

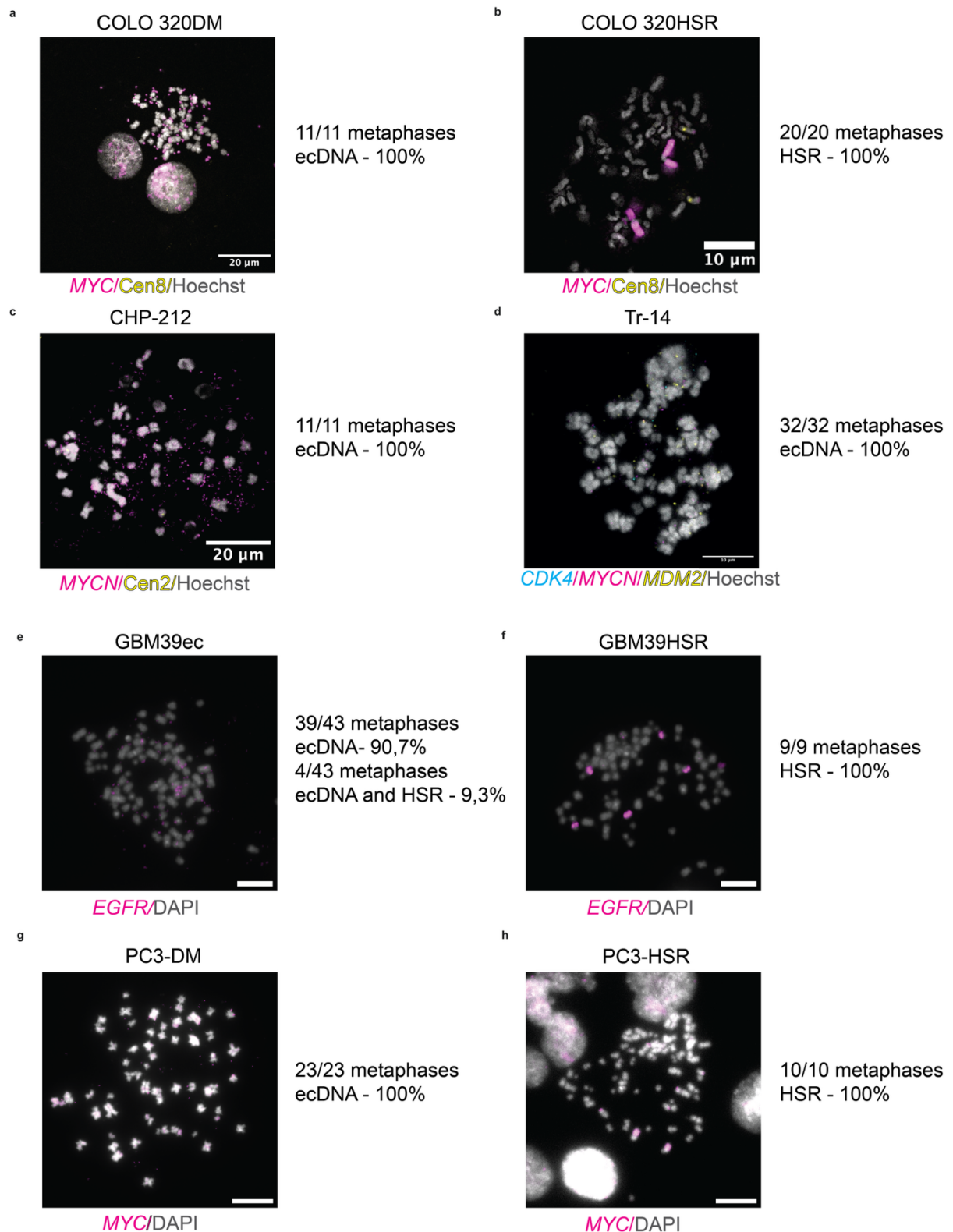
Micronuclear ecDNA constrains tumor cell fitness and is associated with favorable patient survival

In the format provided by the
authors and unedited

Supplementary information

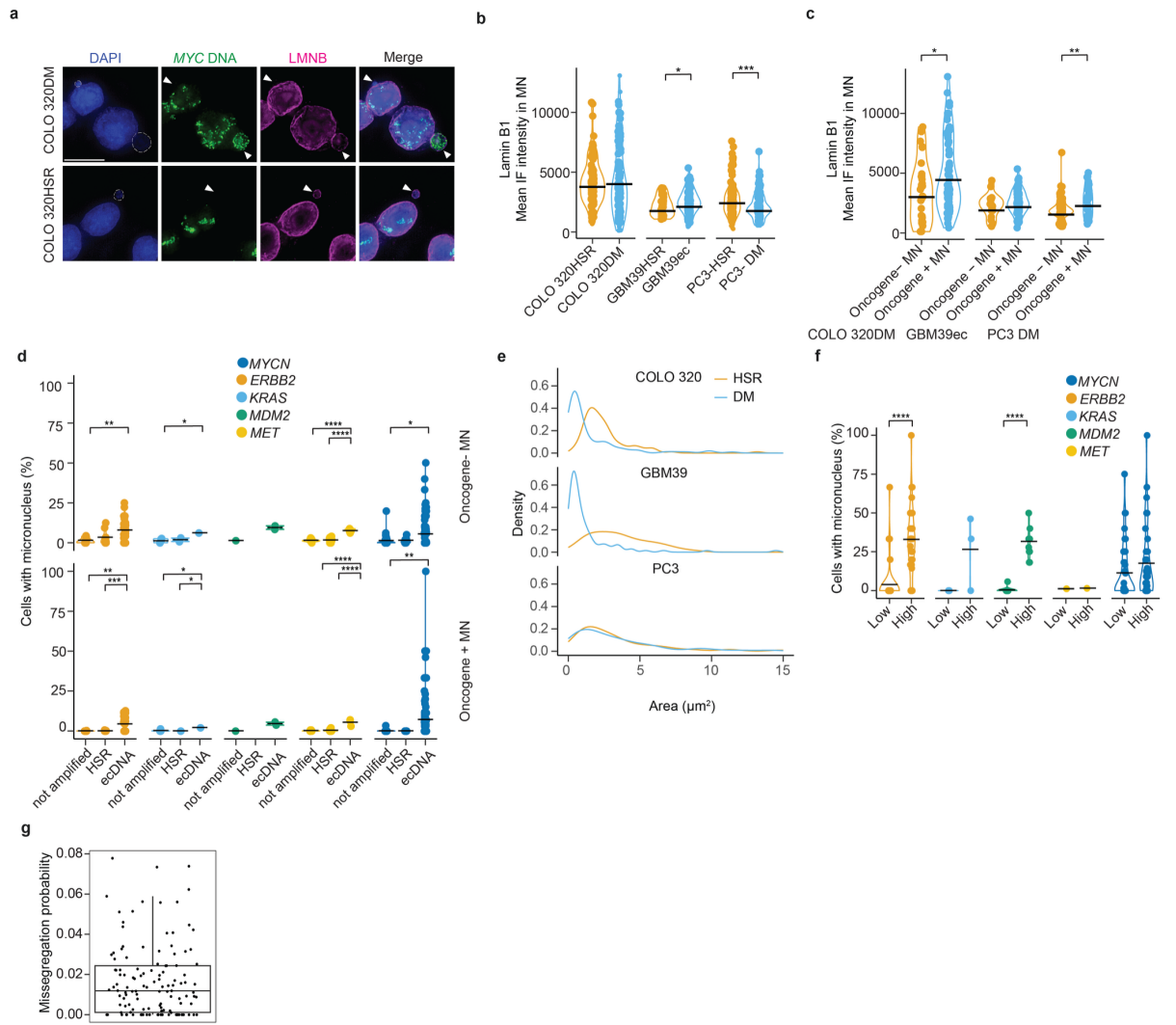
Table of contents:

pp. 2-16: Supplementary Figures 1-14
pp. 17-29: Supplementary Note: Methods
p. 30: Supplementary videos captions
pp. 31-33: Unprocessed Gels
pp. 34-35: References



Supplementary Fig. 1: Confirmation of amplification status of the major cell line models used in this study

a. Exemplary metaphase spread and quantification of amplification status at early (p= 5) passage of the COLO 320DM single cell clone. Scale bar: 20 μm.
b. Exemplary metaphase spread and quantification of amplification status of the COLO 320HSR cell line. Scale bar: 10 μm.
c. Exemplary metaphase spread and quantification of amplification status of the CHP-212 cell line. Scale bar: 20 μm.
d. Exemplary metaphase spread and quantification of amplification status of the Tr-14 cell line. Scale bar: 10 μm.
e. Exemplary metaphase spread of the GBM39ec cell line. Scale bar: 10 μm.
f. Exemplary metaphase spread of the GBM39HSR cell line. Scale bar: 10 μm.
g. Exemplary metaphase spread of the PC3-DM cell line. Scale bar: 10 μm.
h. Exemplary metaphase spread of the PC3-HSR cell line. Scale bar: 10 μm. For all images n=1, number of metaphases imaged indicated next to the figure.



Supplementary Fig. 2: Cancer cells with high ecDNA content are prone to micronucleation irrespective of the amplified gene

a. Exemplary photomicrographs of COLO 320DM and COLO 320HSR cells with micronucleus (white arrowhead and dash) stained for MYC and lamin B1 by ImmunoFISH. Scale bar: 10 μm

b. Mean lamin B1 staining intensity in micronuclei in three near-isogenic cell line pairs ($n = 32$ -107, two-sided Student's t-test, exact P -values can be found in Supplementary Table 3; * $P < 0.05$; ***, $P < 0.001$). Each dot represents one micronucleus.

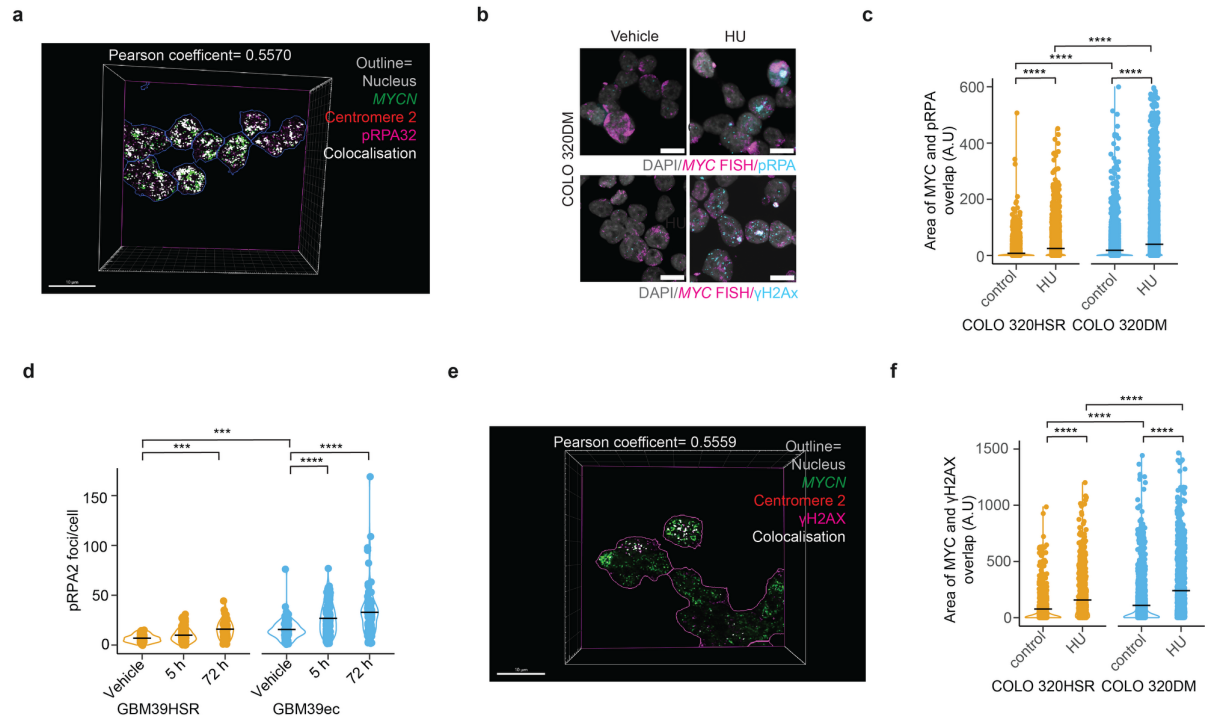
c. Mean lamin B1 staining intensity in ecDNA⁺ MN and ecDNA⁻ MN determined by FISH in three near-isogenic cell line pairs ($n = 22$ -84, two-sided Student's t-test, exact P -values can be found in Supplementary Table 3; * $P < 0.05$; **, $P < 0.01$). Each dot represents one micronucleus.

d. Fraction of cells with micronucleus in patient tumor samples stratified by amplified gene and oncogene status in MN ($n = 1$ -106, exact values can be found on Supplementary Table 2, one-way ANOVA, exact P -values can be found in Supplementary Table 3; * $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$, ****, $P < 0.0001$).

e. Density plot displaying micronuclei size distribution in three near-isogenic cell line pairs: COLO 320DM ($n = 103$), COLO 320HSR ($n = 100$), GBM39EC ($n = 100$), GBM39HSR ($n = 33$), PC3-DM ($n = 103$) and PC3-HSR ($n = 102$).

f. Fraction of cells with micronuclei and low (bottom 30%) vs. high (top 30%) ecDNA content measured in patient tumor samples stratified by amplified gene (j) (two-sided Student's t-test, exact P -values can be found in Supplementary Table 2; ****, $P < 0.0001$). Black lines indicate the mean.

g. Mis-segregation probabilities (the chance that at least one ecDNA micronucleates during mitosis) across 8 cell lines (COLO 320DM, CHP-212, STA-NB10DM, Tr-14, UKF-NB6, Lan5, SMS-KAN, KP-N-YN) and 119 patient samples. Data are presented as box-and-whisker plots. The centre line indicates the median. The box extends from the 25th to the 75th percentile (IQR). The whiskers represent the smallest and largest values within $1.5 \times \text{IQR}$, and points beyond the whiskers are shown as individual outliers.



Supplementary Fig. 3: Replication stress and DNA damage induction on ecDNA

a. Representative 3D reconstruction of fixed untreated CHP212 cells, that contain *MYCN* ecDNA, stained for Hoechst (grey outline), *MYCN* (green, FISH), centromere 2 (red, FISH) and pRPA32 (magenta, immunofluorescence). Areas where *MYCN* ecDNA (green) and pRPA32 (S4/S8) (magenta) colocalise, were computed using Imaris (v10.2) and are depicted in white. Scale bar: 10μm.

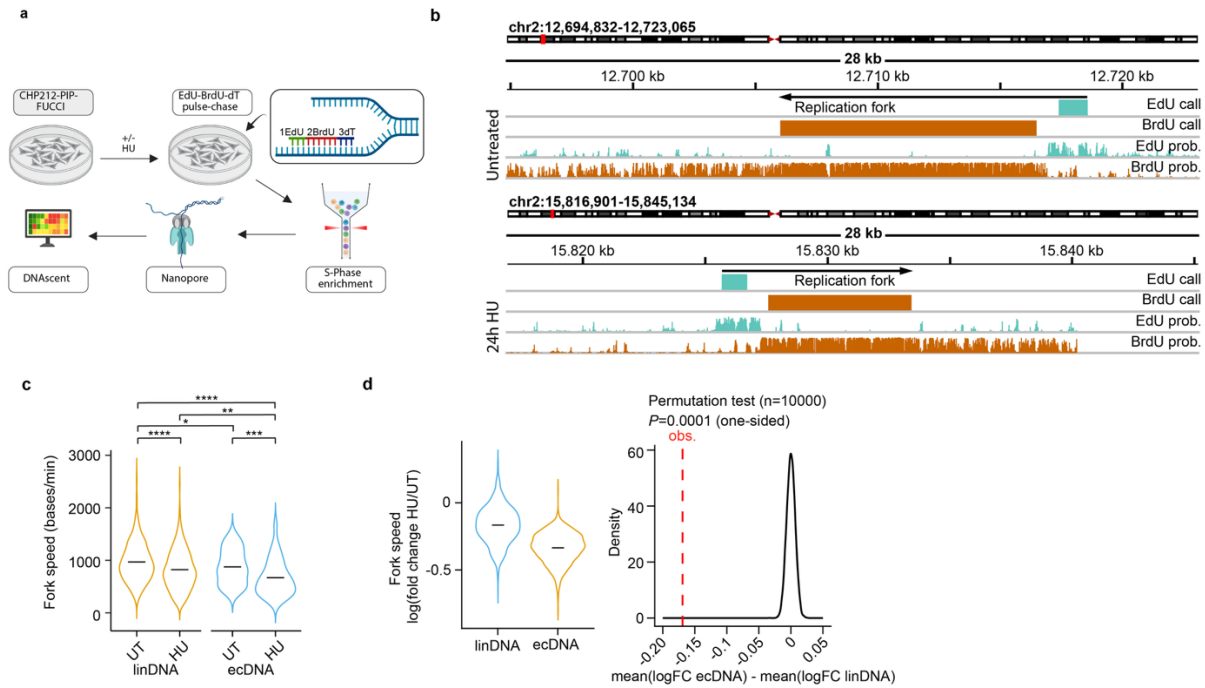
b. Exemplary photomicrographs of fixed COLO 320DM cells in the presence of HU (50 μM, 72h) or vehicle, stained for *MYC* (magenta, FISH) and pRPA32 (blue, immunofluorescence) (top row) or *MYC* (magenta, FISH) and γH2AX (blue, immunofluorescence) (bottom row). Scale bar: 10μm.

c. Quantification of colocalization between pRPA2 and *MYC* FISH in COLO 320DM and COLO 320HSR to indicate replication stress on the amplicon. (n = 3602–8424; heteroscedasticity-robust two-way ANOVA, exact *P*-values can be found in Supplementary Table 3; ****, *P*<0.0001). Each dot denotes one cell.

d. Quantification of pRPA2 foci per cell in isogenic glioblastoma cell lines at different timepoints of 100μM HU treatment (n= 68-175, Two-way ANOVA, exact *P*-values can be found in Supplementary Table 3; ***, *P*<0.001; ****, *P*<0.0001). Each dot represents one cell.

e. Representative 3D reconstruction of ImmunoFISH stained cells of *MYCN* and γH2AX in untreated CHP-212 cells. Areas of colocalisation are indicated in white. Colocalisation was calculated between Channel 2 (*MYCN*, green) and Channel 4 (γH2AX, magenta).

f. Quantification of γH2AX foci colocalised with *MYC* in an isogenic colon cancer cell line pair after 50 μM HU treatment for 3 days (n = 596- 1648; heteroscedasticity-robust two-way ANOVA, exact *P*-values can be found in Supplementary Table 3; ****, *P*<0.0001). Each dot denotes one cell.



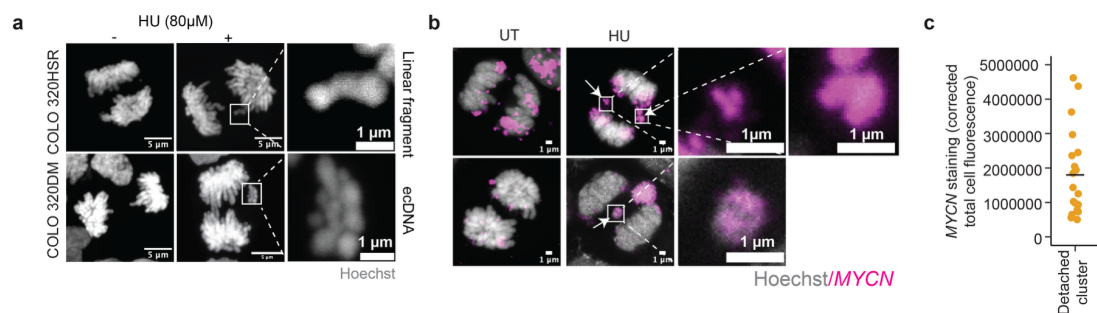
Supplementary Fig. 4: Hydroxyurea induces replication fork slowing on ecDNA.

a. Schematic of the experimental design. CHP212-PIP-FUCCI cells were incubated in the presence or absence of HU (80 μ M) for 24h and subsequently treated with EdU, BrdU and dT. Cells were enriched for S-phase using FACS and sequenced using Nanopore (ONT). DNAscent¹ was used to infer replication fork dynamics. Figure partially created in BioRender. Xu, R. (2026) <https://BioRender.com/8k2z6r7>.

b. Representative IGV snapshot of EdU and BrdU tracks on ecDNA in CHP212-PIP-FUCCI cells in the presence or absence of HU (80 μ M, 24h). The top 2 rows denote EdU and BrdU calls from DNAscent forksense. The bottom 2 rows denote the EdU and BrdU probability called by DNAscent detection. Arrows show the movement direction of the exemplary replication forks.

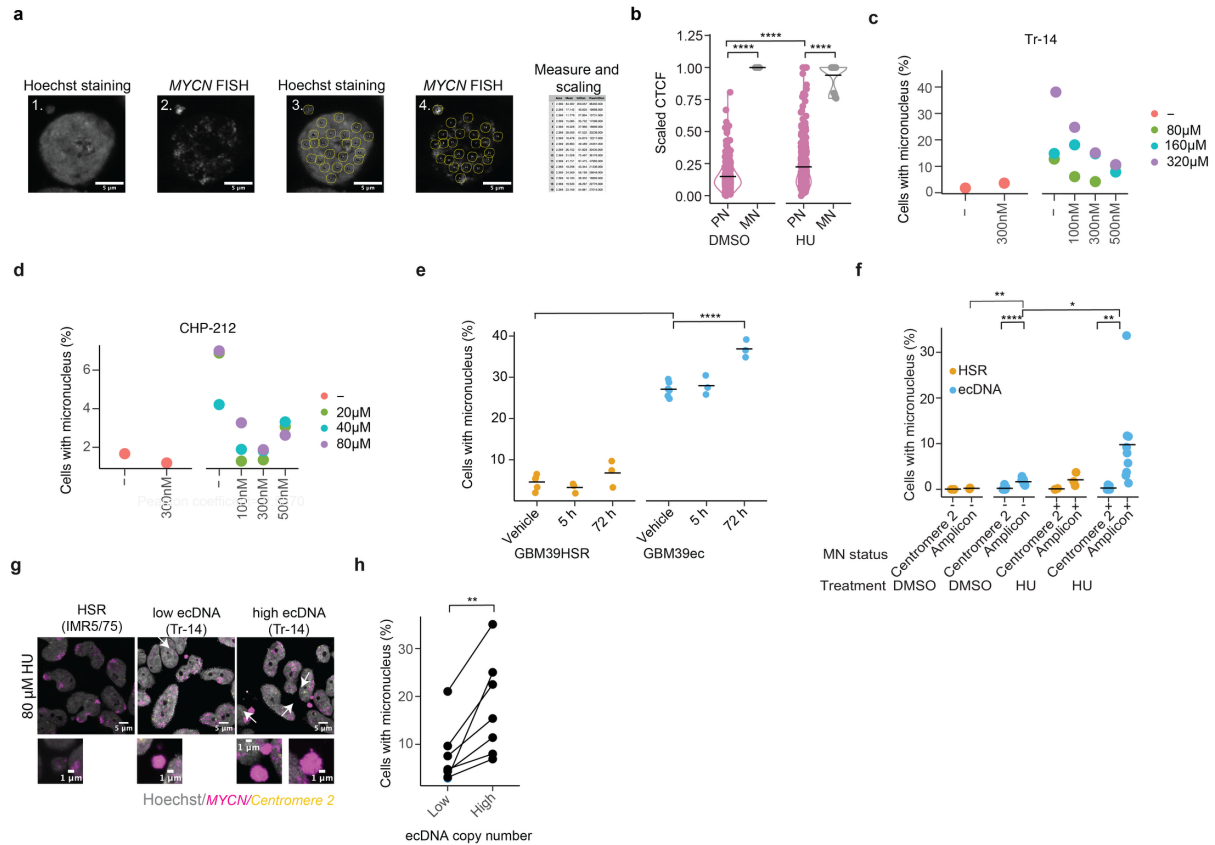
c. Quantification of replication fork speed, as computed by replication fork length divided by treatment time (bases/min) on linear DNA (linDNA), or ecDNA in the presence or absence of HU (80 μ M, 24h). (UT-linDNA, n = 598; HU-linDNA, n = 535; UT-ecDNA, n = 73; HU-ecDNA, n = 63; two-sided Welch's t-test; Exact *P*-values can be found in Supplementary Table 3; *, *P*<0.05; **, *P*<0.01, ***, *P*<0.001, ****, *P*<0.0001).

d. Low dose hydroxyurea (80 μ M, 24h) has a differential effect on ecDNA. Left: Log foldchanges of HU-treated vs. untreated forkspeeds on linear DNA (linDNA) and ecDNA, as computed by subsampling 30 fork speeds per treatment and represented as log foldchange between the median fork speeds between HU treatment and untreated fork speeds (n=1000 repetitions per condition). Right: Comparison of the differential mean fork speed log foldchanges between linear and ecDNA (exact *P*=0.0001, n permutations=10000, Permutation test, one sided). Red line indicates observed mean differential log foldchange from the left panel.



Supplementary Fig. 5: Detached DNA clusters in CHP-212 are positive for amplicon staining

- a. Exemplary photomicrographs of representative fixed COLO 320DM (contains *MYC* as ecDNA, bottom) and COLO 320HSR (contains *MYC* as HSR, top) cells in anaphase that were incubated in the presence or absence of HU (80μM) for 3 days and stained with Hoechst. Scale bar: 5 μm and 1 μm (zoom). White box indicates enlarged region (n=3).
- b. Exemplary photomicrographs of fixed CHP212 cells during anaphase without (first column) and with HU treatment (second column) (80 μM, 24h), stained for Hoechst (grey) and *MYCN* FISH (magenta). Arrows indicate detached DNA cluster. White boxes indicate enlarged regions. Scale bar: 1 μm.
- c. Quantification of the corrected total cell fluorescence of *MYCN* FISH signal on detached clusters during anaphase in fixed CHP212 cells after HU treatment (80 μM, 24h, n=100 anaphases).



Supplementary Fig. 6: Induction of replication stress induces ecDNA micronucleation

a Schematic setup of the image analysis approach, quantifying fluorescent intensity and calculating the corrected total cell fluorescence in the micronuclei and primary nuclei (same size area). Scaling was done to account for differential background signal in cells (n=1).

b Quantification of the scaled corrected total cell fluorescence of equal size areas in the primary nucleus (PN) and micronucleus (MN) of CHP-212 cells (n=12 cells with micronuclei, two way ANOVA, exact *P*-values can be found in Supplementary Table 3; ****, *P*<0.0001). Each dot represents one area.

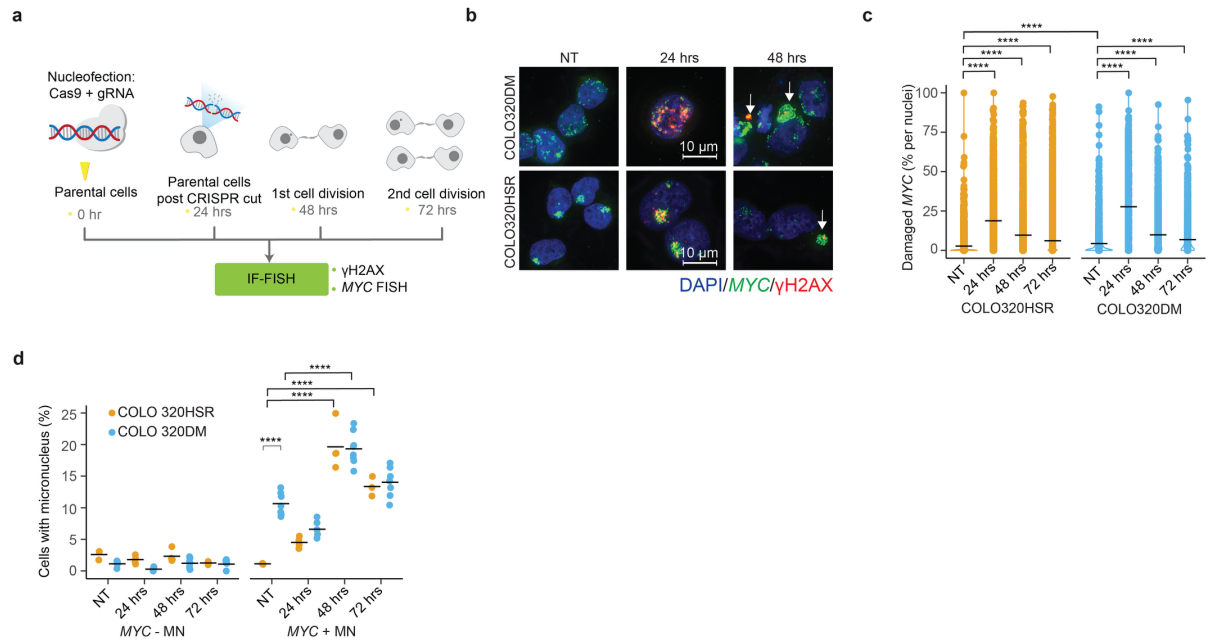
c-d. Percentage of cells with micronuclei in 2 ecDNA containing cell lines (c: Tr-14, d: CHP-212) in the presence of different concentrations of dNTPs and at different concentrations of HU after 3 days. Each dot represents the average per experiment (n= 100-221 cells per experiment).

e. Percentage of cells with micronuclei in the isogenic glioblastoma cell line pair GBM39HSR (contains *EGFR*vIII on HSR, left) and GBM39ec (contains *EGFR*vIII on ecDNA, right) after different lengths of HU treatment (80μM) (n= 5 for vehicle and n= 3 for each treatment group, Two-way ANOVA, exact *P*-values can be found in Supplementary Table 3; ****, *P*<0.0001). Each dot indicates the average per replicate.

f. Percentage of cells with micronuclei stratified by the presence of either centromere 2 or the respective amplicon in cell lines with ecDNA(ecDNA: COLO 320DM, CHP-212, STA-NB10DM, Tr-14, UKF-NB6, Lan5, SMS-KAN, KP-N-YN, n=8) or without (HSR: COLO 320HSR, NGP, IMR5/75, n=3) and by the presence or absence of HU (80μM, 72h), as quantified by FISH for centromere 2 or the respective amplicon (Two-way ANOVA, exact *P*-values can be found in Supplementary Table 3; * *P*<0.05; ** *P*<0.01; ****, *P*<0.0001). Each dot represents the average per cell line.

g. Exemplary photomicrographs of fixed cells with HSR, low ecDNA content or high ecDNA content and micronuclei (white arrowhead) in the presence of HU (80 μM, 72h).

h. Fraction of cells with micronuclei with low (bottom 30%) and high (top 30%) ecDNA content after HU treatment (COLO 320DM, CHP-212, STA-NB10DM, Tr-14, UKF-NB6, Lan5, SMS-KAN, n=1 experimental replicate, n= 98-506, exact numbers can be found in Supplementary Table 2, cells per replicate, two-sided paired Student's t-test, exact *P*-values can be found in Supplementary Table 3; **, *P*<0.01;). Each paired dot represents one ecDNA cell line.



Supplementary Fig. 7: Nucleofection with Cas9 induces DNA damage and micronucleation

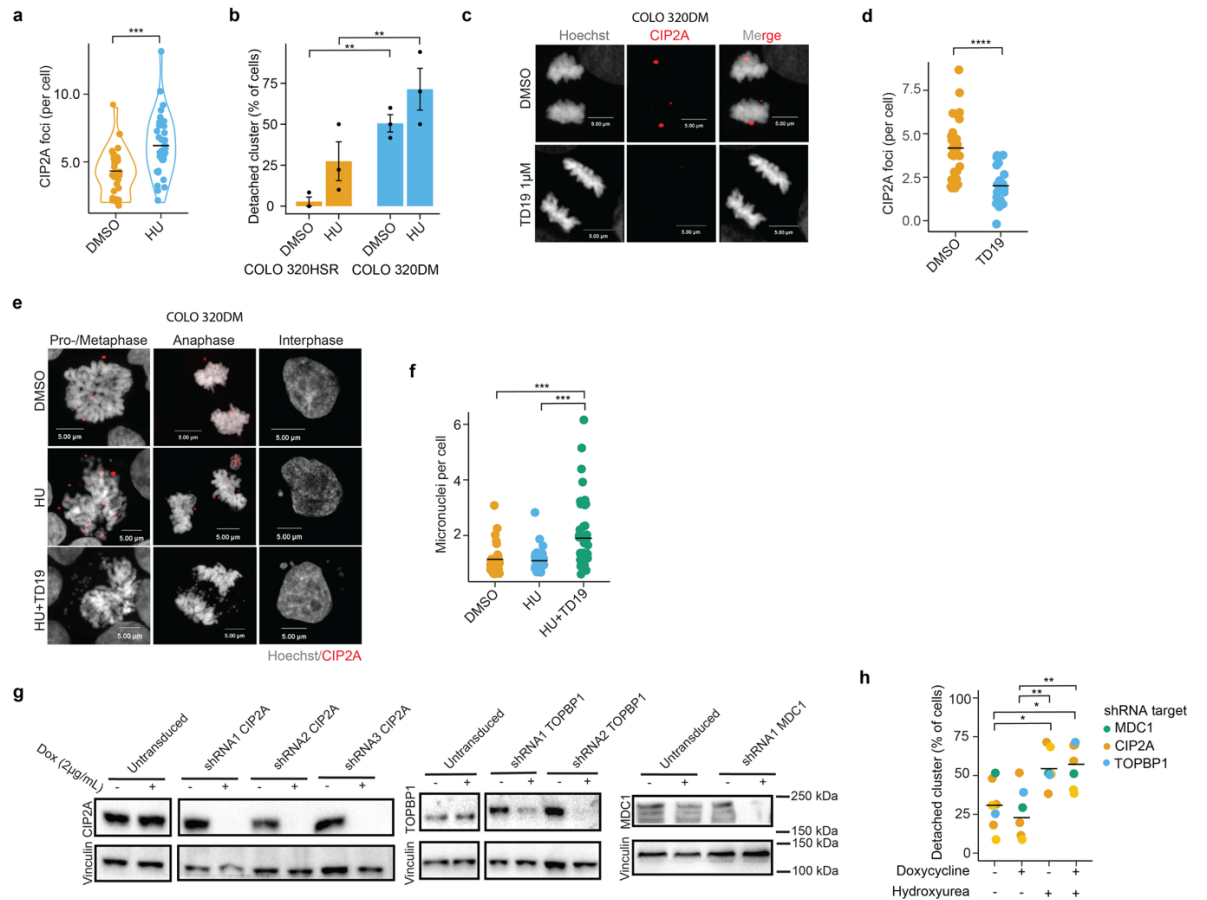
a. Schematics to depict samples collected at different time points. Figure partially created in BioRender. Xu, R. (2026)

<https://BioRender.com/8k2z6r7>.

b. Exemplary photomicrographs in COLO 320DM/COLO 320HSR cells to show DNA damage induced at amplicon⁺ micronuclei formation. White arrows indicate micronuclei. Scale bar: 10 μ m.

c. Quantification of damaged MYC percentage in COLO 320DM and COLO 320HSR at different timepoints after nucleofection of Cas9+ sgRNA. (n= 2741-6696, one-way ANOVA, exact *P*-values can be found in Supplementary Table 3; ****, *P*<0.0001). Each dot denotes one cell.

d. Percentage of cells with MYC⁺ or MYC⁻ micronuclei at different timepoints after nucleofection with guides targeting the amplicon in the isogenic cell line pair COLO 320HSR (contains MYC on HSR, orange) and COLO 320DM (contains MYC on ecDNA, blue) (n= 1919-4441, exact numbers can be found in Supplementary Table 2 ,One-way ANOVA, exact *P*-values can be found in Supplementary Table 3; ****, *P*<0.0001). Each dot represents one image.



Supplementary Fig. 8: Perturbation of the CIP2A-TOPBP1 complex reduces clustering of detached ecDNA

a. Quantification of immunofluorescent CIP2A foci in COLO 320DM anaphase cells after incubation in the presence of HU (100 μ M) for 24hrs compared to untreated cells ($n = 31-33$ anaphases; two-sided Student's t-test; $P=0.0001243$, ***, $P<0.001$). Each dot denotes one anaphase.

b. Quantification of the percentage of cells with detached cluster presented as mean values $\pm$ SEM (two-way ANOVA with Šidák's multiple comparisons test; DMSO- $P=0.0063$, HU- $P=0.0097$; **, $P<0.01$)

c. Exemplary photomicrographs of a fixed cell with ecDNA during anaphase treated with and without the CIP2A inhibitor TD19 (1 μ M), stained with Hoechst (grey) and CIP2A immunofluorescence (red). Scale bar: 5 μ m.

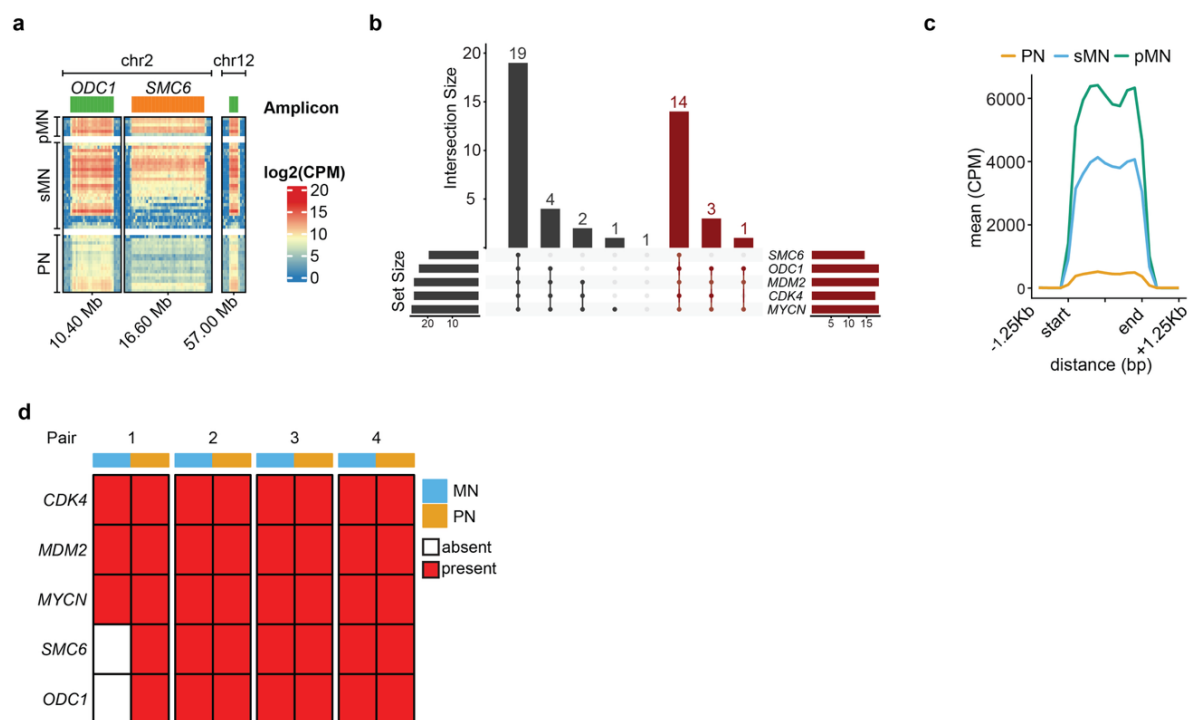
d. Quantification of CIP2A foci in COLO 320DM cells during anaphase without (DMSO) and with the CIP2A inhibitor TD19 (1 μ M) (DMSO, $n=31$; TD19, $n=24$; two-sided Student's t-test; exact P -value can be found in Supplementary Table 3; ****, $P<0.0001$). The mean is indicated as line. Each dot represents one cell.

e. Exemplary photomicrographs of fixed COLO 320DM cells during pro-/metaphase, anaphase and interphase after treatment with DMSO, HU (80 μ M) and HU (80 μ M) and TD19 (1 μ M) for 24h, stained for Hoechst (grey) and CIP2A immunofluorescence (red). Scale bar: 5 μ m.

f. Number of micronuclei per cell in COLO 320DM cells after treatment with DMSO, HU (80 μ M) or HU (80 μ M) and TD19 (1 μ M) for 24h. (DMSO, $n=43$; HU, $n=43$; HU+TD19, $n=40$; One-way ANOVA, exact P -values can be found in Supplementary Table 3, ***, $P<0.001$). Line indicates the mean.

g. Western blot validation of the inducible knockdown of CIP2A, TOPBP1 and MDC1 after 2 days of 2 μ g/mL doxycycline treatment in COLO 320DM after transduction with different shRNA constructs.

h. Quantification of detached clusters in COLO 320DM cells after shRNA-mediated inducible knockdown of CIP2A, TOPBP1 and MDC1 ($n = 10-12$ anaphases per condition, exact numbers and P -value can be found in Supplementary Table 2 and Supplementary Table 3; one-way ANOVA, *, $P<0.05$; **, $P<0.01$) with or without the addition of doxycycline and HU (Dox: 2 μ g/mL, 2d; HU: 80 μ M, 1d). The mean is indicated as line.



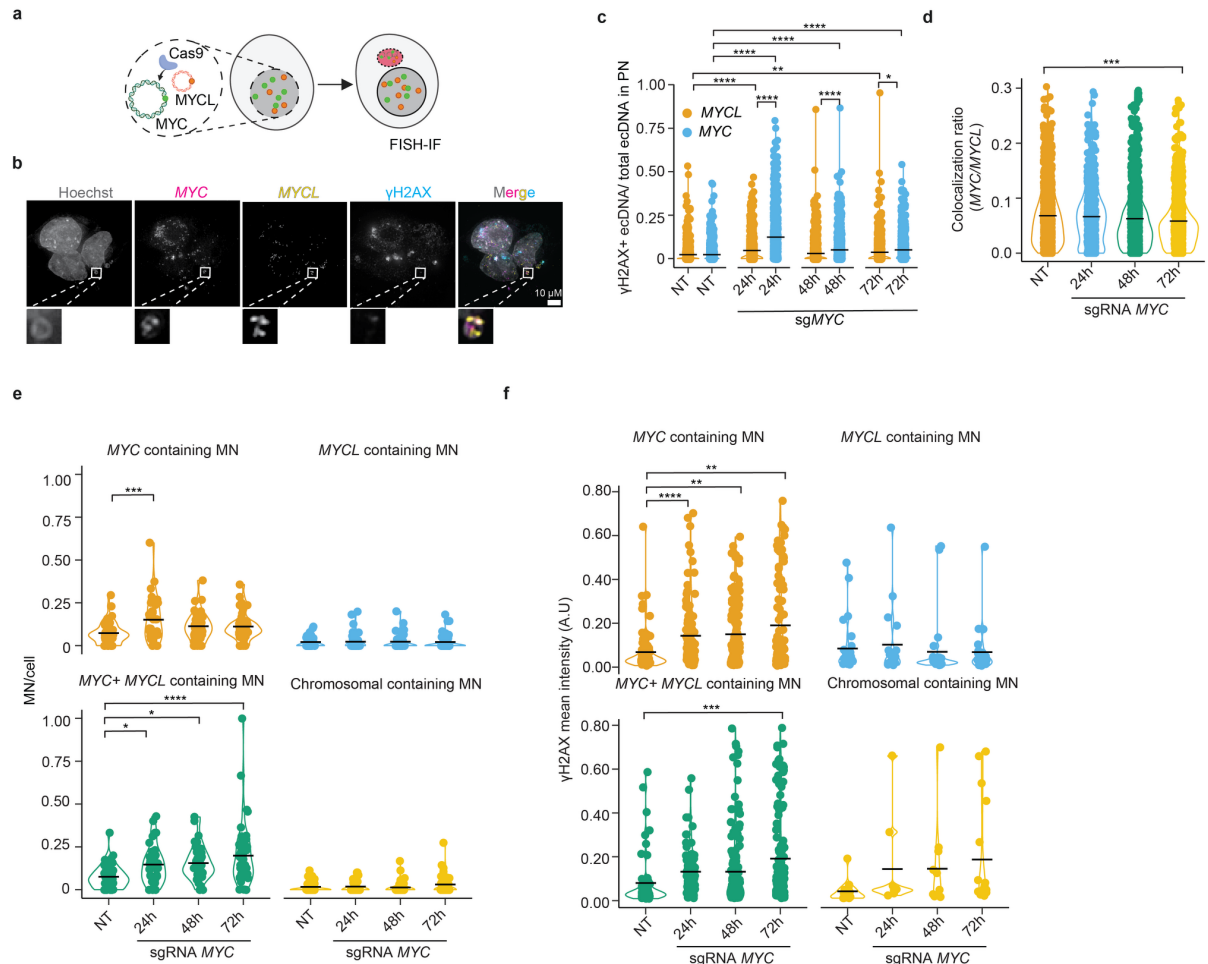
Supplementary Fig. 9: Single micronucleus sequencing reveals collective entrapment of different ecDNA species.

a. Laser microdissection directed sequencing of primary nuclei (PN) and micronuclei (MN) after incubation of TR-14 cells, containing 5 different ecDNA species, in the presence of HU (80 μ M, 72h). Counts per million (CPM), normalized and log2 transformed read coverage (25kb bins) across ecDNA (*ODC1*, green; *SMC6*, orange) and their flanking regions (100 kb; primary nucleus, PN, n = 18; single micronucleus, sMN, n = 27; pooled micronuclei, pMN, n = 6).

b. UpSet plot of the co-occurrence of 5 ecDNA species in single micronuclei (black) and primary nuclei (red).

c. Signal plot of the mean coverage over length-scaled ecDNA in single primary nuclei (PN, n = 18, orange), single micronuclei (sMN, n = 27, blue) and pooled micronuclei (pMN, n = 6, green). Counts per million (CPM) across all ecDNA aligned and length normalized.

d. Co-occurrence of 5 ecDNA species in 4 pairs of collected micronuclei (MN) and their corresponding primary nucleus (PN) in the same cell (n=4).



Supplementary Fig. 10: Cutting of one ecDNA species induces co-micronucleation of the other ecDNA species

a. Schematic of experimental design. DMS273 cells, containing *MYC* and *MYCL* on different ecDNA species are subjected to CRISPR-Cas9 mediated cutting of the *MYC* ecDNA. FISH for *MYC* and *MYCL* combined with immunofluorescence against γ H2AX in fixed cells at different timepoints was used to measure the colocalisation of DNA damage and *MYC* and *MYCL* ecDNA. Scale bar: 10 μ m. Figure partially created in BioRender. Xu, R. (2026) <https://BioRender.com/8k2z6r7>.

b. Exemplary photomicrograph of fixed DMS273 cells, containing *MYC* and *MYCL* on different ecDNA species, after *sgMYC* transfection, stained for Hoechst (grey), *MYC* FISH (magenta), *MYCL* FISH (yellow) and γ H2AX immunofluorescence (blue).

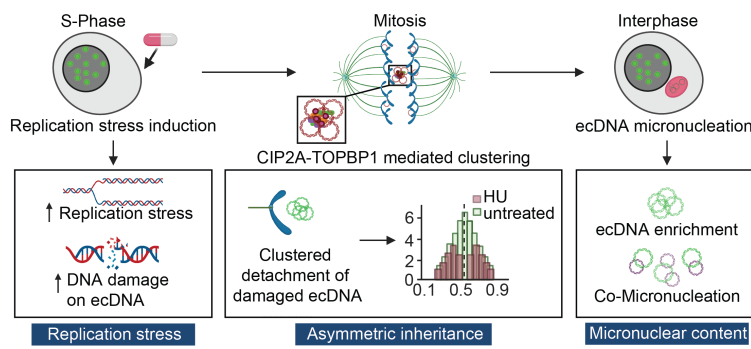
c. Quantification of the fraction of γ H2AX⁺ *MYC* ecDNA (blue) or γ H2AX⁺ *MYCL* ecDNA (orange) in DSM273 cells after *MYC* sgRNA transfection at different timepoints, as measured by *MYC*, *MYCL* FISH and γ H2AX immunofluorescence (n= 655-1092, Two-way ANOVA, exact *P*-values can be found in Supplementary Table 3; **, *P*<0.01; ****, *P*<0.0001). Each dot represents one cell.

d. Quantification of colocalisation of *MYCL* and *MYC* ecDNA in fixed DMS273 cells after *MYC* sgRNA transfection at different timepoints, measured by *MYC* and *MYCL* FISH (n=656-1091, One-way ANOVA; exact *P*-values can be found in Supplementary Table 3; ***, *P*<0.001). Each dot represents one cell.

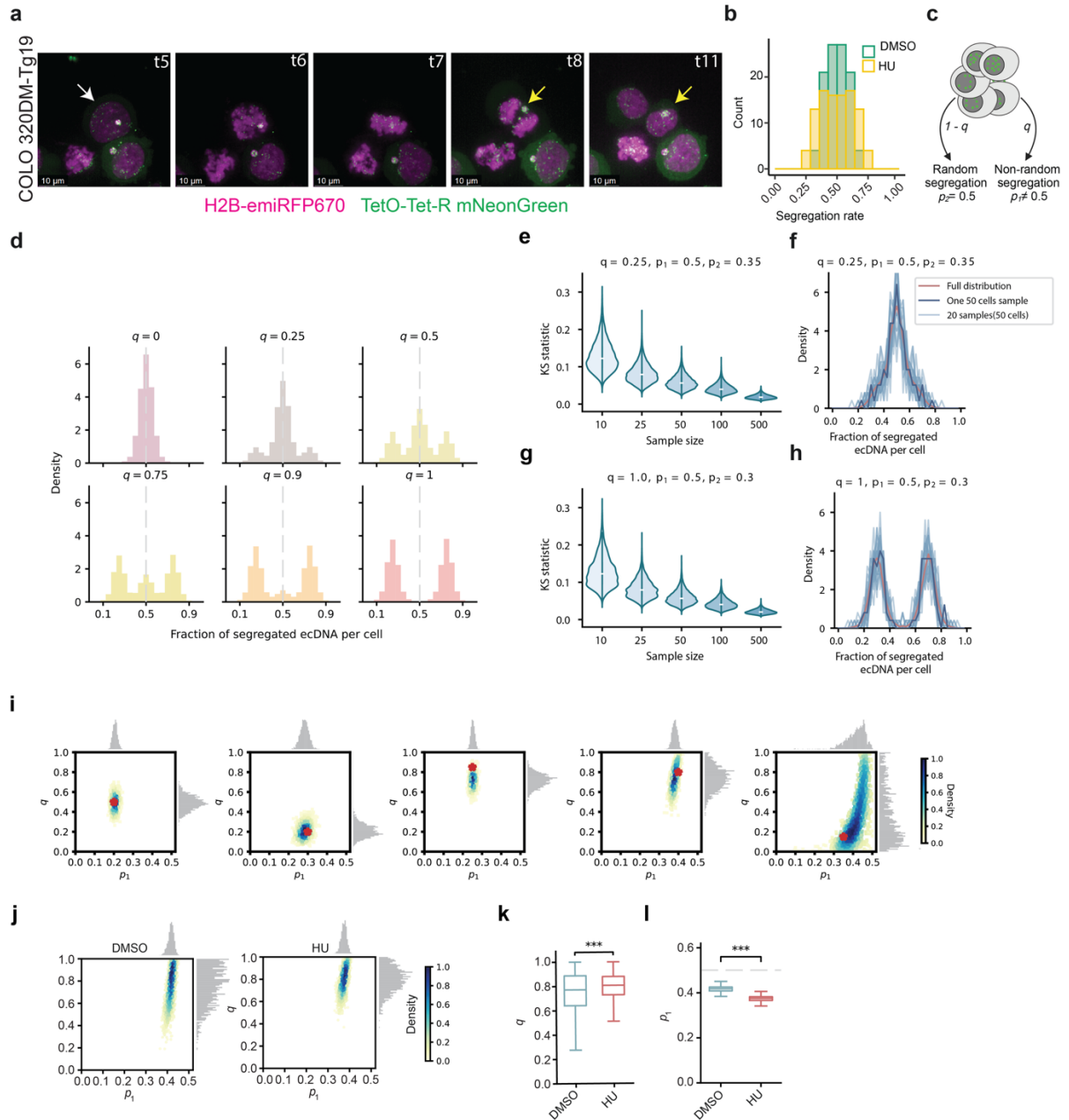
e. Fraction of micronuclei, normalized by cell number, containing *MYC* ecDNA, *MYCL* ecDNA, *MYC* and *MYCL* ecDNA or other chromosomes in fixed DSM273 cells after *MYC* sgRNA transfection, quantified by *MYC* and *MYCL* FISH (n=38-54, One-way ANOVA, exact *P*-values can be found in Supplementary Table 3; * *P*<0.05; **, *P*<0.01; ****, *P*<0.0001). Each dot represents one image.

f. Quantification of γ H2AX mean intensity in micronuclei in fixed DMS273 cells after *MYC* sgRNA transfection stratified by micronucleus content, measured by *MYC*, *MYCL* FISH and γ H2AX immunofluorescence (n= 92-128, One-way ANOVA, exact *P*-values can be found in Supplementary Table 3; ** *P*<0.001; ***, *P*<0.001; ****, *P*<0.0001). Each dot represents one micronucleus.

a



Supplementary Fig. 11: Schematic model illustrating the pathway from replication stress induction to the formation of ecDNA-containing micronuclei. Figure partially created in BioRender. Xu, R. (2026) <https://BioRender.com/8k2z6r7>.



Supplementary Fig. 12: Asymmetric inheritance of ecDNA⁺ MN by live cell imaging and a biased random-segregation model.

a. Live-cell imaging of the COLO 320DM-Tg19 cell line engineered to visualize MYC ecDNA (TetO-TetR-mNeonGreen) during mitotic transition. Representative snapshots of a mitotic cell (white arrow) carrying an ecDNA-containing micronucleus (yellow arrow) getting asymmetrically inherited into one of the daughter cells upon mitotic exit (n=3). Scale bar: 10 μ m.

b. Histograms of ecDNA fractions in the CHP212 cell line in DMSO treated cells without visible lagging clusters (left), DMSO treated cells (middle) and HU treated cells (HU), showing the distribution of ecDNA segregation across daughter cells (n= 50 anaphases untreated, n= 50 anaphases after 24h HU treatment). Each count represents one daughter cell.

c. Overview and simulation results of the biased random segregation model. Figure partially created in BioRender. Xu, R. (2026) <https://BioRender.com/8k2z6r7>.

d. Simulated distributions of ecDNA fractions under varying q values (proportion of biased segregation), with $p_1 = 0.25$ (extent of biased segregation).

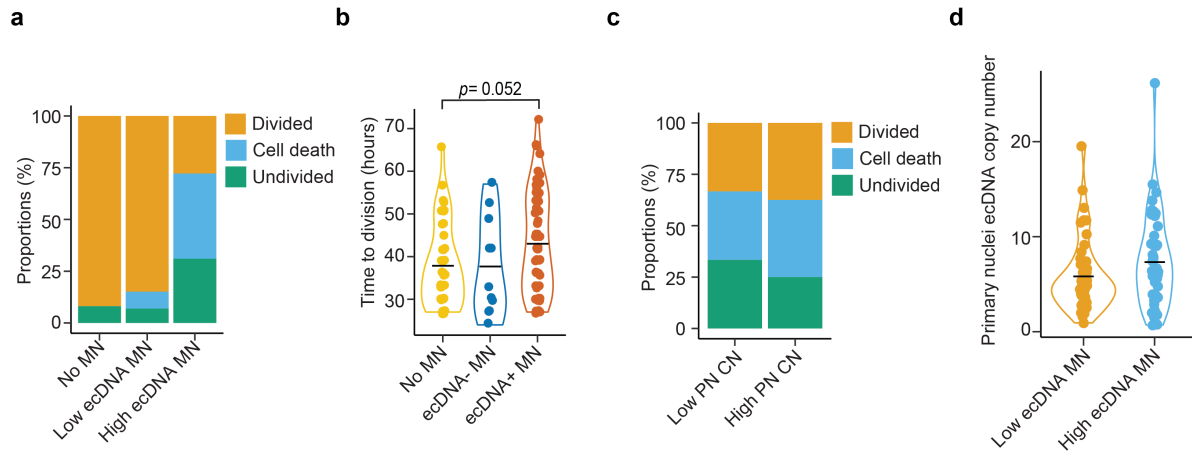
e-h. Comparison of ecDNA fraction distributions under different sample sizes. (e) Violin plot showing the Kolmogorov distances between the sampled distributions and the full distribution across varying sample sizes. (f) Comparison of ecDNA fraction distributions: the red line represents the full distribution generated from 10^7 cell divisions, while the blue line represents the distribution based on 50 sampled ecDNA fractions. (g, h) Same analyses as shown in (e) and (f), respectively, but generated using different parameters as indicated in the figure.

i. Density scatter plot of inferred parameters p_1 and q using synthetic data (50 pairs of daughter cells) for different combinations p_1 and q . The marginal distributions of p_1 and q are displayed above and to the right, respectively. Density values are normalized to 1. The red star indicates the true parameter values.

j. Density scatter plot of inferred parameters p_1 (extent of biased segregation) and q (proportion of biased segregation), with untreated (DMSO) cells on the left and HU-treated cells on the right. The marginal distributions of p_1 and q are shown above and to the right, respectively. Density values are normalized to 1.

k. The inferred proportion of cells exhibiting biased segregation is significantly increased in drug-treated cells compared to untreated cells. Statistical significance was determined using the two-sided Mann-Whitney U test (n= 50 anaphases per condition, $P < 0.001$; ***, $P < 0.001$). Data are presented as box-and-whisker plots. The centre line indicates the median. The box extends from the 25th to the 75th percentile (IQR). The whiskers represent the smallest and largest values within $1.5 \times$ IQR, and points beyond the whiskers are shown as individual outliers.

l. The inferred p_1 is significantly reduced in drug-treated cells compared to untreated cells. Statistical significance was determined using the two-sided Mann–Whitney U test ($n=50$ anaphases per condition $P<0.001$; ***, $P<0.001$). Data are presented as box-and-whisker plots. The centre line indicates the median. The box extends from the 25th to the 75th percentile (IQR). The whiskers represent the smallest and largest values within $1.5 \times \text{IQR}$, and points beyond the whiskers are shown as individual outliers.



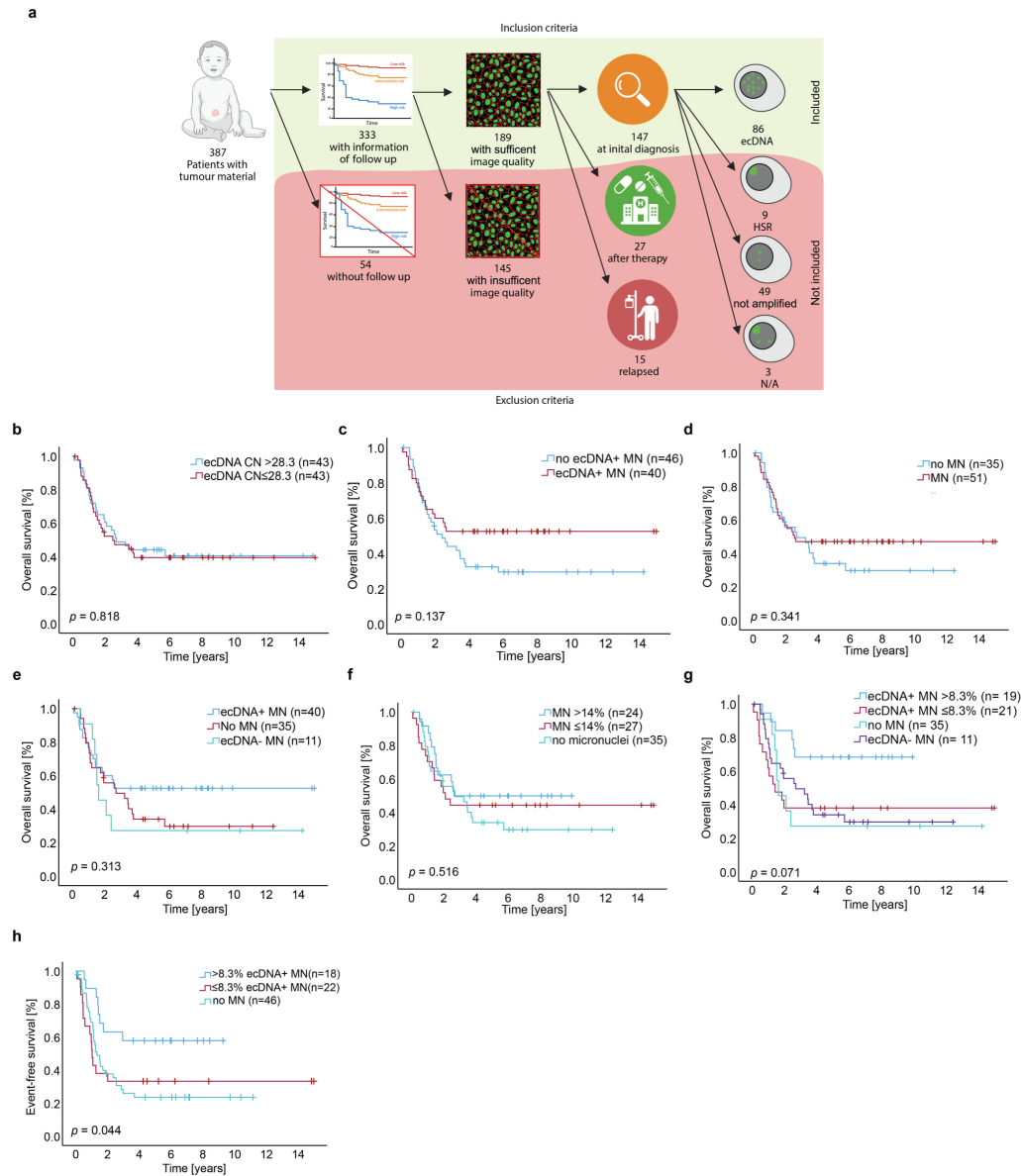
Supplementary Fig. 13: Fitness in cells with ecDNA⁺ MN is correlated with micronucleus size, but not the ecDNA copy number in primary nucleus

a. Quantification of COLO 320DM-Tg19 cell fates as determined using live cell imaging in the absence of HU for cells without micronuclei (No MN, n=17), with MN containing low ecDNA content (n=18), or with MN containing high ecDNA content (n=13). Cell fates include cell division, cell death, or remaining undivided.

b. Time to the next cell division determined by live cell imaging in COLO 320 DM-Tg19 cells pre-treated with hydroxyurea for 24 hours (no MN n=49, ecDNA⁻ MN n=13, ecDNA⁺ MN n= 58). Data is collected from three independent experiment (One-way ANOVA).

c. Fate of daughter cells stratified by low or high ecDNA content in PN, as determined by whether the ecDNA content in PN is below (low, n=55) or above (high, n=37) the mean of ecDNA content in all PN. Outcomes include continued division, cell death, or remaining undivided. Each bar plot depicts an individual experiment.

d. Copy number of ecDNA in the primary nuclei in cells with low (below median, n=46) or high (above median, n= 46) ecDNA content in micronuclei.



Supplementary Fig. 14: High ecDNA micronucleation improves patient survival

a. Schematic showing the inclusion and exclusion criteria for patient selection in the following analysis. Figure partially created in BioRender. Xu, R. (2026) <https://BioRender.com/8k2z6r7>.

b. Kaplan–Meier survival curves for patients with above or below median copy number (CN) in primary nucleus determined by FISH (OS, log-rank test).

c. Kaplan–Meier survival curves for patients with or without ecDNA⁺ MN, as determined by FISH signal in the MN (OS, log-rank test).

d. Kaplan–Meier survival curves for patients with or without MN (OS, log-rank test).

e. Kaplan–Meier survival curves for patients with ecDNA⁺ MN, ecDNA⁻ MN or without MN, as determined by FISH signal in the MN (OS, log-rank test).

f. Kaplan–Meier survival curves for patients with above or below median MN frequency or no MN (OS, log-rank test).

g. Kaplan–Meier survival curves for patients with >8.3% ecDNA⁺ MN, <8.3% ecDNA⁺ MN, ecDNA⁻ MN or no MN, as determined by FISH signal in the MN (OS, log-rank test).

h. Kaplan–Meier survival curves for patients with >8.3% ecDNA⁺ MN, <8.3% ecDNA⁺ MN, or no MN, as determined by FISH signal in the MN (Event-free survival, log-rank test).

Supplementary Note

Methods

Cell lines

Human tumour cell lines were obtained from the American Type Culture Collection (ATCC) or a gift from collaborative laboratories. The identity of all cell lines was verified by short tandem repeat genotyping (Eurofins Genomics). The absence of *Mycoplasma* sp. contamination was determined using a Lonza MycoAlert system (Lonza). For this study, 11 ecDNA amplified cell lines (DMS273, STA-NB10DM, COLO320 DM, TR-14, CHP-212, Lan5, UKF-NB6, KP-N-YN, PC3-DM, SiMA, SMS-KAN), 5 HSR amplified cell lines (STA-NB10HSR, COLO 320HSR, PC3-HSR, NGP, IMR5/75), 2 non-amplified cancer cell line (ShEP, HeLa) and 1 non-cancerous cell line (RPE-hTERT) were used. HEK293T cells were used for virus production. Near isogenic cell line pairs GBM39ec, GBM39HSR, COLO320 DM, COLO320 HSR, PC3-DM, PC3-HSR cells were extensively characterized before².

Cell culture

Most neuroblastoma and colon cancer cell lines were cultured in RPMI-1640 medium (Gibco) supplemented with 1 % of penicillin, streptomycin, and 10 % of fetal calf serum (FCS) (Thermo Fisher). STA-NB10DM and HSR cells were cultured in RPMI-1640 medium (Gibco) supplemented with 1 % of penicillin, streptomycin, 10 % of fetal calf serum (FCS) (Thermo Fisher), 1 % Glutamax (Thermo Fisher), 1 % Sodium Pyruvate (Thermo Fisher) and 10 mM HEPES (Sigma). UKF-NB6 was cultured in IMDM medium (Gibco) supplemented with 1 % of penicillin, streptomycin, and 10 % of FCS. RPE and HEK293T cells were cultured in DMEM (Gibco) supplemented with 1 % of penicillin, streptomycin, and 10 % of FCS. GBM39ec and GBM39HSR cell lines were cultured in Dulbecco's Modified Eagle's Medium/F12 (Gibco, catalogue no. 11320-033) supplemented with 1× B27 (Gibco, catalogue no. 17504-01), 20 ng/ml epidermal growth factor (Sigma, catalogue no. E9644), 20 ng/ml fibroblast growth factor (Peprotech, catalogue no. AF-100-18B), 5 µg/ml heparin (Sigma, catalogue no. H3149) and 1× GlutaMAX (Gibco, catalogue no. 35050-061). COLO 320DM, COLO 320HSR, PC3-DM, PC3-HSR and COLO 320DM-Tg19 (live-cell imaging line) cells were cultured in DMEM (Corning, #10-013-CV) with 1% PSQ and 10% FBS. To assess the number of viable cells, cells were trypsinized (Thermo Fisher), resuspended in medium, and sedimented at 300 g for 5 minutes. STA-NB10 cells were split using Accutase (Thermo

Fisher). HeLa cells were cultured in Advanced DMEM-F12 (Gibco) with addition of 10% FCS, 1% Penicillin-Streptomycin (Gibco) and 1% GlutaMax (Gibco). Cells were then resuspended in medium, mixed in a 1:1 ratio with 0.02 % trypan blue (Thermo Fisher), and counted with a Bio-Rad TC20 cell counter. CHP212-PIP-FUCCI (Nanopore-based replication fork dynamics) cells were generated by lentiviral transduction of pLenti-PGK-Hygro-PIP-FUCCI (Addgene # 138714). Transduced cells were selected by FACS sorting 10 days after transduction and kept under 100 µg/mL Hygromycin B selection. aNSC were maintained in 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, 07980).

In drug experiments, cells were treated for 24 or 72h with 50µM or 80 µM or 100µM hydroxyurea or a combination of 80 µM hydroxyurea together with 1 µM TD-19 (Merck Millipore, 532912) as indicated in figure legends respectively. Control cells were treated with an equivalent volume of DMSO or H₂O. Drug concentrations were selected based on cell viability as assessed using the CellTiter-Glo Assay (Promega G7571). GBM39ec/GBM39HSR cells were treated with 100 µM hydroxyurea for 5 and 72 hrs, EdU (10 µM) was added for 30 mins before fixation. dNTP supplementations were conducted at the indicated concentrations at a 1:1 ratio with Lipofectamine 3000 (ThermoFisher, L3000001) according to manufactures protocol.

Retrovirus production

pBABE-Puro-C-Myc plasmid was generated by site-directed mutagenesis from pBABE-Puro-C-Myc T58A (Addgene #53178) by removing the T to A mutation in the mouse coding region. For retrovirus production, HEK-293T cells were seeded in high-glucose Dulbecco's modified Eagle medium (DMEM, 4.5 g l⁻¹; Thermo Fisher Scientific, 11995065) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Sigma-Aldrich, F2442) without antibiotics in 100 mm petri dishes. The next day, cells at approximately 60–70% confluence were transfected with 20 µg of retroviral vector carrying the mouse C-Myc cDNA or a control vector, 5 µg of packaging plasmid and 1 µg of envelope plasmid. After 24 h, the medium was replaced with DMEM with antibiotics. Cells were incubated for 48 h with two consecutive collections of the medium containing the retroviral particles at 24 and 48 h. Medium collected was filtered using a 0.45 µm filter unit and used to transduce aNSC cells by incubating cells with viral supernatant supplemented with polybrene (0.2 µl ml⁻¹; Sigma, TR-1003G).

Genetically induced ecDNA cell lines

aNSC overexpressing Myc transgene were generated by infecting p53fl/fl aNSC with pBABE-Puro-Myc retrovirus. Mycec/+; p53fl/fl and p53fl/fl aNSC were also infected with pBABE-Puro retrovirus. Upon infection, cells were selected with puromycin (1 µg ml⁻¹; Sigma-Aldrich, P9620) for 3-4 days. Selected cells were then infected with AdCre (ViraQuest, VQ-Ad-CMV-Cre; 1 × 10¹² particles per milliliter; 091317) to promote p53 loss and Myc-containing ecDNA accumulation as described in Pradella et al.³

Imaging of aNSCs

aNSC were fixed with ice-cold methanol for 20 min. Briefly, fixed slides were pretreated with Pepsin (Sigma P-6887), dehydrated and counterstained with DAPI (Thermo, D1306) and mounted with Prolong Diamond Antifade Mounting medium (Thermo, P36965). Images were acquired with a REVOLVE R4 (Echo Laboratories) microscope through a ×63 objective lens. Images were analyzed in the built-in Echo app. Each nucleus was evaluated for the presence of micronuclei and Myc amplification using DNA FISH.

Immunoblotting

Whole-cell protein lysates were prepared by lysing cells in 10x Cell lysis buffer (Cell signaling) supplemented with cOmplete (Roche) and PhosStop (Roche) phosphatase inhibitors. Protein concentrations were assessed by bicinchoninic acid assay (BCA, Santa Cruz Biotechnology). For 5 minutes, 15 µg of protein was denatured in Laemmli buffer at 90 °C. Samples were run on NuPage 10 % polyacrylamide, 1 mm Tris-Glycine Protein Gels (Thermo Fisher Scientific) and transferred to PVDF membranes (Roche). Membranes were blocked with 5 % dry milk in TBS with 0.1 % (v/v) Tween-20 (Carl Roth). Membranes were probed with primary antibodies overnight at 4 °C and then with secondary antibodies conjugated to horseradish peroxidase for 1 hour at room temperature. Details on antibodies used and dilutions can be found in Supplementary Table 4. Chemiluminescent detection of proteins was carried out using Pierce ECL Western blotting substrate (Thermo Fisher) and the ChemiDoc (Bio-Rad). Densitometry was performed using ImageJ (NIH).

aNSC cells were lysed in Laemmli buffer, supplemented with protease (cOmplete; Roche, COEDTAF-RO) and phosphatase (EDTA-free Protease Inhibitor Cocktail; Roche, PHOSS-RO) inhibitors. Proteins were separated by SDS polyacrylamide gel electrophoresis and analyzed by western blotting using standard procedures. After protein transfer, the

nitrocellulose membranes (Bio-Rad, catalogue no. 1704271) were blocked by incubation with LICOR Intercept (TBS) blocking buffer. Images were acquired using an Odyssey Imaging System (LICOR).

Preparation of metaphase spreads

Cells were grown to 80% confluency in a 15 cm dish and metaphase-arrested by adding KaryoMAX™ Colcemid™ (10 µL/mL, Gibco) for 1-2 hours. Cells were washed with PBS, trypsinized and centrifuged at 200 g for 10 min. We added 10 mL of 0.075 M KCl preheated at 37 °C, one mL at a time and vortexing at maximum speed in between. Afterwards, cells were incubated for 20 min at 37 °C. 5 mL of ice-cold 3:1 MeOH/acetic acid (kept at -20 °C) were added, one mL at a time followed by resuspension of the cells by flicking the tube. The sample was centrifuged at 200 g for 5 min. Addition of the fixative followed by centrifugation was repeated four times. Two drops of cells within 200 µL of MeOH/acetic acid were dropped onto prewarmed slides from a height of 15cm. Slides were incubated overnight.

Combined FISH and immunofluorescence (ImmunoFISH)

Interphase cells were plated on Poly-D-lysine-coated coverslips (Neuvitro Corporation) in 24-well plates. Cells were fixed with 4% paraformaldehyde for 10 minutes at room temperature, followed by three washes with 125 mM glycine in PBS (5 minutes each). Permeabilization was performed using 0.5% Triton X-100 in PBS for 10 minutes at room temperature, followed by three washes with PBS. Cells were then blocked in 3% bovine serum albumin (BSA) for 30 minutes and incubated overnight at 4 °C with the primary antibody diluted in 3% BSA. After three additional PBS washes, cells were incubated with the secondary antibody (diluted in 3% BSA) for 2 hours at room temperature. Details on antibodies used and dilutions can be found in Supplementary Table 4. Post-incubation, cells were re-fixed in 4% paraformaldehyde for 20 minutes at room temperature. DNA was subsequently denatured by treating cells with 0.7% Triton X-100 and 0.1 M HCl on ice, followed by additional chemical denaturation in 1.9 M HCl for 30 minutes at room temperature. The slides were then washed with 125 mM glycine in PBS, followed by a 5-minute wash in 2× saline-sodium citrate (SSC) buffer. Dehydration and probe hybridization was conducted following the standard FISH protocol described above.

Triple FISH

Cells were cultured on poly-D-lysine-coated coverslips (Neuvitro Corporation) placed in 24-well plates and fixed in methanol/acetic acid (3:1, v/v) for 20 min at -20 °C. Coverslips were

equilibrated in 2× SSC and subsequently dehydrated through a graded ethanol series (70%, 85%, and 100%, 2 min each). For hybridization, 2 µl of each probe specific for *CDK4*, *MYCN*, and *MDM2* (myTags®, Arbor Biosciences) conjugated to Alexa 488, Atto 550, and Atto 633, respectively, were mixed with 10 µl hybridization buffer (50% formamide, 10% dextran sulfate, 40 ng/µl RNase A in SSCT). A 5 µl aliquot of this mixture was applied per coverslip. Hybridization was carried out at 80 °C for 10 min followed by incubation at 37 °C overnight. Post-hybridization washes were performed according to the standard protocol for single-gene DNA FISH. For ecDNA copy number estimation in triple FISH experiments, foci were detected using Radial Symmetry-FISH (RS-FISH)⁴. Parameters were set manually for each channel and applied to all images from the same coverslip.

Microscopy, image acquisition and analysis

Confocal images were acquired using a Leica Stellaris 8 confocal microscope controlled by Las X software and equipped with an HC PL APO CS2 63×/1.40 OIL objective lens or ZEISS LSM 880 inverted confocal microscope using ZEN (black v.2.3).

For imaging individual anaphase cells, the acquisition settings were as follows: 512 × 512 pixels frame size, 6× zoom, 0.79 µs pixel dwell time, and a final pixel size of 0.06 µm × 0.06 µm. Z-sampling was set to 0.23 µm to capture the entire cell in three dimensions.

For groups of interphase cells, images were acquired with a 2048 × 2048 pixels frame size, 2× zoom, 3.16 µs pixel dwell time, and a final pixel size of 0.05 µm × 0.05 µm.

Fluorophores were excited using diode laser lines at 561 nm (Alexa Fluor 568), 488 nm (Alexa Fluor 488), and 405 nm (Hoechst). Fluorescence emissions were detected at 566–620 nm (Alexa Fluor 568), 500–550 nm (Alexa 488), and 420–480 nm (Hoechst) using a pinhole set to 1 Airy unit (AU).

To ensure comparability across samples, all images were acquired under identical instrument settings.

The “Find Maxima” function was employed to quantify extrachromosomal DNA (ecDNA) copy numbers in FISH images. In anaphase cells, lagging DNA was defined as chromosomal material visibly separated from the primary segregating chromosomes by a minimum distance of 0.3 µm across the Z-stack, consistent with the lateral resolution limit of the imaging system. Widefield imaging was performed on a Leica DMI8 widefield microscope controlled by Las X software (v.3.8.2.27713) using a ×63 oil objective. Z-stacks were taken for each field of view with image pre-processed using Small Volume Computational Clearing, and a max projection was performed by ImageJ (v.1.53t) before proceeding for further image analysis. Individual

split channel images were semi-automatically segmented and analyzed by CellProfiler⁵ (v.4.2.1). Micronuclei were scored manually and defined as round structures separated from the primary nucleus, with a diameter up to one third of the size of the main nucleus, as previously described⁶.

Colocalization analysis was done using Huygens software (deconvolution, v24.10), FIJI (tile stitching, v2.14) and Imaris (colocalization analysis, v10.2).

Nanopore-based replication fork dynamics

CHP212-PIP-FUCCI cells were seeded into T-300 flasks without Hygromycin B. At 70% confluency, cells were treated with or without 80 μ M hydroxyurea for 24h. CHP212-PIP-FUCCI cells were then successively treated with 50 μ M EdU for 3 minutes, 50 μ M BrdU for 6 minutes, 100 μ M Thymidine for 20 minutes, with 3 PBS washes in between each step. 80 μ M hydroxyurea was maintained during the whole time, if treated. Cells were trypsinized, washed with PBS and sorted for S-phase (mCherry+, mVenus-) using a BD FACSAria Fusion. Sorted cells were pelleted, flash frozen in liquid nitrogen and stored at -80°C until DNA extraction.

High molecular weight DNA was extracted using the NEB Monarch HMW DNA Extraction Kit for Tissue (NEB, T3060). DNA libraries for nanopore sequencing were prepared using the Ultra-Long DNA sequencing KIT (ONT, SQK-ULK001) according to the manufacturer's instructions. Libraries were sequenced on R9.4.1 flow cells on the MinION platform.

Reads were basecalled using Guppy (v6.4.6, dna_r9.4.1_450bps_hac model, ONT) and aligned to hg38 using minimap2⁷ (v2.30). We used DNAscent (v3.1.2)⁸ to investigate replication fork dynamics. We applied the DNAscent detect function to detect EdU and BrdU incorporation on nanopore reads in batches of around 4000 reads. The forkSense function was then used to classify replication forks based on the order EdU and BrdU incorporation.

To determine replication fork speeds, we only retained complete forks (forks running off the end of a read were excluded) from the forkSense output. We then overlapped the remaining forks with the Decoil amplicon reconstruction for CHP212 and classified them as either circular (overlapping with ecDNA regions) or linear (overlapping with any other region). To compare ecDNA forks with linear forks of similar chromatin context, we further filtered forks which overlapped with active regions, by H3K27ac ChIPseq peaks. Bed files for peak calls were taken from⁹ and downloaded from GEO (GSE90683) and converted from hg19 to hg38 using the UCSC liftover tool (<https://github.com/ucscGenomeBrowser/kent>). Fork speeds were calculated by dividing the length of each detected fork by the treatment time beginning with

the start EdU treatment until the start of the thymidine chase out (Untreated: 15 min; HU: 16.4 min).

Analysis of single micronucleus sequencing data

Amplicon reconstruction in TR-14

Four distinct ecDNA elements were reconstructed for TR14 using Decoil¹⁰ 1.1.2 with parameters `–min-vaf 0.1 –min-cov-alt 10 –min-cov 8 –fragmentmin-cov 10 –fragment-min-size 1000 –filter-score 35` or `–min-vaf 0.01 –min-cov-alt 10 –min-cov 10 –max-explog-threshold 0.01 –fragment-min-cov 10 –fragment-min-size 500`, with the reference genome GRCh38/hg38 and annotation GENCODE V42, (as in¹⁰). Additionally, a de novo assembly contig containing CDK4 for TR14 was assembled using Shasta¹¹ (v0.6.0) with adjusted parameters `minReadLength = 10000, k = 14, consensus caller Bayesian:SimpleBayesianConsensusCaller-10.csv`. The contig was aligned to GRCh38/hg38 using minimap2⁷ with parameters `-x asm10`. From the PAF alignment the overlapping genomic coordinates were extracted as BED file. Using the genomic regions of all amplicons found in TR14 a BED file and a custom reference genome was computed, which includes (1) a masked GRCh38/hg38 genome of the regions overlapping the amplicons and (2) the FASTA sequence of all the ecDNA elements as additional contigs.

Identification of micronucleated sequences

Reads were quality and adapter-trimmed using Trim Galore¹² (v0.6.10) and aligned to the human reference assembly GRCh38 with added EBV contig (NC_007605.1) using BWA-MEM¹³ (v 0.7.18) with default parameters. Duplicate reads were removed using Picard¹⁴ (v3.3.0) MarkDuplicates. We collected all samples over a total of 2 experiments which were sequenced on different flowcells. Samples from the run with higher sequencing depth were downsampled to match the median number unique reads from the other run using samtools view with the `-s` parameter. BigWig files were generated using Deeptools2¹⁵ (v3.5.5) bamCoverage with a binsize of 25kb and CPM normalization.

Candidate window identification

For each sample, we identify candidate windows, which represent genomic regions covered by reads significantly over background, detected by performing the following 4-steps: (1) The genome was divided into 25kb bins and the number of reads and number of bases covered by at least one read were counted with bedtools coverage. (2) For each sample, we fitted a negative

binomial distribution on the per-bin read counts, similar to¹⁶. Distribution parameters were estimated using Maximum Likelihood Estimation as implemented in the `fitdistrplus`¹⁷ (v1.2.1) R-Package. The threshold RC_{th} for per-bin read counts was derived from the first read count value of the cumulative mass function, where the probability of it being larger than threshold RC_{th} was <0.001 , as defined as $>1-P$, where P denotes the probability of a value being less or equal than threshold RC_{th} . (3) To remove bins with focal read accumulation, thus being bins with high read counts but low coverage, we further sorted all bins based on the percentage of bases covered. We defined a coverage threshold C_{th} as the 99th percentile. (4) Only bins with a read count $> RC_{th}$ and coverage $> C_{th}$ were retained as candidate windows.

To identify enriched sequences under the assumption that DNA sequestered into micronuclei, which are detectable by light microscopy, derive from chromosomes, chromosomal fragments or other longer chromosomal units (such as ecDNA) of a certain length, we further refine the candidate window set by removal of any windows, which do not have any adjacent candidate windows or are of small size. With this approach we limit the smallest detectable unit size, but make sure to further remove regions, which are most likely derived from amplification and/or mapping noise.

We pivot through each candidate window and search for adjacent candidate windows up and downstream within a radius r . A radius of $r=4$ was used in this study. If a candidate window has at least one other candidate window upstream and downstream within the radius r , this candidate window is further classified as a body segment, while candidate windows with only at least one other candidate window within the radius r on either side, are classified as edge segment. If a candidate window does not have any other candidate windows within its radius, this window is removed from analysis. We further align edge and body segments into closed segments, where body segments must be flanked by edge segments to be retained, thus making an edge-body-edge alignment the smallest detectable unit, which for this experiment is a fragment size of 75kb.

Copy number calling

Copy number profiles for each sample were computed similarly to CNVkit¹⁸ using a custom python script utilizing the Pysam API (<https://github.com/pysam-developers/pysam>). First, the genome was tiled into 500kb bins. For each bin, per base coverages were summed up and divided by the bin size to get a per bin mean coverage, which was then log2 transformed.

Visualization and segmentation of copy number profiles were computed using a custom R script (<https://github.com/henssen->

lab/ecDNA_micronucleus/tree/main/single_micronucleus_seq/copy_number). First, only canonical chromosomes (autosomes and sex chromosomes) were retained and telomeric and centromeric regions filtered out. GC content for each bin was calculated using the R package Biostrings¹⁹ (v.2.66.0). We performed GC correction for the log2 per bin mean coverage values using the loess function from the stats R-package with a smoothing span of 0.3 and performed median centering based on all autosomal bins. Segmentation was performed using circular binary segmentation, as implemented in the DNACopy²⁰ R-package. After smoothing and outlier removal using the smooth.CNA function, segments were called using the segment function with following parameters: undo.splits = "sdundo", undo.SD = 1.5. Log2 per bin coverages and segments were visualized with ggplot2²¹.

(Micro-)nuclear sequence analysis

Quality control and filtering

Single and pooled micronuclei samples were retained for downstream analysis if they contained more than 20 called windows and >100000 uniquely mapping reads. We additionally added a second quality control layer, where we inspected the copy number profiles and excluded samples, which showed high degrees of background amplification. Primary nuclei samples were retained if they had more than 50 called windows.

EcDNA consensus sequence calling in single micronuclei

After visual inspection of called regions in single micronuclei, we observed an enrichment of sequences overlapping the amplicon regions, as described above. To identify the consensus region between enriched sequences in single micronuclei and ecDNA reconstructions for further quantification, we generated a set of all called windows from single micronuclei passing quality control. We retained each genomic region, which was called in at least 2 samples. We overlapped these regions with the custom reconstructions and were able to retain the majority of ecDNA regions in single micronuclei. The only region we could not capture was a short region in chromosome 1, whose length fell below the smallest detectable unit size of 75kb as described above. The consensus ecDNA regions were used for subsequent analysis.

Circular read enrichment estimation (log2 coverage)

To quantify read enrichment over the consensus ecDNA regions, we further normalized the absolute per-bin read counts to counts per million to account for different library size. To estimate the enrichment of circular reads, we divided the mean coverage of all consensus

ecDNA regions by the winsorized mean coverage (top and bottom 5% values replaced) of all non-ecDNA regions, as implemented in the Winsorize function from the R-Package DescTools²² (v0.99.58). The coverage ratio was further log2 transformed.

To quantify the signal over each connected consensus ecDNA region over their flanking region, we obtained region-scaled coverage values using Deeptools2 computeMatrix scale-regions with parameters: -m 250000 -b 125000 -a 125000 -bs 25000 --missingDataAsZero --outFileNameMatrix.

Co-segregation and relative ecDNA fraction analysis

To assess ecDNA co-segregation heterogeneity, we counted the presence or absence for each amplicon in all single micronuclei and primary nuclei samples. For each sample, we overlapped all called windows with the amplicon reconstructions. An amplicon was deemed present, if it had at least 1 overlapping called window.

To evaluate the relative copy number differences between different amplicons within one sample we computed amplicon fractions within all circular reads. Read counts for each amplicon were first divided by its respective amplicon length to normalize counts for amplicon size. The length normalized counts for each amplicon were further divided by the sum of the length normalized counts of all amplicons to obtain a value, which represents the fraction of reads belonging to each respective amplicon within all circular reads.

Linear chromatin enrichment analysis

To identify chromosomal fragments, which are enriched in single micronuclei, we first accumulated per chromosome read counts, excluding all reads mapping to the amplicon regions, and normalised them to chromosome size. Enriched chromosomes have high read accumulation over background mapping. We therefore identified enriched chromosomes as extreme outliers, if the normalised read count was larger than: $Q3$ (3rd quartile) + $10 * IQR$ (interquartile range). The fragment length of the enriched chromosome was determined by summing up the number of bins, where the per bin read count was larger than the 90th percentile of all non-zero bins within the identified enriched chromosome.

Preparation of samples for Chromatin immunoprecipitation (ChIP)-sequencing

Hydroxyurea- and DMSO-treated COLO320DM cells were washed 2x with prewarmed 1X PBS. Cells were fixed by addition of 10mL ice-cold 1% FA in 10% FCS-RPMI-1640 and incubated for 10 min on a shaker. The fixation reaction was quenched by addition of 550uL

ice-cold 2.5M Glycine in PBS. Cells were washed 1x with ice-cold 1X PBS before scraping down in 10 mL ice-cold 1X PBS. Fixed cells were centrifuged at 500 x g for 5 min at 4°C and resuspended in ice-cold 1X PBS for a total of 2 washes. Cells were pelleted and snap frozen in liquid nitrogen and stored at -80 C for no more than 1 week.

Cell pellets were resuspended in 5 mL ice-cold lysis buffer 1 (140 mM NaCl; 50 mM HEPES-KOH pH 7.5; 1 mM EDTA; 10% Glycerol (v/v); 0.5% IGEPAL CA-630 (v/v); 0.25% Triton X-100 (v/v); 1x Protease inhibitor (Roche, 04693132001)) and incubated on a roller at 4°C for 10 min. Lysed cells were centrifuged at 1357 x g for 5 min, resuspended in 4 mL lysis buffer 2 (20 mM NaCl; 10 mM Tris-HCl pH 8.0; 2 mM EDTA; 0.5M EGTA; 1x Protease inhibitor) and incubated on a roller for 10 min at room temperature. The lysate was centrifuged at 1357 x g for 5 min at 4°C and resuspended in 900 µL sonication buffer (100 mM NaCl; 10 mM Tris-HCl pH 8; 1 mM EDTA; 0.5 EGTA; 0.1% Na-deoxycholat (w/v); 0.5% N-Laroylsarcosine (w/v); 1x Protease inhibitor). Lysed chromatin was sheared to a fragment size of 200-500 base pairs (bp) on a Covaris S220 (PIP 140, DF 5, CPB 200) for 12 min. Sheared chromatin was clarified by addition of 1/10 volume 10% Triton X-100 (v/v) and centrifugation at maximum speed at 4°C.

Protein-DNA-complexes were immunoprecipitated by overnight incubation with the respective primary antibody on a rotator at 4°C. 10-15 µg total chromatin was used per sample. 30 µL Protein G beads (Invitrogen, 10003D) were washed 3x in 0.25% BSA in PBS (w/v) and added to each sample. Samples were incubated overnight on a rotator at 4°C.

Antibody-bead - complexes were washed in RIPA buffer (50 mM HEPES-KOH pH 7.5; 1mM EDTA; 1% IGEPAL CA-630 (v/v); 0.7% Na-deoxycholat (w/v); 500 mM LiCl; 1x Protease inhibitor) for a total of 7 times followed by 1 wash in TE buffer (10 mM Tris-HCl pH 8.0; 1mM EDTA; 50 mM NaCl; 1x Protease inhibitor). Beads were eluted in 200 µL Elution Buffer (50 mM Tris-HCl pH 8.0; 10 mM EDTA, 1% SDS (v/v)). 20 µL 5M NaCl and 5 µL Proteinase K (NEB, P8107S) were added to each sample and DNA was de-crosslinked at 65°C overnight, followed by RNA digestion with 2 uL RNase A (Invitrogen, R1253) at 37°C for 30 minutes. DNA was purified using the QIAquick PCR Purification Kit (Qiagen, 28104) according to the manufacturer's instructions. ChIP DNA libraries were constructed using the NEBNext Ultra II DNA Library Prep Kit (NEB, E7645S) and barcoded for Illumina sequencing using the NEBNext Multiplex Oligos for Illumina kit (NEB, E6444). Libraries were sequenced on a NovaSeq X Plus sequencer (Illumina) in 100 nt single-end mode.

Preparation of samples for RNA-sequencing

Three replicates of Hydroxyurea- and DMSO-treated COLO320DM cells were washed with 1x PBS, trypsinized (Gibco) and centrifuged at 300xg for 5 min. Total RNA was immediately extracted from the cell pellets using the RNeasy Mini Kit (Qiagen, 74104) following the manufacturer's instructions. Genomic DNA was removed by DNase treatment using the RNase-Free DNase Set (Qiagen, 79254). The mRNA library was constructed using the TruSeq Stranded mRNA Library Prep kit (Illumina) at the Berlin Institute of Health (BIH) Genomics Technology Platform. Libraries were sequenced on a NovaSeq X Plus sequencer (Illumina) in 100 nt single-end mode.

Estimation of ecDNA mis-segregation probability

To estimate the mis-segregation probability of ecDNA in a patient or cell line, we adapted a binomial model of micronucleation²³. According to this model, the frequency of cells without micronuclei is given by $P_0(p_{MN}, C) = (1 - \frac{1}{2}p_{MN})^C$, where p_{MN} is the mis-segregation probability of an ecDNA element during cell division and C is the amount of ecDNA elements in the cells. Per patient and cell line where we have more than five cells imaged, we then use the average observed ecDNA copy number by FISH per cell and frequency of non-micronucleated cells to calculate p_{MN} .

Biased Random Segregation Model

We developed a model to simulate ecDNA segregation during cell division. A single cell begins with a random ecDNA copy number n , sampled from a uniform distribution. Prior to cell division, ecDNA is amplified to $2n$ copies, which are then partitioned between two daughter cells. For cells undergoing random segregation, ecDNA partitioning follows a binomial distribution $B(2n, p_2)$, where $p_2 = 0.5$. In contrast, for cells exhibiting biased random segregation, partitioning follows a binomial distribution $B(2n, p_1)$, where $p_1 < 0.5$ reflects asymmetric segregation. The probability of biased segregation is denoted as q , while $1 - q$ represents the probability of random segregation. The ecDNA copy numbers in the two daughter cells are denoted as n_1 and $n_2 = 2n - n_1$, respectively. The ecDNA fractions f_1 and f_2 of the two daughter cells can be calculated as: $f_1 = \frac{n_1}{2n}$, $f_2 = \frac{n_2}{2n}$. By repeating this process, we generate the distribution of ecDNA fractions across daughter cells.

Reconstructing ecDNA Segregation Distribution under Finite Sampling

We first simulated ecDNA segregation in 10^7 cells to establish the “true” or “full” distribution. We then sampled subsets of cells at sample sizes of 10, 25, 50, 100, and 500, repeating each sampling 10^5 times. To quantify the deviation between sampled and true distributions, we calculated the Kolmogorov distances.

Parameter Estimation Using Approximate Bayesian Computation (ABC)

We applied Approximate Bayesian Computation (ABC) to infer the parameters p_1 and q in the model. For the simulated data, 50 pairs of daughter cells were randomly drawn using predefined parameters. During inference, p_1 and q were sampled from prior uniform distributions: $p_1 \sim U(0, 0.5)$ and $q \sim U(0, 1)$. 50 parent cells were randomly selected, with their ecDNA copy numbers drawn from a uniform distribution $U(10, 100)$. For each set of parameters, we ran the model to obtain the ecDNA segregation fraction distribution. To compare the similarity between simulated and target distributions, we computed the Wasserstein distance. A threshold was applied to determine acceptance, generating the posterior distributions of p_1 and q . For experimental data, the same procedure was followed, with the sample size adjusted according to the specific experimental conditions.

Supplementary videos captions

Supplementary Video 1

3d reconstruction from live cell imaging of the COLO 320DM-Tg19 cell line, showing a second smaller-sized volume to the primary nuclei. Reconstruction corresponds to Fig.2b (t8). Green: TetO-Tet-RmNeonGreen; Magenta: H2B-emiRFP670.

Supplementary Video 2

3d reconstruction from live cell imaging of the COLO 320DM-Tg19 cell line, showing a second spatially separate volume to the primary nuclei . Reconstruction corresponds to Extended Fig. 12a (t5). Magenta: H2B-emiRFP670.

Supplementary Video 3

3d reconstruction from live cell imaging of the COLO 320DM-Tg19 cell line, showing a second spatially separate volume to the primary nuclei. Reconstruction corresponds to Extended Fig. 12a (t11). Magenta: H2B-emiRFP670.

Supplementary Video 4

3d reconstruction from live cell imaging of the COLO 320DM-Tg19 cell line in the presence of HU (50 μ M), showing a second smaller-sized volume to the primary nuclei. Reconstruction corresponds to Fig. 5b (t=33hrs). Green: TetO-Tet-RmNeonGreen; Magenta: H2B-emiRFP670.

Supplementary Video 5

3d reconstruction from live cell imaging of the COLO 320DM-Tg19 cell line in the presence of HU (50 μ M), showing a second smaller-sized volume to the primary nuclei. Reconstruction corresponds to Fig. 5b (t=42hrs; F2_1,F2_2). Green: TetO-Tet-RmNeonGreen; Magenta: H2B-emiRFP670.

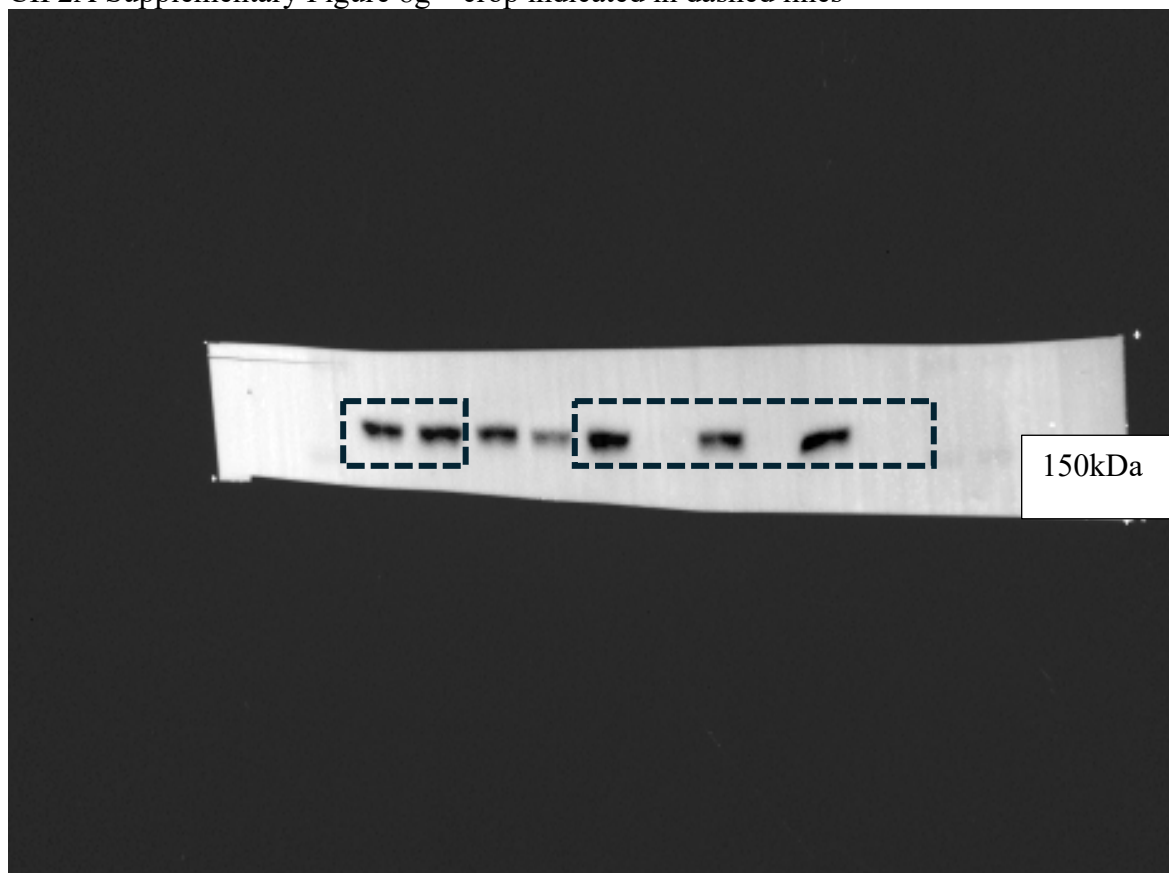
Supplementary Video 6

3d reconstruction from live cell imaging of the COLO 320DM-Tg19 cell line in the presence of HU (50 μ M), showing a second spatially separate volume to the primary nuclei. Reconstruction corresponds to Fig. 5b (t=57 hrs). Green: TetO-Tet-RmNeonGreen; Magenta: H2B-emiRFP670.

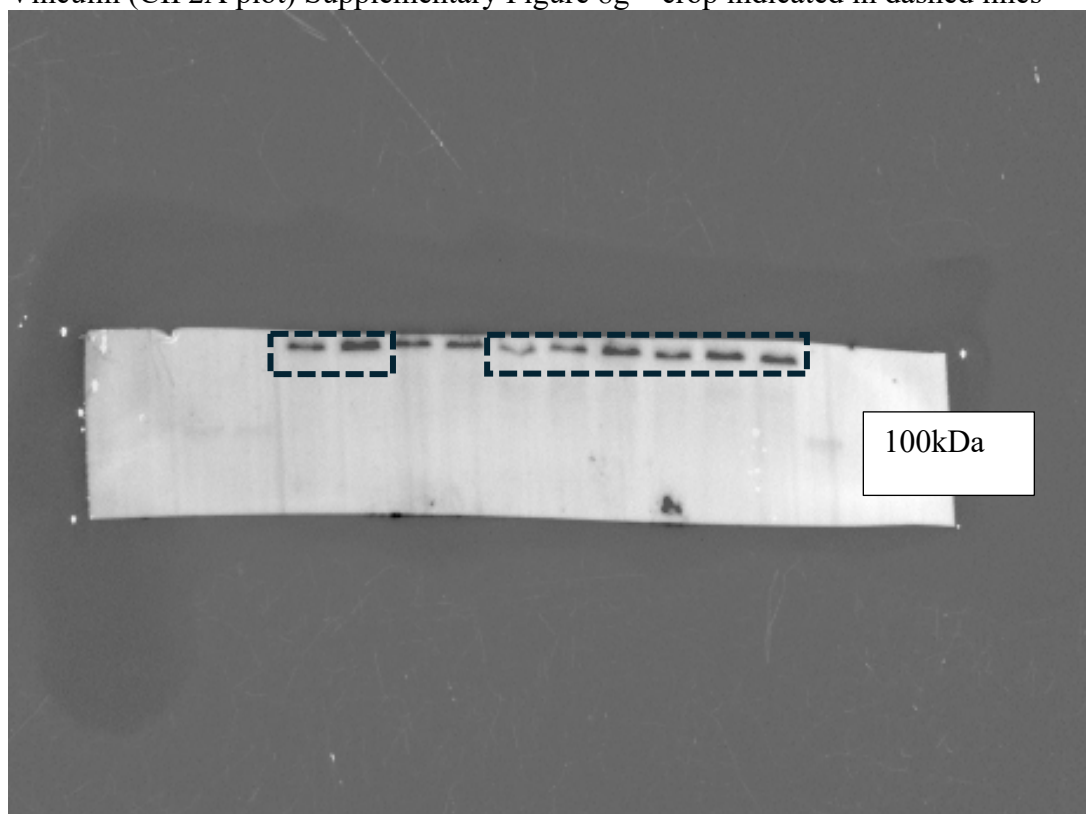
Unprocessed Gels

All loading controls and proteins were run on the same gel

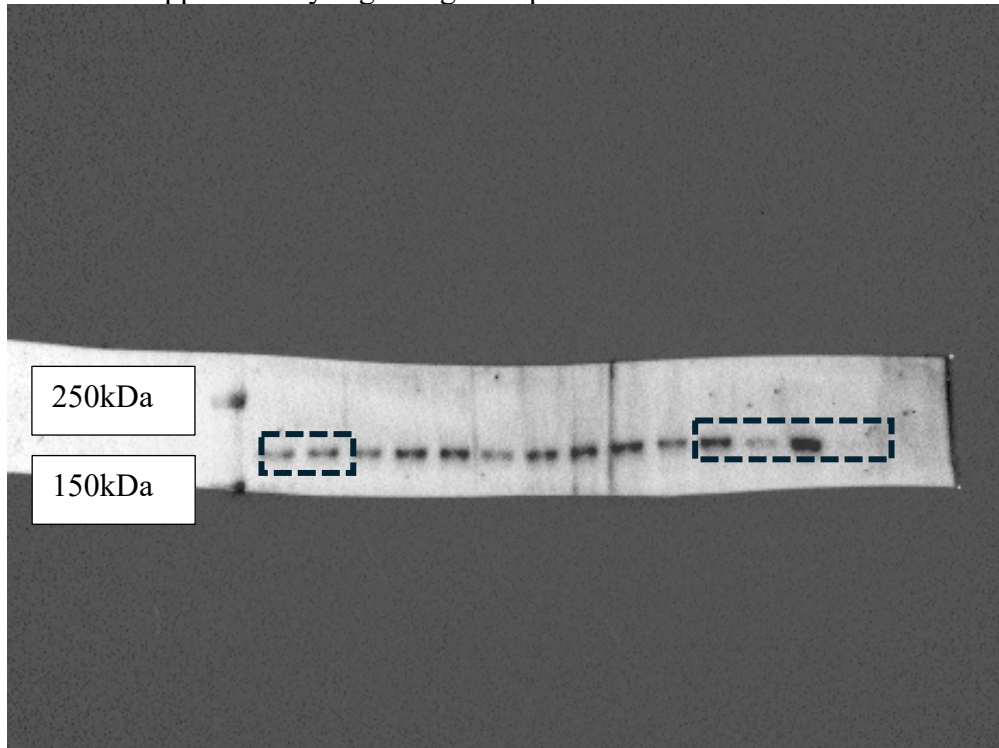
CIP2A Supplementary Figure 8g – crop indicated in dashed lines



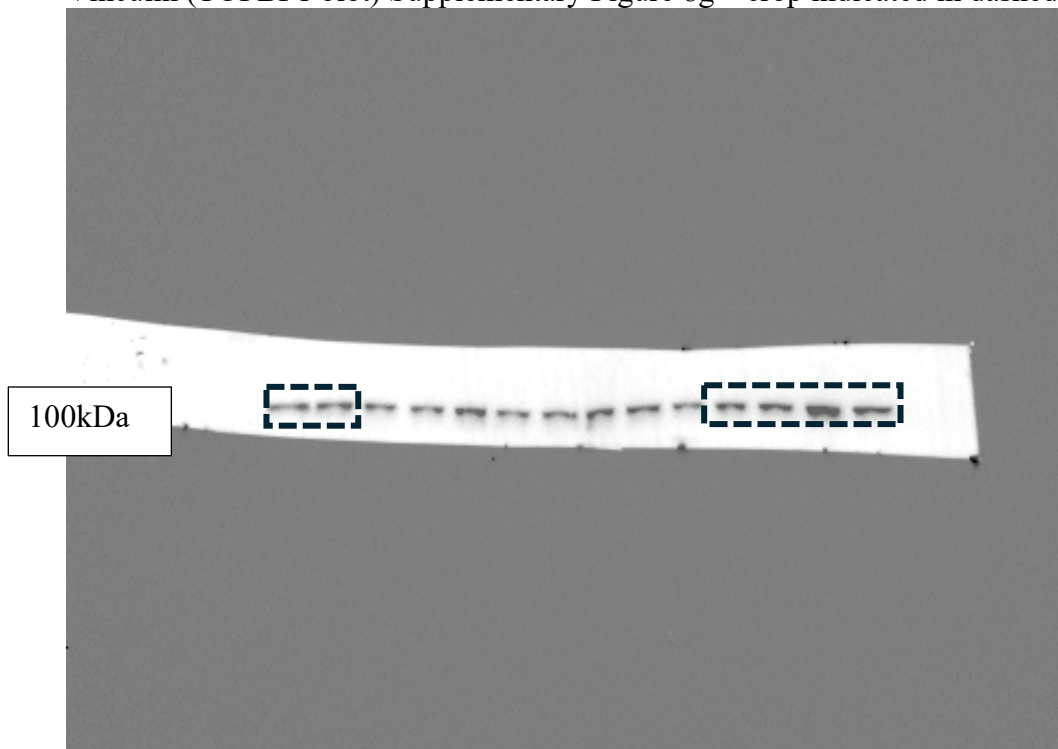
Vinculin (CIP2A plot) Supplementary Figure 8g – crop indicated in dashed lines



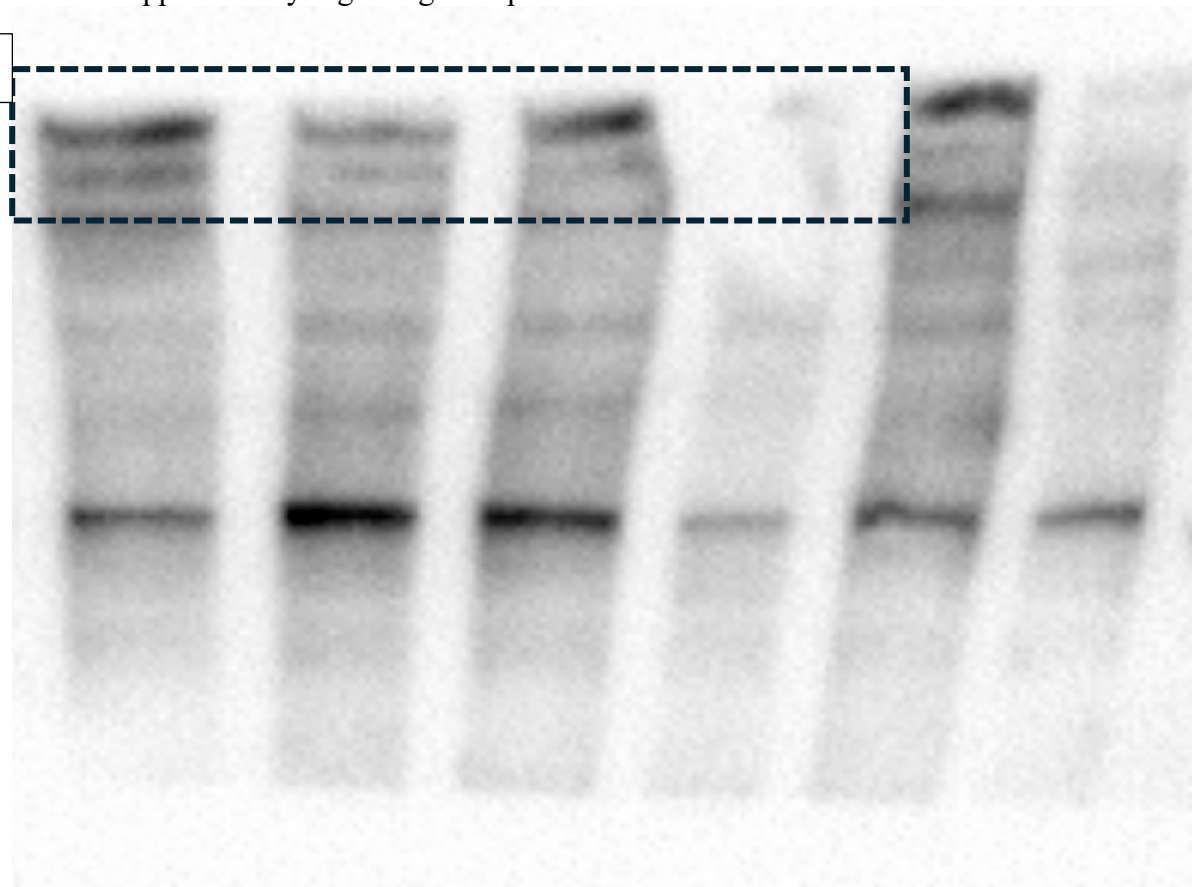
TOPBP1 Supplementary Figure 8g – crop indicated in dashed lines



Vinculin (TOPBP1 blot) Supplementary Figure 8g – crop indicated in dashed lines



MDC1 Supplementary Figure 8g – crop indicated in dashed lines



Vinculin (MDC1) Supplementary Figure 8l) – crop indicated in dashed lines



References (Supplementary Information)

- 1 Boemo, M. A. DNAscent v2: detecting replication forks in nanopore sequencing data with deep learning. *BMC Genomics* 22, 430, doi:10.1186/s12864-021-07736-6 (2021).
- 2 Tang, J. *et al.* Enhancing transcription–replication conflict targets ecDNA-positive cancers. *Nature* 635, 210-218, doi:10.1038/s41586-024-07802-5 (2024).
- 3 Pradella, D. *et al.* Engineered extrachromosomal oncogene amplifications promote tumorigenesis. *Nature* 637, 955-964, doi:10.1038/s41586-024-08318-8 (2025).
- 4 Bahry, E. *et al.* RS-FISH: precise, interactive, fast, and scalable FISH spot detection. *Nat Methods* 19, 1563-1567, doi:10.1038/s41592-022-01669-y (2022).
- 5 Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* 22, 433, doi:10.1186/s12859-021-04344-9 (2021).
- 6 Sabharwal, R. *et al.* Emergence of micronuclei as a genomic biomarker. *Indian J Med Paediatr Oncol* 36, 212-218, doi:10.4103/0971-5851.171541 (2015).
- 7 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).
- 8 Jones, M. J. K. *et al.* A high-resolution, nanopore-based artificial intelligence assay for DNA replication stress in human cancer cells. *Nat Commun* 16, 7732, doi:10.1038/s41467-025-63168-w (2025).
- 9 Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet* 49, 1408-1413, doi:10.1038/ng.3921 (2017).
- 10 Giurgiu, M. *et al.* Reconstructing extrachromosomal DNA structural heterogeneity from long-read sequencing data using Decoil. *Genome Res* 34, 1355-1364, doi:10.1101/gr.279123.124 (2024).
- 11 Shafin, K. *et al.* Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol* 38, 1044-1053, doi:10.1038/s41587-020-0503-6 (2020).
- 12 Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. TrimGalore v.0.6.10. <https://zenodo.org/records/7598955> (2021).
- 13 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio.GN]* (2013).
- 14 Picard toolkit. *Broad Institute, GitHub repository* <https://broadinstitute.github.io/picard/> (2019).
- 15 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44, W160-165, doi:10.1093/nar/gkw257 (2016).
- 16 Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543, 519-524, doi:10.1038/nature21411 (2017).
- 17 Delignette-Muller, M. L. & Dutang, C. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software* 64, 1 - 34, doi:10.18637/jss.v064.i04 (2015).
- 18 Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* 12, e1004873, doi:10.1371/journal.pcbi.1004873 (2016).
- 19 Pagès H, A. P., Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings version 2.66.0. <https://bioconductor.org/packages/Biostrings> (2022).
- 20 Seshan VE, O. A. DNACopy: DNA copy number data analysis version 1.72.3. <https://bioconductor.org/packages/DNACopy> (2023).

- 21 H, W. *ggplot2: Elegant Graphics for Data Analysis.*, (Springer-Verlag New York., 2016).
- 22 Signorell, A. DescTools: Tools for Descriptive Statistics. <https://CRAN.R-project.org/package=DescTools>, doi:10.32614/CRAN.package.DescTools (2024).
- 23 Savage, J. R. A comment on the quantitative relationship between micronuclei and chromosomal aberrations. *Mutat Res* 207, 33-36, doi:10.1016/0165-7992(88)90008-5 (1988).