Article

MMEJ repair of breaks at TA repeats maintains ecDNA and cancer fitness

https://doi.org/10.1038/s41586-026-11048-8

Received: 30 June 2025

Accepted: 14 August 2026

Published online: 23 September 2026

Open access

Check for updates

David Billing[1,2,14], Monica Selvaraj[1,14], Megan E. Kelley[1], Gabriel Matos-Rodrigues[3,11], Benton Bailey[1], Finnja Becker[4,5], Gang Zhen[3], Matthew A. Myers[6,7], Alessandra Brambati[1,12], Oluchi Ezekwenna[1], Ashley Nichols[1], Yi-Zhen Jiang[1], Lucia Wang[1], Shih-Chun Wang[1], Marie-Claude Mathieu[8], Davide Pradella[9,13], Hugo Poirier[8], Elias Rodriguez-Fos[4,5], Stephen J. Morris[8], Michal Zimmermann[8], Elsa Callen[3], Michael Zinda[10], Andrea Ventura[9], John Maciejowski[1], Andrew McPherson[6,7], Sohrab P. Shah[6,7], Anton George Henssen[4,5], André Nussenzweig[3] & Agnel Sfeir[1]✉

Extrachromosomal DNA (ecDNA) comprises megabase-sized circular DNA elements that frequently carry oncogene amplifications, driving aggressive tumour phenotypes, therapeutic resistance and poor clinical outcomes across many cancers[1–4]. Although ecDNA is thought to arise from canonical double-strand break repair, the pathways that maintain it remain unclear. Here we show that inhibition of microhomology-mediated end joining, but not non-homologous end joining or homologous recombination, selectively depletes ecDNA, induces ecDNA-specific damage and promotes its sequestration into micronuclei, compromising the fitness of cancer cells that depend on ecDNA-driven oncogene amplification. Mechanistically, TA-rich loci on ecDNA are hotspots for DNA damage and breakage[5,6]. The DNA translocase FANCM suppresses break formation at these sites, while breaks that escape FANCM surveillance are cleaved by ERCC1–ERCC4 and channelled into microhomology-mediated end joining for repair. Single-cell whole-genome sequencing shows that disrupting FANCM or polymerase θ (Polθ) in COLO320DM cells causes structural instability characterized by deletions and small duplications, with breakpoints enriched at TA-rich regions. This fragility is recapitulated in human tumours, in which ecDNA rearrangements are enriched at TA repeats. Collectively, our findings reveal TA repeat fragility as an intrinsic vulnerability of circular DNA and identify Polθ inhibition as a promising strategy to potentially destabilize ecDNA and sensitize ecDNA-driven tumours to therapeutic intervention.

ecDNA comprises circular, megabase-sized DNA elements that frequently carry oncogene amplifications and are found in approximately 17% of all human cancers[1]. Because ecDNA segregates randomly during cell division, it drives substantial intratumoural heterogeneity[2,3]. Serving as hubs for high-level gene expression and rapid genomic evolution, ecDNAs confer an adaptive advantage to cancer cells under therapeutic and environmental pressure. Their presence is associated with aggressive tumour behaviour, resistance to targeted therapies and poor clinical outcomes[4].

Mechanistic studies have established non-homologous end joining (NHEJ) as a central pathway driving the formation of ecDNA, particularly following chromothripsis, in which it ligates shattered chromosomal fragments into circular, oncogenic elements[7,8]. CRISPR–Cas9-based screens and genetic studies have further emphasized the critical role of core NHEJ factors, including *LIG4*, as well as homologous recombination (HR) components such as *BRCA1*, in catalysing ecDNA formation[9,10]. Yet how these elements are maintained remains largely unknown. ecDNAs are subject to persistent replication stress from disorganized replication and increased transcription, often exceeding that of linear chromosome amplification[11]. This drives transcription–replication conflicts and the accumulation of DNA breaks during S phase[12]. Exogenous DNA damage induced by hydroxyurea or ionizing radiation further exacerbates ecDNA instability, promoting its sequestration into micronuclei and subsequent elimination or reintegration into chromosomes[13,14]. Despite this vulnerability, ecDNAs are stably maintained and expand in tumours, suggesting the engagement of DNA repair

[1]Molecular Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [3]Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA. [4]Department of Pediatric Oncology and Hematology, Charité–Universitätsmedizin Berlin, Berlin, Germany. [5]Experimental and Clinical Research Center (ECRC), Max Delbrück Center and Charité–Universitätsmedizin Berlin, Berlin, Germany. [6]Halvorsen Center for Computational Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [7]Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [8]Repare Therapeutics, Montréal, Québec, Canada. [9]Cancer Biology and Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [10]Repare Therapeutics, Cambridge, MA, USA. [11]Present address: CNRS-UMR9019, Gustave Roussy, Université Paris-Saclay, Villejuif, France. [12]Present address: Department of Pharmacology, Anschutz Medical Campus, University of Colorado School of Medicine, Aurora, CO, USA. [13]Present address: Italian Institute for Genomic Medicine, Candiolo, Italy. [14]These authors contributed equally: David Billing, Monica Selvaraj. ✉e-mail: sfeira@mskcc.org

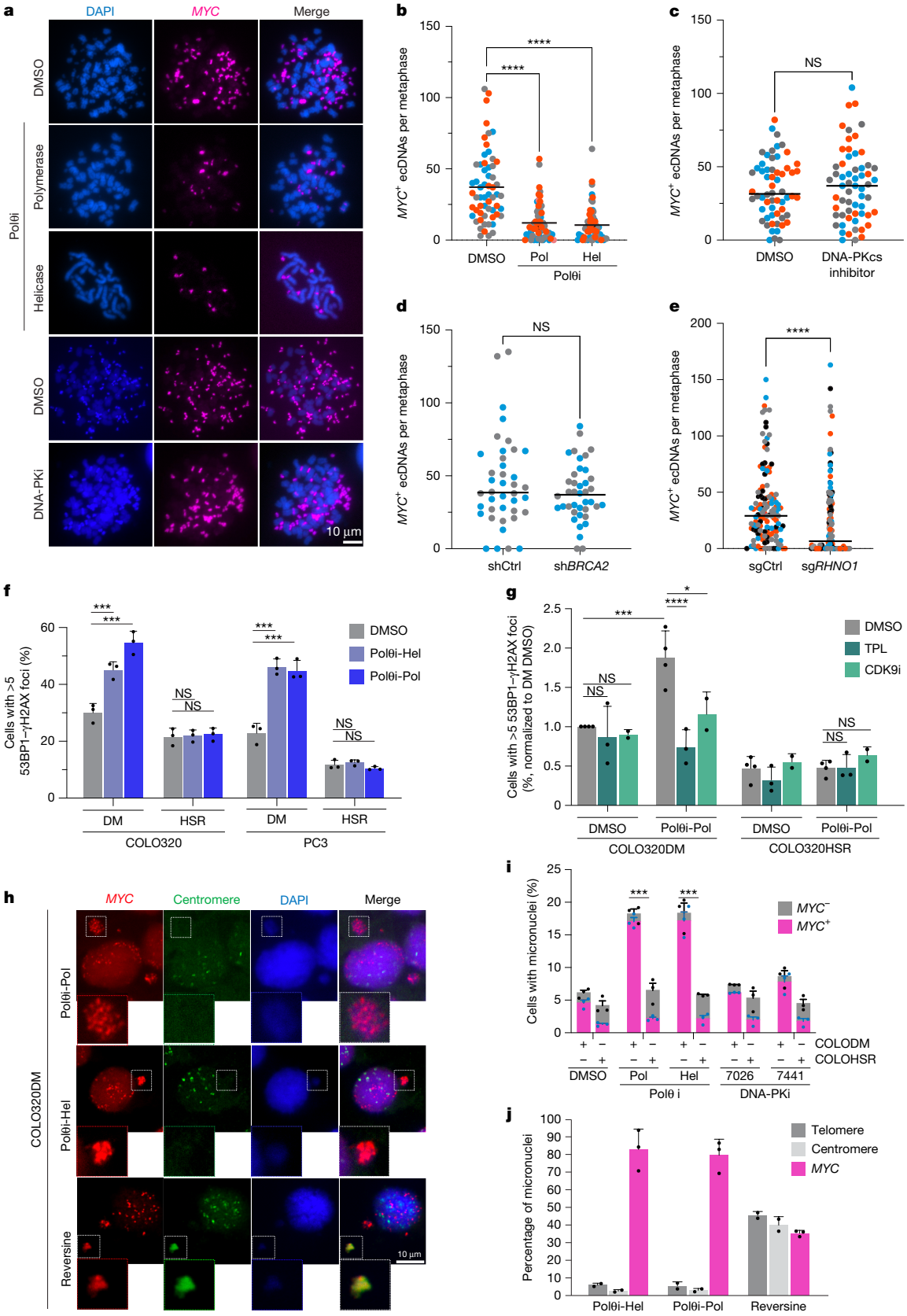


Fig. 1 | See next page for caption.

pathways that can resolve breaks and preserve ecDNA integrity under ongoing stress.

The error-prone microhomology-mediated end joining (MMEJ) pathway of DNA double-strand break (DSB) repair has emerged as a key regulator of extrachromosomal genetic elements[15]. MMEJ facilitates the circularization and replication of retrotransposons such as *HMS-Beagle* in flies and intracisternal A-particle (IAP) in mouse cells, controls the chromosomal integration of T-DNA (transferred DNA)

Fig. 1 | Inhibition of MMEJ, but not NHEJ or HR, selectively destabilizes ecDNA in cancer cells. a, Representative metaphase FISH images of COLO320DM cells, probed for *MYC* after the indicated treatments. DNA-PKi, DNA-PKcs inhibition. **b**, Quantification of *MYC*$^+$ ecDNA per metaphase spread after 1 week of treatment with vehicle, RP-6685 (Polθi-Pol) or RP-2119 (Polθi-Hel). **c**,**d**, Quantification of *MYC*$^+$ ecDNA in cells treated with DNA-PKcs inhibitor (NU7441) (**c**) or following stable knockdown of *BRCA2* using shRNA (sh*BRCA2*) (**d**). shCtrl, control shRNA. **e**, Quantification of *MYC*$^+$ ecDNA in cells nucleofected with sgRNA targeting *RHNO1* (sg*RHNO1*). Data in **b**–**e**: $n = 3$ biological replicates with at least 20 metaphases quantified per replicate; each dot represents the number of ecDNA in an individual metaphase. One-way ANOVA with multiple comparisons in **b**; unpaired two-tailed *t*-test in **c**–**e**. **f**, Cells with five or more co-localized 53BP1–γH2AX foci after Polθ inhibition in COLO320 and PC3 DM and HSR cells versus DMSO. Mean ± s.e.m. $n = 3$ biological replicates, unpaired two-tailed *t*-test. **g**, DNA damage in COLO320DM and COLO320HSR cells treated with DMSO, triptolide (TPL, 1 μM) and CDK9 inhibitor (CDK9i, 200 nM), with or without Polθi-Pol. Percentage of cells with more than five 53BP1–γH2AX foci normalized to DMSO. Mean ± s.d. $n = 3$ biological replicates, one-way ANOVA with multiple comparisons. **h**, Representative *MYC* FISH, centromere FISH and DAPI staining in COLO320DM cells treated with Polθ inhibitors and reversine. **i**, *MYC*$^+$ and *MYC*$^-$ micronuclei in cells treated with DMSO, Polθi-Pol, Polθi-Hel or DNA-PKcs inhibitors (NU7026 and NU7441), as determined by FISH. **j**, Percentage of micronuclei positive for *MYC*, centromere or telomeric signals following Polθ inhibition or reversine treatment. Data in **i**,**j**, $n = 3$ biological replicates. Two-way ANOVA with multiple comparisons. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; NS, not significant ($P \geq 0.05$).

in plants and plasmid DNA in mammalian cells, and contributes to the accumulation of microDNA in cancer cells. MMEJ has also been implicated in the oncogenic integration of the hepatitis B virus into human genomes (reviewed in ref. 15). Finally, MMEJ serves as a backup during ecDNA formation, driving circularization following excision induced by two widely spaced CRISPR–Cas9-generated breaks[9]. These roles underscore the central role of MMEJ in non-chromosomal DNA processing, stabilization and integration. In this study, we demonstrate that MMEJ components, unlike those of canonical NHEJ or HR, are essential for maintaining ecDNA levels in cancer cells. This dependency is underpinned by the accumulation of DNA breaks at TA-rich sequences in circular genomes, which are predisposed to forming stable secondary cruciform structures and may serve as hotspots for rearrangement[16–18]. While FANCM unwinding of secondary structure suppresses TA repeat fragility, structures that escape resolution are cleaved by the ERCC1–ERCC4 endonuclease complex, generating DSBs that are repaired by Polθ-mediated MMEJ. Finally, single-cell whole-genome sequencing and pan-cancer whole-genome sequencing analysis converge on the same conclusion: TA repeat fragile sites are recurrent hotspots for structural rearrangements within ecDNA. Together, our findings reveal that TA repeat instability drives structural rearrangements on ecDNA.

MMEJ maintains ecDNA in cancer cells

To evaluate the role of MMEJ in ecDNA maintenance, we compared a pair of isogenic cell lines derived from the same patient: COLO320DM cells, which carry *MYC*-ecDNA (also known as double minutes (DM)), and the control COLO320HSR cells, which contain a chromosomally integrated *MYC* amplification within a homogeneously staining region (HSR)[19]. The cell lines were exposed to two selective small-molecule inhibitors of Polθ, a key effector of the MMEJ pathway[20,21]. RP-6685 (Polθi-Pol) targets the polymerase domain of Polθ and blocks MMEJ at micromolar concentrations[22]. RP-2119 (Polθi-Hel) is a more potent inhibitor, targeting the helicase domain of Polθ and disrupting its ATPase activity with nanomolar efficacy[23] (Extended Data Fig. 1a–d). Following a 7-day Polθ inhibitor treatment, metaphase spreads and fluorescence in situ hybridization (FISH) staining revealed a significant reduction in *MYC*-ecDNA in COLO320DM cells (Fig. 1a,b), whereas chromosomally amplified *MYC* in COLO320HSR cells remained unaffected (Extended Data Fig. 1e). These findings were replicated across multiple ecDNA-positive cancer cell lines, including PC3-DM prostate cancer (Extended Data Fig. 1f), SNU-16 gastric cancer (Extended Data Fig. 1g) and NCI-H716 colorectal cancer cells (Extended Data Fig. 1h), all of which showed significant reductions in ecDNA copy number upon Polθ inhibition and are in agreement with a previous report implicating MMEJ in ecDNA stability[24]. To isolate the role of MMEJ in ecDNA maintenance from its potential role in formation, we used an established neural stem cell line in which transient Cre expression generates *MYC* ecDNA while permanently excising the chromosomal *MYC* locus, precluding any further de novo ecDNA formation[25]. Polθ inhibitors were applied only after Cre-mediated recombination was complete, ensuring that all ecDNA present was pre-existing and no new copies could arise. This resulted in a marked and progressive depletion of ecDNA over time (Extended Data Fig. 1i–l), establishing that MMEJ is required to sustain the existing ecDNA pool independently of its formation.

To assess the contribution of other DSB repair pathways in maintaining ecDNA, we inhibited NHEJ using a DNA-PKcs inhibitor (1 μM NU7441) and suppressed HR by short hairpin RNA (shRNA)-mediated depletion of *BRCA2* (Extended Data Fig. 1m–o). Analysis of *MYC*-FISH staining on metaphase spreads revealed that neither treatment significantly reduced ecDNA levels (Fig. 1a,c,d and Extended Data Fig. 1f,g,o), supporting a specific and non-redundant role for MMEJ in maintaining ecDNA levels. Given the well-established synthetic lethality between MMEJ and the two canonical DSB repair pathways[20,26], we could not determine whether the residual ecDNA observed following Polθ inhibition was maintained through compensatory repair by NHEJ or HR. To corroborate these findings, we deleted *RHNO1* (coding for RHINO), which promotes Polθ recruitment to DNA lesions[27]. Deletion of *RHNO1* in COLO320DM cells resulted in a significant reduction in ecDNA copy number (Fig. 1e and Extended Data Fig. 1p), consistent with a role for MMEJ in maintaining ecDNA.

Polθ inhibition also led to the accumulation of DNA damage in ecDNA-containing cells. COLO320DM cells exhibited higher baseline levels of DNA damage compared with COLO320HSR cells, as evidenced by increased accumulation of 53BP1 and γH2AX foci[12], and the DNA damage was further increased by Polθ inhibition. We observed similar results in PC3-DM cells, which showed increased DNA damage foci upon Polθ inhibition compared with PC3-HSR cells (Fig. 1f and Extended Data Fig. 2a,b). ecDNAs are known to exhibit high transcriptional activity, leading to frequent replication–transcription conflicts that can generate DNA breaks[12]. Pharmacological inhibition of transcription using triptolide or a CDK9 inhibitor rescued the elevated DNA damage observed upon Polθ inhibition, implicating MMEJ in resolving damage inflicted by transcription (Fig. 1g and Extended Data Fig. 2c–f). To confirm that the detected DNA damage occurred specifically at ecDNA loci, we performed combined 53BP1 immunofluorescence and *MYC*-FISH staining, which revealed significant colocalization of DNA damage foci with ecDNA (Extended Data Fig. 2g,h), indicating that Polθ inhibition induces DNA damage specifically at extrachromosomal elements.

Polθ inhibition sequesters ecDNA in micronuclei

Faithful inheritance of ecDNA during cell division is driven by its tethering to chromosomal DNA during mitosis, a mechanism that ensures co-segregation and inheritance in daughter cells. Specifically, inhibition of BRD4 with JQ1 or loss of the long non-coding RNA PVT1 triggered extensive, acute ecDNA untethering within 6–24 h of perturbation[3]. More broadly, CpG-rich promoters have been shown to act as retention elements that facilitate ecDNA tethering[28]. To investigate whether MMEJ inhibition disrupts ecDNA tethering, we performed live-cell imaging in

COLO320DM cells with TetO-tagged ecDNA and co-expressing TetR–eGFP and H2B–mCherry[3,29]. Unlike BRD4 inhibition, which causes rapid and severe ecDNA untethering[3,29], Polθ inhibition had no detectable effect at 48 h and resulted in only a modest increase in ecDNA detachment after 72 h (Extended Data Fig. 3a,b). This suggests that any effect of MMEJ loss on chromosomal tethering is indirect, and likely to be a secondary consequence of accumulated DNA damage and ecDNA instability rather than direct regulation of tethering.

We next tested whether MMEJ prevents ecDNA elimination via micronuclear sequestration, a process previously linked to DNA damage in ecDNA[12,13,30]. Interphase FISH revealed that Polθ inhibition substantially increased *MYC*-positive micronuclei in COLO320DM cells but not in COLO320HSR cells (Fig. 1h,i). Inhibition of DNA-PKcs did not induce micronucleation, consistent with NHEJ being dispensable for ecDNA maintenance (Fig. 1i). By contrast, co-treatment of cells with hydroxyurea, which induces replication stress, and Polθ inhibitors further increased *MYC*-positive micronuclei (Extended Data Fig. 3c,d). Notably, whereas micronuclei in Polθ-inhibited COLO320DM cells were enriched for *MYC*-positive signal, they showed no enrichment for centromeric or telomeric sequences (Fig. 1h–j and Extended Data Fig. 3e,f). These results suggest that ecDNA accumulation in micronuclei is not driven by piggybacking on missegregated chromosomes due to Polθ inhibition but rather reflects selective sequestration of ecDNA.

Context-dependent MMEJ vulnerabilities

Having established that Polθ inhibition reduces ecDNA copy number and leads to the accumulation of DNA damage factors at extrachromosomal elements, we next investigated whether these effects translate into a fitness cost. Despite reducing ecDNA levels and increasing the accumulation of DNA damage, Polθ inhibition alone did not impair COLO320DM proliferation (Fig. 2a), alter cell cycle (Fig. 2b and Extended Data Fig. 4a–d) or confer a competitive disadvantage in co-culture assays (Fig. 2c and Extended Data Fig. 4e,f). These observations suggest that, in the absence of strong selective pressure for high levels of oncogenes or drug-resistance genes encoded on ecDNA, cells can tolerate a reduction in ecDNA abundance without apparent fitness cost. Moreover, the lack of growth inhibition indicates that, although Polθ inhibition elicits detectable DNA damage, this burden does not reach the threshold required to compromise cellular proliferation.

We therefore explored whether increasing the overall DNA damage load might reveal a proliferative defect in cells carrying ecDNA. It has been established that treatment with Chk1 inhibitors leads to increased DNA damage in ecDNA-containing tumour cells and compromised growth[12]. We therefore treated COLO320DM and COLO320HSR cells with Chk1 inhibitors alone or in combination with Polθi-Hel. Whereas Chk1 inhibition alone had a small but notable effect on cellular growth, co-treatment with Polθi-Hel produced a marked increase in DNA damage and a cellular growth defect in COLO320DM but not COLO320HSR cells (Extended Data Fig. 5a–f).

We next tested whether Polθ inhibition could selectively compromise cells whose survival depends on ecDNA-driven gene amplification. To address this, we first utilized a methotrexate (MTX) selection system in which engineered HeLa cells rely on high-copy *DHFR* amplification on ecDNA (HeLa-DM) or HSRs (HeLa-HSR) to withstand MTX-induced cytotoxicity[7] (Fig. 2d). Cells were pre-treated with Polθi-Hel for one week, followed by combined MTX and Polθi treatment for an additional 2 weeks. Metaphase FISH confirmed a marked reduction in *DHFR* ecDNA in HeLa-DM cells (Fig. 2e and Extended Data Fig. 5g). Notably, whereas HeLa-HSR cells continued to proliferate, HeLa-DM cells exhibited a significant growth defect upon combination treatment (Fig. 2f,g), suggesting that MMEJ inhibition selectively compromises cells that depend on ecDNA-driven oncogene amplification for survival under selective pressure. To corroborate these findings in a clinically relevant context of oncogene amplification, we examined SNU-16 gastric cancer cells, which exhibit *FGFR2* amplification on ecDNA. We treated cells with a sub-saturating dose of the FGFR inhibitor (6.25 nM) that had no discernible effect on proliferation as a single agent. Notably, combining this low dose with Polθ inhibition was sufficient to suppress cell growth (Fig. 2h and Extended Data Fig. 5h). Together, these results demonstrate that MMEJ inhibition does not broadly impair proliferation but instead selectively exposes vulnerabilities in cells carrying ecDNA, both when DNA damage is exacerbated and when survival depends on the dosage of ecDNA-encoded oncogenes or drug resistance genes.

END-seq maps ecDNA breaks to TA repeats

To understand the mechanistic basis for the observed dependency of ecDNA maintenance on MMEJ, we considered three distinct but potentially overlapping scenarios: (1) a global rewiring in DNA repair pathway preference favouring MMEJ over HR or NHEJ; (2) locus-specific alterations in repair pathway usage at ecDNA amplifications compared with their chromosomal counterparts; and (3) accumulation of endogenous DNA lesions uniquely within ecDNA that necessitate MMEJ for resolution (Fig. 3a).

To assess whether global DNA repair pathway utilization is altered in ecDNA-positive cells, we applied *AAVS1*-Repair-seq. This sequencing-based assay quantitatively profiles the relative contributions of HR, NHEJ and MMEJ following CRISPR–Cas9-mediated cleavage at the *AAVS1* safe harbour locus (Fig. 3a). In this system, HR events are identified by incorporating a co-transfected donor oligonucleotide bearing a traceable sequence marker. NHEJ is characterized by small insertion–deletion mutations (indels) or blunt-end repair events. Finally, MMEJ is distinguished by specific repair events that exhibit microhomology at junctions, are characterized by deletions or insertions, and depend on Polθ[22,23]. Comparing the spectrum of repair events between COLO320DM and COLO320HSR cells revealed no significant differences in the relative contributions of each repair pathway upon DNA break induction at the *AAVS1* locus (Fig. 3b,c). More specifically, quantification of MMEJ-dependent repair junctions (Extended Data Fig. 6a–c) revealed comparable frequencies in ecDNA-positive and ecDNA-negative cell lines, suggesting that the accumulation of ecDNA does not broadly alter the landscape of DSB repair pathway usage in cells (Fig. 3b,c). Similar results were obtained when a Repair-seq assay was used to compare repair events at the *MYC* locus in COLO320DM and COLO320HSR cells (Fig. 3b,c and Extended Data Fig. 6d). Together, these data argue against both a global shift in repair preference and a locus-specific change in pathway choice as the primary explanation for heightened MMEJ dependence in ecDNA-positive cells.

We therefore considered the third possibility: that ecDNA itself is an endogenous source of DNA lesions dependent on MMEJ for resolution. To test this, we performed END-seq, a high-resolution technique for genome-wide mapping of DSBs[31] in COLO320DM and COLO320HSR cells treated with Polθ inhibitors. Both cell types showed similar read coverage across the regions within the ecDNA locus, including on chromosome 8, a 1.58-Mb region encompassing the amplified *MYC* locus, chromosome 6 (126 kb) and chromosome 13 (168 kb) (Fig. 3d and Extended Data Fig. 6e). Furthermore, CRISPR–Cas9-induced breaks at the *MYC* locus were detected in both lines, validating assay sensitivity (Extended Data Fig. 6f). Of note, cells carrying ecDNA exhibited more breaks at the *MYC* locus following CRISPR-mediated cleavage, possibly reflecting increased chromatin accessibility on the circular DNA, which has been shown to lack canonical chromatin features and higher-order organization[30,32].

Analysis of END-seq reads revealed recurrent endogenous DSBs at multiple discrete loci within the *MYC* ecDNA in COLO320DM cells, whereas the corresponding regions in COLO320HSR cells were less prone to breakage (Fig. 3e and Extended Data Fig. 6g). These breaks were not randomly distributed; motif analysis revealed a pronounced enrichment for TA dinucleotide repeats at ecDNA break sites, a pattern that was absent at breaks mapping to linear chromosomal DNA (Fig. 3e–g

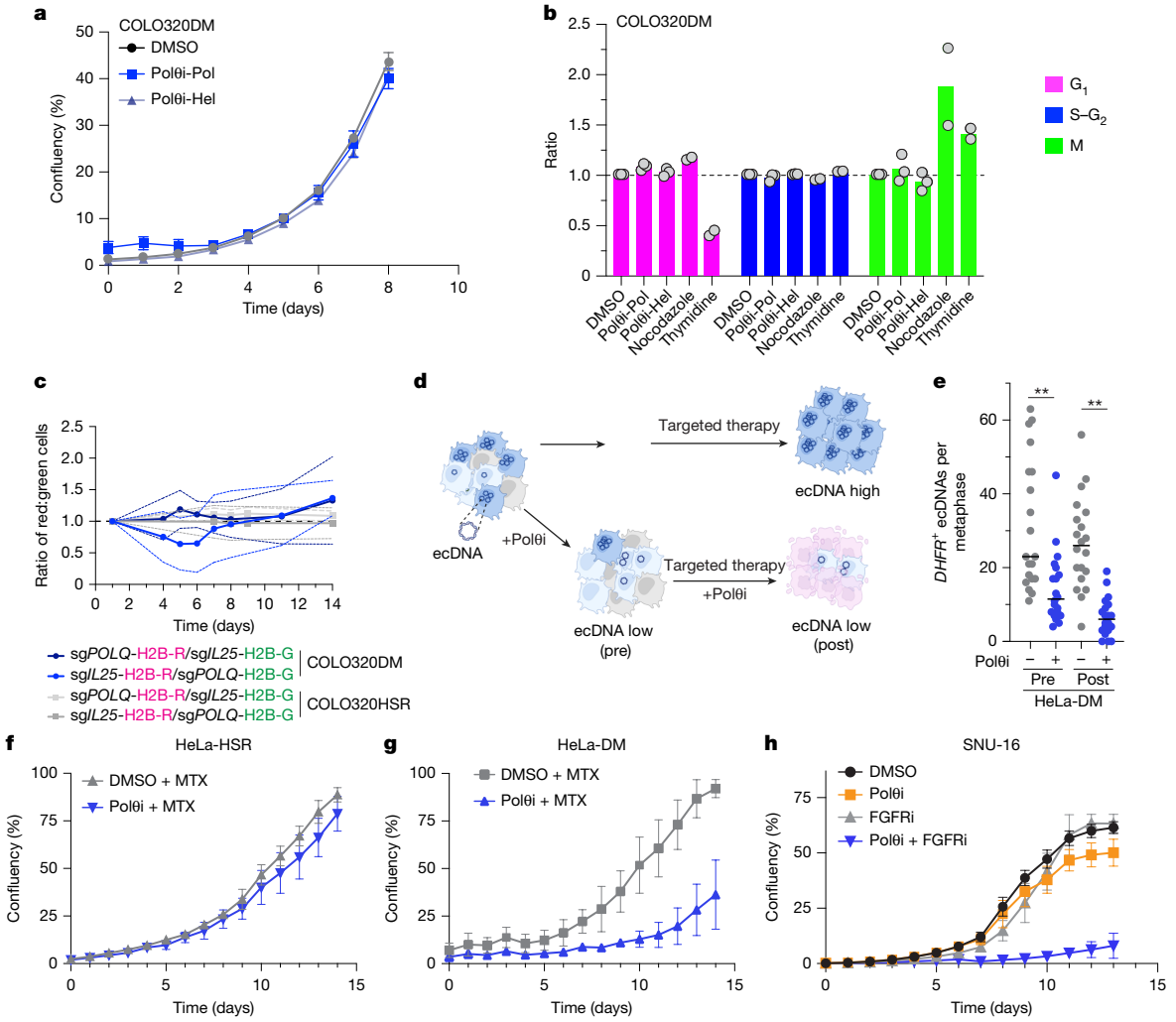


Fig. 2 | Cells that rely on ecDNA-encoded gene dosage are selectively vulnerable to Polθ inhibition. a, Proliferation (IncuCyte confluency over time) of COLO320DM cells treated with vehicle (DMSO), Polθi-Pol or Polθi-Hel. Mean ± s.e.m. of $n = 3$ biological replicates per condition and 3 technical replicates per experiment. **b**, Cell cycle distribution of COLO320DM cells treated with vehicle, Polθ inhibitors or control compounds, by FUCCI (fluorescent ubiquitination-based cell cycle indicator) reporter analysis. Mean ± s.d. $n = 2$–3 biological replicates. Values are normalized to the DMSO control for each phase. **c**, Growth competition assay comparing COLO320DM or COLO320HSR cells after sgRNA targeting of *POLQ* (sg*POLQ*). Plotted as the ratio of red-labelled to green-labelled nuclei over time; mean ± s.d. (dashed line), with $n \geq 3$ biological replicates per condition and 3 technical replicates per experiment. H2B-G, H2B–GFP; H2B-R, H2B–mCherry; sg*IL25*, sgRNA targeting *IL25*. **d**, Schematic of the experimental design in HeLa cells carrying *DHFR* on ecDNA and SNU-16 carrying *FGFR2* on ecDNA. Cells were pre-treated (pre) with Polθ inhibitor before methotrexate or infigratinib (FGFRi), followed by continued MTX with or without Polθi (post) or FGFRi with or without Polθi treatment, to assess ecDNA maintenance and drug resistance. Created in BioRender; Sfeir, A. https://biorender.com/cvwpf9d (2026). **e**, *DHFR*$^+$ ecDNA per metaphase spread in HeLa-DM cells before and after MTX treatment, in the presence or absence of Polθ inhibition. Each dot represents one metaphase; bars denote mean. Ordinary one-way ANOVA with multiple comparisons. **f**, Proliferation of HeLa cells carrying *DHFR* on chromosomal HSRs cultured in MTX with vehicle or Polθ inhibitor. **g**, Proliferation of HeLa cells with *DHFR* amplification on ecDNA (HeLa-DM) cultured in MTX with vehicle or Polθ inhibitor. Data in **f**,**g** represent the mean ± s.d. of $n = 3$ replicates. **h**, Proliferation of SNU-16 cells carrying *MYC* and *FGFR2* amplifications on ecDNA cultured with vehicle, Polθ inhibitor, low-dose FGFR inhibitor or Polθ inhibitor plus FGFRi. Mean ± s.d. of $n = 3$ replicates.

and Extended Data Fig. 6h,i). Specifically, 11 of the 20 END-seq peaks within the chromosome 8 amplicon of COLO320DM cells overlapped with TA repeat tracts spanning at least 6 repeat units, a proportion well above that seen at other genomic loci or in COLO320HSR cells (Fig. 3e–h and Extended Data Fig. 6j). These TA-rich break sites were observed across independent Polθ inhibitor conditions and single guide RNA (sgRNA) treatments. They were consistently detected in COLO320DM but not COLO320HSR cells, indicating that they reflect an intrinsic susceptibility to breakage on *MYC*-containing ecDNA rather than a treatment artefact (Fig. 3e,h and Extended Data Figs. 6j–l and 7). Furthermore, stratification of END-seq signal by TA repeat length revealed a positive correlation between repeat length and DSB burden in COLO320DM cells (Extended Data Fig. 6l). Together, these findings establish TA repeats as hotspots of intrinsic DNA fragility on ecDNA.

FANCM and MMEJ guard ecDNA TA repeats

TA repeats are intrinsically predisposed to adopt non-B DNA conformations, such as cruciform, making them well-established hotspots of genome instability[5]. TA repeats can function as DNA unwinding elements to promote double helix melting during replication[6]. Studies in *Escherichia coli* and yeast have shown that transcription-induced negative supercoiling is sufficient to drive cruciform extrusion at TA-rich sequences[33–35], supporting the notion that the transcriptional hyperactivity of ecDNA may potentiate the structural vulnerability of its TA-rich loci. Consistent with this, END-seq peaks at TA loci 1–3 overlap with the highly transcribed *MYC–PVT1* region of the chromosome 8 ecDNA amplicon (Fig. 3e), possibly linking break formation to active transcription. Mammalian cells have evolved opposing enzymatic

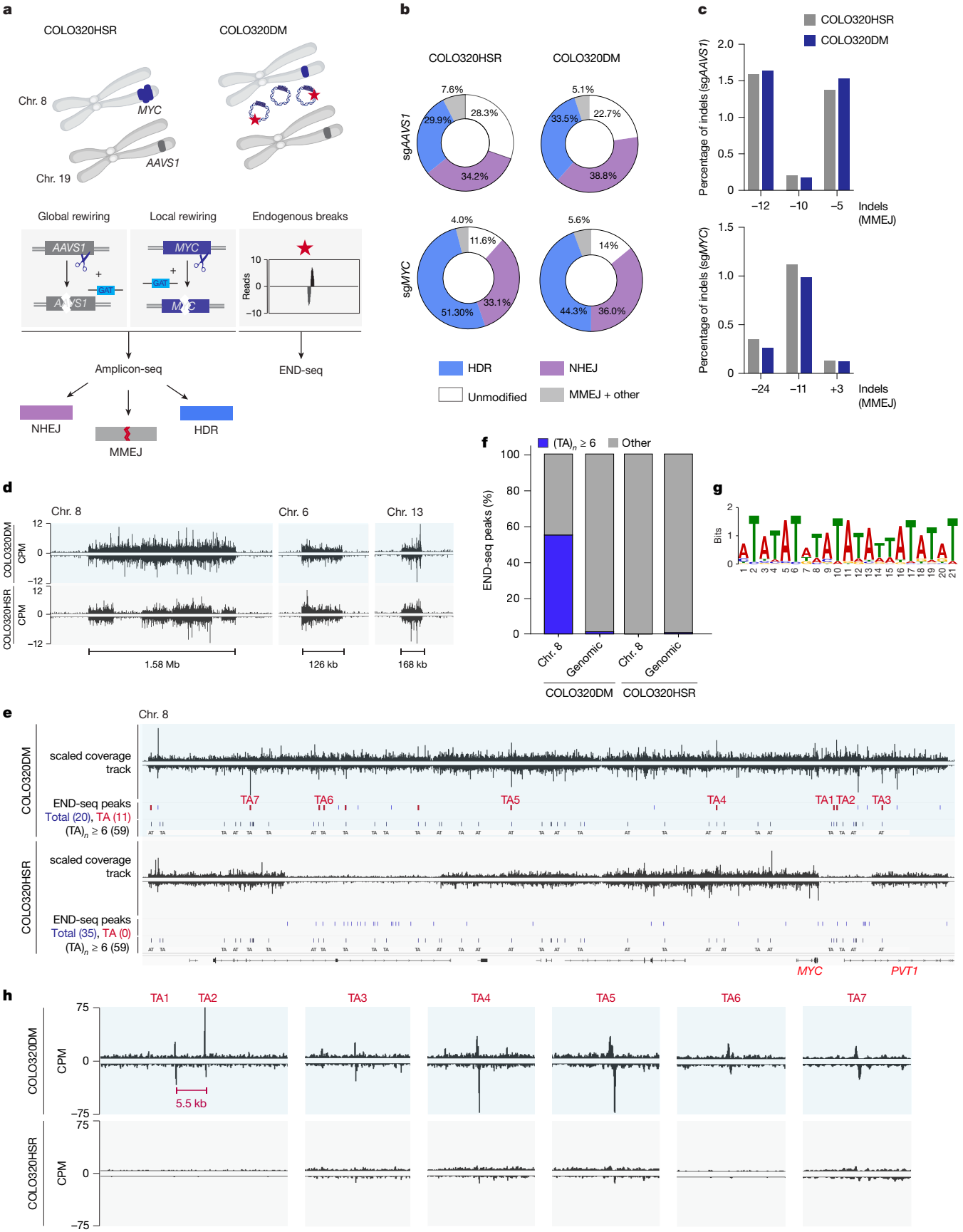


Fig. 3 | See next page for caption.

Fig. 3 | ecDNA accumulates endogenous breaks at TA repeats without global DSB repair rewiring. a, Schematic of the experimental framework dissecting MMEJ dependency in ecDNA-containing cells using isogenic COLO320 models. Three non-mutually exclusive scenarios were tested: (1) a global shift in repair pathway preference toward MMEJ over HR or NHEJ; (2) locus-specific rewiring of repair at amplified *MYC* depending on chromosomal or ecDNA context; and (3) accumulation of ecDNA-specific endogenous breaks (red stars) requiring MMEJ. DSBs were introduced at *AAVS1* and *MYC* by CRISPR–Cas9 and outcomes quantified by amplicon sequencing; END-seq mapped spontaneous lesions genome-wide. Created in BioRender; Sfeir, A. https://biorender.com/wxtnj4o (2026). Chr., chromosome; HDR, homology-directed repair. **b**, Pie charts of HR, NHEJ, unedited and MMEJ outcomes at the *AAVS1* and *MYC* target site in COLO320DM and COLO320HSR cells. **c**, Quantification of representative MMEJ-associated indels at *AAVS1* (top) and *MYC* (bottom) in COLO320HSR and COLO320DM cells. **d**, Genome browser views of normalized END-seq signal in COLO320DM and COLO320HSR cells across the amplified regions on chromosomes 8, 6 and 13 (hg19). The highly transcribed *MYC–PVT1* locus maps to chromosome 8; plus- and minus-strand reads are shown in grey, and each amplicon size is indicated. CPM, counts per million. **e**, Scaled END-seq coverage across the amplified chromosome 8 region in COLO320DM and COLO320HSR cells, with peaks below the tracks (blue). Peaks enriched for TA repeats ($(TA)_n \geq 6$) are shown in red (11 loci in COLO320DM, 0 loci in COLO320HSR); the bottom track marks all 59 TA repeat elements ($(TA)_n \geq 6$) in the chromosome 8 amplicon. **f**, Fraction of END-seq peaks overlapping TA repeats ($(TA)_n \geq 6$) versus other sequences at the chromosome 8 amplicon in COLO320DM and COLO320HSR cells. **g**, Motif enrichment at END-seq break sites in COLO320DM cells. **h**, Zoomed-in browser views of END-seq peaks (from **e**) at seven representative TA repeat-containing loci (TA1–TA7) in COLO320DM and COLO320HSR cells, with the distance between TA1 and TA2 indicated. TA1, TA2 and TA3 are in the *MYC–PVT1* region, which is highly transcribed.

activities to manage secondary structures forming at TA-rich regions: translocases such as FANCM and SMARCAL1, and the helicase WRN, act to suppress cruciform accumulation, while the structure-specific endonucleases ERCC1–ERCC4 and MUS81 resolve them through cleavage, generating DSBs in the process[5,36–38]. The balance between these activities, therefore, determines whether cruciform structures are safely resolved or converted into DNA breaks that must be repaired.

To dissect the relative contributions of these factors, we targeted each genetically in COLO320DM and COLO320HSR cells. Loss of FANCM or SMARCAL1, but not WRN, elevated DNA damage, an effect further compounded by Polθ inhibition, with a similar trend observed upon co-depletion of RHINO and FANCM. Conversely, depletion of ERCC1 or ERCC4 rescued the DNA damage induced by Polθ inhibition (Fig. 4a–c and Extended Data Fig. 8a–d). Our data also showed that ecDNA-positive micronucleation induced by Polθ inhibition was further increased by FANCM depletion and suppressed by ERCC1 or ERCC4 depletion (Fig. 4d,e). Consistent with damage occurring specifically on ecDNA, FANCM depletion increased colocalization of 53BP1 with *MYC* (Extended Data Fig. 8e) and reduced ecDNA expression (Extended Data Fig. 8f), supporting a model in which nucleolytic cleavage at TA cruciform drives ecDNA destabilization. Critically, transcriptional inhibition with triptolide or CDK9i rescued the DNA damage caused by FANCM loss (Fig. 4f), linking the transcriptional hyperactivity of ecDNA to its structural vulnerability at TA repeat fragile sites.

To confirm that these perturbations converge specifically on TA repeat-containing regions at ecDNA, we performed END-seq in COLO320DM and COLO320HSR cells depleted of FANCM or treated with Polθ inhibitors and normalized the TA repeat-associated END-seq signal to total read coverage across the amplified *MYC* locus. Both FANCM depletion and Polθ inhibition independently increased DSBs specifically at TA repeat-containing loci within ecDNA, and their combination produced an additive increase in break accumulation at these sites. By contrast, no such enrichment was observed in COLO320HSR cells (Fig. 4g). Together, our findings posit a two-tiered mechanism that protects the highly transcribed ecDNA from the inherent fragility of TA repeat-containing sequences: FANCM resolves transcription-induced DNA secondary structures, likely to be cruciforms, before they can be cleaved by ERCC1–ERCC4, while Polθ-mediated MMEJ repairs the DSBs that escape this first line of defence (Fig. 4h). The functional importance of both tiers is underscored by the observation that, whereas FANCM loss and Polθ inhibition were individually tolerated, the combination selectively impaired proliferation in COLO320DM, but not COLO320HSR, cells (Extended Data Fig. 9).

TA repeat fragility rearranges ecDNA

Computational modelling and reconstruction of ecDNA structures from long-read sequencing have revealed that ecDNAs are highly complex and dynamically evolving, often comprising heterogeneous and co-existing species with extensive rearrangements and overlapping genomic footprints[39,40]. To functionally assess whether TA repeat-associated fragility contributed to structural rearrangements in ecDNA, as has been suggested in linear DNA[18], we performed single-cell whole-genome sequencing using direct library preparation (DLP+) in COLO320DM cells following long-term depletion with CRISPR-mediated knockout (sg*FANCM*) or treatment with Polθi-Hel[40]. Structural variants on ecDNA, including deletions, duplications and inversions, were called from single-cell DLP+ sequencing data, and sample-specific structural variants were compared across parental ($T = 0$), control (DMSO; $T = 2$ months), sg*FANCM* ($T = 1$ month) and Polθi-Hel ($T = 2$ months) conditions (Fig. 5a). Both sg*FANCM* and Polθi-Hel treatments shifted the deletion length distribution toward larger deletions in ecDNA relative to parental and control cells. At the same time, sg*FANCM* additionally drove the emergence of small duplications (less than 5 kb in length) (Fig. 5b and Extended Data Fig. 10a). To determine whether these structural variants arose at sites of recurrent ecDNA breakage, we examined their distribution relative to four regions with the highest END-seq signal on the chromosome 8 ecDNA amplicon, in which the strongest peaks map to TA repeats (Extended Data Fig. 10b). In both sg*FANCM* and Polθi-Hel-treated cells, large deletions showed significant overlap with these four END-seq-rich regions (Fig. 5c,d and Extended Data Fig. 10c). Similarly, small duplications observed in sg*FANCM*-treated cells were enriched within these fragile regions (Fig. 5e,f). The near absence of small duplications following Polθ inhibition alone suggests that when FANCM surveillance fails, MMEJ resolves the resulting breaks through a repair mode that inherently favours the templated insertion of short duplications.

To investigate whether these observations extend to human tumours more broadly, we analysed the breakpoint junctions of focal amplifications from publicly available ecDNA reconstructions in the Amplicon Repository, which integrates AmpliconArchitect and AmpliconClassifier data across diverse cancer genome datasets[41,42]. Focal amplification breakpoint junctions from both ecDNA and non-ecDNA amplicons were annotated for overlap with TA repeats. This analysis uncovered a statistically significant enrichment of TA repeats at ecDNA amplification breakpoints relative to linear focal amplification breakpoints (Fig. 5g–i and Extended Data Fig. 10d–f), supporting the model whereby TA-rich sequences are more prone to structural fragility in the context of circular DNA.

Discussion

ecDNA has emerged as a powerful engine of tumour evolution[43], yet the mechanisms that enable its persistence despite its unique structural vulnerabilities remain incompletely understood. Here, we demonstrate that ecDNA maintenance reflects an active and specialized dependence on MMEJ. Although canonical repair by NHEJ and HR is essential for ecDNA formation[7,10] and for resolving Cas9-induced breaks on ecDNA,

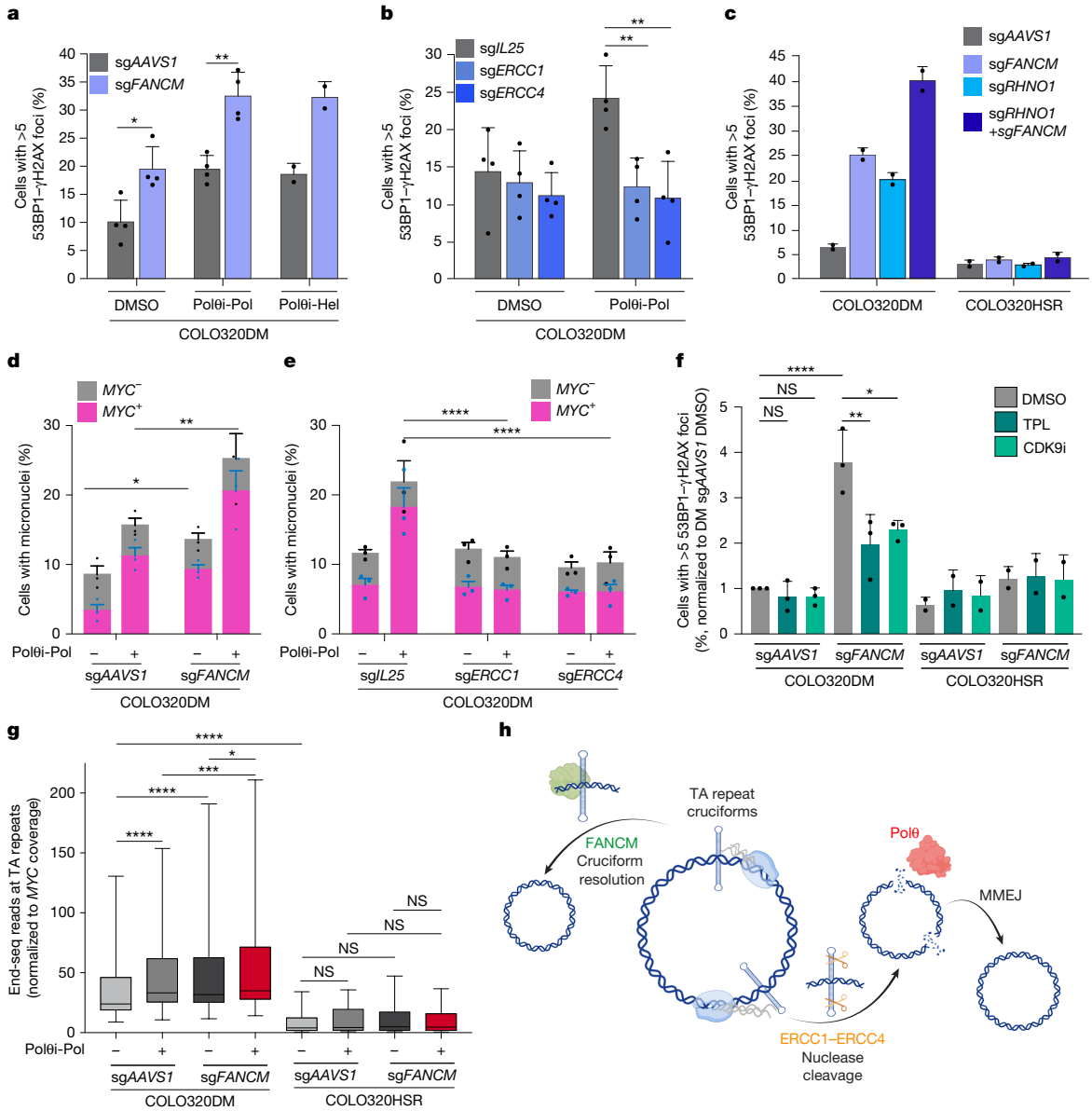


Fig. 4 | FANCM and Polθ-mediated MMEJ cooperate to protect ecDNA from transcription-induced damage at TA repeats. a, DNA damage in COLO320DM cells following CRISPR–Cas9-mediated targeting of *FANCM* or *AAVS1* (control) and treatment with DMSO, Polθi-Hel and Polθi-Pol. Results are reported as percentage of cells containing more than five co-localized 53BP1–γH2AX foci per nucleus. Mean ± s.d. of $n = 4$ biological replicates (more than 250 cells per sample). **b**, DNA damage foci in COLO320DM cells after deletion of *ERCC1*, *ERCC4* or *IL25* (control), with or without Polθi-Pol treatment. Mean ± s.d. of $n = 4$ biological replicates (more than 200 cells per sample). **c**, Percentage of cells expressing sgRNAs targeting *FANCM*, *RHNO1*, or both with more than five co-localized 53BP1–γH2AX foci, versus sg*AAVS1* controls. Mean ± s.d. of $n = 2$ biological replicates per condition (more than 250 cells per sample). **d**,**e**, *MYC*⁺ and *MYC*⁻ micronuclei in COLO320DM cells after sgRNA-targeting of *FANCM* (**d**) and *ERCC1* or *ERCC4* (**e**) versus controls, with or without Polθi-Pol. Mean ± s.e.m. of $n = 3$ biological replicates (more than 250 cells per sample). Two-tailed *t*-tests in **a**–**e**. **f**, DNA damage in COLO320DM and COLO320HSR cells treated with DMSO, triptolide or CDK9i, with and without FANCM depletion. Mean ± s.d. of $n = 3$ biological replicates. One-way ANOVA with multiple comparisons. **g**, END-seq signal at TA repeat loci within the chromosome 8 ecDNA amplicon in COLO320DM and COLO320HSR cells after sg*AAVS1* or sg*FANCM* treatment with or without Polθi-Pol. CPM were normalized to total *MYC* locus coverage. Boxes show 25th–75th percentiles, whiskers extend to minimum and maximum values and the centre line denotes the median of $n = 1$ biological replicate with 19 TA repeat loci breaks. Paired two-tailed *t*-tests with multiple comparisons. **h**, Model of two-tiered protection against TA repeat fragility on ecDNA. Transcription-induced negative supercoiling drives cruciform extrusion at TA repeats. FANCM resolves these structures, but when resolution fails, ERCC1–ERCC4 cleavage generates DSBs that Polθ-dependent MMEJ repairs to maintain ecDNA integrity. Created in BioRender; Sfeir, A. https://biorender.com/lrat2x1 (2026).

our findings show that sustaining ecDNA copy number depends critically on MMEJ activity. This reframes ecDNA as a distinct genomic entity whose structural and transcriptional challenges diverge fundamentally from those of linear chromosomal DNA. Our data support a model in which rampant transcriptional activity in the context of ecDNA generates unwinding at TA-rich loci and the formation of secondary DNA structures, such as cruciforms, that serve as recurrent sites of breakage and rely on MMEJ for repair (Fig. 4h). While transcription-driven cruciform extrusion at TA repeats is well-documented in plasmids, the organization of ecDNA into topological domains and the accumulation of R-loops at its highly transcribed loci can trap negative supercoiling locally[44], even within these megabase-scale molecules, providing sufficient torsional stress to drive cruciform extrusion and subsequent cleavage at TA-rich sequences.

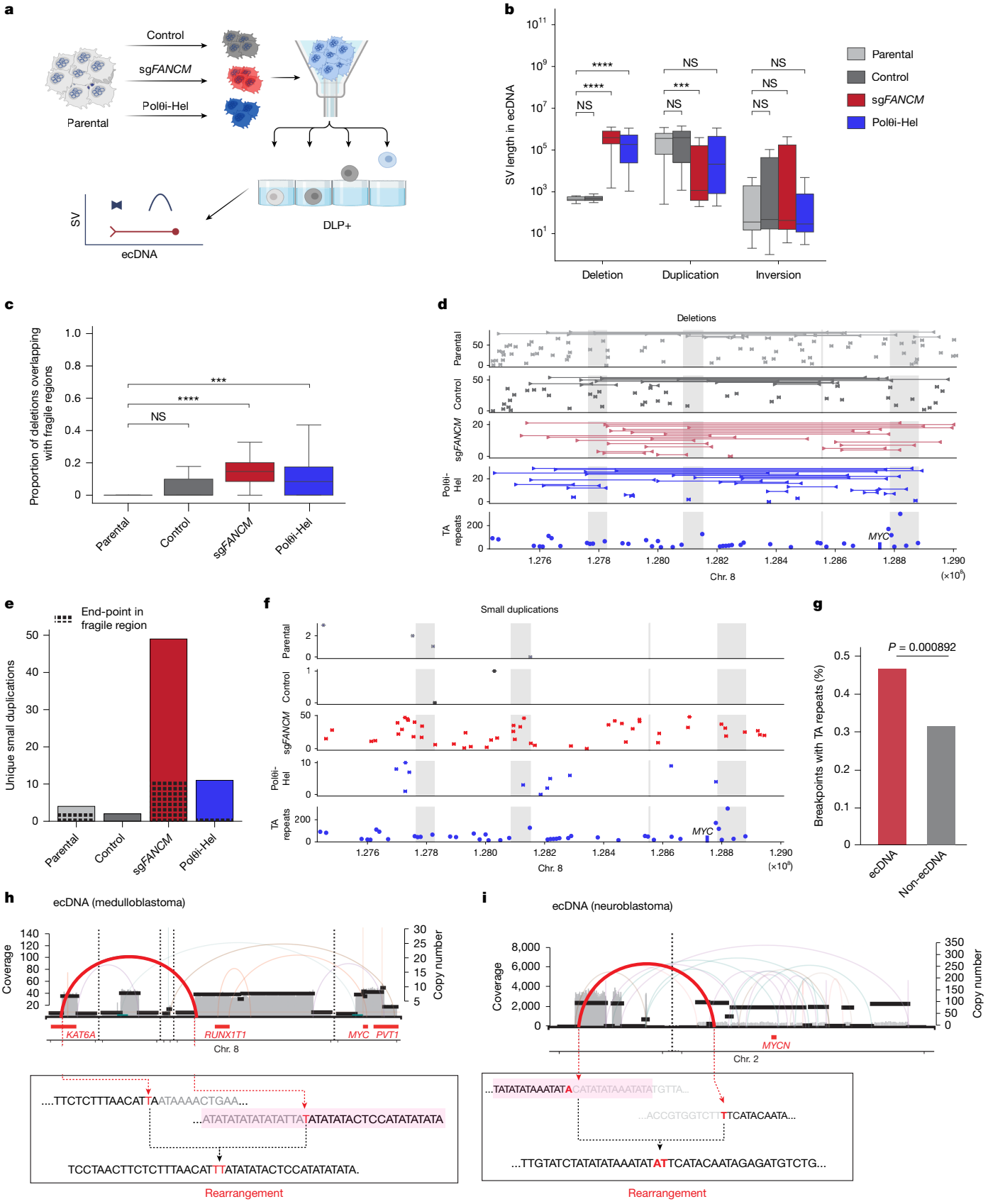


Fig. 5 | See next page for caption.

HR and NHEJ do not resolve breaks at TA-rich regions in ecDNA, leaving MMEJ as the sole competent pathway. The cleavage of a cruciform might generate end structures that are refractory to Ku loading on DNA ends[45], or prevent the long-range resection necessary for HR, while the TA repeat itself provides the microhomology that licenses MMEJ. An alternative model, based on the central role of RHINO in ecDNA

Fig. 5 | TA repeat fragility drives structural rearrangements on ecDNA. **a**, COLO320DM cells profiled by single-cell whole-genome sequencing using DLP+ after CRISPR depletion of FANCM (sg*FANCM*, 1 month) or Polθ helicase inhibition (Polθi-Hel, 2 months); structural variants (SVs) were called alongside parental (T0) and control conditions. Created in BioRender; Sfeir, A. https://biorender.com/d9fh3n6 (2026). **b**, Sample-specific structural variant lengths (deletions, duplications and inversions) across parental, control, sg*FANCM* and Polθi-Hel conditions. **c**, Proportion of sample-specific deletions overlapping END-seq fragile regions on the chromosome 8 ecDNA amplicon across T0, DMSO, Polθi-Hel and sg*FANCM* conditions. In **b**,**c**, boxes delineate 25th and 75th percentiles, whiskers extend to points within 1.5× interquartile range of the lower and upper quartile and the centre line indicates median; Mann–Whitney *U* tests (one-sided). **d**, Genomic map of individual deletion sizes across the chromosome 8 ecDNA amplicon in parental, control, sg*FANCM* and Polθi-Hel treated cells. Grey regions mark four END-seq-enriched and TA-rich fragile regions (Extended Data Fig. 10b). Bottom, TA-rich loci are highlighted in blue. **e**, Number of unique small duplications per condition. Hatched bars indicate duplications with at least one end-point in a fragile region. **f**, Genomic map of small duplications across the chromosome 8 ecDNA amplicon. Shaded regions mark END-seq fragile sites (Extended Data Fig. 10b). Small duplications are enriched with sg*FANCM*, but not in control or Polθi-Hel-treated cells. **g**, Fraction of focal amplification junctions overlapping TA repeats in ecDNA versus non-ecDNA regions across human tumour datasets (PCAWG (Pancancer Analysis of Whole Genomes), Hartwig (Hartwig Medical Foundation), PedPanCan (tumour samples from St Jude and PBTA), GLASS (Glioma Longitudinal Analysis) and CUGA (Chinese Urothelial Carcinoma Genome Atlas)) ($P = 0.000892$, Pearson's chi-squared test). **h**, AmpliconArchitect reconstruction of a *MYC*-containing ecDNA from a medulloblastoma, with read coverage above and arcs marking structural rearrangements. The boxed junction (red arc) shows the reference sequence flanking the breakpoint on linear chromosome 8 (top) and the rearranged ecDNA junction (bottom), with the TA repeat highlighted in pink at the breakpoint. **i**, AmpliconArchitect reconstruction of a *MYC*-containing ecDNA from a neuroblastoma, with a rearrangement junction overlapping a TA repeat highlighted in pink.

maintenance and the suppression of HR and NHEJ during mitosis[27,46], is that these breaks might remain unresolved until mitosis. We speculate that TA-rich breaks could persist into the M phase, when MMEJ becomes the only available means of restoring ecDNA integrity, although we cannot rule out a contribution of MMEJ during $S–G_2$.

The central role of MMEJ in ecDNA maintenance extends its emerging function in stabilizing and propagating extrachromosomal genomes more broadly, including viral episomes and transposable elements[15]. Although the specific mechanisms of propagation may differ across these distinct DNA species, a consistent principle appears to hold: circular or episomal DNA elements rely on MMEJ to persist within the host genome. Our findings further implicate DNA secondary structure as a central driver of ecDNA fragility. While TA repeats are recognized as sources of translocations on linear chromosomes[18], their instability is profoundly amplified on ecDNA, where the combination of exceptionally high transcription and closed-circular topology creates conditions uniquely permissive for cruciform extrusion, converting otherwise stable loci into recurrent hotspots for DNA breakage and rearrangement. We did not observe increased breakage at TA repeats on HSR amplifications, which have also been observed to have high levels of transcription, suggesting that the circular topology of ecDNA amplifications contributes to the TA repeat-associated fragility. A recent study showed that ecDNA replication is associated with ATM-mediated DNA damage responses involving TOP1 and TOP2B, with MMEJ contributing to subsequent repair[24]. Together with our findings, these observations support a model in which ecDNA is subject to multiple, potentially complementary sources of DNA damage that converge on a shared reliance on MMEJ for resolution. Here we identify TA repeat-associated secondary structures as an intrinsic, sequence-encoded source of fragility and delineate the mechanism by which the resulting lesions are repaired.

Specifically, we uncover a two-tiered protection mechanism that buffers this instability. FANCM, known for its ability to remodel branched DNA structures[47], acts upstream to suppress or resolve transcription-induced cruciform structures at TA-rich sequences, thereby preventing their cleavage by the ERCC1–ERCC4 endonuclease complex and limiting DNA damage (Fig. 4). Breaks that escape this surveillance are subsequently channelled into Polθ-mediated MMEJ for repair. Consistent with this model, analysis of single-cell whole-genome sequencing data alongside large-scale whole-genome sequencing datasets from human tumours reveals that rearrangement breakpoints on ecDNA are significantly enriched at TA dinucleotide repeats (Fig. 5). This enrichment suggests that these sequence elements not only represent intrinsic structural vulnerabilities but also serve as recurrent substrates for the ongoing rearrangements that drive ecDNA remodelling and diversification in vivo.

From an evolutionary perspective, the TA fragility that we describe may not simply be a liability; it may instead facilitate ecDNA plasticity. Sequence analyses of ecDNA suggest an accumulation of DNA repair signatures after ecDNA formation, particularly SBS3, which has been associated with HR deficiency[48], suggesting that ecDNA exists in a state of ongoing repair activity. The TA repeat-associated breaks that necessitate MMEJ repair may also provide hotspots for ongoing structural rearrangement, enabling rapid remodelling of ecDNA architecture. In this light, MMEJ may have a dual role: preserving ecDNA integrity in the short term while enabling structural diversification over longer evolutionary timescales. Finally, our observations carry direct therapeutic implications. Polθ inhibition reduces ecDNA levels, but its effect on cell fitness is selectively evident when cells depend on ecDNA-encoded oncogenes, under targeted therapy, or in the context of acquired drug resistance. This context dependence suggests a rational combination strategy: oncogene-directed therapies that heighten reliance on ecDNA for survival may synergize with MMEJ inhibition to destabilize these elements and overcome resistance.

Online content

Any methods, additional references, Nature Portfolio 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/s41586-026-11048-8.

1. Wu, S., Bafna, V., Chang, H. Y. & Mischel, P. S. Extrachromosomal DNA: an emerging hallmark in human cancer. *Annu. Rev. Pathol.* **17**, 367–386 https://doi.org/10.1146/annurev-pathmechdis-051821-114223 (2022).
2. Hung, K. L. et al. Coordinated inheritance of extrachromosomal DNAs in cancer cells. *Nature* **635**, 201–209 https://doi.org/10.1038/s41586-024-07861-8 (2024).
3. Nichols, A. et al. Chromosomal tethering and mitotic transcription promote ecDNA nuclear inheritance. *Mol. Cell* **85**, 2839–2853.e2838 https://doi.org/10.1016/j.molcel.2025.06.013 (2025).
4. Luebeck, J. et al. Extrachromosomal DNA in the cancerous transformation of Barrett's oesophagus. *Nature* **616**, 798–805 https://doi.org/10.1038/s41586-023-05937-5 (2023).
5. van Wietmarschen, N. et al. Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* **586**, 292–298 https://doi.org/10.1038/s41586-020-2769-8 (2020).
6. Kowalski, D. & Eddy, M. J. The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *Escherichia coli* replication origin. *EMBO J.* **8**, 4335–4344 https://doi.org/10.1002/j.1460-2075.1989.tb08620.x (1989).
7. Shoshani, O. et al. Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**, 137–141 https://doi.org/10.1038/s41586-020-03064-z (2021).
8. Engel, J. L. et al. The Fanconi anemia pathway induces chromothripsis and ecDNA-driven cancer drug resistance. *Cell* **187**, 6055–6070.e6022 https://doi.org/10.1016/j.cell.2024.08.001 (2024).
9. Rose, J. C. et al. Disparate pathways for extrachromosomal DNA biogenesis and genomic DNA repair. *Cancer Discov.* **15**, 69–82 https://doi.org/10.1158/2159-8290.Cd-23-1117 (2025).
10. Chung, O. W. et al. BRCA1-A and LIG4 complexes mediate ecDNA biogenesis and cancer drug resistance. *Proc. Natl Acad. Sci. USA* **123** e2530443123 https://doi.org/10.1073/pnas.2530443123 (2026).
11. Jaworski, J. J. et al. ecDNA replication is disorganized and vulnerable to replication stress. *Nucleic Acids Res.* **53**, gkaf711 https://doi.org/10.1093/nar/gkaf711 (2025).
12. Tang, J. et al. Enhancing transcription–replication conflict targets ecDNA-positive cancers. *Nature* **635**, 210–218 https://doi.org/10.1101/2024.03.29.586681 (2024).

13. Oobatake, Y. & Shimizu, N. Double-strand breakage in the extrachromosomal double minutes triggers their aggregation in the nucleus, micronucleation, and morphological transformation. *Genes Chromosomes Cancer* **59**, 133–143 https://doi.org/10.1002/gcc.22810 (2020).
14. Schoenlein, P. V. et al. Radiation therapy depletes extrachromosomally amplified drug resistance genes and oncogenes from tumor cells via micronuclear capture of episomes and double minute chromosomes. *Int. J. Radiat. Oncol. Biol. Phys.* **55**, 1051–1065 https://doi.org/10.1016/s0360-3016(02)04473-5 (2003).
15. Sfeir, A., Tijsterman, M. & McVey, M. Microhomology-mediated end joining chronicles: tracing the evolutionary footprints of genome protection. *Annu. Rev. Cell Dev. Biol.* **40**, 195–218 https://doi.org/10.1146/annurev-cellbio-111822-014426 (2024).
16. Matos-Rodrigues, G. et al. S1-END-seq reveals DNA secondary structures in human cells. *Mol. Cell* **82**, 3538–3552.e3535 https://doi.org/10.1016/j.molcel.2022.08.007 (2022).
17. McClellan, J. A. & Lilley, D. M. A two-state conformational equilibrium for alternating (A-T)$_n$ sequences in negatively supercoiled DNA. *J. Mol. Biol.* **197**, 707–721 https://doi.org/10.1016/0022-2836(87)90477-3 (1987).
18. Inagaki, H. et al. Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. *Genome Res.* **19**, 191–198 https://doi.org/10.1101/gr.079244.108 (2009).
19. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E. & Bishop, J. M. Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-*myc*) in malignant neuroendocrine cells from a human colon carcinoma. *Proc. Natl Acad. Sci. USA* **80**, 1707–1711 https://doi.org/10.1073/pnas.80.6.1707 (1983).
20. Mateos-Gomez, P. A. et al. Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254–257 https://doi.org/10.1038/nature14157 (2015).
21. Ceccaldi, R. et al. Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature* **518**, 258–262 https://doi.org/10.1038/nature14184 (2015).
22. Bubenik, M. et al. Identification of RP-6685, an orally bioavailable compound that inhibits the DNA polymerase activity of Polθ. *J. Med. Chem.* **65**, 13198–13215 https://doi.org/10.1021/acs.jmedchem.2c00998 (2022).
23. Mochirian, P. et al. The discovery of RP-2119: a potent, selective, and orally bioavailable Polθ ATPase inhibitor. *J. Med. Chem.* **68**, 19726–19745 https://doi.org/10.1021/acs.jmedchem.5c02103 (2025).
24. Kang, X. et al. Extrachromosomal DNA replication and maintenance couple with DNA damage pathway in tumors. *Cell* **188**, 3405–3421.e27 https://doi.org/10.1016/j.cell.2025.04.012 (2025).
25. Pradella, D. et al. Engineered extrachromosomal oncogene amplifications promote tumorigenesis. *Nature* **637**, 955–964 https://doi.org/10.1038/s41586-024-08318-8 (2025).
26. Wyatt, D. W. et al. Essential roles for polymerase essential roles for polymerase θ-mediated end joining in the repair of chromosome breaks. *Mol. Cell* **63**, 662–673 https://doi.org/10.1016/j.molcel.2016.06.020 (2016).
27. Brambati, A. et al. RHINO directs MMEJ to repair DNA breaks in mitosis. *Science* **381**, 653–660 https://doi.org/10.1126/science.adh3694 (2023).
28. Sankar, V. et al. Genetic elements promote retention of extrachromosomal DNA in cancer cells. *Nature* **649**, 152–160 https://doi.org/10.1038/s41586-025-09764-8 (2026).
29. Lange, J. T. et al. The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nat. Genet.* **54**, 1527–1533 https://doi.org/10.1038/s41588-022-01177-x (2022).
30. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 https://doi.org/10.1038/s41586-019-1763-5 (2019).
31. Canela, A. et al. DNA breaks and end resection measured genome-wide by end sequencing. *Mol. Cell* **63**, 898–911 https://doi.org/10.1016/j.molcel.2016.06.034 (2016).
32. Kim, H. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 https://doi.org/10.1038/s41588-020-0678-2 (2020).
33. Haniford, D. B. & Pulleyblank, D. E. Transition of a cloned d(AT)n-d(AT)n tract to a cruciform in vivo. *Nucleic Acids Res.* **13**, 4343–4363 https://doi.org/10.1093/nar/13.12.4343 (1985).
34. Dayn, A., Malkhosyan, S. & Mirkin, S. M. Transcriptionally driven cruciform formation in vivo. *Nucleic Acids Res.* **20**, 5991–5997 https://doi.org/10.1093/nar/20.22.5991 (1992).
35. Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl Acad. Sci. USA* **84**, 7024–7027 https://doi.org/10.1073/pnas.84.20.7024 (1987).
36. Feng, S. et al. Profound synthetic lethality between SMARCAL1 and FANCM. *Mol. Cell* **84**, 4522–4537.e4527 https://doi.org/10.1016/j.molcel.2024.10.016 (2024).
37. Lu, S. et al. Short inverted repeats are hotspots for genetic instability: relevance to cancer genomes. *Cell Rep.* **10**, 1674–1680 https://doi.org/10.1016/j.celrep.2015.02.039 (2015).
38. Mengoli, V. et al. WRN helicase and mismatch repair complexes independently and synergistically disrupt cruciform DNA structures. *EMBO J.* **42**, e111998 https://doi.org/10.15252/embj.2022111998 (2023).
39. Zhu, K. et al. CoRAL accurately resolves extrachromosomal DNA genome structures with long-read sequencing. *Genome Res.* **34**, 1344–1354 https://doi.org/10.1101/gr.279131.124 (2024).
40. Lee, J. J.-K. et al. Evolution of oncogene amplification across 86,000 cancer cell genomes. Preprint at *bioRxiv* https://doi.org/10.64898/2026.02.12.705658 (2026).
41. Kim, H. et al. Mapping extrachromosomal DNA amplifications during cancer progression. *Nat. Genet.* **56**, 2447–2454, https://doi.org/10.1038/s41588-024-01949-7 (2024).
42. Luebeck, J. et al. AmpliconSuite: an end-to-end workflow for analyzing focal amplifications in cancer genomes. Preprint at *bioRxiv* https://doi.org/10.1101/2024.05.06.592768 (2024).
43. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 https://doi.org/10.1038/nature21356 (2017).
44. Chedin, F. & Benham, C. J. Emerging roles for R-loop structures in the management of topological stress. *J. Biol. Chem.* **295**, 4684–4695 https://doi.org/10.1074/jbc.REV119.006364 (2020).
45. Inagawa, T. et al. C-terminal extensions of Ku70 and Ku80 differentially influence DNA end binding properties. *Int. J. Mol. Sci.* **21**, 6275 https://doi.org/10.3390/ijms21186725 (2020).
46. Deng, L. et al. Mitotic CDK promotes replisome disassembly, fork breakage, and complex DNA rearrangements. *Mol. Cell* **73**, 915–929.e916 https://doi.org/10.1016/j.molcel.2018.12.021 (2019).
47. Fielden, J. et al. Comprehensive interrogation of synthetic lethality in the DNA damage response. *Nature* **640**, 1093–1102 https://doi.org/10.1038/s41586-025-08815-4 (2025).
48. Bailey, C. et al. Origins and impact of extrachromosomal DNA. *Nature* **635**, 193–200 https://doi.org/10.1038/s41586-024-08107-3 (2024).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

Methods

Cell culture

COLO320DM, COLO320HSR and SNU-16 cells (ATCC CCL-220, CCL-220.1 and CRL-5974, respectively) were maintained in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), 100 U ml^{-1} penicillin-streptomycin, 1% non-essential amino acids, and 2 mM L-glutamine. HEK293T cells (ATCC CRL-3216), PC3-DM, and PC3-HSR (gifts from the P. Mischel laboratory) were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% bovine calf serum, 100 U ml^{-1} penicillin-streptomycin, 1% non-essential amino acids, and 2 mM L-glutamine. NCI-H716 cells (ATCC CCL-251) were cultured in RPMI-1640 with the same supplements. All cells were incubated at 37 °C in a humidified atmosphere containing 5% CO_2. All cells were tested for *Mycoplasma* by PCR.

HeLa cells were cultured as previously described[7]. In brief, cells were grown in DMEM supplemented with 10% FBS (Omega), 100 U ml^{-1} penicillin, 100 U ml^{-1} streptomycin, and 2 mM L-glutamine. For MTX experiments, dialysed FBS was used. Specific HeLa-derived clones were used at the following MTX concentrations: PD29424h (H, DM+, 60 nM MTX) and PD29426a (HSR+, 40 nM MTX).

Adult neuronal stem cells (aNSCs) from $Myc^{ec/+}$; $Trp53^{fl/fl}$ (ecMyc) or control $Myc^{+/+}$; $Trp53^{fl/fl}$ (control) mice were isolated and cultured in adherent condition on laminin-coated (Sigma-Aldrich, L2020) dishes in NeuroCult Stem Cell Basal Media with NeuroCult Proliferation Supplement (Mouse & Rat) (Stem Cells Technologies, 05702), 20 ng ml^{-1} EGF (Stem Cells Technologies, 78006), 10 ng ml^{-1} bFGF (Stem Cells Technologies, 78003), and 2 µg ml^{-1} heparin (Stem Cell Technologies), ecMyc and control aNSCs were treated by Adeno-Cre (from ViraQuest (VQ-Ad-CMV-Cre; 1 × 1012 particles per ml; 091317)) to induce engineered ecDNA formation in ecMyc cells. Cells were maintained in culture for 3 weeks to accumulate *Myc*-containing ecDNAs. All cells were negative for mycoplasma contamination. Cells were maintained in a humidified 5% CO_2 atmosphere at 37 °C.

For proliferation assays, cells were seeded at clonal density in 12-well plates. The following day, cells were treated with a drug (Polθ inhibitors, MTX, CHK1 inhibitor or FGFR inhibitor). For HeLa experiments, MTX concentrations were applied as described above. In both the HeLa and SNU-16 experiments, cells were pre-treated with Polθi or vehicle for 1 week prior to replating for growth assessment. Medium with fresh drug was replaced every 72 h. Cell proliferation was monitored using the IncuCyte Live-Cell Imaging System (Essen BioScience/Sartorius), with confluency measurements collected every 24 h.

For proliferation assays involving *FANCM* loss and Polθ inhibition, *FANCM*-knockout cells were generated using the triple-guide CRISPR–Cas9 strategy described below, with *AAVS1*-targeted cells serving as controls. Two days after knockout, cells were plated at 2,000 cells per well in 48-well plates in three technical replicates for each condition, and Polθi-Pol treatment was initiated. Cell proliferation was monitored using the IncuCyte Live-Cell Imaging System with confluence measurements acquired every 12 h.

Drug treatments

The following compounds were used in this study: Polθ inhibitors: RP-2119 (Polθi-Hel; Repare Therapeutics; 100 nM), RP-6685 (Polθi-Pol; Repare Therapeutics; 10 µM); DNA-PKcs inhibitors: NU7441 (Selleck Chemicals, S2638; 1 µM), NU7026 (Selleck Chemicals, S2893; 10 µM); CHK1 inhibitor (CCT245737; Selleck Chemicals, S8253; 750 nM); methotrexate (MTX; Sigma-Aldrich, 454126; concentrations varied by cell line and clone as noted above); hydroxyurea (Sigma-Aldrich, H8627; 100 µM); PARP inhibitor (Olaparib; Selleck Chemicals, S1060; 1 µM); thymidine (Sigma-Aldrich, T1895; 2 mM); nocodazole (Sigma-Aldrich, M1404; 100 ng ml^{-1}); FGFR inhibitor (Infigratinib/BJ398; Selleck Chemicals, S2183; 6.25 nM); triptolide (Millipore, 645900; 1 µM); and CDK9 inhibitor (AZD4573; Selleck Chemicals, S8719; 200 nM).

For FISH experiments and proliferation assays, the media were replaced every 72 h with fresh medium containing the indicated drugs. Drug concentrations were maintained consistently throughout the assay.

CRISPR–Cas9 ribonucleoprotein complex assembly and nucleofection

CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) (Integrated DNA Technologies, IDT) were each resuspended in IDTE buffer to a final concentration of 100 µM. Equimolar amounts were mixed to yield a 50 µM crRNA:tracrRNA (sgRNA) duplex, which was annealed by heating to 95 °C for 5 min, followed by cooling to room temperature. For ribonuclear protein (RNP) complex formation, purified Cas9 protein (Berkeley Facility; https://macrolab.qb3.berkeley.edu/) and sgRNA were mixed at a 1:1.2 molar ratio and incubated at room temperature for 20 min before use. In experiments using three sgRNAs simultaneously to knock out a gene, sgRNAs were designed to be between 35 and 300 bp apart. Each sgRNA was resuspended in IDTE buffer and pooled to a final concentration of 40 µM per guide. RNP complexes were assembled by mixing 104 pmol of Cas9 with 120 pmol of the combined sgRNA mix (1:1.15 molar ratio) and incubating at room temperature for 20 min. Target cells were prepared using the SF Cell Line 4D-Nucleofector X Kit (Lonza) according to the manufacturer's instructions. Nucleofection was performed using the Lonza 4D-Nucleofector System with program CM158. Knockouts were validated by PCR amplifying the region around the sgRNA target site, followed by Sanger sequencing and TIDE[49] or ICE CRISPR Analysis (EditCo Bio., v3.0). For simultaneous RHINO and FANCM knockouts, a triple-guide CRISPR–Cas9 strategy was used for each gene, and RNP complexes targeting both genes were combined at a 1:1 ratio before nucleofection. Cells were resuspended in SF solution, mixed with the combined RNP complexes, and nucleofected as described above. For the corresponding single-knockout controls, *IL25*-targeting RNPs were included to match the total guide load in the double-knockout condition. Sequences for sgRNA in Supplementary Table 1. PCR primers for TIDE/EditCo ICE and CRISPR analysis in Supplementary Table 2.

Lentiviral production and transduction

Lentivirus was produced in HEK293T cells using a standard four-plasmid transfection system. For each transfection, 5 µg of RRE, 3 µg of VSV-G, 2.5 µg of REV, and 20.5 µg of a BFP-expressing *BRCA2* shRNA transfer plasmid were combined with 62 µg ml^{-1} polyethylenimine (PEI) and 150 mM NaCl. The DNA-PEI complexes were incubated at room temperature for 15 min before being added to 10-cm dishes of HEK293T cells at ~70–80% confluence. Cells were incubated overnight at 37 °C in a humidified 5% CO_2 incubator. The following day, the culture medium was replaced, and after a 6–8 h recovery period, the first viral supernatant was collected. This process was repeated twice at 24-h intervals to collect the second and third viral supernatants. Fresh medium was added after each collection.

The *BRCA2* knockdown construct targeted exon 16 of the *BRCA2* coding sequence and was designed using the Broad Institute's RNAi design tool. The hairpin sequence (5′-GCGTTTCTAAACATTGCATAA-3′) was cloned into the pLKO.1-puro vector (Addgene #8453). Target cells were transduced with the lentiviral supernatant and selected with puromycin to generate stable knockdown lines. Functional validation of BRCA2 depletion was assessed by measuring cell proliferation in response to 1 µM olaparib. Cell growth was monitored using the IncuCyte Live-Cell Imaging System (Essen BioScience/Sartorius), with confluency measurements recorded every 24 h.

Generation of *RHNO1*$^{-/-}$ cells

COLO320DM cells were stably infected with pLentiCRISPRV2 Cas9 – blast (Addgene #98293) and selected with blasticidin (5 µg ml^{-1}). Blasticidin-resistant cells were then stably infected with pLenti-GFP-2A-Puro-gRHINO1_A (Addgene #211613) and selected with puromycin at

3 μg ml^{-1}. *RHNO1* knockout was confirmed with TIDE analysis (forward: TTTTGCTTGGTGGTTGTAGG, reverse: TATCTGGCATTCTCACCCAG).

Generation of Fucci4 cells

Lentivirus was prepared containing one of each of the following plasmids, gifts from M. Lin, to generate Fucci4 cells[50]: pLL3.7m-Clover-Geminin(1-110)-IRES-mKO2-Cdt(30-120) (Addgene #83841), pLL3.7m-mTurquoise2-SLBP(18-126)-IRES-H1-mMaroon1 (Addgene #83842), mKO2-SLBP(18-126) (Addgene #83914) and Clover-Geminin(1-110) (Addgene #83915). COLO320DM and COLO320HSR cells were transduced with lentiviral supernatant containing either mTurquoise2-SLBP/H1-mMaroon1, mKO2-SLBP, or Clover-Geminin. Transgene-positive cells were isolated by FACS and used as colour controls in experiments. A split of mTurquoise2-SLBP/H1-mMaroon1-positive cells was transduced with lentiviral supernatant containing Clover-Geminin/mKO2-Cdt, and transgene-positive, fully labelled Fucci4 cells were isolated by FACS.

Cell cycle analysis of Fucci4 cells by flow cytometry

COLO320DM or COLO320HSR Fucci4 cells were plated and treated for 72 h with either DMSO (0.1–0.3% as required per control), Polθi-Pol, Polθi-Hel, DNA-PKcs inhibitors, Chk1i, Chk1i plus Polθi-Pol, or Chk1i plus Polθi-Hel. Cell cycle controls included a double thymidine block using 2 mM thymidine or an overnight treatment with 100 ng ml^{-1} nocodazole. For the experiment, CRISPR–Cas9-mediated sgRNA targeting of either *AAVS1* or *FANCM* was performed in COLO320DM Fucci4 cells. Knockout cells were plated and treated for 72 h with either 0.1% DMSO, Polθi-Pol, or Polθi-Hel. Cells were trypsinized, collected in RPMI medium, washed with DPBS, pelleted, resuspended in DPBS with 3% FBS, and filtered using a 35-μm mesh into polystyrene tubes (Falcon; 38030) for flow cytometry. Cells were then analysed on an Aurora full-spectrum analyser (Cytek Biosciences), with 100,000 to 150,000 events collected per condition, and quantification was performed using FlowJo (v10.10.1). Single or dual-colour controls (mTurquoise2–SLBP/H1–mMaroon1, mKO2–SLBP or Clover–Geminin) and unstained cells were used to determine the gating strategy for G_1-, S–G_2- and M-phase cells (Supplementary Fig. 1).

Generation of H2B–GFP and H2B–RFP cells

Lentivirus was prepared containing one of each of the following plasmids, gifts from E. Fuchs: H2B–GFP (Addgene #25999) and H2B–RFP (Addgene #26001). COLO320DM and COLO320HSR cells were transduced with H2B–GFP and H2B–RFP separately, and transgene-positive cells were isolated by FACS. Labelling was confirmed by live-cell microscopy in 35 mm no. 1.5 glass-bottomed dishes (Mattek; P35G-1.5) using a Nikon Ti2 inverted microscope equipped with a TokaiHIT STX stage-top incubator and objective heating collar (37 °C, 5% CO_2), a Yokogawa CSU-W1 spinning disc confocal unit, and dual Hamamatsu C14440-20UP sCMOS cameras (Dual Fusion configuration). The system was controlled using NIS-Elements AR software version 6.10.02. Cells were imaged using an Apo TIRF 60× 1.49 NA oil immersion objective (Nikon). Four fluorescence channels were acquired sequentially using laser excitation from an LUN-F multi-laser unit and 300 ms exposures. Excitation wavelengths and laser powers were as follows: 405 nm (100 mW) at 30% power, 488 nm (100 mW) at 10% power, and 561 nm (100 mW) at 30% power. A quad-band dichroic mirror (Di01-T405/488/568/647) was used with the following emission filters: 455/50 nm (DAPI), 526/36 nm (GFP), and 605/52 nm (dsRed–mCherry), split across two detection paths using a 561 nm long-pass dichroic (DM A561LP). Images were saved in Nikon ND2 format and subsequently processed using Fiji (ImageJ version 2.16.0)[51].

Fluorescence-based competition assays

CRISPR–Cas9-mediated targeting of *IL25* or *POLQ* (see 'CRISPR–Cas9 ribonucleoprotein complex assembly and nucleofection') was performed in COLO320DM and COLO320HSR cells labelled with either H2B–GFP or H2B–RFP. Each cell type was mixed 1:1 in the following pairings and plated in triplicate at 2,000 cells per well in a 48-well tissue culture plate (Corning Costar; 3548):

H2B label	Knockout	H2B label	Knockout
RFP	sg*IL25*	GFP	sg*POLQ*
RFP	sg*IL25*	GFP	sg*IL25*
RFP	sg*POLQ*	GFP	sg*IL25*

1:1 mixtures were imaged on an IncuCyte Live-Cell Imaging System (Essen BioScience/Sartorius) for up to 2 weeks. Knockouts were confirmed during the experiments using DNA extraction with the Direct-PCR Lysis Reagent (Viagen; 301-C), PCR amplification of the knockout regions, Sanger sequencing and EditCo's ICE Analysis tool (v3). Cell confluency, the number of GFP-positive cells, and the number of RFP-positive cells were quantified using the IncuCyte Analysis software. Technical replicates were averaged, and the number of RFP- or GFP-positive cells per condition was normalized first to timepoint 0 (day 1), and then to the H2B–RFP (sg*IL25*) or H2B–GFP (sg*IL25*) control condition.

Live-cell imaging of mitotic timing

COLO320DM or COLO320HSR cells expressing H2B–GFP were plated on 35 mm no. 1.5 glass-bottomed dishes (Mattek; P35G-1.5) and treated for 48 h with either 0.1% DMSO or Polθi-Hel. At 48 h, cell medium was changed to RPMI without phenol red (supplemented with respective drugs or treatments). Time-lapse images were acquired on a Nikon Ti2 inverted microscope equipped with a TokaiHIT STX stage-top incubator and objective heating collar (37 °C, 5% CO_2), a Lumencor Sola II light engine, and a Hamamatsu C14440-20UP sCMOS camera (SN:000740). The system was controlled using NIS-Elements AR software version 5.21.03. Cells were imaged using a Plan Apo λ 100× 1.45 NA oil immersion objective (Nikon). Two fluorescence channels were acquired at each time point: GFP (ET-GFP; Nikon; 96366; ET470/40X; ET525/50 M; 200 ms exposure) and DIC. Time-lapse imaging was performed with a 3-min interval between frames, and the Perfect Focus System (PFS) was engaged. Images were saved in Nikon ND2 format and subsequently processed in Fiji (ImageJ version 2.16.0)[51] to manually screen for mitotic events.

Polθ ATPase enzymatic assay

In a 384-well assay plate (Corning 3572), recombinant full-length human POLQ (3 nM) was combined with a 10-point concentration range of Polθi-Hel (highest concentration 0.1 mM, threefold serial dilutions) in a buffer containing 50 mM Tris-Cl pH 7.5, 10% glycerol, 5 mM DTT, 10 mM MgCl2, and 0.1 mg ml^{-1} BSA. The plate was incubated at room temperature for 15 min, after which the reaction was initiated by adding a DNA substrate (20 nM) and ATP (100 μM). Plates were incubated for 90 min at room temperature, and ATPase activity was subsequently detected using an ADP-Glo assay kit (Promega) according to the manufacturer's protocol. In brief, 15 μl of ADP-Glo reagent was added to each well, followed by a 1 h incubation at room temperature. 30 μl of Detection reagent was then added per well, the plates were incubated for 45 min at room temperature, and luminescence was read on an Envision plate reader (Revvity). The forked DNA substrate was prepared as follows: oligos 1–3 were mixed (1:1:1) in a buffer containing 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 1 mM EDTA, heated to 95 °C for 5 min, and cooled to room temperature.

Oligo 1: 5′-GCACTGGCCGTCGTTTTACGGTCGTGACTGGGAAAACCCTGGCG-3′; oligo 2: 5′-TTTTTTTTTTTTTTTTTTTTTTCCAAGTAAAACGACGGCCAGTGC-3′; oligo 3: 5′-TTGGAAAAAAAAAAAAAAAAAAAA-3′.

Expression and purification of recombinant full-length Polθ

Four 1-l cultures of 293-6E cells were grown in F17 medium supplemented with 0.1% Pluronic F-68, 4 mM GlutaMAX, and 25 μg ml^{-1}

G418 in 2-l shake flasks. Cultures were transfected with 1 mg l^{-1} of the phCMV1-2m6h-POLQ plasmid (lab of S. Doublié) using PolyPlus PEIPro at a 1:1 (w/v) DNA:PEI ratio. Transfected cultures were maintained at 37 °C with 5% CO_2 and shaken at 135 rpm. Cells were collected 4 days post-transfection by centrifugation at 1,000*g* for 5 min at 4 °C. For lysis, the resulting cell pellet was resuspended in 250 ml of lysis buffer (20 mM Tris-HCl, 250 mM NaCl, 250 mM KCl, 0.01% NP-40, 10% glycerol, pH 8.0) supplemented with 5 mM β-mercaptoethanol and EDTA-free protease inhibitors (Roche). Cells were lysed by freeze–thaw cycles followed by processing through a microfluidizer. The lysate was stirred at 4 °C for 30 min, then clarified by centrifugation at 20,000*g* for 60 min at 4 °C.

The soluble fraction was batch-bound to 5 ml of amylose resin (New England Biolabs) in buffer A (20 mM Tris-HCl, 200 mM NaCl, 200 mM KCl, 0.01% NP-40, 10% glycerol, 5 mM β-mercaptoethanol, pH 8.0) for 1.5 h at 4 °C with gentle nutation. The resin was first collected in a 5-cm Econo-column, transferred to a 2.5-cm column, and washed with 30 column volumes of buffer A. Elution was carried out using buffer B (buffer A supplemented with 50 mM maltose). The eluate was diluted threefold with pre-heparin dilution buffer (10 mM HEPES, pH 7.5), then loaded at 2 ml min^{-1} onto a 5 ml HiTrap Heparin Sepharose HP column (Cytiva, 17040703) pre-equilibrated in buffer D (20 mM Tris-HCl, 500 mM NaCl, 500 mM KCl, 0.01% NP-40, 5% glycerol, 5 mM β-mercaptoethanol, pH 8.0). After loading, the column was washed with 20 column volumes of buffer C (20 mM Tris-HCl, 100 mM NaCl, 0.01% NP-40, 5% glycerol, 5 mM β-mercaptoethanol, pH 8.0), and 5 ml fractions were collected. The protein was eluted using a linear 0–100% gradient of buffer D over 20 column volumes at 2 ml min^{-1}. The peak fractions (heparin pool 1) were concentrated to ~0.8 ml using a Vivaspin6 centrifugal device (10 kDa MWCO, PES membrane), then subjected to size-exclusion chromatography using a Superose 6 Increase column (Cytiva) equilibrated in destination buffer (20 mM Tris-HCl, 200 mM NaCl, 100 mM KCl, 2% glycerol, 3 μM sucrose monolaurate, 0.5 mM TCEP, pH 8.0). Fractions (0.5 ml) were collected over 1.1 column volumes and analysed by SDS–PAGE. Protein-containing fractions were pooled and concentrated using a Vivaspin6 centrifugal device (10 kDa MWCO, PES membrane) for downstream applications.

Metaphase FISH for detection of *MYC*, *FGFR2* and *DHFR* ecDNA

Cells were seeded to reach approximately 50% confluence 36–48 h before collection. To enrich for mitotic cells, Colcemid (0.1 μg ml^{-1}) was added for 6 h, followed by trypsinization. Cells were pelleted by centrifugation (5 min at 1,000 rpm), gently resuspended in 5 ml of pre-warmed 0.075 M KCl, and incubated at 37 °C for 15–30 min with occasional mixing to induce swelling. After centrifugation, residual KCl was retained, and fixative (3:1 methanol:acetic acid) was added dropwise with gentle mixing. The total volume was adjusted to 10 ml with fixative, and samples were stored at 4 °C overnight or longer. For slide preparation, fixed cells were pelleted and resuspended in a minimal volume of fresh fixative. Cold, wet microscope slides were prepared by briefly soaking in water, and cells were dropped from a height onto tilted slides. Slides were immediately placed on a humidified 80 °C heating block for 1 min to promote chromosome spreading, then air-dried overnight. Slide quality was assessed by light microscopy.

To detect extrachromosomal *MYC*, FISH was performed using a locus-specific probe targeting the *MYC* genomic region (Empire Genomics: *MYC* FISH probe). Slides were rehydrated in PBS for 5 min, fixed in 4% formaldehyde for 2 min, and washed 3 times in PBS (5 min each), followed by three washes in 2× SSC (5 min each). Slides were then denatured in 70% formamide/2× SSC at 72 °C for 2 min, immediately dehydrated through an ethanol series (70%, 90% and 100%), and air-dried. The *MYC* probe, resuspended in hybridization buffer, was applied to the slide, denatured at 80 °C for 5 min, and hybridized overnight at 37 °C in a humidified chamber. The following day, excess probe was removed by sequential washes: 0.4× SSC at 72 °C for 5 min, followed by 2× SSC with 0.05% NP-40 at room temperature. Additional washes were performed in 50% formamide/2× SSC at 42 °C, 2× SSC, and 0.1× SSC. Slides were counterstained with DAPI and mounted using ProLong Gold antifade reagent. Fluorescence microscopy was used to visualize *MYC*-positive ecDNA, which appeared as distinct fluorescent foci located outside the main chromosomal arms in metaphase spreads.

Detection of additional ecDNA species was performed as follows. For *FGFR2*, a probe targeting the *FGFR2* genomic region (Empire Genomics: *FGFR2* FISH probe) was used; FISH was carried out as described above. For *DHFR*, a probe targeting the *DHFR* genomic region (Empire Genomics: DHFR FISH probe) was used; FISH was performed using the same protocol.

Detection of *MYC*-containing micronuclei by interphase FISH

To detect *MYC*-containing micronuclei, cells were seeded on poly-L-lysine-coated glass coverslips 96 h prior to fixation. Polθ inhibitor or vehicle was added 72 h before collection, followed by hydroxyurea or PBS 24 h before fixation. Cells were fixed in 4% paraformaldehyde (PFA) for 5 min, washed three times with PBS for 5 min each, and permeabilized with 0.5% Triton X-100 in PBS for 10 min. After two additional PBS washes (5 min each), coverslips were rinsed three times in 2× SSC for 5 min and then dehydrated through a graded ethanol series (70%, 90% and 100%). For FISH, the *MYC* probe (Empire Genomics) was mixed with hybridization buffer, applied to the coverslip, and denatured at 80 °C for 5 min. Hybridization was carried out overnight in a humidified chamber. The following day, coverslips were fixed in freshly prepared methanol: acetic acid (3:1) for 10 min at room temperature, air-dried, and incubated in 2× SSC. Denaturation was performed in 70% formamide/2× SSC at 72 °C for 2 min, followed by rapid dehydration in ethanol (70%, 90%, 100%). The fluorescently labelled *MYC* probe was then applied and hybridized overnight at 37 °C. The excess probe was removed by post-hybridization washes in 50% formamide/2× SSC at 42 °C, followed by rinses in 2× SSC and 0.1× SSC. A final wash was performed in 0.4× SSC at 72 °C for 5 min, followed by a rinse in 2× SSC containing 0.05% NP-40 at room temperature. Nuclei were counterstained with DAPI and mounted using ProLong Gold antifade reagent. Imaging was performed by fluorescence microscopy. *MYC*-positive micronuclei were identified as small, DAPI-stained structures that were spatially separated from the primary nucleus and displayed distinct *MYC* FISH signals.

DNA telomere FISH

Cells (2×10^5) were plated onto poly-L-lysine-treated coverslips and treated with Polθi-Pol, Polθi-Hel, or 0.1% DMSO for 72 h or 0.5 μM Reversine for 24 h, prior to fixation. Cells on coverslips were washed 2 × 5 min with 1× DPBS, then fixed with 4% PFA for 10 min, followed by 3 × 5 min washes in 1× DPBS. Cells were then permeabilized with 0.5% Triton X-100 for 10 min and washed with DPBS, 2 × 5 min each. Coverslips were dehydrated in 70%, 95%, and 100% ethanol for 2 min each, then air-dried. Cells were incubated in 30 μl hybridization buffer (70% formamide, 1 mM Tris-Cl, and Roche blocking reagent at 1 mg ml^{-1}) and 100 nM TelC-Cy3 (PNAbio) on slides marked with a wax pen before denaturation at 80 °C for 5 min. Cells were incubated in the dark at room temperature for 2 h, followed by 4×15 min formamide washes (70% formamide, 10 mM Tris-Cl in H_2O). Cells were washed 3× 10 min in PBST (0.1% Tween-20), with 5 μg ml^{-1} DAPI added to the second wash. Coverslips were air-dried before being mounted with ProLong Glass antifade.

DNA pan-centromere FISH

Cells (2×10^5) cells were plated onto poly-L-lysine-treated coverslips and treated with Polθi-Pol, Polθi-Hel, or 0.1% DMSO for 72 h or 0.5 μM Reversine for 24 h, prior to fixation. Cells on coverslips were washed 2 × 5 min with 1× DPBS, then fixed with 4% PFA for 10 min, followed by 3 × 5 min washes in 1× DPBS. Cells were then permeabilized with 0.5% Triton-X for 10 min and washed with DPBS, 2 × 5 min each. Coverslips were washed 3 × 5 min in 2× SSC, then dehydrated in 70%, 85% and 100% ethanol for 2 min each, then dried completely. Coverslips were

incubated with XCE pan-cen (MetaSystems), diluted 1:5 in hybridization buffer (Empire Genomics), and then sealed with CytoBond (SciGene) before denaturation at 80 °C for 10–15 min. After overnight incubation in a humidified chamber at 37 °C, the CytoBond sealant was removed from the coverslips, which were washed in 0.4× SSC (heated to 72 °C) for 5 min, then washed for 30 s in 2× SSC + 0.05% Tween-20. Coverslips were rinsed briefly with ddH_2O, then air-dried, and washed 3× 10 min in DPBS (with 5 μg ml^{-1} DAPI added to the second wash). Coverslips were air-dried before being mounted with ProLong Glass antifade.

Immunofluorescence for the detection of 53BP1 and gH2AX

Cells were seeded on poly-L-lysine-coated glass coverslips and fixed 72 h later following the indicated treatments. Cells were fixed in 4% PFA for 15 min at room temperature, then permeabilized with 0.5% Triton X-100 for 10 min. Cells were then blocked for 1 h at room temperature using a blocking buffer consisting of 1 mg ml^{-1} BSA, 3% goat serum, 0.1% Triton X-100, and 1 mM EDTA in PBS (pH 8.0). After blocking, cells were incubated for 90 min at room temperature with primary antibodies diluted in the same blocking buffer: rabbit anti-53BP1 (Novus NB100-304, 1:1,000) and mouse anti-γH2AX (Millipore JBW301, 1:1,000). Following three PBS washes, cells were incubated with Alexa Fluor–conjugated secondary antibodies (donkey anti-rabbit 568 and donkey anti-mouse 488, 1:500, invitrogen) for 30 min at room temperature in the dark. Coverslips were then washed three times with PCB and mounted using ProLong Gold Antifade Mountant with DAPI (Thermo Fisher). DNA damage foci were visualized using fluorescence microscopy (Nikon Si) and quantified based on the number of 53BP1 and γH2AX foci per nucleus.

For immunofluorescence experiments involving transcription inhibition, transcription was acutely inhibited for 3 h with triptolide or CDK9 inhibitor either 3 days after Polθ inhibitor or vehicle pretreatment, or 5 days after CRISPR–Cas9-mediated knockout of *FANCM* or *AAVS1*. For immunofluorescence experiments assessing DNA damage following CRISPR–Cas9-mediated targeting of the indicated nucleases and helicases, cells were treated with Polθ inhibitor or vehicle beginning 2 days after nucleofection and collected on day 5. For experiments involving CRISPR–Cas9-mediated co-targeting of *RHNO1* and *FANCM*, cells were nucleofected and collected on day 5 for immunofluorescence.

Combined 53BP1 immunofluorescence and *MYC*-FISH

Cells were seeded onto glass coverslips and allowed to adhere overnight. The following day, cells were fixed in 4% PFA for 15 min at room temperature, then permeabilized with 0.5% Triton X-100 for 10 min. Cells were then blocked for 1 h at room temperature using a blocking buffer consisting of 1 mg ml^{-1} BSA, 3% goat serum, 0.1% Triton X-100, and 1 mM EDTA in PBS (pH 8.0). After blocking, cells were incubated for 90 min at room temperature with primary antibody diluted in the same blocking buffer: rabbit anti-53BP1 (Novus NB100-304, 1:1,000). Following three PBS washes, cells were incubated with Alexa Fluor–conjugated secondary antibodies (donkey anti-rabbit 488 1:500, Invitrogen) for 30 min at room temperature in the dark. Coverslips were then washed three times with PBS. Cells were then fixed with 4% PFA at room temperature for 20 min, followed by re-permeabilization using 1× PBS containing 0.7% Triton X-100 and 0.1 M HCl on ice for 10 min. Coverslips were then subjected to acid denaturation using 1.9 M HCl for 30 min at room temperature, followed by washes in PBS and 2× SSC. Coverslips were dehydrated through an ethanol series (70%, 90% and 100%) and air-dried. The *MYC* probe, resuspended in hybridization buffer, was applied to the coverslip, denatured at 80 °C for 5 min, and hybridized overnight at 37 °C in a humidified chamber. The following day, excess probe was removed by sequential washes: 0.4× SSC at 72 °C for 5 min, followed by 2× SSC with 0.05% NP-40 at room temperature. Additional washes were performed in 50% formamide/2× SSC at 42 °C, 2× SSC, and 0.1× SSC. Slides were counterstained with DAPI and mounted using ProLong Gold antifade reagent.

Colony formation assay

COLO320DM cells were plated at clonal density in triplicate in 6-well plates with 1 μM NU7441 or DMSO. 48 h after plating, cells were then treated with 2 Gy of ionizing radiation. Cells were then allowed to grow for an additional 10 days, after which they were fixed in 4% PFA and stained with crystal violet. Colonies were counted with the Fiji ImageJ Colony Area plugin[52].

Immunoblot to detect Cre recombinase expression

Cells were lysed in Laemmli buffer, supplemented with protease and phosphatase inhibitors (cOmplete and EDTA-free Protease Inhibitor cocktail; Roche, COEDTAF-RO and PHOSS-RO). Proteins were separated in NuPage Bis-Tris gels (Invitrogen, NP0322BOX), transferred on a nitrocellulose membrane (0.2μm; Bio-Rad), and blocked by incubation with Intercept (TBS) Blocking buffer (LI-COR: 927-60001). The following primary antibodies were used: anti-Cre recombinase (1:1,000; Cell Signaling, 15036) and anti-α-tubulin (1:5,000; Cell Signaling, 2144). The following secondary antibodies were used: IRDye 800 anti-rabbit (LI-COR, 926-32213) and IRDye 680 anti-mouse (LI-COR, 926-68072). Images were acquired using an Odyssey Imaging System (LI-COR). Full uncropped blots with molecular weight markers are available in the supplementary data (Supplementary Fig. 2).

***AAVS1* and *MYC* Repair-seq**

Cells were nucleofected using the Lonza 4D-Nucleofector System with the SF Cell Line Kit- following the manufacturer's protocol. Cas9 RNP complexes were assembled by incubating purified Cas9 protein (Berkeley Facility; https://macrolab.qb3.berkeley.edu/) with synthetic sgRNAs (IDT) at a 1:2 molar ratio for 10–15 min at room temperature. sgRNAs were designed to target the *AAVS1* locus and exon 2 of the *MYC* gene. Approximately 1×10^6 cells were nucleofected per condition using program CM158 and collected 72 h post-nucleofection to allow sufficient time for editing events and genomic DNA recovery. Genomic DNA was extracted using the ZymoResearch Quick-DNA Genomic Kit according to the manufacturer's instructions. DNA concentration and quality were assessed using a Qubit fluorometer and agarose gel electrophoresis. A ~450 bp region surrounding each CRISPR cut site was amplified by PCR using locus-specific primers (primer sequences provided in Supplementary Table 2). PCRs were performed using Taq DNA Polymerase (NEB) with the following cycling conditions: initial denaturation at 95 °C for 3 min; 35 cycles of 95 °C for 15 s, 53 °C for 15 s, 68 °C for 30 s; final extension at 68 °C for 5 min. Amplicons were quantified using Qubit and pooled in equimolar ratios. The IGO facility core of MSKCC prepared Pooled PCR products for sequencing. Libraries were sequenced on the Illumina MiSeq platform using 250 bp paired-end reads (MiSeq v2 PE250 kit). All primer pairs were pre-validated to produce amplicons of similar length and efficiency, ensuring uniform sequencing coverage. Sequence quality was monitored across all samples, and depth uniformity was confirmed during post-run analysis. For experiments evaluating MMEJ, a minimum of 700,000 paired-end reads per sample was obtained. For homology-directed repair quantification, a minimum of 30,000 paired-end reads per sample was achieved. Sequencing reads were demultiplexed and aligned with a reference sequence. The CRISPResso2 toolkit was used to quantify editing outcomes[53], characterize insertions and deletions, and reconstruct repair events. For each sample, allele fractions for these events were calculated by counting the number of reads with respective mutational signatures identified by CRISPResso2 and dividing the count by the total reads.

Live-cell imaging of ecDNA tethering

Live-cell imaging to assess ecDNA tethering to mitotic chromosomes was performed using COLO320DM TetO-EGFP cells stably expressing H2B–mCherry and TetR–eGFP. Cells were seeded into four-well

chamber slides (ibidi) 48 h prior to imaging and treated with DMSO, Polθi-Pol, or Polθi-Hel for 24 or 72 h. Imaging was conducted in a humidified, 37 °C chamber with 5% CO_2 using a Nikon SoRa Spinning Disk Confocal system equipped with a Borealis microadapter, Perfect Focus 4, motorized turret and encoded stage, and a 5-line laser launch (405 nm (100 mW), 445 nm (45 mW), 488 nm (100 mW), 561 nm (80 mW) and 640 nm (75 mW)). A PRIME 95B Monochrome Digital Camera and CFI Apo TIRF 60× 1.49 NA objective lens (W.D. 0.12 mm) were used in super-resolution mode. Laser power was set to 5% with a 600 ms exposure time. Images were acquired using NIS-Elements Advanced Research Software on a dual-Xeon imaging workstation and denoised using the default settings. Time-lapse imaging of mitotic cells was initiated at metaphase or anaphase and continued at 5-min intervals until reformation of daughter nuclei. *z*-stacks were processed as maximum intensity projections, and image adjustments (brightness and contrast) were performed using Fiji. To quantify cytosolic mis-segregation events, $TetR^+$ ecDNA foci were tracked throughout mitosis; foci excluded from daughter nuclei post-mitosis were classified as mis-segregated ecDNA.

END-seq

To induce a control DSB at the *MYC* locus, COLO320DM and COLO320-HSR cells were transfected with a Cas9 RNP complex targeting the *MYC* gene. The RNP complex was assembled by mixing purified Cas9 protein (Berkeley MacroLab) with an sgRNA targeting exon 2 of *MYC* (sequence: GCCGTATTTCTACTGCGACG; IDT), to a final concentration of 20 pmol Cas9 and 25 pmol sgRNA. The mixture was incubated at room temperature for 10–15 min. Cells were nucleofected using the Lonza 4D-Nucleofector System with the SF Cell Line Kit and program CM158. After overnight recovery, cells were treated with either the Polθi-Pol or vehicle control for 72 h, then collected for END-seq. For sgRNA-mediated depletion of *AAVS1* or *FANCM*, cells were nucleofected with the triple-guide knockout strategy described above and treated 72 h later with Polθ-Pol inhibitor or vehicle (DMSO). Cells were collected 72 h after treatment for END-seq. END-seq was performed as previously described[54]. In brief, COLO320 cells were resuspended in cell suspension buffer (10 mM Tris-HCl, pH 7.2, 50 mM EDTA, 2 mM NaCl) and embedded in 0.75% low-melting-point agarose plugs using the Bio-Rad CHEF Mammalian Genomic DNA Plug Kit. Approximately 5×10^6 cells were used per plug, and two plugs were prepared per condition. Plugs were incubated in freshly prepared lysis buffer (10 mM Tris-HCl, pH 8.0, 50 mM EDTA, 150 mM NaCl, 1% SDS) containing Proteinase K (Qiagen) at 50 °C for 1 h, followed by overnight incubation at 37 °C. After lysis, plugs were rinsed and washed with plug wash buffer (10 mM Tris-HCl, pH 8.0, 50 mM EDTA) and TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), followed by treatment with RNase A (Qiagen) at 37 °C for 1 h. Additional washes in plug wash buffer were performed to remove residual enzymes. DNA DSB ends were blunted in-plug using Exonuclease VII and Exonuclease T, A-tailed with Klenow fragment (3′→5′ exo−), and ligated to a biotinylated hairpin adaptor. Plugs were then melted, digested with β-agarase I, and DNA was fragmented by sonication to an average size of ~175 bp. Biotinylated fragments were captured with streptavidin-coated beads, end-repaired, A-tailed, ligated to a second adaptor, treated with USER enzyme, and amplified by PCR with barcoded Illumina TruSeq primers. Libraries were purified with AMPure XP beads, size-selected by gel extraction, quantified, and sequenced on an Illumina platform[55].

END-seq analysis

Raw sequencing reads were trimmed using Trimmomatic (v0.39)[56]. Filtered reads were aligned to the human reference genome (hg19) using Bowtie2 (v2.2.5.1)[57]. Reads mapping to the mitochondrial genome or ENCODE blacklisted regions were excluded from downstream analysis. BAM files were converted to BED format using the bamtobedfunction in bedtools (v2.31.1)[58]. The ecDNA region was defined as chromosome 8:127433703–129010006 based on the strong enrichment of END-seq signal across this interval. Peaks were called using MACS2 (v2.2.7.1)[59] with the parameters–keep-dup all–nomodel–shift −50–extsize 100 -q 0.001–fe-cutoff 5. Annotations for simple repeats were downloaded from the UCSC Genome Browser and overlaps between repeat elements and END-seq peaks were identified using the intersect function in bedtools[60]. For visualization and downstream analysis, scaled coverage tracks normalized to reads per million were generated using BAMscale[61].

Direct library preparation

COLO320DM cells were used for DLP+ experiments. Four populations were prepared: the parental starting population (T0), a DMSO-treated control population cultured for 2 months, a population cultured in parallel for 2 months in the presence of a Polθi-Hel, and a *FANCM*-knockout population cultured for 1 month. The *FANCM* knockout was generated using CRISPR–Cas9-mediated targeting, and knockout efficiency was monitored weekly by genomic DNA extraction, PCR amplification of the targeted cut site, Sanger sequencing, and EditCo's ICE Analysis tool (v3). Single-cell whole-genome libraries were generated using DLP+, as described previously[62]. In this approach, individual cells are dispensed into nanowell chips containing preloaded primers and processed directly for single-cell library construction. Cells were subjected to lysis and protease treatment, followed by heat lysis, transposition-based fragmentation and adapter insertion, reaction neutralization and indexed PCR amplification. The resulting single-cell libraries were recovered, pooled, size-selected with AMPure XP beads, and sequenced. Single-cell DNA sequencing data was analysed using the publicly available Mondrian pipeline (https://github.com/mondrian-scwgs/mondrian) within the Isabl platform[63]. This included alignment using bwa-mem[64] and structural variant calling using DeStruct[65]. DeStruct was run jointly on all four samples. Analysis was restricted to those structural variants that were identified exclusively in a single sample, as these were thought to be newly arising somatic events. For events outside of the amplicon-implicated chromosomes (6, 8 and 13), only those events with at least 3 supporting reads were included.

Sequence analysis of the breakpoint of rearrangement junctions in human tumours

Focal amplifications in human tumour samples from PCAWG (Pancancer Analysis of Whole Genomes), Hartwig (Hartwig Medical Foundation)[41], PedPanCan (tumour samples from St Jude and PBTA), GLASS (Glioma Longitudinal AnalySiS) and CUGA (Chinese Urothelial Carcinoma Genome Atlas), detected and classified by Amplicon Architect and Amplicon Classifier, publicly available through the Amplicon Repository (https://ampliconrepository.org/), were analysed[41,42,66,67]. Only amplifications from whitelisted primary tumour samples were retained for the analysis.

To identify amplicon junctions occurring within TA repeat regions, junction coordinates were intersected with TA/AT-rich repeats (for example, $(TA)_n$, $(TATA)_n$, etc., including reverse complements), extracted from the UCSC Table Browser RepeatMasker track. The intersection was performed using BEDTools intersect v2.31.1. Each dataset was processed individually with the TA-rich repeat annotation from their corresponding reference genome: hg19 (PCAWG, GLASS) and hg38 (CUGA, PedPanCan).

A two-sided Pearson's chi-squared test (R base function prop.test v4.3.2) was used to determine whether there is a significant difference between the fraction of junctions overlapping TA repeats on ecDNA vs other focal amplification types.

MYC copy number assay using TaqMan

MYC gene amplification was evaluated using TaqMan copy number variation assays on a QuantStudio 6 Flex real-time PCR system.

Myc copy number in mouse neuronal stem cells was assessed using the TaqMan Copy Number Assay (probe Mm00734221_cn) and TaqMan Copy Number Reference Assay (Tfrc, 4458367). *MYC* copy number in human COLO320(DM and HSR) cells was assessed using the TaqMan Copy Number Assay (probe Hs01764918_cn) and TaqMan Copy Number Reference Assay (RnaseP, 4403326). Genomic DNA was isolated using the Zymo Quick-DNA Miniprep kit and amplified using the TaqPath ProAmp Master Mix (Applied Biosystems, A30865), following the supplier's instructions. Relative copy number variations were calculated using the $\Delta\Delta C_t$ method.

RT–qPCR to detect *MYC* expression

Total RNA was isolated using the NucleoSpin RNA Clean-up kit (Macherey-Nagel) following the manufacturer's instructions. Columns were treated with DNase I to degrade genomic DNA. 1 µg of purified RNA was reverse transcribed using the iScript gDNA Clear cDNA Synthesis Kit (Bio-Rad) following the suppliers' instructions. *MYC* expression was assessed using a 1:10 dilution of cDNA with PowerUp SYBR Green Master Mix (Applied Biosystems) and standard cycling conditions on a QuantStudio 6 Flex real-time PCR system. *MYC* expression was normalized to *ACTB* and quantified using the $\Delta\Delta C_t$ method.

RT–qPCR to detect *cre* recombinase expression

Total RNA was isolated using RNeasy Mini Kit (QIAGEN, 74106) following the manufacturer's instructions. After treatment with DNAse I (Ambion, AM2222), 0.5 µg of purified RNA was retro-transcribed with random hexamers by using SuperScript IV First-strand System (Invitrogen), *cre* recombinase expression was assessed using a 1:20 dilution of cDNA with PowerUp SYBR Green Master Mix (Applied Biosystems) and standard cycling conditions on a QuantStudio 6 Flex real-time PCR system. Cre recombinase expression was normalized to GAPDH and quantified using the $\Delta\Delta C_t$ method. Quantitative PCR with reverse transcription (RT–qPCR) primers are provided in Supplementary Table 3.

Reporting summary

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

Data availability

Sequencing data generated in this study have been deposited at the NCBI Sequence Read Archive under BioProject accessions PRJNA1505691 (DLP+ single-cell whole-genome sequencing of COLO320DM cells following *FANCM* knockout or DNA polymerase θ inhibition) and PRJNA1507177 (END-seq analysis of DNA DSBs in COLO320DM and COLO320HSR cells). Source data are available at Mendeley Data https://doi.org/10.17632/2vh4pg9y6n.1 (ref. 68).

Code availability

Custom code used for the DLP+ single-cell whole-genome sequencing analysis is available at Zenodo (https://doi.org/10.5281/zenodo.21725678 (ref. 69)). No other custom code was generated for this study. All remaining analyses were performed using publicly available software, and all scripts used for analysing sequencing data are available on GitHub as part of previously published studies. Specifically, END-seq data were processed as described previously[54] (https://github.com/Gang-Zhen/ENDseq). Repair-seq data were analysed using CRISPResso2 (https://github.com/pinellolab/CRISPResso2). Breakpoint junctions of ecDNA and non-ecDNA amplicons were obtained using AmpliconArchitect (https://github.com/AmpliconSuite/AmpliconSuite-pipeline) and AmpliconClassifier (https://github.com/AmpliconSuite/AmpliconClassifier). Amplicon junction coordinates were intersected with TA/AT-rich repeat annotations using BEDTools v2.31.1 (https://github.com/arq5x/bedtools2).

49. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 https://doi.org/10.1093/nar/gku936 (2014).
50. Bajar, B. T. et al. Fluorescent indicators for simultaneous reporting of all four cell cycle phases. *Nat. Methods* **13**, 993–996 https://doi.org/10.1038/nmeth.4045 (2016).
51. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 https://doi.org/10.1038/nmeth.2019 (2012).
52. Guzman, C., Bagga, M., Kaur, A., Westermarck, J. & Abankwa, D. ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. *PLoS ONE* **9**, e92444 https://doi.org/10.1371/journal.pone.0092444 (2014).
53. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 https://doi.org/10.1038/s41587-019-0032-3 (2019).
54. Wong, N., John, S., Nussenzweig, A. & Canela, A. in *Homologous Recombination* (eds Aguilera, A. & Carreira, A.) 9–31 https://doi.org/10.1007/978-1-0716-0644-5_2 (Springer 2021).
55. Tubbs, A. et al. Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **174**, 1127–1142.e1119 https://doi.org/10.1016/j.cell.2018.07.011 (2018).
56. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 https://doi.org/10.1093/bioinformatics/btu170 (2014).
57. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 https://doi.org/10.1038/nmeth.1923 (2012).
58. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 https://doi.org/10.1093/bioinformatics/btq033 (2010).
59. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 https://doi.org/10.1186/gb-2008-9-9-r137 (2008).
60. Casper, J. et al. The UCSC Genome Browser database: 2026 update. *Nucleic Acids Res.* **54**, D1331–D1335 https://doi.org/10.1093/nar/gkaf1250 (2026).
61. Pongor, L. S. et al. BAMscale: quantification of next-generation sequencing peaks and generation of scaled coverage tracks. *Epigenet. Chromatin* **13**, 21 https://doi.org/10.1186/s13072-020-00343-x (2020).
62. Laks, E. et al. Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. *Cell* **179**, 1207–1221.e1222 https://doi.org/10.1016/j.cell.2019.10.026 (2019).
63. Medina-Martínez, J. S. et al. Isabl Platform, a digital biobank for processing multimodal patient data. *BMC Bioinform.* **21**, 549 https://doi.org/10.1186/s12859-020-03879-7 (2020).
64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 https://doi.org/10.1093/bioinformatics/btp324 (2009).
65. McPherson, A., Shah, S. & Sahinalp, S. C. deStruct: accurate rearrangement detection using breakpoint specific realignment. Preprint at *bioRxiv* https://doi.org/10.1101/117523 (2017).
66. Grobner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 https://doi.org/10.1038/nature25480 (2018).
67. Lv, W. et al. Spatial-temporal diversity of extrachromosomal DNA shapes urothelial carcinoma evolution and tumor-immune microenvironment. *Cancer Discov.* **15**, 1225–1246 https://doi.org/10.1158/2159-8290.CD-24-1532 (2025).
68. Sfeir, A. MMEJ repair of breaks at TA repeats maintains ecDNA and cancer fitness. *Mendeley Data* https://doi.org/10.17632/2vh4pg9y6n.1 (2026).
69. Myers, M., McPherson, A., Sfeir, A. & Selvaraj, M. Code to generate figures from DLP+ data, supporting "MMEJ Repair of Breaks at TA Repeats Maintains ecDNA and Cancer Fitness". *Zenodo* https://doi.org/10.5281/zenodo.21725678 (2026).
70. Ciszewski, W. M., Tavecchio, M., Dastych, J. & Curtin, N. J. DNA-PK inhibition by NU7441 sensitizes breast cancer cells to ionizing radiation and doxorubicin. *Breast Cancer Res. Treat.* **143**, 47–55 https://doi.org/10.1007/s10549-013-2785-6 (2014).

Acknowledgements We thank D. Cleveland, O. Shoshani and H. Chang for support and providing reagents; Sfeir laboratory members for comments on the manuscript; and staff at the National Cancer Institute/Center for Cancer Research Genomics Core for help with sequencing. The computational resources of the NIH High-Performance Computational Biowulf cluster were used for data analysis. The contributions of the NIH authors are considered works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services.

Author contributions D.B. and A.S. conceived the study and designed the experiments. D.B. and M.S. performed the majority of the experimental work, with help from M.E.K., B.B., O.E., Y.-Z.J., S.-C.W. and L.W. M.S. performed END-seq experiments together with G.M.-R., G.Z. and E.C. A. Nussenzweig supervised and interpreted the END-seq analyses. A.B. performed and analysed the Repair-seq experiments. D.P. analysed *MYC* ecDNA levels in genetically engineered cells, supervised by A.V. A. Nichols performed live-cell ecDNA imaging experiments, supervised by J.M. F.B. and E.R.-F. analysed ecDNA breakpoints from whole-genome sequencing of human tumours, supervised by A.G.H. M.-C.M., H.P., S.J.M., M. Zimmermann and M. Zinda designed, performed and interpreted the experiments to develop RP-2119. M.A.M. performed the single-cell whole-genome sequencing (DLP+) analysis, with A.M. and S.P.S. contributing to analysis design and interpretation. A.S. supervised the study and acquired funding. A.S. wrote the manuscript, with input and feedback from all authors.

Funding This work is directly supported by grants from the NIH/NCI (R01CA304441) for A.S. and NIH/NCI (F32CA298730) for M.E.K. Work on MMEJ in the A.S. lab is supported by the V Foundation (AST2025-001), NIH/NCI (R01CA294696), and Breast Cancer Alliance grants. Work in the J.M. laboratory is supported by NIH/NCI grants R37CA261183 and R01CA270102. Work in the S.P.S. laboratory is supported by Halvorsen Center for Computational Oncology and NIH/NCI grants RM1HG011014, R01CA281928 and U24CA264028, the Susan G. Komen Foundation and the Breast Cancer Research Foundation. We acknowledge the use of the

Integrated Genomics Operation Core, funded by the NCI Cancer Center Support Grant (CCSG, P30 CA08748) and the Flow Cytometry Core Facility, both funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748 (RRID: SCR_021105). This work also utilized resources from the High-Performance Computing Group at Memorial Sloan Kettering Cancer Center. A.V. is supported by the Cancer Grand Challenges partnership funded by Cancer Research UK (CGCATF-2021/100025) and the National Cancer Institute (OT2CA301085), NIH/NCI R01CA282913, The Mark Foundation for Cancer Research ASPIRE II grant, and the American Cancer Society Discovery Boost grant. Work in the A.G.H. laboratory is supported by the Cancer Grand Challenges partnership funded by Cancer Research (CGCATF-2021/100017), the National Cancer Institute (OT2CA278644), the Deutsche Krebshilfe (German Cancer Aid) Mildred Scheel Professorship programme–70114107, and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the Collaborative Research Center CRC1588, project number 493872418. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 949172). This research in the A. Nussenzweig laboratory was supported in part by the Intramural Research Program of the National Institutes of Health (NIH) (National Cancer Institute contract HHSN2612015000031). F.B. is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) RTG2424/CompCancer, project number 377984878.

Competing interests A.S. is a co-founder, consultant, and shareholder for Repare Therapeutics. M.-C.M., H.P., S.J.M., M. Zimmermann and M. Zinda are current or former employees of Repare Therapeutics and receive salary and equity compensation. S.P.S. receives research funding from AstraZeneca and Bristol Myers Squibb, unrelated to this work. A.G.H. is a founder, consultant and shareholder for Econic Biosciences. The other authors declare no competing interests.

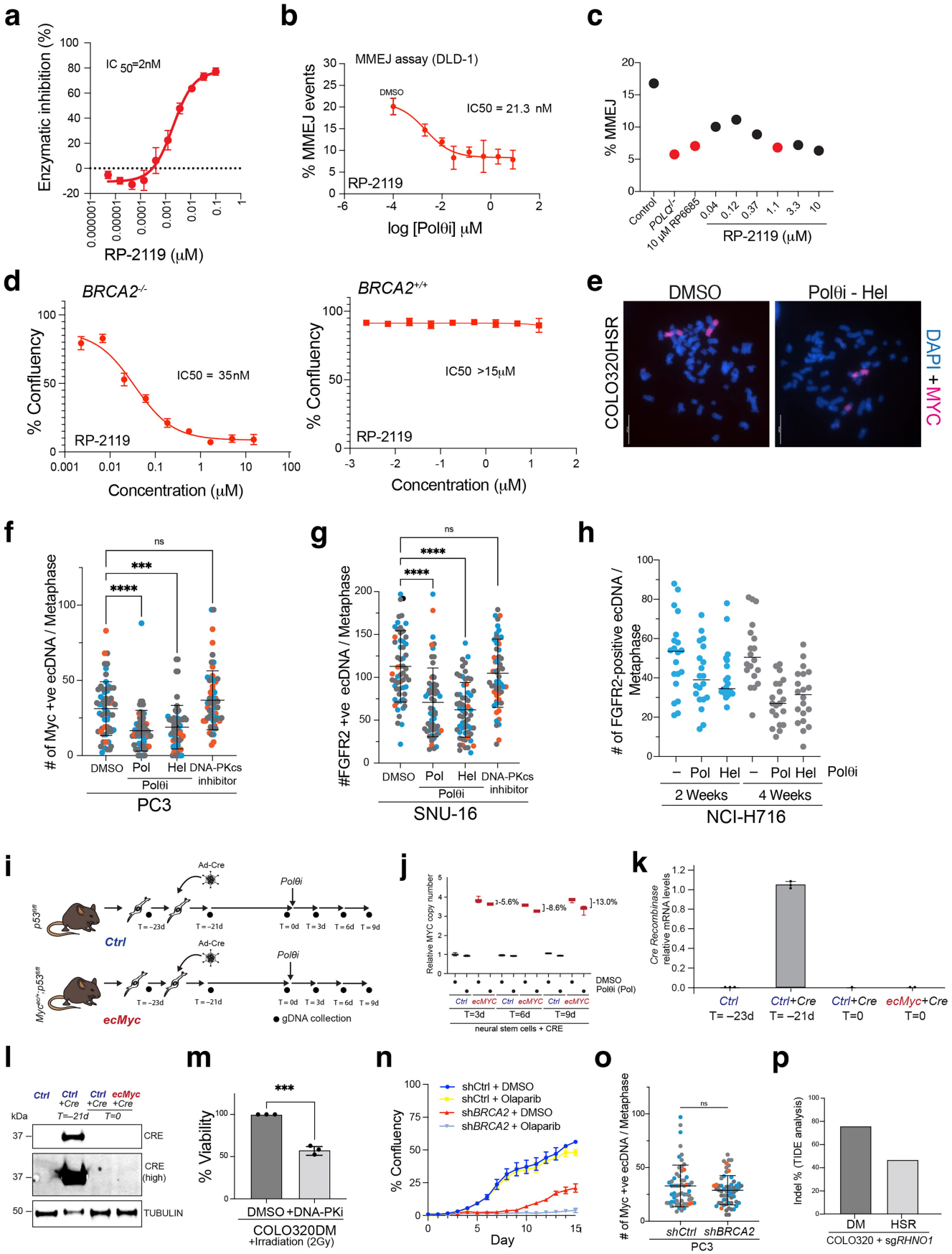
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-11048-8.

Correspondence and requests for materials should be addressed to Agnel Sfeir.

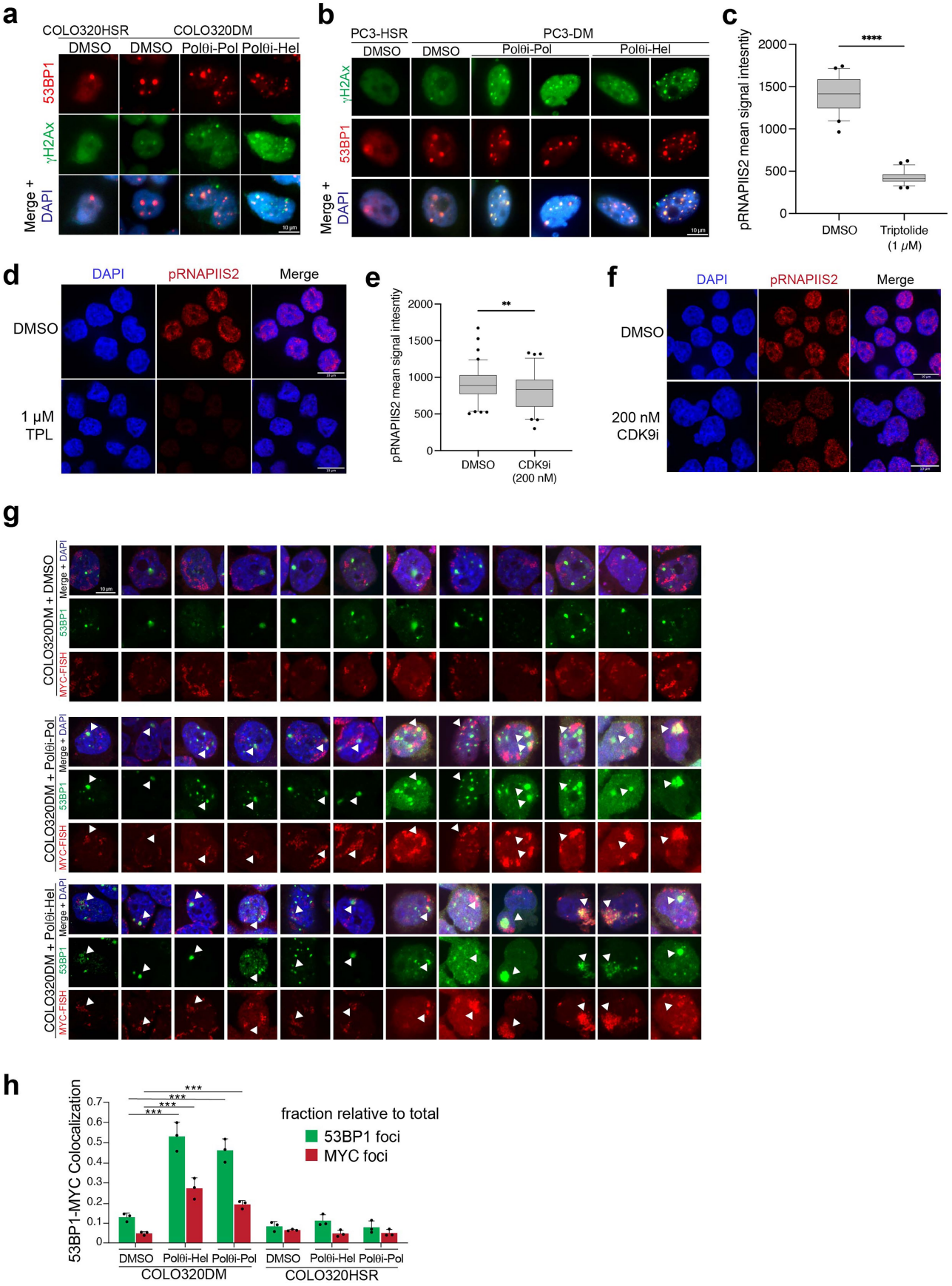
Peer review information *Nature* thanks Jan Korbel who co-reviewed with Nikolaus Watson; Sergei Mirkin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.



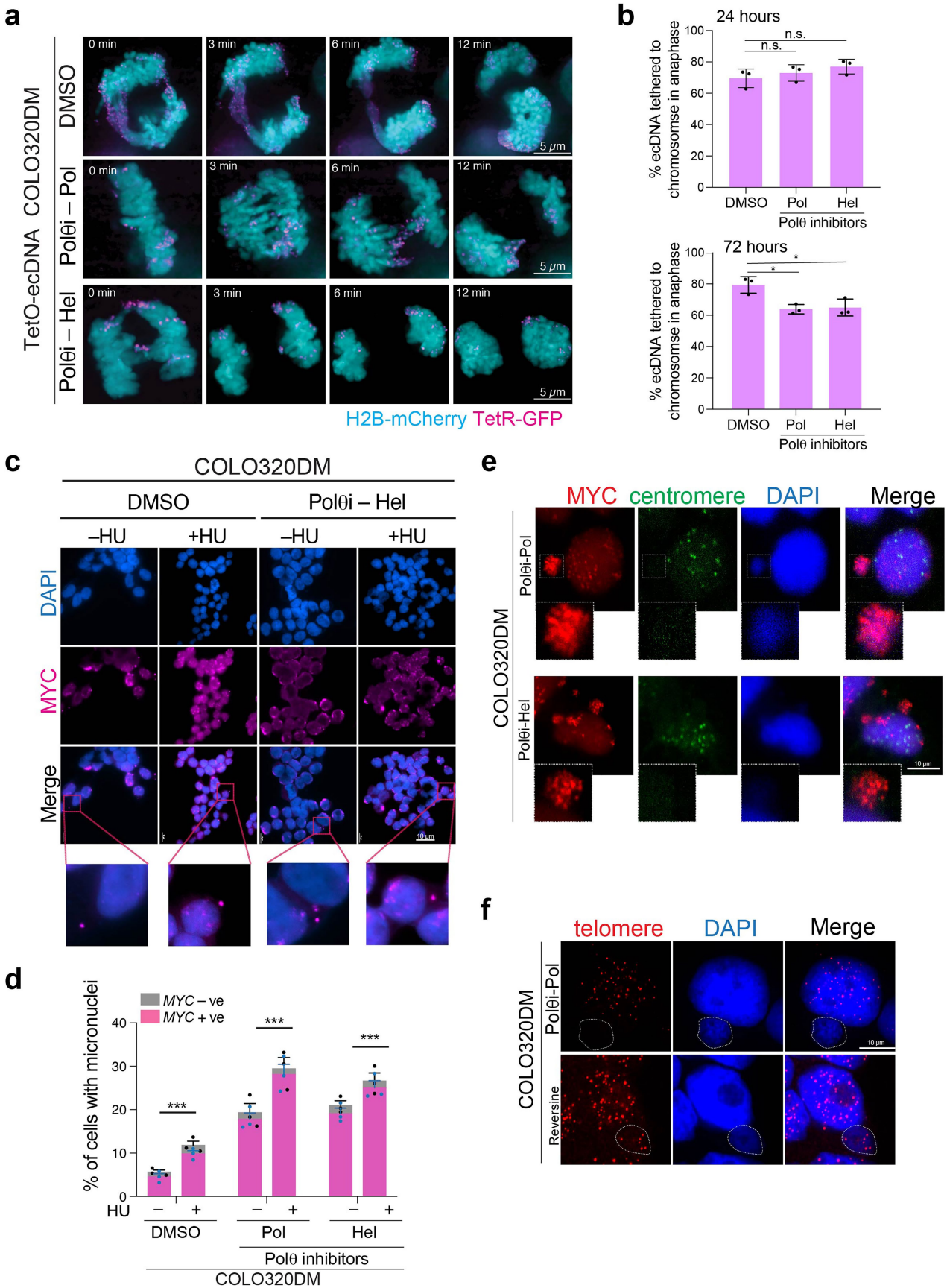
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | The effect of Polθ inhibition on ecDNA maintenance. **a**. In vitro dose-response curve showing enzymatic inhibition of Polθ ATPase activity by RP-2119. IC_{50} = 2 nM, based on nonlinear regression analysis; mean of n = 3 biological experiments +/− s.d. **b**. Percentage of MMEJ events measured in DLD-1 reporter cells as a function of RP-2119 concentration (log scale); mean of n = 3 independent experiments +/− s.d. IC_{50} = 21.3 nM. **c**. Quantification of MMEJ activity in DLD-1 cells treated with increasing concentrations of RP-2119, compared to *POLQ*$^{-/-}$ cells and cells treated with the previously characterized Polθ inhibitor RP-6685 (10 μM). Data from n=1 biological replicate. **d**. Selective sensitivity of BRCA2-deficient cells to RP-2119. Left, dose-response curve showing reduced viability of *BRCA2*$^{-/-}$ DLD-1 cells upon RP-2119 treatment (IC_{50} = 35 nM). Right, *BRCA2*$^{+/+}$ isogenic control cells exhibit minimal sensitivity (IC_{50} > 15 μM). Cell confluency was monitored over a 7-day period; mean of n = 3 independent experiments +/− s.d. **e**. Representative metaphase FISH image from COLO320HSR cells treated with DMSO or Polθi-Hel for one week, with *MYC* signal shown in pink and chromosomes counterstained with DAPI (blue). **f**. Quantification of *MYC*-positive ecDNA per metaphase in PC3 cells treated with DMSO, Polθi-Pol, Polθi-Hel, or DNA-PKcs inhibitor (NU7441). **g**. Quantification of *FGFR2*-positive ecDNA per metaphase in SNU-16 cells following the indicated treatment for one week. **h**. Quantification of *FGFR2*-positive ecDNA foci per metaphase in NCI-H716 cells treated with Polθi-Pol or Polθi-Hel for 2 or 4 weeks. Data in f-h represent *n* = 1-3 biological replicates with ≥20 metaphases quantified per replicate, each dot represents the number of ecDNA in an individual metaphase, horizontal bars indicate mean values and statistical significance determined by ordinary one-way ANOVA with multiple comparisons (**** $p < 0.0001$, *** $p < 0.001$) **i**. Experimental schematic illustrating the inducible ecDNA model used to assess the effect of Polθ inhibition on ecDNA maintenance after de novo ecDNA formation in engineered neuronal stem cells. In this system, a conditional *MYC* allele flanked by loxP sites is excised upon Cre recombinase activation by adenoviral delivery (Ad-Cre), generating extrachromosomal circular *MYC* and deleting the chromosomal Cre locus. Polθ inhibitor treatment was initiated > 3 weeks after Ad-Cre treatment, and genomic DNA was collected at the indicated time points. This design allows the effect of Polθ inhibition to be assessed specifically on ecDNA generated by Cre-mediated excision. **j**. qPCR-based quantification of relative *MYC* copy number in control and ecDNA-containing engineered neuronal stem cells[25] following Polθ inhibitor treatment across days 3, 6, and 9. The percent reduction relative to the control is noted. mean of *n* = 3 biological replicates +/− s.d. **k**. RT–qPCR quantification of Cre recombinase mRNA levels in *MYC*$^{F/F}$ neuronal stem cells at the time of Polθ inhibitor treatment initiation (T = 0); mean of n = 3 biological replicates +/− s.d. **l**. Immunoblot confirming the absence of Cre recombinase protein at the time of Polθ inhibitor treatment initiation. Tubulin serves as a loading control. High- and low-exposure images are shown. **m**. Validation of DNA-PKcs inhibition efficacy. NU7441 is known to sensitize cells to ionizing radiation[70]. Survival was normalized to the vehicle-treated cells; mean of n = 3 biological replicates +/− s.e.m **n**. Proliferation curves of COLO320DM cells with control or BRCA2 knockdown treated with DMSO or Olaparib. Cell confluency was measured over time using IncuCyte live-cell imaging. Mean of n = 3 biological replicates with three technical replicates per experiment, +/− s.e.m. **o**. Quantification of *MYC*-positive ecDNA per metaphase in PC3-DM cells expressing control or *BRCA2* shRNA, mean of *n* = 3 biological replicates, with ≥20 metaphases quantified per replicate, each dot represents an individual ecDNA count +/− s.d. Statistical significance determined by unpaired two-tailed *t*-test (ns, not significant). **p**. TIDE analysis quantifying genome editing efficiency following *RHNO1* targeting in COLO320DM and COLO320HSR cells.



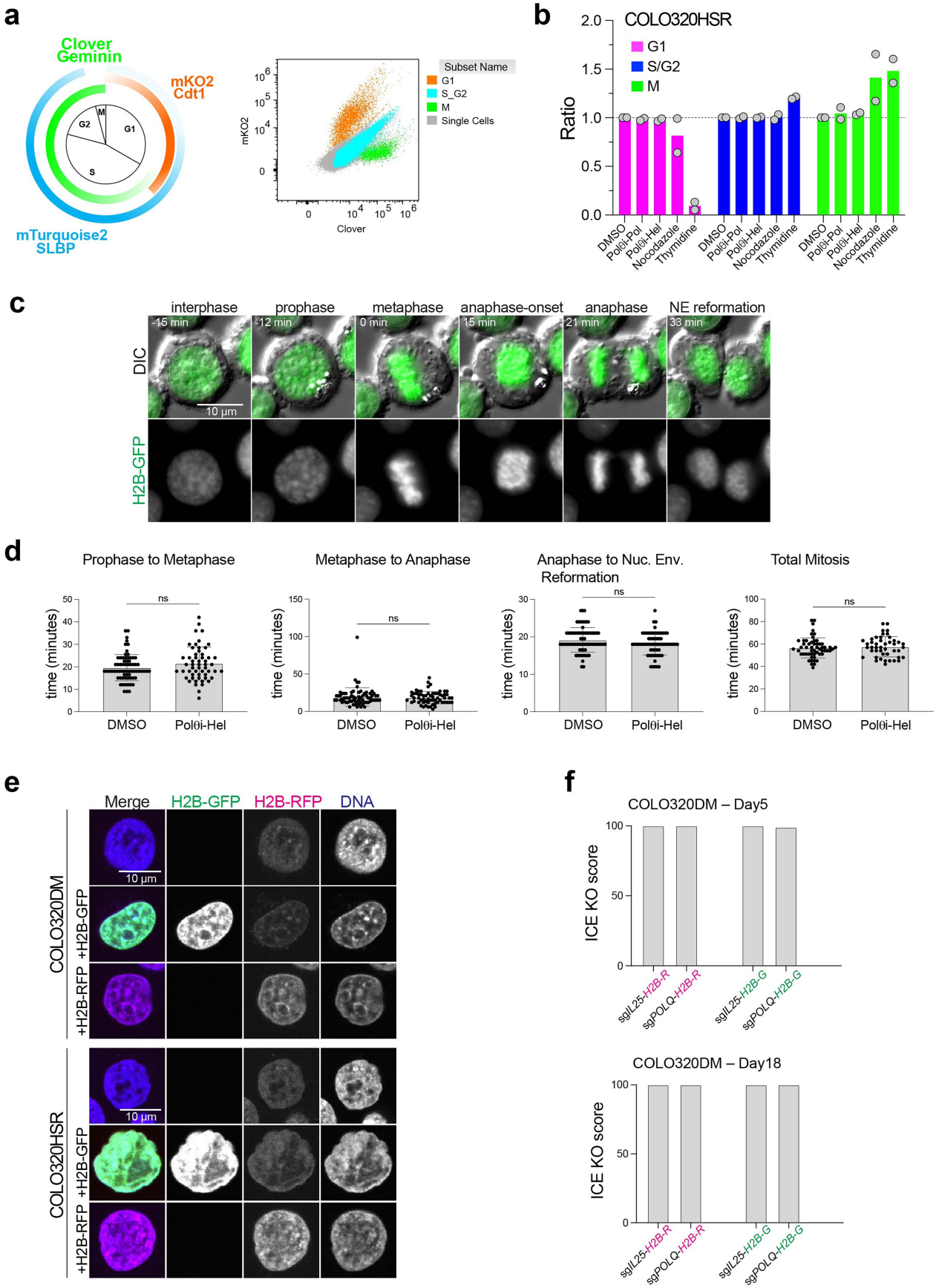
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Polθ inhibition increases DNA damage signaling at ecDNA. a. Representative immunofluorescence (IF) images of interphase COLO320DM and COLO320HSR cells treated with DMSO, Polθi-Pol, or Polθi-Hel, stained for 53BP1 (red), γH2AX (green), and DAPI (blue) **b**. Representative IF images of PC3-DM and PC3-HSR cells treated with DMSO, Polθi-Pol, or Polθi-Hel. Cells were stained for γH2AX (green) and 53BP1 (red); nuclei are counterstained with DAPI (blue). Merged images are shown in the bottom row. **c**. Quantification of pRNAPII S2 mean signal intensity in COLO320DM cells treated with DMSO or triptolide (1 μM). **d**. Representative IF images of cells treated with DMSO or triptolide (1 μM), stained for pRNAPII S2 (red) and DAPI (blue). **e**. Quantification of pRNAPII S2 mean signal intensity in cells treated with DMSO or CDK9i (200 nM). Data in c and e are from n=1 biological replicate (>40 nuclei per condition) presented as box-and-whisker plots. Boxes delineate 25th and 75th percentiles, whiskers extend from 5th to 95th percentiles and the centre line indicates the median. Statistical significance was determined using Welch's *t*-test (**** $p < 0.0001$, ** $p < 0.01$). **f**. Representative IF images of cells treated with DMSO or CDK9i (200 nM), stained for pRNAPII S2 (red) and DAPI (blue). **g**. Representative immuno-FISH images of COLO320DM cells treated with DMSO, Polθi-Pol, or Polθi-Hel, showing *MYC* ecDNA detected by FISH (red) and 53BP1 immunofluorescence (green). Nuclei are counterstained with DAPI (blue). Representative images are shown for each condition. **h**. Quantification of 53BP1 focus colocalization with *MYC* signal in COLO320DM and COLO320HSR cells under the indicated treatment conditions. The number of colocalizing foci was normalized separately to the total number of 53BP1 foci (green bars) and to the total number of *MYC* foci (red bars), providing a measure of the fraction of DNA damage events occurring at ecDNA loci and the fraction of ecDNA loci marked by DNA damage, respectively. Bars depict mean ± s.e.m. from $n = 3$ biological replicates (>100 nuclei per replicate). Statistical significance was assessed by unpaired *t*-tests (*** $p < 0.001$).



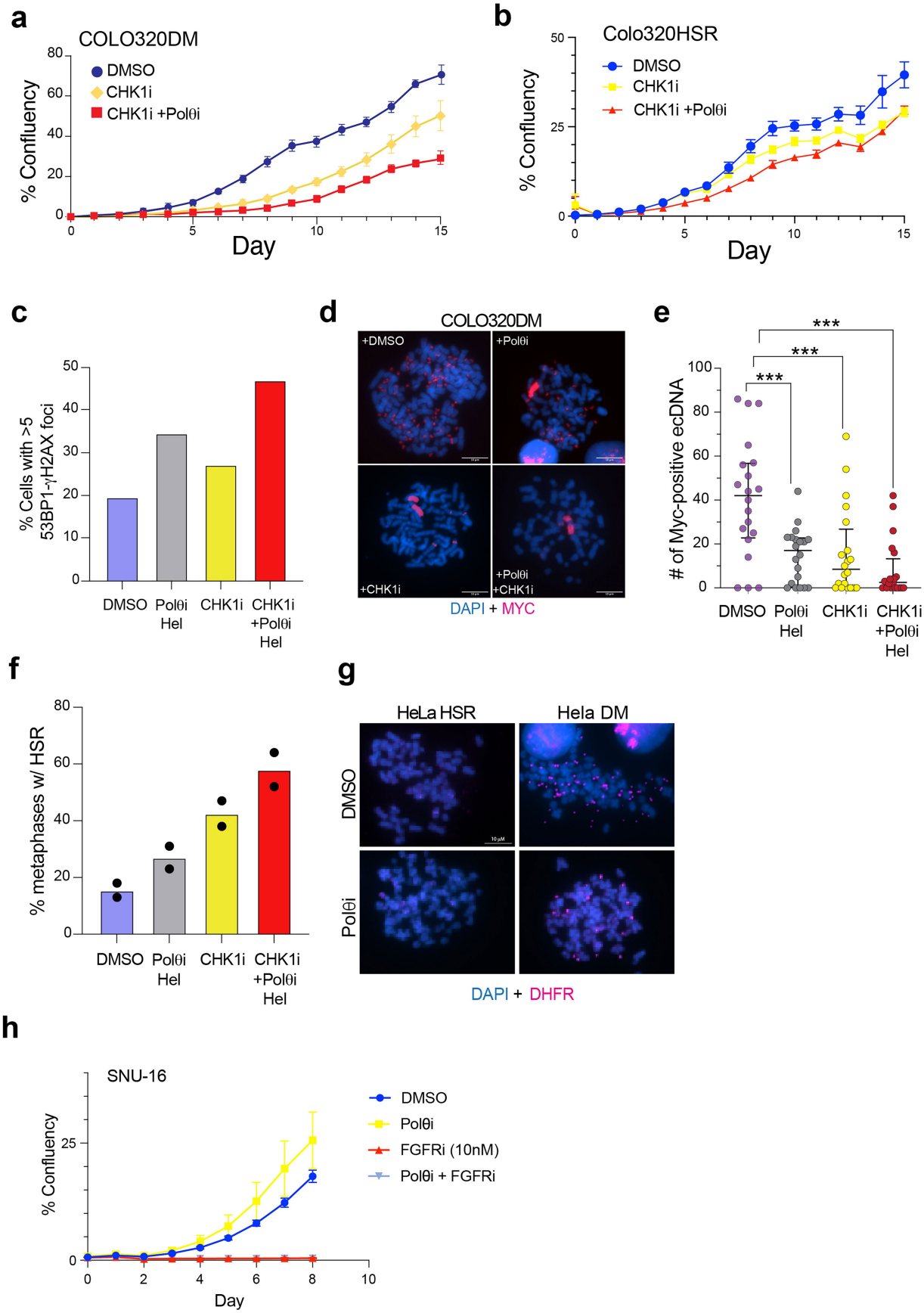
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Polθ inhibition promotes ecDNA sequestration into micronuclei. a. Time-lapse imaging of ecDNA dynamics in COLO320DM cells expressing TetO-tagged ecDNA and H2B-mCherry. Cells were treated with DMSO, Polθi-Pol, or Polθi-Hel and imaged during mitosis. Representative frames are shown at the indicated time points. Scale bars, 5 µm. **b**. Quantification of ecDNA tethering to chromosomes during anaphase at 24 h (top) and 72 h (bottom) following treatment with DMSO, Polθi-Pol, or Polθi-Hel. Data are presented as mean ± s.d. from three independent experiments. One-way ANOVA with multiple comparisons (* $p < 0.05$; ns, not significant). **c**. Representative interphase FISH images of COLO320DM cells treated with DMSO or Polθi-Hel, in the absence or presence of HU (added 24 h before harvest). *MYC* ecDNA is shown in magenta, and nuclei are counterstained with DAPI (blue). Insets highlight *MYC*-positive micronuclei. **d**. Quantification of micronucleated cells in COLO320DM cultures treated with DMSO or Polθi-Hel, with or without HU. Micronuclei were scored for the presence (MYC^+) or absence (MYC^-) of *MYC* signal. Data represent mean ± s.d. from $n = 3$ biological replicates. Statistical significance determined by two-way ANOVA (*** $p < 0.001$). **e**. Representative interphase images of COLO320DM cells treated with Polθi-Pol or Polθi-Hel, stained for *MYC* FISH (red), centromere FISH (green), and DAPI (blue), showing *MYC*-positive micronuclei lacking centromeric signal. **f**. Telomere FISH analysis of COLO320DM cells treated with Polθi-Pol, showing telomere signal (red) and DAPI (blue). Dashed outlines indicate micronuclei; *MYC*-positive micronuclei lack detectable telomeric sequences.



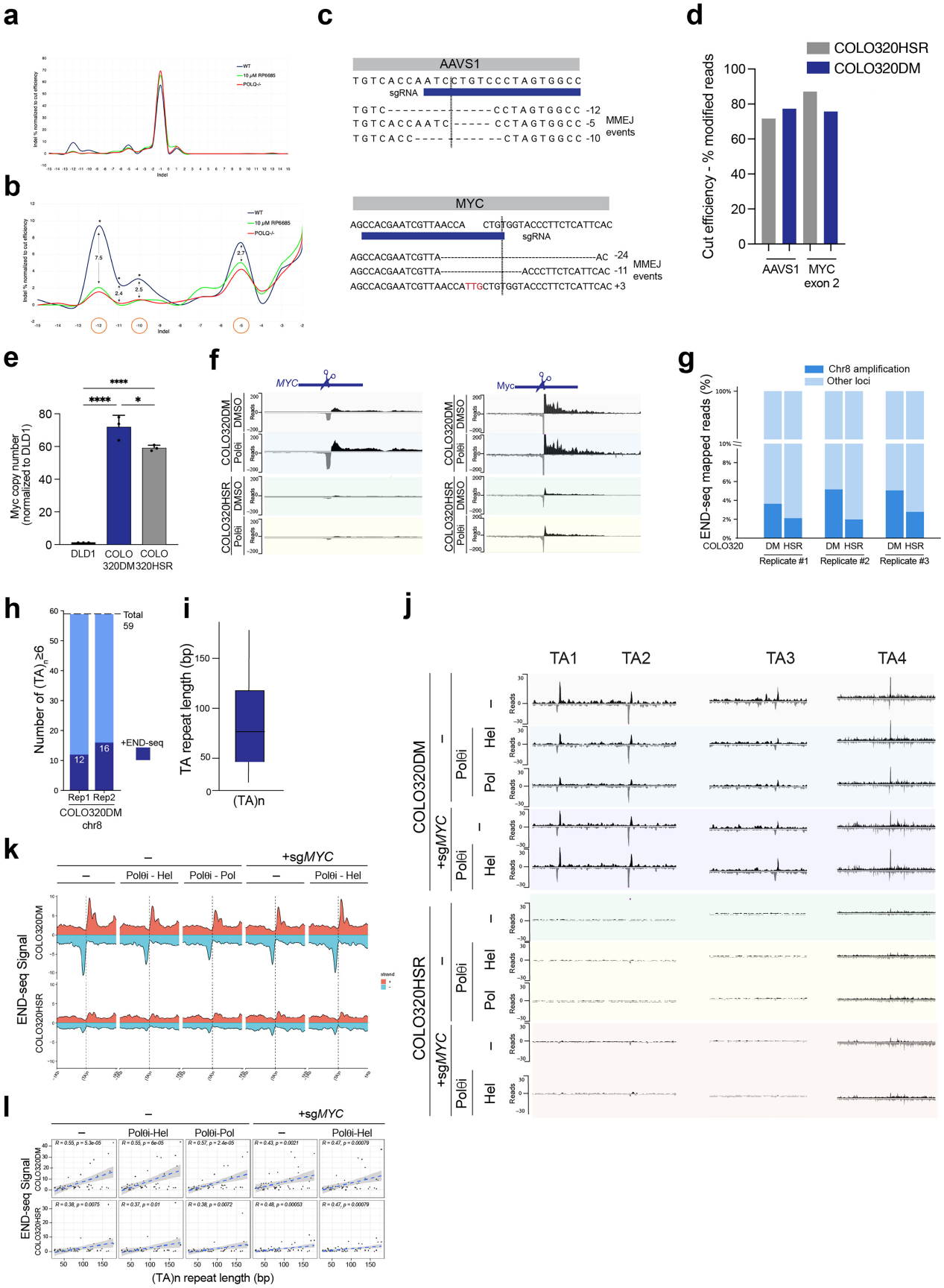
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Cell cycle analysis and competitive growth following Polθ inhibition. a. Schematic of the Fucci4 cell-cycle reporter system, left, and a representative flow cytometry dot plot showing the respective cell-cycle phases (G1, S/G2, and M). Uncategorized cells in gray. **b**. Cell-cycle distribution of COLO320HSR cells following treatment with vehicle, Polθ inhibitors, or control compounds, determined by FUCCI cell-cycle reporter analysis. Bars are mean ± s.d. with $n = 2$ biological replicates per condition, normalized to the DMSO control for each cell-cycle phase. **c**. Representative time-lapse images of COLO320DM cells expressing H2B-GFP progressing through mitosis, showing prophase, metaphase, anaphase-onset, anaphase, and nuclear envelope reformation. Scale bar, 10 µm. **d**. Quantification of mitotic timing in COLO320DM cells treated with DMSO or Polθi-Hel, including prophase-to-metaphase, metaphase-to-anaphase, anaphase-to-nuclear envelope reformation, and total mitosis duration. Each dot represents one cell (≥58 cells per condition); mean ± s.d.; $n = 2$ independent experiments per condition. Distributions were not normal by the Shapiro-Wilk test, and statistical significance was therefore evaluated using a two-tailed Mann-Whitney test (ns, not significant). **e**. Representative images of COLO320DM and COLO320HSR cells expressing H2B-GFP or H2B-RFP for competitive co-culture with sg*POLQ* or sg*IL25*, used to assess relative fitness. **f**. ICE KO scores quantifying CRISPR/Cas9 editing efficiency of *POLQ* in COLO320DM cells at day 5 (top) and day 18 (bottom) following sgRNA transfection.



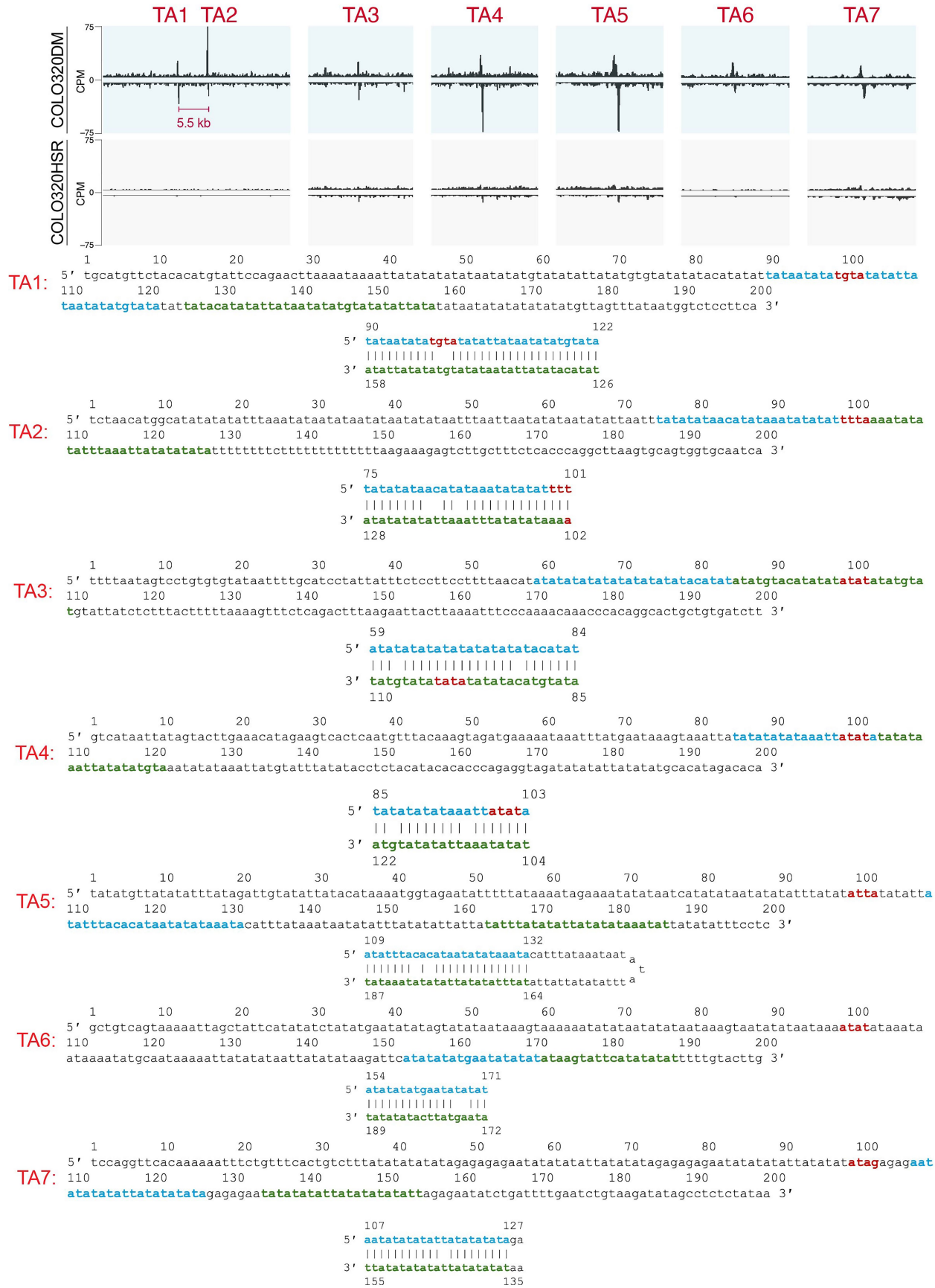
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Combined Polθ and Chk1 inhibition selectively impairs the growth of ecDNA-harboring cells. a. Proliferation of COLO320DM cells treated with DMSO, Chk1 inhibitor (Chk1i), or Chk1i in combination with Polθi-Hel, measured by IncuCyte live-cell imaging. **b**. Proliferation of COLO320HSR cells treated with DMSO, Chk1i, or Chk1i combined with Polθi-Hel, measured by IncuCyte live-cell imaging. Data in a,b represent the mean ± s.e.m. of n=3 biological replicates. **c**. Quantification of COLO320DM cells containing ≥ 5 co-localized 53BP1–γH2AX foci following treatment with DMSO, Polθi-Hel, Chk1i, or the combination for 72 h. Data from n=1 biological replicate. **d**. Representative metaphase spreads from COLO320DM cells treated with DMSO, Polθi-Hel, Chk1i, or the combination for one week, with *MYC* ecDNA detected by FISH (red) and chromosomes counterstained with DAPI (blue). **e**. Quantification of *MYC*-positive ecDNA foci per metaphase in COLO320DM cells under the conditions shown in d. Each dot represents one metaphase; horizontal bars indicate mean values. n = 3 biologic replicates; statistical significance determined by one-way ANOVA with multiple comparisons; (*** $p < 0.001$). **f**. Quantification of the percentage of metaphases containing HSRs in COLO320DM cells following treatment with DMSO, Polθi-Hel, Chk1i, or the combination for one week. Data represent n=2 replicates. **g**. Representative metaphase FISH images of HeLa-HSR and HeLa-DM cells stained for *DHFR* (red) and DAPI (blue), showing the effect of Polθi-Hel on *DHFR* ecDNA levels. **h**. Proliferation of SNU-16 cells treated with DMSO, Polθi-Hel, FGFR inhibitor infigratinib (10 nM), or infigratinib combined with Polθi-Hel, measured by IncuCyte live-cell imaging, n = 3 biological replicates per condition and three technical replicates per experiment.



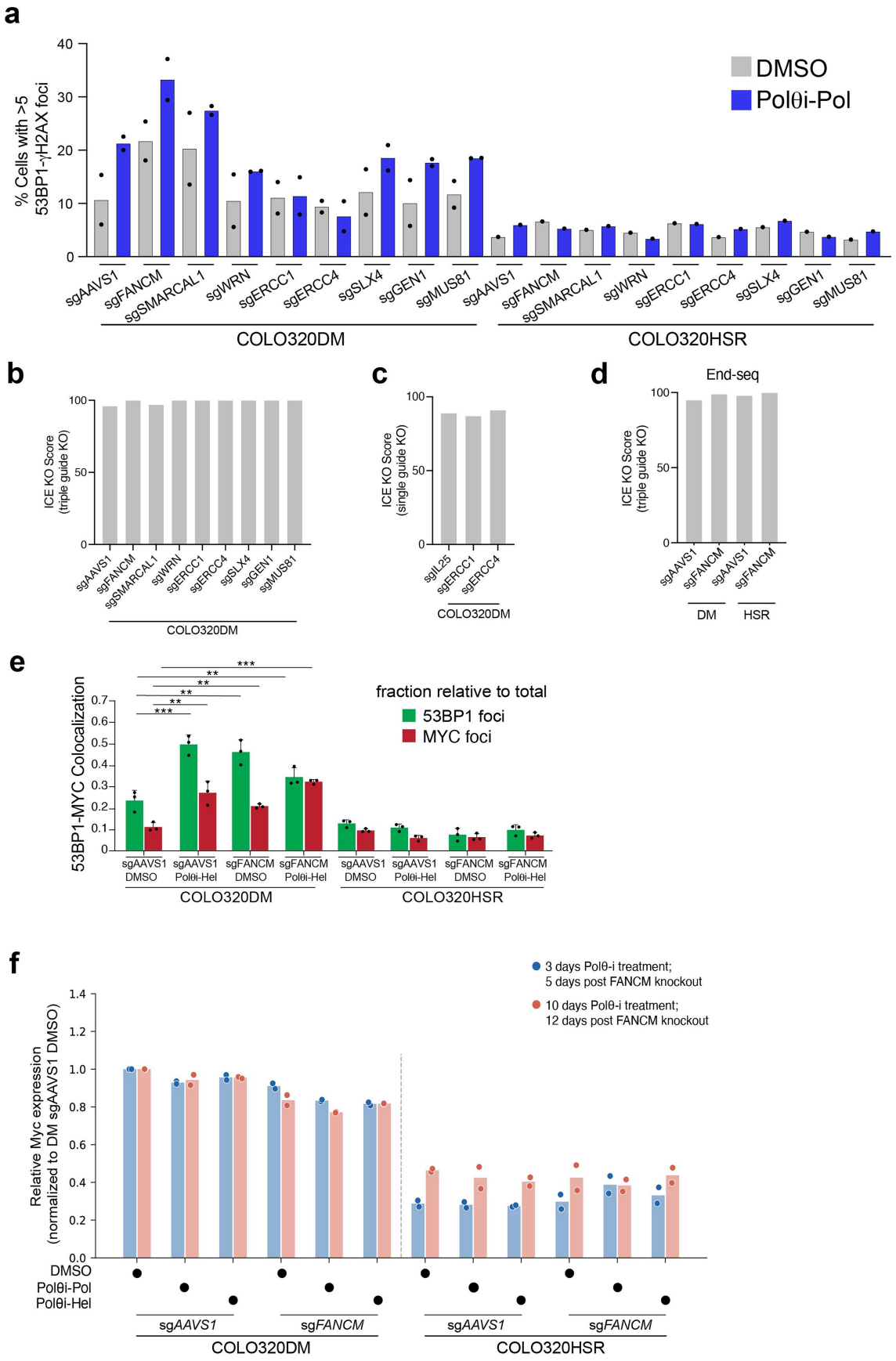
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Break distribution and MMEJ signatures at AAVS1 and ecDNA-associated *MYC* in ecDNA versus HSR cells. a. Frequency distribution of indels centered on the *AAVS1* CRISPR-Cas9 cut site in DLD1 wild-type (WT) and *POLQ*$^{-/-}$ cells, or in WT cells treated with Polθi-Pol. Indel profiles were derived from amplicon sequencing and normalized to editing efficiency. Positive x-axis values indicate insertions and negative values indicate deletions. **b**. Magnified view of deletions ranging from 2 to 15 bp, highlighting deletions enriched in WT cells (blue) but diminished in *POLQ*-deficient (red) and Polθi-Pol-treated (green) samples. Amplicons lost in the *POLQ*-null setting were classified as MMEJ-associated signatures at the *AAVS1* locus and are marked by black asterisks and circles. **c**. Representative alignments of MMEJ-characteristic indels at the *AAVS1* and *MYC* exon 2 loci in reads from COLO320DM and COLO320HSR cells. The sgRNA target site is shown above in blue, dashed lines indicate deletions, and insertions are shown in red. **d**. Quantification of the percentage of modified reads at the *AAVS1* and *MYC* exon 2 loci in COLO320DM and COLO320HSR cells, with comparable CRISPR–Cas9 cutting efficiencies across loci and cell lines. **e**. *MYC* copy number measured by qPCR in DLD-1, COLO320DM, and COLO320HSR cells, normalized to DLD-1. Data represent mean ± s.d. from n=3 biological replicates; ** $p < 0.01$, ns, not significant, ordinary one-way ANOVA with multiple comparisons. **f**. END-seq tracks at the *MYC* locus in COLO320DM and COLO320HSR cells following CRISPR/Cas9 targeting of *MYC* with DMSO or Polθ inhibitor treatment, confirming assay sensitivity. Data shown for two independent biological replicates. **g**. Quantification of END-seq mapped reads at the chr8 amplicon versus other genomic loci across three independent biological replicates in COLO320DM and COLO320HSR cells. **h**. Quantification of TA repeats [$(TA)_n \geq 6$] within the chr8 amplicon that overlap with END-seq peaks in two biological replicates. In sum, of the 59 TA-rich loci, up to 16 regions overlap with END-seq peaks. **i**. Boxplot representing the distribution of TA repeat tract lengths (bp) within the chr8 amplicon. n = 2 biological replicates. **j**. Zoomed-in genome browser views of END-seq signal at four individual TA repeat-containing loci (TA1–TA4) within the chr8 amplicon in COLO320DM and COLO320HSR cells, across DMSO, Polθi-Hel, and Polθi-Pol treatment conditions, with or without sg*MYC*. **k**. Aggregate END-seq signal across all TA-rich loci within the chr8 amplicon, computed by summing END-seq peaks at the TA repeat-containing sites in COLO320DM and COLO320HSR cells, with or without sg*MYC*, Polθi-Hel, or Polθi-Pol treatment. **l**. Scatter plots showing the correlation between TA repeat tract length and END-seq signal intensity in COLO320DM and COLO320HSR cells across the indicated treatment conditions, with and without sg*MYC*.



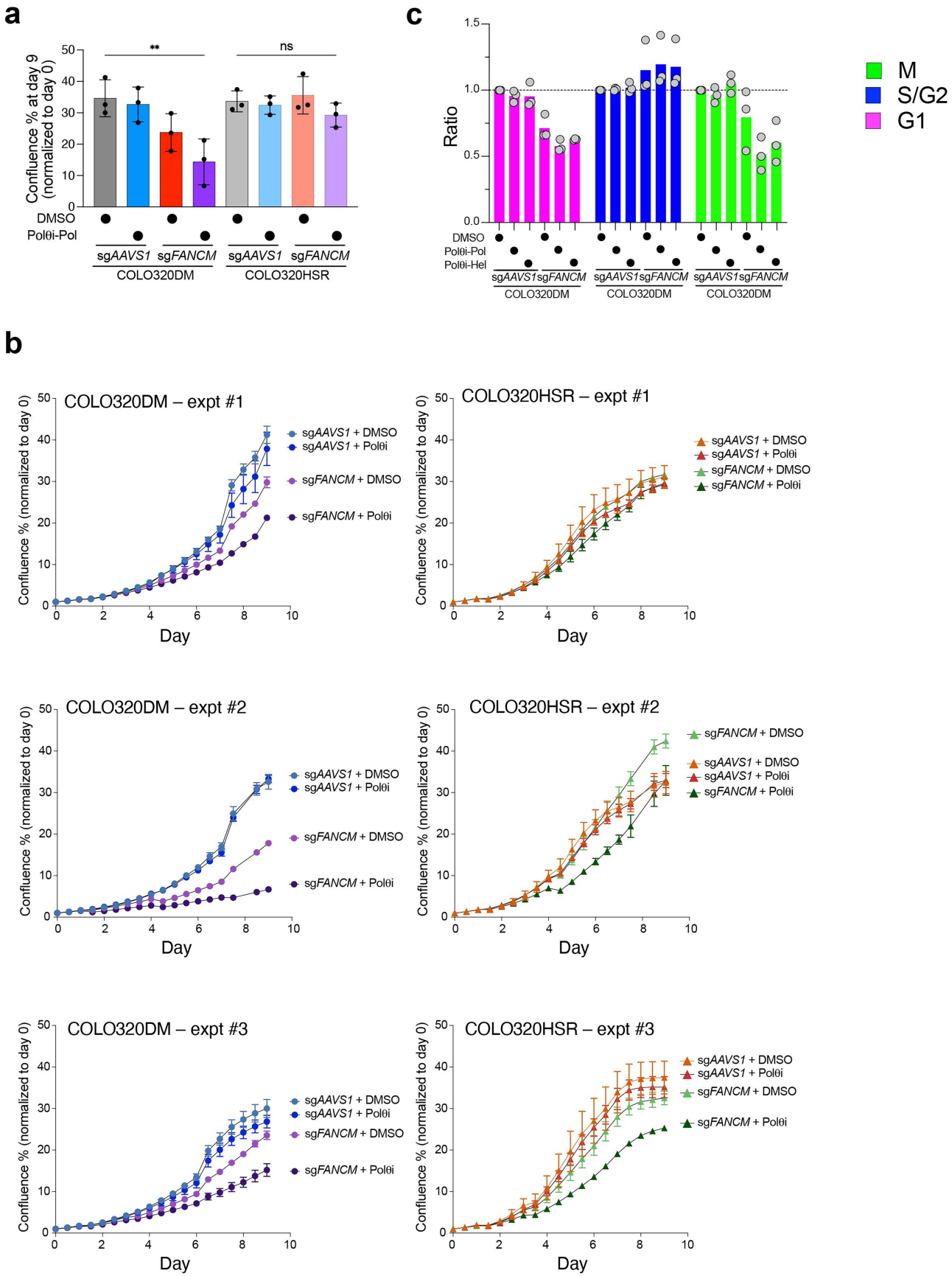
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Sequence composition of TA repeats at fragile END-seq peaks. Zoomed-in views of regions TA1 to TA7 highlighted in Fig. 3e-h, showing the END-seq signal and the underlying TA repeat sequences at each break-enriched locus. Within the sequence, the approximate END-seq peak is highlighted in red. Inverted repeats that facilitate secondary structure formation within the 100 bp upstream and downstream of the END-seq peak were identified using the EMBOSS palindrome inverted repeat tool (https://www.bioinformatics.nl/cgi-bin/emboss/palindrome). The predicted sequence that forms the cruciform is shown below each TA repeat in blue (5′ sequence) and green (3′ sequence) as identified by the EMBOSS palindrome IR tool. TA1, TA2, and TA3 are within the MYC-PVT1 region that is highly transcribed in the context of ecDNA.



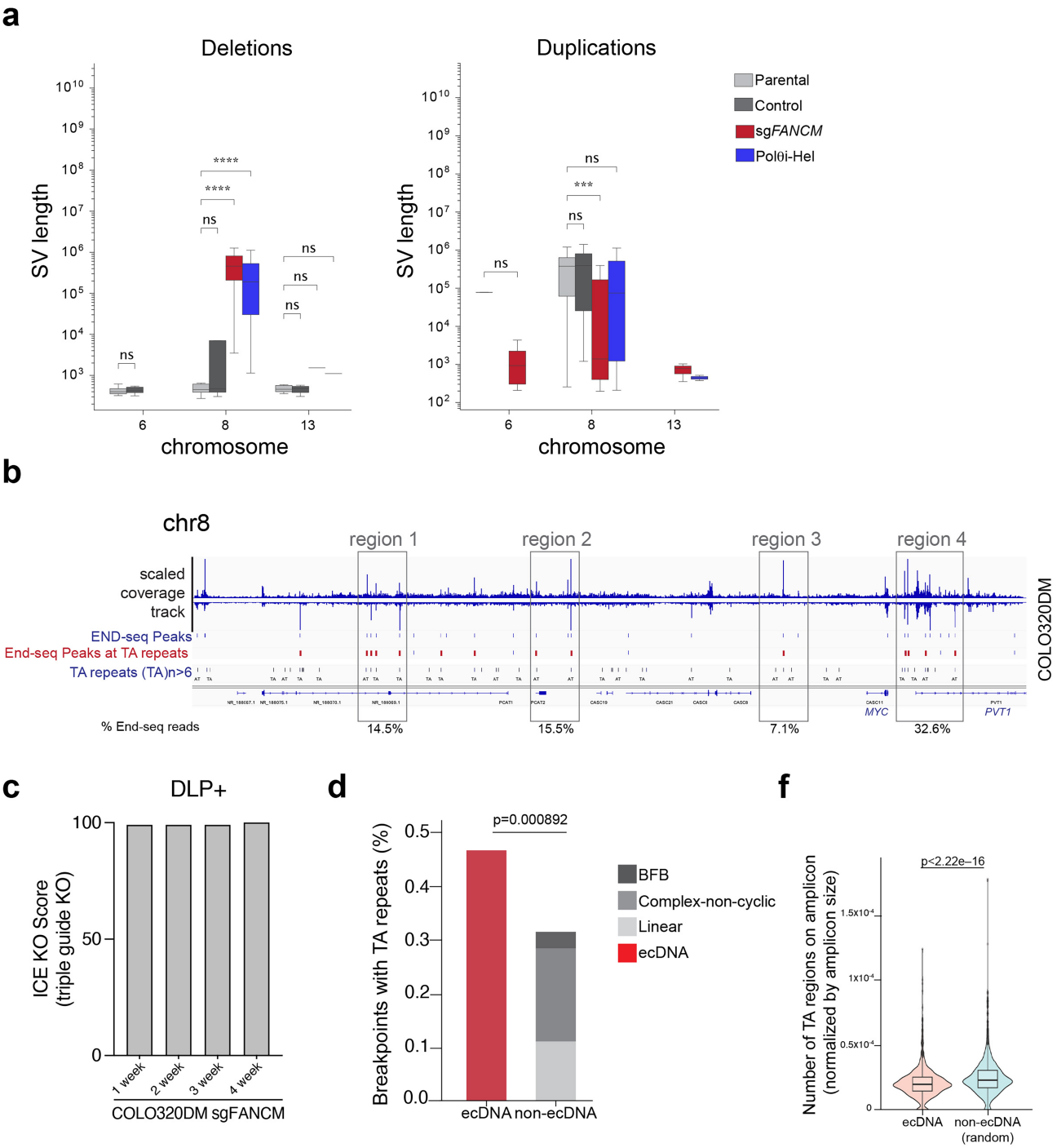
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | FANCM depletion exacerbates DNA damage and ecDNA loss upon Polθ inhibition. a. Quantification of DNA damage in COLO320DM cells following CRISPR/Cas9 targeting of the indicated helicases (*FANCM*, *SMARCAL1*, *WRN*), endonucleases (*ERCC1*, *ERCC4*, *GEN1*, *SLX4*, *MUS81*), or *AAVS1* as a control, with cells treated with DMSO (gray) or Polθi-Pol (blue). DNA damage was scored as the percentage of cells containing more than five co-localized 53BP1–γH2AX foci per nucleus. $n = 2$ biological replicates. **b**. ICE analysis of editing efficiency at the indicated CRISPR/Cas9 target sites in COLO320DM cells. **c**. Corresponding ICE analysis for the targeted loci shown in Fig. 4b. **d**. ICE analysis of editing efficiency in the COLO320DM and COLO320HSR cells used for END-seq in Fig. 4g. **e**. Quantification of 53BP1 focus colocalization with *MYC* signal in COLO320DM and COLO320HSR cells with AAVS1 of FANCM knockout under the indicated treatment conditions. The number of colocalizing foci was normalized separately to the total number of 53BP1 foci (green bars) and to the total number of *MYC* foci (red bars), providing a measure of the fraction of DNA damage events occurring at ecDNA loci and the fraction of ecDNA loci marked by DNA damage, respectively. Bars depict mean ± s.e.m. from $n = 3$ biological replicates (>100 nuclei per replicate). Statistical significance was assessed by unpaired *t*-tests (*** $p < 0.001$). **f**. Quantification of relative *MYC* expression in COLO320DM and COLO320HSR cells expressing sg*AAVS1* or sg*FANCM* and treated with DMSO or Polθi-Pol at two time points: 3 days of Polθ inhibitor treatment and 5 days after *FANCM* knockout (blue), or 10 days of Polθ inhibitor treatment and 12 days after *FANCM* knockout (red). Values were normalized to COLO320DM sg*AAVS1* DMSO. Data represent the mean of $n = 2$ biological replicates.



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

Extended Data Fig. 9 | FANCM loss and Polθ inhibition synergize to impair the proliferation of ecDNA-harboring cells. a. Endpoint confluency measurements at day 9 for COLO320DM and COLO320HSR cells transfected with sg*AAVS1* or sg*FANCM*, treated with DMSO or Polθi-Pol. Data represent mean ± s.d. from $n = 3$ biological replicates. Statistical significance was determined by ordinary one-way ANOVA with multiple comparisons (**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns, not significant). **b**. Proliferation curves for COLO320DM (left) and COLO320HSR (right) cells expressing sg*AAVS1* or sg*FANCM*, treated with DMSO or Polθi-Pol, from three independent biological replicates (experiments 1–3). Cell confluency was normalized to day 0 and monitored over 9 days by IncuCyte live-cell imaging. **c**. Cell-cycle distribution of sg*AAVS1* or sg*FANCM* COLO320DM cells treated with DMSO, Polθi-Pol, or Polθi-Hel, determined by FUCCI cell-cycle reporter analysis. Bars are mean ± s.d. with $n = 3$ biological replicates per condition, normalized to the DMSO control for each cell-cycle phase.



e

Classification	Total BKPs	TA repeat breakpoints
BFB	5134	23
Complex-non-cyclic	28422	84
Linear	18040	56
ecDNA	28682	134

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

Extended Data Fig. 10 | Structural variant landscape in COLO320DM and human tumors. a. Sample-specific Structural variant (SV) length for deletions (left) and duplications (right) on chromosomes 6, 8, 13 (the makeup of ecDNA) across T0, DMSO, sg*FANCM*, and Polθi-Hel conditions. Statistical comparisons by Mann-Whitney U-test (**** $p < 0.0001$, *** $p < 0.001$, ns, not significant). **b**. Genome browser view of the chr8 ecDNA amplicon in COLO320DM cells showing scaled coverage, END-seq peaks, End-seq peaks overlapping TA repeats, and TA repeat density $(TA)_n \geq 6$. Four regions with the highest End-seq signal (regions 1–4) are highlighted, with region 4 coinciding with the *MYC-PVT1* locus. The percentage of End-seq reads in each region relative to the total peaks in the ecDNA is noted below each region. **c**. ICE knockout scores in COLO320DM sg*FANCM* cells over 1 month of long-term culture. DLP+ was performed on the 4-week sample. **d**. Percentage of breakpoints overlapping TA repeats in ecDNA versus non-ecDNA amplicons, stratified by structural variant class (BFB, complex-non-cyclic, linear, ecDNA). ecDNA breakpoints are significantly enriched at TA repeats relative to all non-ecDNA classes ($p = 0.000892$, Pearson's chi-squared test). **e**. Raw numbers of total breakpoints and TA repeat breakpoints in the graph shown in (d) **f**. Number of TA repeat regions per amplicon normalized to amplicon size, comparing ecDNA and non-ecDNA random amplicons. ecDNA amplicons carry a significantly lower density of TA repeat regions than non-ecDNA amplicons ($p < 2.22e\text{-}16$, one-sided *t*-test).

nature portfolio

Corresponding author(s): SFEIR, AGNEL

Last updated by author(s): Aug 2, 2025

Reporting Summary

Nature Portfolio 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 Portfolio 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	All data was generated for this study.
Data analysis	END-seq reads were trimmed using Trim Galore and aligned to the human genome (hg19) using Bowtie2. Quality assessment, sorting, and indexing of the resulting BAM files were performed using Samtools. PCR duplicates were removed using Picard. For visualization and downstream analysis, scaled coverage tracks (normalized to reads per million) were generated using BAMscale. Peaks corresponding to DNA double-strand break sites were identified using MACS3. For mutational signature at AAVS1 and Myc, The CRISPResso2 toolkit was used to quantify editing outcomes, characterizing insertions and deletions, and reconstruct repair events. For each sample, allele fractions for these events were calculated by counting the number of reads with respective mutational signatures identified by CRISPResso2 and dividing the count by the total reads.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequencing data for AAVS1- and MYC- Repair-Seq and END-seq will be deposited in NCBI's Sequence Read Archive (SRA) under BioProject (TBD). Requests for materials and/or questions regarding any of the constructs, cell lines, or other data described should be addressed to the corresponding authors. Source data are provided in this paper, and raw data will be available in Mendeley. All scripts used for analyzing sequencing data will be made available on GitHub.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	does not apply
Reporting on race, ethnicity, or other socially relevant groupings	does not apply
Population characteristics	does not apply
Recruitment	does not apply
Ethics oversight	does not apply

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

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	does not apply
Data exclusions	does not apply
Replication	does not apply
Randomization	does not apply
Blinding	does not apply

Behavioural & social sciences study design

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

Study description	does not apply
Research sample	does not apply
Sampling strategy	does not apply
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper,*

Data collection	*computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used

rabbit anti-53BP1 (Novus NB100-304, 1:1000)
mouse anti-γH2AX (Millipore JBW301, 1:1000)
donkey anti-rabbit 568 and donkey anti-mouse 488, 1:500, invitrogen
anti-BRCA2 antibody (Sigma-Aldrich, Cat# OP95)

Validation

these antibodies have been validated in the last decade as indicated on the website of the manufacturor and used in numerous studies

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

COLO320DM and COLO320HSR cells (ATCC CCL-220 and CCL-220.1, respectively)
NCI-H716 cells (ATCC CCL-251)
HEK293T cells (ATCC CRL-3216)
HelaDM and HelaHSR from Don Cleveland lab

Authentication

STR ever year

Mycoplasma contamination

tested for mycoplasma ever 6 months and confrimed negative.

Commonly misidentified lines
(See ICLAC register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.