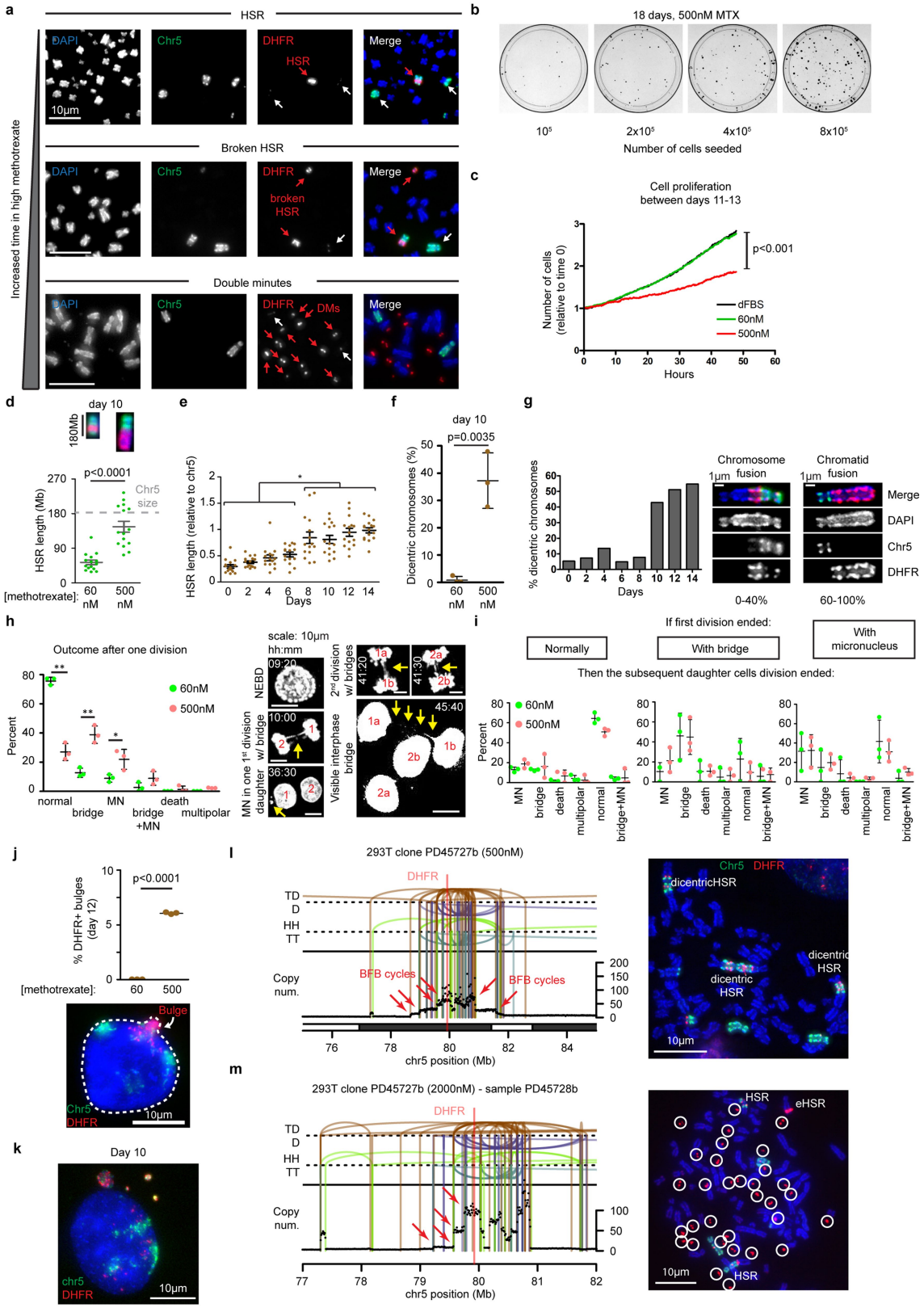
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Increased selective pressure drives the transition of intra- to extrachromosome amplification through chromothripsis. **a**, Experimental outline of a parental HSR+ clone treated with increasing methotrexate concentrations. **b**, Representative FISH image of 23 metaphase spreads prepared from the methotrexate-resistant HSR+ clone PD29426a hybridized with chromosome 5 paint (green) and *DHFR* locus (red) probes. Inset shows the *DHFR* repeat located on the 5q arm of the derivative 3p5q chromosome. **c**, DNA copy number, rearrangement, and allelic ratio profile of clone PD29426a showing two copy number jumps flanked by head-to-head and tail-to-tail inversions corresponding to two BFB cycles. **d**, Schematic depicting the order of events during the two break-fusion-bridge (BFB) cycles that led to the formation of the HSR in clone PD29426a. **e**, RNA expression presented on the linear chromosome 5 map showing increased gene expression from the HSR region. **f**, In situ Hi-C map of chromosome 5 (65-110 Mb) of mitotic cells from methotrexate-resistant clone PD29426a, showing the HSR in which there is higher interaction within the regions with copy number jumps. **g**, In situ Hi-C map of chromosome 5 (65-110 Mb) of asynchronized cells from HSR+ clone PD29426a. TADs are visible throughout this region and increased interactions are observed in the HSR region (increased red coloration, top left quadrant). **h**, Cell cycle analysis using propidium iodide staining showing cell cycle distribution in asynchronized cells and mitotic cells of clone PD29426a. For gating strategy see Supplementary Information Fig. 2. **i**, Representative FISH image (of two independent clones, total of 26 spreads imaged) of metaphase spreads prepared from a DM subclone (PD29426h) derived from the HSR+ clone PD29426a. Spreads were hybridized with chromosome 5 paint (green) and *DHFR* locus (red) probes. Inset shows the *DHFR* positive DMs. **j**, DNA copy number, rearrangement, and allelic ratio profile of clone PD29426h. The DM is composed of fragments with varying copy number states derived from the original HSR region, connected by multiple rearrangements.



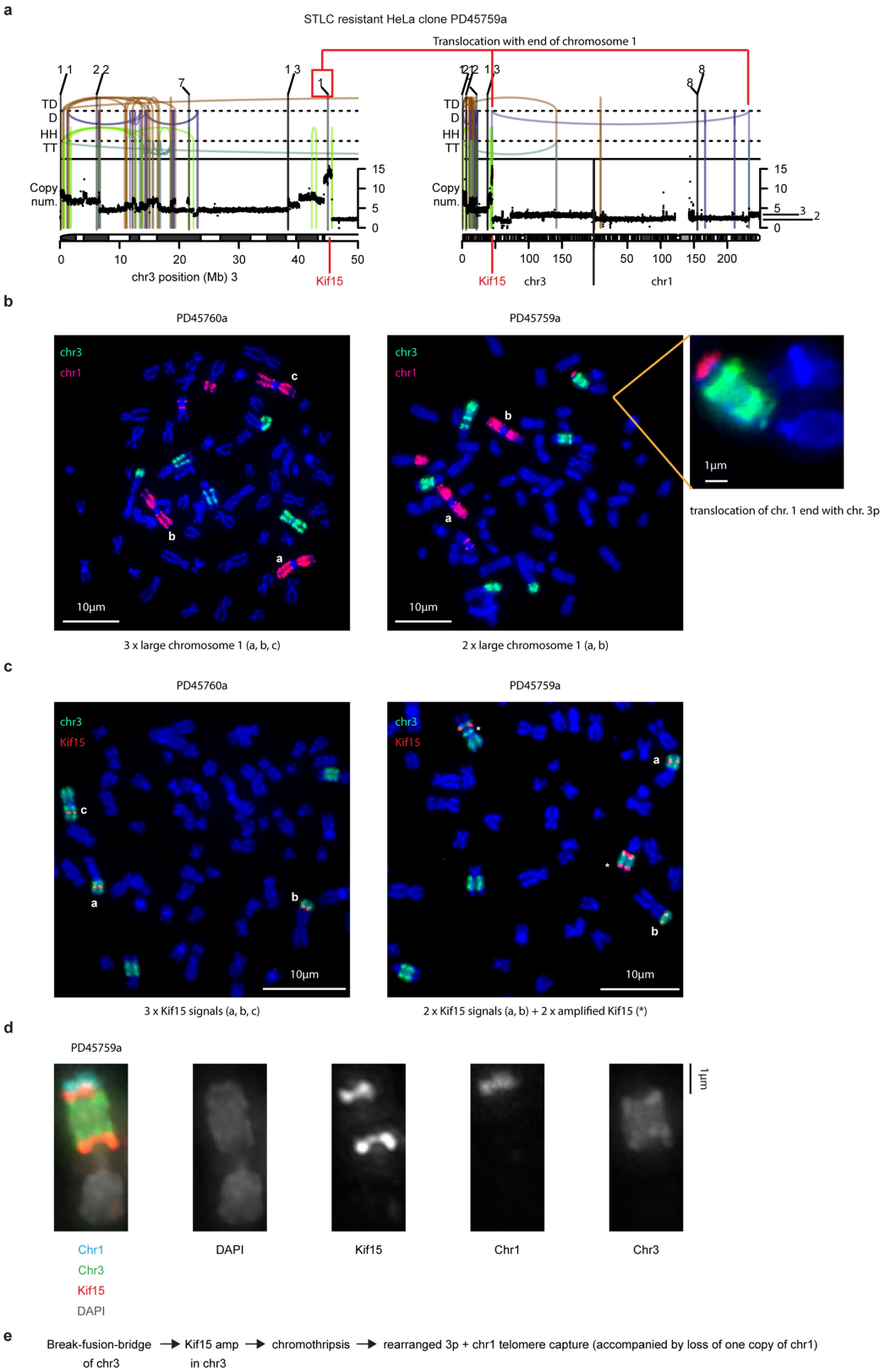
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Chromothripsis and kataegis in gene amplification. **a**, Representative FISH image (of 12 spreads imaged) of metaphase spreads prepared from clone PD29427p and hybridized with chromosome 5 paint (green) and *DHFR* locus probes, revealing the existence of a DHFR+ HSR. **b**, Top: Copy number profile of a methotrexate-resistant clone (PD29427p) presenting an HSR profile resulting from 3 BFB cycles with overlaid complex rearrangements. Rearrangements are presented on top: TD – tandem duplication, D – deletion, HH – head-to-head inversion, TT – tail-to-tail inversion. Inter-mutation distance (i.m.d) and allelic ratio (bottom, blue/red dots) are also presented. Bottom: RNA expression profile of clone PD29427p, relative to control naive HeLa cells, presented on the linear map of chromosome 5 and matching the DNA copy number plotted above. **c**, Representative FISH image (of 6 spreads imaged) of metaphase spreads prepared from clone PD29427p and hybridized with chromosome 5 paint (green) and BAC probe 314L7 (see chromosomal location indicated in panel **b**) probes. Inset shows probe 314L7 signal is flanking the DHFR signal within the HSR. **d**, Rainfall plots showing the inter-mutation distance (i.m.d.) within chromosome 5 of each DM clone. **e**, Structural variations and inter-mutation distance (i.m.d.) of three kataegis positive DM clones.



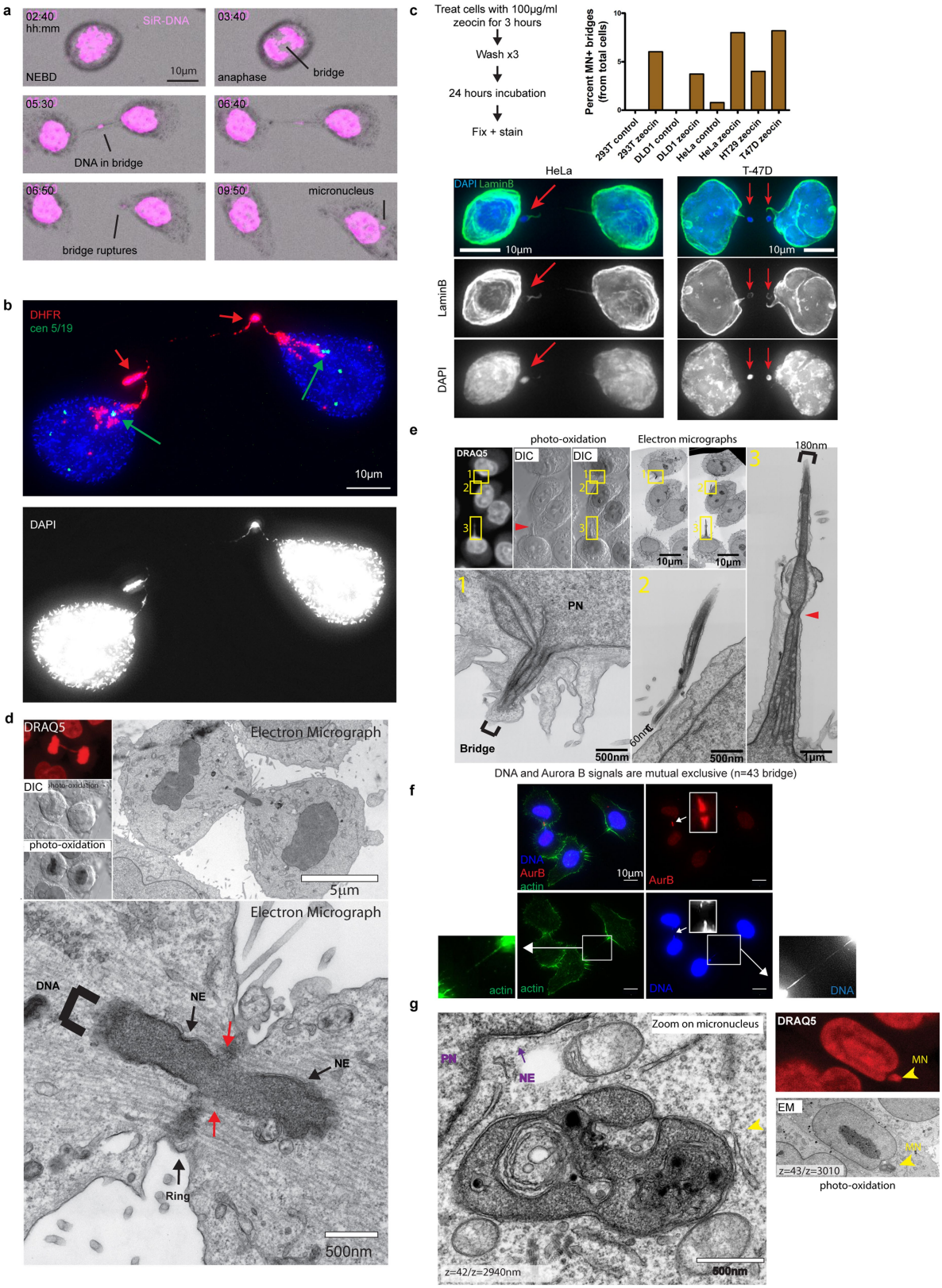
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Characterization of steps leading to HSR fragmentation and DM formation. a, Representative FISH images (of 464 spreads imaged from 8 different time points as shown in Fig. 4a) showing chromosome 5 (green) and *DHFR* (red) probes in HSR+ clone PD29428e at basal methotrexate concentration (60nM, top), and following treatment with increased methotrexate (500 nM, middle and bottom) in which HSR fragments of variable sizes and DMs can be detected. **b**, Colony assay of clone PD29428e (HSR+) resistant to 60 nM and treated with 500 nM methotrexate. Numbers of cells seeded are indicated and colonies were fixed and stained at day 18. **c**, Number of cells counted (automatically) using the CQ1 microscope in 10 min interval during 48 h of filming of PD29428e cells treated with 60 nM or 500 nM or without (dFBS) methotrexate. Cell numbers were normalized to the initial number present at time point 0. P-value calculated using one-way ANOVA. **d**, Measurement of the HSR length in clone PD29428e after ten days of exposure to higher methotrexate concentration. The length of the HSR was normalized to the length of the normal chromosome 5 from the same spread. Mean ± s.e.m. of $n = 15$ per group and p-value calculated using two-tailed *t*-test are presented. Insets showing length of the HSR at day 10 after exposure to basal (left) or increased (right) methotrexate concentrations (scale is 180 Mb, equivalent to chromosome 5 size). **e**, Measurement of the HSR length in clone PD29428e during a two-week exposure to higher methotrexate concentration. The length of the HSR was normalized to the length of the normal chromosome 5 from the same spread. Means ± s.e.m. of 15 HSR lengths per each time point (except day 14 – 16 HSRs scored). p-values calculated using one-way ANOVA and * denotes significant (range of $P < 0.05$-0.001) change in HSR length between each point of 8-14 days and each point of 0-6 days. **f**, Percentage of spreads containing dicentric HSR+ chromosomes in clone PD29428e exposed to basal or increased methotrexate concentration for 10 days. Mean ± s.d. of three independent experiments (60nM: $n = 42, 30, 33$; 500nM: 23, 33, 29) and p-value calculated using two-tailed *t*-test are presented. **g**, Percentage of spreads containing dicentric HSR+ chromosomes in clone PD29428e after increasing methotrexate concentration for the indicated times. day 0: $n = 98$; day 2: $n = 41$; day 4: $n = 39$; day 6: $n = 63$; day 8: $n = 53$; day 10: $n = 55$; day 12: $n = 66$; day 14: $n = 123$. Inset shows representative image (from 538 spreads imaged) of a chromatid-type dicentric chromosome with one fused end, and chromosome type fusion with two distinct centromeres. Percentages indicate frequency range of each dicentric type throughout the two-week experiment. **h**, Abnormalities identified in PD29428e cells treated without or with increased methotrexate concentration (60 nM and 500 nM, respectively) for 11 days during 48 h filming. Mean ± s.d. of three independent experiments and p-values ($*P < 0.01$ and $*P < 0.001$) calculated using two-way ANOVA are presented. Division scored: 60nM: $n = 115, 108, 95$; 500nM: $n = 51, 50, 42$. Bridges (anaphase bridges), micronuclei (MN), bridges+MN, death, and multipolar divisions were scored. Insets from Supplementary Video 1 showing anaphase bridges, micronuclei, and also interphase bridges are presented. **i**, Analysis as performed in (h) looking at events occurring during the division of the daughter cells generated from the parental cells scored in **h**. Division scored: 60nM: $n = 100, 90, 66$; 500nM: $n = 85, 76, 66$. **j**, Top: Percentage of nuclei presenting *DHFR*+ nuclear bulges indicative of bridge rupture in PD29428e HSR+ clone treated with increased methotrexate (500nM) over the course of 12 days. Mean ± s.d. of three independent experiments (60nM: $n = 144, 221, 204$; 500nM: $n = 163, 218, 214$) and p-value calculated using two-tailed *t*-test are presented. Bottom: Representative image of *DHFR*+ (red) labelled nuclear bulge. **k**, Representative image of the data in Fig. 4d. **l**, Copy number profile of a methotrexate-resistant clone (PD45727b) presenting an HSR profile resulting from multiple BFB cycles (marked with red arrows) with overlaid complex rearrangements. A representative DNA-FISH image (of 20 spreads imaged) showing dicentric HSRs is presented. DMs were observed in low frequency under this condition. **m**, Copy number profile of clone PD45727b subjected to increased methotrexate concentration (sample PD45728b) presenting additional rearrangements and formation of multiple DMs (as seen in the representative DNA-FISH image from 10 spreads imaged, white circles). TD – tandem duplication, D – deletion, HH – head-to-head inversion, TT – tail-to-tail inversion.



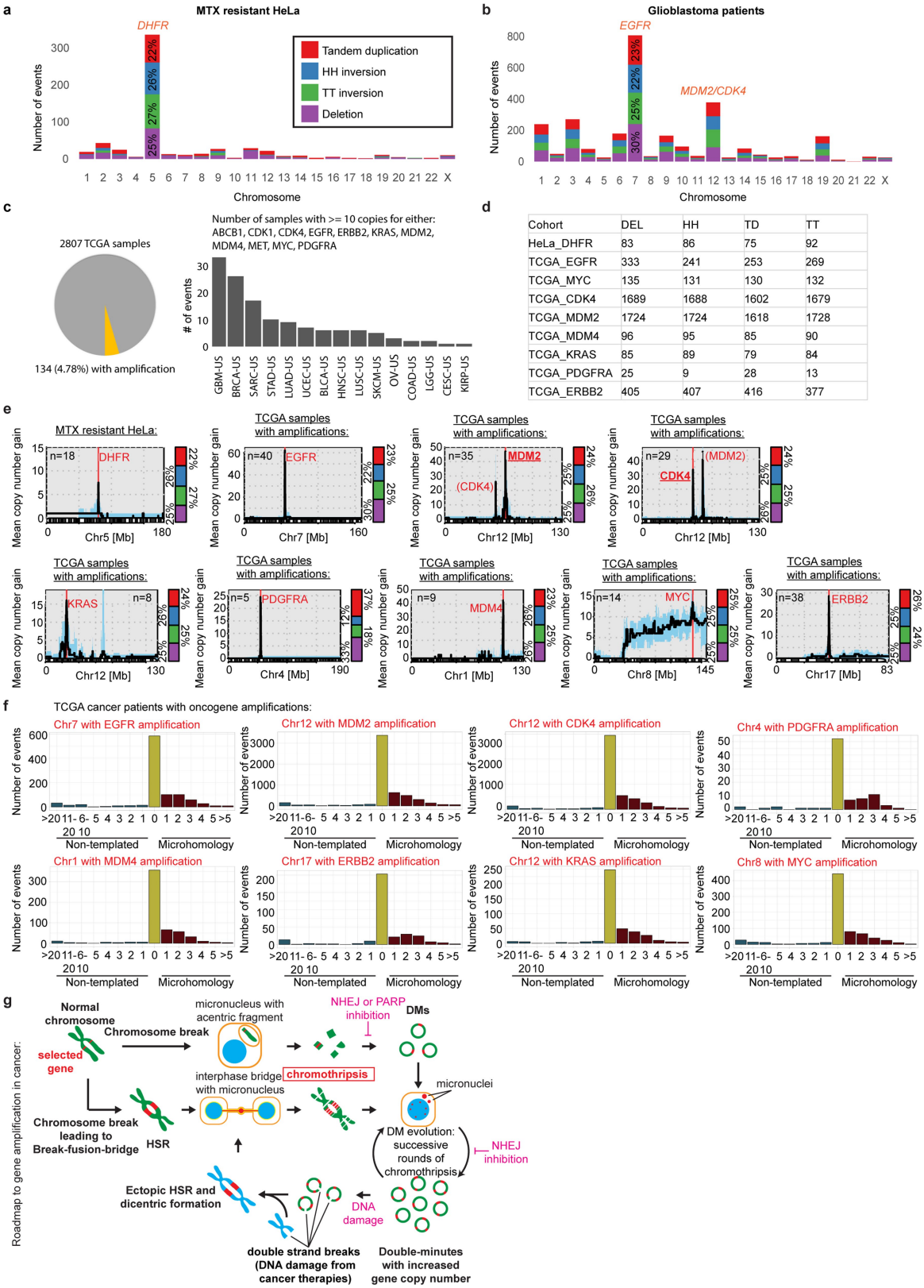
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Chromothripsis drives resolution of *KIF15* intrachromosomal amplification through telomere capture. a, Copy number profile of chromosome 3 showing a rearranged 3p region with *KIF15* amplification (copy number = 15) and a translocation of a telomeric fragment from a lost chromosome 1. **b**, Representative DNA-FISH images (of 33 spreads imaged) from STLC resistant clones with (PD45759a) or without (PD45760a) *KIF15* amplification showing the translocation of the telomeric chromosome 1 fragment to the end of a chromosome 3 arm. **c**, Representative DNA-FISH images (of 20 spreads imaged) for chromosome 3 (green) and *KIF15* (RP11-659N22, red) showing normal *KIF15* in PD45760a (STLC resistant without *KIF15* amplification, 3 *KIF15* copies labelled with a, b, and c) and in PD45759a (STLC resistant with *KIF15* amplification found on two rearranged chromosomes marked with * and originating from chromosome c as seen in PD45760a). **d**, Representative DNA-FISH image (from 25 spreads imaged) showing from PD45759a showing the rearranged chromosome containing *KIF15* amplification and telomeric region of chromosome 1 at the tip of the chromosome. **e**, Proposed order of events leading to *KIF15* amplification and resolution of the dicentric state through a chromothriptic event.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Structure of the interphase chromosome bridge. **a**, Snapshots from live-cell imaging (Supplementary Video 2) of PD29428e cells exposed to increased methotrexate concentration (500nM) for 11 days, labelled with SiR-DNA label, and filmed for 48 h (of 944 cell divisions filmed, as shown in Extended Data Fig. 8h-i). Frames are at 10-min intervals. An initial DNA+ bridge forms from which a DNA fragments is maintained as a micronucleus following bridge rupture. **b**, Representative image of PD29428e cells exposed to increased methotrexate concentration (500nM) for 10 days and analysed using DNA-FISH probes for chromosome 5 centromere (green) and DHFR (red), showing micronuclei within interphase DNA bridges (representative of 34 bridges). **c**, Immunofluorescent image analysis of zeocin treated cancer cell lines showing formation of micronuclei within interphase DNA bridges (representative images of 85 bridges with micronuclei from the four cell lines tested. Number of cells scored: 293T control: $n = 99$; 293T zeocin: $n = 116$; DLD1 control: $n = 88$; DLD1 zeocin: $n = 107$; HeLa control: $n = 126$; HeLa zeocin: $n = 412$; HT29 zeocin: $n = 25$; T47d zeocin: $n = 73$). **d**, **e**, CLEM images (of a total of 5 images taken) of an anaphase bridge (**d**, red arrow marks aberrant nuclear envelope at midbody region) and an interphase bridge (**e**, red arrow marks constriction in DNA bridge at midbody site). NE – nuclear envelope. **f**, Representative immunofluorescence images (of 43 different bridges) showing AuroraB localization in interphase DNA bridges. **g**, CLEM of an abnormally shaped micronucleus found in PD29428e cells exposed to increased methotrexate concentration (500 nM) for 10 days in which micronuclei form following interphase bridge rupture (representative of 2 micronuclei imaged).



Cohort	DEL	HH	TD	TT
HeLa_DHFR	83	86	75	92
TCGA_EGFR	333	241	253	269
TCGA_MYC	135	131	130	132
TCGA_CDK4	1689	1688	1602	1679
TCGA_MDM2	1724	1724	1618	1728
TCGA_MDM4	96	95	85	90
TCGA_KRAS	85	89	79	84
TCGA_PDGFRA	25	9	28	13
TCGA_ERBB2	405	407	416	377

Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Similarities between DHFR and oncogene amplifications. a, **b**, Summary of the overall structural variant events found across all samples analysed in the methotrexate-resistant HeLa cohort (**a**, 18 clones) and in patients with glioblastoma cancer from the TCGA database (**b**, 41 patients). **c**, Amplifications (> = 10 copy numbers) of either of 11 indicated genes in patients found in the TCGA database. **d**, Distribution of rearrangements in chromosomes containing specific amplifications. Near equal distribution of rearrangements suggest events are random as expected in chromothripsis. TD – tandem duplication, D – deletion, HH – head-to-head inversion, TT – tail-to-tail inversion. **e**, Copy number plots showing the mean copy number gain presented on the linear chromosome map in methotrexate-resistant HeLa cells (chromosome 5) and TCGA samples presenting oncogene amplifications. The percentage of each type of rearrangement found in chromosomes with amplifications is presented. Red – tandem duplications, Blue – head-head inversions, Green – tail-tail inversions, Purple – deletions. **f**, Microhomology (red), non-templated (teal) sequences, or direct end-joining (yellow), found at breakpoints in chromosomes with specific oncogene amplifications from TCGA patients. The number of homologous or non-homologous bases at breakpoints are shown on the *x*-axis. The number of breakpoints showing each kind of junction is shown on the *y*-axis. **g**, Schematic illustrating a roadmap to gene amplification in cancer cells.

nature research

Corresponding author(s): Don W. Cleveland

Last updated by author(s): Oct 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Paired-end HiC sequencing reads were mapped using BWA-MEM to the reference genome (hg19) in single-end mode with default parameter setting for each of the two ends separately.

For RNAseq, quality of sequencing reads was assessed using FastQC (Babraham Bioinformatics, version 0.11.2) and aligned to a reference genome (hg19, UCSC Genome Browser) using TopHat. Sequencing yielded on average 46.3 million reads per sample with a ~71% mapping rate to generate ~33 million reads per sample uniquely mapped to the reference genome.

For whole genome sequencing, mapping to the human genome was performed using the BWA algorithm BWA-MEM (Version 0.7.8) and genome build hs37d5.

Data analysis

For HiC sequencing, the independently mapped ends were then paired-up and the read pairs were kept if both ends uniquely mapped to the genome (MQAL>10). Next, read pairs were further removed if either end was mapped more than 500bp apart away from the closest MboI cutter site. Read pairs were next sorted based on genomic coordinates followed by PCR duplicate removal using samtools rmdup. Processed Hi-C data is converted to .hic file and is visualized using juicebox (version 1.5.2) without normalization.

For RNAseq, Cufflinks was used to generate transcript abundance as fragments per kilobase of transcript per million mapped reads (FPKM), and statistical analysis of FPKM values was calculated using R (version 3.5.0).

For whole genome sequencing, copy number was called using the ascatNgs algorithm. The variant calling pipeline of the Cancer Genome Project, Wellcome Sanger Institute was used to call somatic mutations. The following algorithms with standard settings and no additional post-processing were used: CaVEMan for substitutions and the BRASS algorithm for rearrangements.

Statistical analyses were performed with GraphPad Prism software (version 4.0)
Live imaging software: CQ1 software version 1.05.01.02

ImageJ version 1.51f used for image analysis
FACS Diva version 8.0.1 for flow cytometry analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data reported in this study are available upon request.

Paired-end whole genome sequencing data is available at ENA (European Nucleotide Archive), accession number ERP107458 https://www.ebi.ac.uk/ena/data/view/PRJEB25535

In-situ Hi-C sequencing data is available at GEO (Gene Expression Omnibus), accession number GSE119825..

RNA sequencing data is available at GEO (Gene Expression Omnibus), accession number GSE119979.

TCGA database can be accessed in https://portal.gdc.cancer.gov/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were chosen based upon previous standards in the field.
Data exclusions	No data was excluded from analysis.
Replication	All experiments were independently conducted and reproduced multiple times, as described in the figure legends. All P-values were derived from measurements obtained from experiments conducted independently at least three times.
Randomization	No experiments requiring randomization were performed (e.g. animal experiments). Human samples analyzed were chosen due to their amplification of BRAF.
Blinding	Whole genome sequencing was done blinded (using barcoded tubes). Data was not blinded for other experiments as per the standards in the field.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: Aurora B (Abcam, ab2254), Lamin B (Proteintech, 12987-1-AP).

Validation: Validation was done using immunofluorescence showing the expected cell localization and appearance.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s):
HeLa S3 cervical adenocarcinoma cancer cells (Cleveland lab stock)
Colo320DM-GFP colorectal adenocarcinoma containing double minute labeled with GFP (gift from Prof. Noriaki Shimizu)
HT-29 colorectal cancer cells (ATCC® HTB-38™)
293T Human embryonic kidney cells (Cleveland lab stock)
DLD-1 (ATCC)
T-47D (ATCC)

Authentication:
Cells purchased through ATCC were not authenticated.
HeLa cells authenticated using our sequencing and cytogenetic analysis (showing identical genomic events to previous reports)
293T cells authenticated by their culture behavior (loosely attached), morphology under the microscope, and by their efficiency in transfection and virus production.

Mycoplasma contamination:
All cell lines used in this study were confirmed to be free of mycoplasma contamination, by routine (every 6 months) testing and by frequent DAPI staining.

Commonly misidentified lines (See ICLAC register):
No commonly misidentified cell lines were used in this study

Human research participants

Policy information about studies involving human research participants

Population characteristics:
Patient data was previously reported in: Yaeger, R. et al. Mechanisms of Acquired Resistance to BRAF V600E Inhibition in Colon Cancers Converge on RAF Dimerization and Are Sensitive to Its Inhibition. Cancer Res 77, 6513-6523, doi:10.1158/0008-5472.CAN-17-0768 (2017).

Recruitment:
Patient data was previously reported in: Yaeger, R. et al. Mechanisms of Acquired Resistance to BRAF V600E Inhibition in Colon Cancers Converge on RAF Dimerization and Are Sensitive to Its Inhibition. Cancer Res 77, 6513-6523, doi:10.1158/0008-5472.CAN-17-0768 (2017).

Ethics oversight:
Memorial Sloan Kettering (MSK) Cancer Center Institutional Review Board/Privacy Board biospecimen protocol 19-323 was used to analyze samples. Patients provided consent and signed MSK's tissue banking protocol (06-107). Samples were assigned a sample number with no patient identifying information before submission for DNA extraction and genomic analysis. No one performing genomic analysis had access to the patients' identifying information.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Non synchronized HeLa cells, or mitotic HeLa cells (obtained by mitotic shake-off following 4 hours treatment with 0.1ug/ml nocodazole) were ethanol fixed, stained with 10µg/ml propidium iodide, and treated with 50µg/ml RNase A.
Instrument	BD LSR II instrument (BD Biosciences)
Software	BD cellquest (version 5.1)
Cell population abundance	10,000 cells were analyzed. Non synchronized cells contained 20% G2/M cells. Mitotic enriched samples contained 75-88% G2/M cells.
Gating strategy	Dead cells were excluded using SSC/FSC gating. G1, S, and G2/M populations were gated using a histogram plot as shown in the figures.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.