

1 **Principles of ecDNA random inheritance drive rapid genome change and therapy**
2 **resistance in human cancers**

3
4 Joshua T. Lange^{1,2#}, Celine Y. Chen^{3#}, Yuriy Pichugin^{4#}, Liangqi Xie^{5,6}, Jun Tang^{1,2}, King
5 L. Hung⁷, Kathryn E. Yost⁷, Quanming Shi⁷, Marcella L. Erb⁸, Utkrisht Rajkumar⁹, Sihan
6 Wu^{1,2}, Charles Swanton^{10,11,12}, Zhe Liu⁵, Weini Huang^{13,14*}, Howard Y. Chang^{7,15*}, Vineet
7 Bafna^{9*}, Anton G. Henssen^{3,16,17,18*}, Benjamin Werner^{19*}, Paul S. Mischel^{1,2*}

8
9 ¹ Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA
10 ² ChEM-H, Stanford University, Stanford, CA, USA
11 ³ Department of Pediatric Oncology/Hematology, Charité—Universitätsmedizin Berlin, Berlin,
12 Germany
13 ⁴ Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön,
14 Germany
15 ⁵ Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA
16 ⁶ Department of Molecular and Cell Biology, Li Ka Shing Center for Biomedical and Health
17 Sciences, CIRM Center of Excellence, University of California, Berkeley, CA, USA
18 ⁷ Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford,
19 CA, USA
20 ⁸ UCSD Light Microscopy Core Facility, Department of Neurosciences, University of California
21 San Diego, La Jolla, CA, USA
22 ⁹ Department of Computer Science and Engineering, University of California, San Diego, La
23 Jolla, CA, USA
24 ¹⁰ Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK
25 ¹¹ Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer
26 Institute, London, UK
27 ¹² Department of Medical Oncology, University College London Hospitals, London, UK
28 ¹³ Group of Theoretical Biology, The State Key Laboratory of Biocontrol, School of Life Science,
29 Sun Yat-sen University, Guangzhou, China
30 ¹⁴ Department of Mathematics, Queen Mary University of London, London, UK
31 ¹⁵ Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA
32 ¹⁶ Experimental and Clinical Research Center (ECRC), Max Delbrück Center for Molecular
33 Medicine and Charité—Universitätsmedizin Berlin, Berlin, Germany.
34 ¹⁷ German Cancer Consortium (DKTK), partner site Berlin, and German Cancer Research
35 Center DKFZ, Heidelberg, Germany.
36 ¹⁸ Berlin Institute of Health, Berlin, Germany.
37 ¹⁹ Evolutionary Dynamics Group, Centre for Cancer Genomics and Computation Biology, Barts
38 Cancer Institute, Queen Mary University of London, London, UK

39
40 # These authors contributed equally to this work.

41 * These authors jointly supervised this work.

42

43

44

45 **The foundational principles of Darwinian evolution are variation, selection, and**
46 **identity by descent. Oncogene amplification on extrachromosomal DNA (ecDNA)**
47 **is a common event, driving aggressive tumour growth, drug resistance, and**
48 **shorter survival in patients¹⁻⁴. Currently, the impact of non-chromosomal oncogene**
49 **inheritance—random identity by descent—is not well understood. Neither is the**
50 **impact of ecDNA on variation and selection. Here, integrating mathematical**
51 **modeling, unbiased image analysis, CRISPR-based ecDNA tagging, and live-cell**
52 **imaging, we identify a set of basic “rules” for how random ecDNA inheritance**
53 **drives oncogene copy number and distribution, resulting in extensive**
54 **intratumoural ecDNA copy number heterogeneity and rapid adaptation to**
55 **metabolic stress and targeted cancer treatment. Observed ecDNAs obligatorily**
56 **benefit host cell survival or growth and can change within a single cell cycle. In**
57 **studies ranging from well-curated, patient-derived cancer cell cultures to clinical**
58 **tumour samples from patients with glioblastoma and neuroblastoma treated with**
59 **oncogene-targeted drugs, we show how these ecDNA inheritance “rules” can**
60 **predict, *a priori*, some of the aggressive features of ecDNA-containing cancers.**
61 **These properties are entailed by their ability to rapidly change their genomes in a**
62 **way that is not possible for cancers driven by chromosomal oncogene**
63 **amplification. These results shed new light on how the non-chromosomal random**
64 **inheritance pattern of ecDNA underlies poor outcomes for cancer patients.**

65

66 Inheritance, variation, and selection are foundational principles of Darwinian organismal
67 evolution that have been used to explain how cancers evolve⁵⁻⁸. The concept of genetic
68 identity by descent is central to the application of evolutionary theory to cancer,
69 suggesting a physical basis for identity through chromosomal inheritance during cell
70 division – thereby explaining the clonal trajectories commonly seen in tumours⁹⁻¹².
71 However, several issues challenge current models of tumour clonal evolution. First, some
72 aggressive forms of cancer maintain high levels of intratumoural copy number
73 heterogeneity instead of undergoing selective sweeps, as would be predicted¹³. This is
74 especially true for amplified oncogenes, whose cell-to-cell variability remains high,
75 despite the fitness advantage conferred^{2,14-16}. Consequently, the mechanisms

76 maintaining heterogeneous oncogene amplification events have been difficult to
77 establish. Second, the ability of some cancers to rapidly adapt to changing conditions,
78 including treatment, by changing their genomes, especially changing the copy number of
79 amplified oncogenes, isn't well explained by current models of genetic inheritance². Third,
80 the lag time to resistance predicted by the selection for drug resistance-conferring
81 mutations arising in a single cell, or a small number of cells, isn't seen in some cancers,
82 raising questions about whether tumours are undergoing a genetic bottleneck^{2,17}. The
83 presence of extrachromosomal DNA (ecDNA) amplification may explain some of these
84 paradoxical features. Extrachromosomal oncogene amplification on circular particles that
85 lack centromeres is now recognized to be a common event in human cancer that is linked
86 to poor outcome and treatment resistance in patients^{1,3}. It has been suggested that
87 ecDNAs, because they lack centromeres, are unequally segregated to daughter cells
88 during cell division^{18,19}. However, the impact of non-chromosomal oncogene inheritance
89 in cancer—random identity by descent—on intratumoural genetic heterogeneity,
90 accelerated tumour evolution, enhanced ability to withstand environmental stresses, and
91 rapid genome change on therapeutic resistance, is not well understood. Here, we apply
92 a powerful, integrated tool kit, including mathematical modeling, evolutionary theory,
93 unbiased image analysis, CRISPR-based ecDNA tagging with live cell imaging, and
94 longitudinal analyses of patients' tumours, to deduce the “rules” of ecDNA inheritance
95 and to reveal the functional consequences.

96

97 Chromosomal segregation during mitotic cell division ensures that each daughter cell has
98 the same DNA content (red line, Fig. 1a). If ecDNA segregation is random, then we predict
99 a Binomial (approximately Gaussian) distribution in the per-cell content of ecDNA, post-
100 mitotic division (Fig. 1a, Supplementary Information 1.1). Therefore, we developed a
101 method of using unbiased image analysis to quantify ecDNA in daughter cells after cell
102 division, using FISH probes to detect the amplified oncogenes residing on those ecDNAs,
103 and Aurora B Kinase immunostaining to identify the daughter cells post-mitosis²⁰ (Fig.
104 1b). In cancer cell lines of different histological types, including prostate, gastric, colon
105 cancer cells, and glioblastoma cells, carrying different oncogenes on ecDNA, we
106 quantified the ecDNA distribution of approximately 200 post-mitotic daughter cells per cell

107 line, which permits sufficient resolution (Supplementary Information 1.3), revealing a
108 Gaussian distribution that was independent of cancer cell type or the oncogene contained
109 on the ecDNA (Fig. 1b,c). The fraction of segregated ecDNA per daughter cell
110 (histograms) was highly concordant with the theoretical prediction of random segregation
111 (dashed line) (Kolmogorov-Smirnov (KS) test $p > 0.05$) (Fig. 1c, Supplementary
112 Information 1.2, 1.3). In one of the cancer cell line models, SNU16, *MYC* and *FGFR2* are
113 found on separate ecDNAs, revealing that oncogenes on different ecDNA segregated
114 independently and randomly (Fig. 1c), adding an additional layer of genetic diversity to
115 tumour cells.

116
117 To confirm these correlative observations, we designed a live-cell imaging system to
118 visualize ecDNA dynamics during cell division. We used CRISPR-Cas9²¹ to insert a TetO
119 array into the intergenic region between *MYC* and *PVT1* of the ecDNA in PC3 prostate
120 cancer cells (Fig. 1d). Insertion of this array was confirmed by PCR, sanger sequencing,
121 and TetO-MYC dual FISH (Extended Data Fig. 2a-d). Subsequent expression of TetR-
122 GFP, which binds the TetO array enabled tracking of ecDNA throughout the cell cycle
123 (Fig. 1d). Chromatin was detected by a histone H2B-SNAP tag fusion labeled with the
124 newly developed JF₆₆₉ SNAP tag ligand²². Live-cell time-lapse imaging of PC3-TetO cells
125 revealed the random inheritance pattern of ecDNA during cell division (Fig. 1e,
126 Supplementary Video 1).

127
128 Having demonstrated that ecDNA drives random identity by descent through random
129 segregation during cell division, we turned our attention to the other pillars of Darwinian
130 evolution – variance and selection. Intratumoural heterogeneity plays a significant role in
131 therapy resistance and tumour evolution^{23,24}. To better understand the impact of ecDNA
132 on heterogeneity, we generated a theoretical model of the per-cell distribution of ecDNA
133 (Fig 2a, Supplementary Information 2.1), based on the observed pattern of random
134 segregation. Specifically, starting with a single cell with a single ecDNA, let $N_k(t)$ denote
135 the number of cells with k ecDNA at time t . Assuming independent replication and random
136 segregation, the dynamics of $N_k(t)$ are governed by a set of coupled differential
137 equations.

138

$$\frac{dN_k(t)}{dt} = -N_k(t) + 2 \sum_{i=\lfloor \frac{k}{2} \rfloor}^{\infty} N_i(t) \binom{2i}{k} \frac{1}{2^{2i}}$$

139

140 The differential equations can be used to analytically estimate the distribution of ecDNA
141 numbers per cell in a growing tumour population. To test the dynamics, we quantified
142 ecDNA copy number distributions from 6 ecDNA⁺ lines of different cancer types, bearing
143 different amplified oncogenes on ecDNA — two lines contain two distinct species of
144 ecDNA as indicated (Fig. 2b). We observed a wide distribution of copy number in each
145 cancer cell line, with variation primarily dependent on the mean copy number in each cell
146 line model. The observed ecDNA copy number distributions were clearly non-Normal (Fig.
147 2b; Shapiro Wilk $p < 0.05$) and matched the predicted analytical distribution (KS test $p >$
148 0.05), except for inflation at extreme values in a few cell-lines (Supplementary Information
149 1.2). The inflation is likely due to positive selection as described below.

150

151 We next sought to test whether ecDNA heterogeneity can be observed and modeled in
152 patient tumour samples. We received FISH images on patient tumour samples or patient
153 tumour tissue from 6 GBM² and 14 neuroblastoma (NB) patients. These tumours were
154 suspected of having ecDNA amplification of either EGFR or MYCN, respectively, due to
155 their extremely high copy number—copy number greater than 16 has been found to be
156 almost exclusively due to ecDNA amplification¹. We quantified the distribution of ecDNA
157 FISH signals in these patient samples and observed distributions that again showed
158 extreme cell-to-cell variation with a non-Normal distribution (Fig. 2c, Extended Data Fig.
159 3a), strongly suggestive of positive selection in vivo, but remained in strong agreement
160 with the analytic distributions for most samples (KS test $p > 0.05$). Small discrepancies
161 can possibly be attributed to underestimation of counts due to the much more limited
162 resolution and number of cells quantified (Supplementary Information 1.3).

163

164 Importantly, the significant divergence from a normal distribution (Shapiro Wilk test $p <$
165 0.05), is indicative of a power-law tail shift, or overrepresentation of extremely high copy
166 number cells in line with our predicted modeling of ecDNA (Supplementary Information

167 1.2). Furthermore, the shift to high copy ecDNA suggests that there may be an important
168 role for selection.

169

170 To understand whether there is positive selection for ecDNA and to determine how it
171 shapes tumour evolutionary dynamics, we simulated the expansion of a single cell colony
172 with a single ecDNA into a population of 10^5 cells (Supplementary Information 1.1). Due
173 to random segregation, cells with low ecDNA copy number frequently give rise to a
174 daughter cell without ecDNA. Under neutral selection, this cell is not disadvantaged.
175 Consequently, ecDNA prevalence rapidly decays to a small minority of cells, consistent
176 with the rare observation of ecDNA in normal cells (Fig. 3a,b, Supplementary Information
177 4.1)³. In conditions where ecDNA is positively selected, however, the simulations show
178 that ecDNA remains frequent, with a continued presence in a vast majority of the cells
179 (Fig. 3a,b). We compared these simulated data to our empirical measurements of ecDNA
180 prevalence in the cell lines and patient samples measured in Figure 2. In all samples,
181 ecDNA prevalence levels suggested strong positive selection for ecDNA (Fig. 3c).

182

183 To better understand the selection landscape of ecDNA, we modeled the predicted
184 ecDNA copy number under strong positive selection, where cells carrying ecDNA are 3
185 times ($s=3$) more likely to divide compared to cells with no ecDNA. The simulations predict
186 an exponential increase in the average copy number per cell (Fig. 3d). Remarkably, when
187 plotted against the observed copy number averages in GBM, we again saw strong
188 agreement between the predicted model and our observations (Fig. 3d). Interestingly,
189 these samples fit with the modeled tumour growth when the tumour reaches a size
190 reasonable for clinical detection (10^{11} cells), potentially suggesting ecDNA as an early
191 event in the development of these tumours. An additional prediction of our simulations
192 relates to the power-law tail shift (Supplementary Information 2.3,3.1), or
193 overrepresentation of extremely high copy number cells, predicted in ecDNA⁺ populations
194 (Supplementary Information 4.1). We modeled this feature by plotting the distribution of
195 reciprocal ecDNA copy number in simulated populations under either positive selection
196 or neutral evolution (Fig. 3e). When we overlaid the simulated distributions with data from

197 the GBM patient samples, we saw a strong left-shift indicative of strong positive selection
198 (Fig. 3e).

199

200 To complement the evolutionary analyses showing ecDNA selection, we designed a set
201 of CRISPR studies to determine the reliance of tumours on ecDNA and on the oncogenes
202 encoded within the ecDNAs for growth. We designed sgRNAs targeting different genomic
203 regions of COLO320-DM *MYC* ecDNA (intergenic region on ecDNA and *MYC* gene body
204 on ecDNA) and a non-amplified, intergenic region of chromosome 8 (Fig. 3f). We infected
205 the cells with Cas9 and the sgRNAs by lentiviral vectors, quantifying cell proliferation and
206 ecDNA copy number. While Cas9-targeted cutting of chromosome 8 showed minimal
207 impact on cell proliferation, targeting of the ecDNA on an intergenic region, and even
208 more so on *MYC* on the ecDNA, caused an extreme growth deficit (Fig. 3g). When we
209 quantified ecDNA copy number in these cells, we saw a significant decrease in ecDNA 6
210 days after initial infection (Fig. 3h, Extended Data Fig. 3b). These data together confirm
211 that ecDNAs, and the oncogenes contained therein, are under strong selective pressure,
212 which influences the mean ecDNA oncogene copy number and per cell distribution in
213 tumours.

214

215 Having shown that ecDNA contributes to each of the three pillars of Darwinian evolution
216 — inheritance (i.e. random identity by descent), variation, and selection — in a unique
217 fashion relative to chromosomal inheritance, we asked whether these ecDNA features
218 enable more rapid tumour adaptation to stress than possible through chromosomal
219 inheritance (Fig. 4a). We utilized an isogenic cell line pair derived from a GBM patient² to
220 examine the importance of ecDNA in driving rapid adaptation. GBM39-EC is a patient-
221 derived neurosphere model with a mean copy number of approximately 100 copies of
222 *EGFRvIII*, a gain of function *EGFR* mutation residing on ecDNA^{3,4}. GBM39-HSR is an
223 isogenic model, in which all the *EGFRvIII* amplicons reside on chromosomal HSRs, at the
224 same mean copy number with the same DNA sequence (Extended Data Fig. 4a)⁴.
225 Importantly, the heterogeneity of *EGFRvIII* copy number in GBM39-EC correlates with
226 the heterogeneity of EGFRvIII protein expression assessed by flow cytometry (Extended
227 Data Fig. 4b,c). GBM39-EC cells are highly glycolytic². Therefore, we tested the

228 differential effect of glucose restriction on GBM39-EC and GBM39-HSR cells. We
229 withdrew 80% of normal glucose levels from the culture medium and saw a striking
230 difference — the GBM39-HSR cells were exquisitely sensitive to glucose withdrawal,
231 whereas the GBM39-EC showed no significant decrease in cell growth (Fig. 4b). This
232 ability of GBM39-EC cells to adapt to glucose restriction was mirrored by a rapid decrease
233 in the mean level and overall distribution of *EGFRvIII*-containing ecDNAs per cell (Fig.
234 4c). Remarkably, this genomic shift took place within a couple of cell cycles. In contrast,
235 the GBM39-HSR cells, which were highly sensitive to glucose restriction, were not
236 capable of rapidly changing their *EGFRvIII* copy number (Fig. 4c).

237

238 We had previously shown that GBM39-EC cells could become reversibly resistant to the
239 EGFR tyrosine kinase inhibitor (TKI) erlotinib, by lowering ecDNA copy number.
240 Therefore, we examined whether GBM39-EC cells would develop resistance to erlotinib
241 more rapidly than GBM39-HSR cells. Similar to glucose deprivation, GBM39-EC adapted
242 to the changing condition by altering its ecDNA copy number. After initially decreasing in
243 cell number, the GBM39-EC cells became resistant to erlotinib after just two weeks of
244 treatment, shifting their per cell ecDNA distribution in a reversible fashion (Fig 4d,e). In
245 contrast, the GBM39-HSR cells did not shift *EGFRvIII* chromosomal copy number and
246 remained highly sensitive to erlotinib (Fig. 4d,e). We then analyzed two samples taken
247 from GBM patient tumours, as previously described². We compared the primary tumour
248 resection (naïve) to the resected relapse which was treated with EGFR TKI lapatinib for
249 7-10 days prior to resection. We found a significant decrease in mean EGFR copy number
250 and in the ecDNA distribution in these patients' tumours (Fig. 4h). To extend our analysis
251 to other ecDNA-containing cancer types, we studied the effect of vincristine, a
252 chemotherapeutic that antagonizes *MYCN* amplification²⁵. In vitro, neuroblastoma cell
253 lines TR14 and CHP212 with *MYCN* amplified on ecDNA responded to vincristine by left-
254 shifting the ecDNA distribution, (Fig. 4f,g). When we compared treatment-naïve
255 neuroblastoma biopsies with primary tumour resections after treatment including
256 vincristine, we found a similarly significant decrease in the mean copy number and a left-
257 shift in the ecDNA distribution of *MYCN* in both of these patient tumours, in parallel with
258 the cell line data (Fig. 4i). Interestingly, when CHP212 was treated with the CDK4

259 inhibitors Abemaciclib, and to a greater extent Palbociclib, a right-shift in the distribution
260 of CDK4 ecDNA was detected in resistant tumour cells (Extended Data Fig. 4d,e,
261 Extended Data Fig. 5).

262

263 Together, these data indicate a clear pattern in which ecDNA enables high levels of
264 heterogeneity, which enable increased initial resistance to environmental or therapeutic
265 challenges. Further, the ongoing random inheritance of ecDNA-based oncogenes causes
266 rapid adaptation and the formation of resistance, through a mechanism which is
267 impossible in cells driven by chromosomal alterations.

268

269 ecDNA has emerged as a major challenge that forces reconsideration of our basic
270 understanding of cancer. Emerging data demonstrate that the altered topology of ecDNA
271 drives enhanced chromatin accessibility and rewires gene regulation to drive oncogenic
272 transcription⁴. Further, the unique higher-level organization of ecDNA particles into hubs²⁶
273 further contributes to ecDNA-mediated pathogenesis. The findings presented here reveal
274 that ecDNA uniquely shapes each of the foundational principles of Darwinian evolution –
275 random inheritance by descent, enhanced variation through random segregation, and
276 selection, thereby accelerating tumour cell evolution to maximize adaptation. Treating
277 such cancers may require targeting the unique adaptability of ecDNAs in the future.

278

279 **Acknowledgements**

280

281 P.S.M. was supported by a grant from The National Brain Tumour Society. Supported
282 by NIH R35-CA209919 (H.Y.C.). H.Y.C. is an Investigator of the Howard Hughes
283 Medical Institute. BW is supported by a Barts Charity Lectureship (grant MGU045). The
284 UCSD microscopy core is supported by NINDS NS047101.

285

286 **Author Contributions**

287 J.T.L., V. B, B.W., and P.S.M. conceived the project. J.T.L., C.Y.C, L.X., J.T., K.L.H.,
288 K.E.Y., Q.S., and M.L.E., performed experiments. Y.P., W.H., V.B., and B.W. performed
289 computational modelling. J.T.L., C.Y.C., Y.P., L.X., J.T., K.L.H., K.E.Y., Q.S., and U.R.
290 analysed data, guided by S.W., C.S., Z.L., W.H., H.Y.C., V.B., A.G.H., B.W., and P.S.M.
291 J.T.L., W.H., V.B., B.W., and P.S.M. wrote the manuscript with feedback from all
292 authors.

293

294 **Disclosure**

295

296 P.S.M. is co-founder of Boundless Bio, Inc. He has equity and chairs the Scientific
297 Advisory Board, for which he is compensated. V.B. is a co-founder, consultant, SAB
298 member and has an equity interest in Boundless Bio, Inc. and Digital Proteomics, LLC.
299 The terms of this arrangement have been reviewed and approved by UC San Diego in
300 accordance with its conflict-of-interest policies. H.Y.C. is a co-founder of Accent
301 Therapeutics, Boundless Bio, and advisor of 10x Genomics, Arsenal Biosciences, and
302 Spring Discovery.

303

304 **Materials and Correspondence**

305 Correspondence and material requests should be directed to Paul S. Mischel:
306 pmischel@stanford.edu.

307

308

309 **References**

- 310 1 Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and poor
311 outcome across multiple cancers. *Nat Genet* **52**, 891-897, doi:10.1038/s41588-020-0678-2
312 (2020).
- 313 2 Nathanson, D. A. *et al.* Targeted therapy resistance mediated by dynamic regulation of
314 extrachromosomal mutant EGFR DNA. *Science* **343**, 72-76, doi:10.1126/science.1241328 (2014).
- 315 3 Turner, K. M. *et al.* Extrachromosomal oncogene amplification drives tumour evolution and
316 genetic heterogeneity. *Nature* **543**, 122-125, doi:10.1038/nature21356 (2017).
- 317 4 Wu, S. *et al.* Circular ecDNA promotes accessible chromatin and high oncogene expression.
318 *Nature* **575**, 699-703, doi:10.1038/s41586-019-1763-5 (2019).
- 319 5 Merlo, L. M., Pepper, J. W., Reid, B. J. & Maley, C. C. Cancer as an evolutionary and ecological
320 process. *Nat Rev Cancer* **6**, 924-935, doi:10.1038/nrc2013 (2006).
- 321 6 Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306-313,
322 doi:10.1038/nature10762 (2012).
- 323 7 McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and
324 the Future. *Cell* **168**, 613-628, doi:10.1016/j.cell.2017.01.018 (2017).
- 325 8 McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in
326 cancer evolution. *Cancer Cell* **27**, 15-26, doi:10.1016/j.ccell.2014.12.001 (2015).
- 327 9 Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23-28,
328 doi:10.1126/science.959840 (1976).
- 329 10 Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* **376**,
330 2109-2121, doi:10.1056/NEJMoa1616288 (2017).
- 331 11 Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*
332 **545**, 446-451, doi:10.1038/nature22364 (2017).
- 333 12 Barthel, F. P. *et al.* Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* **576**,
334 112-120, doi:10.1038/s41586-019-1775-1 (2019).
- 335 13 Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic
336 heterogeneity in cancer evolution. *Nature* **501**, 338-345, doi:10.1038/nature12625 (2013).

- 337 14 Furnari, F. B., Cloughesy, T. F., Cavenee, W. K. & Mischel, P. S. Heterogeneity of epidermal
338 growth factor receptor signalling networks in glioblastoma. *Nat Rev Cancer* **15**, 302-310,
339 doi:10.1038/nrc3918 (2015).
- 340 15 Watkins, T. B. K. *et al.* Pervasive chromosomal instability and karyotype order in tumour
341 evolution. *Nature* **587**, 126-132, doi:10.1038/s41586-020-2698-6 (2020).
- 342 16 Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-
343 Cell Sequencing. *Cell* **173**, 879-893 e813, doi:10.1016/j.cell.2018.03.041 (2018).
- 344 17 Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature* **575**, 299-309,
345 doi:10.1038/s41586-019-1730-1 (2019).
- 346 18 Lundberg, G. *et al.* Binomial mitotic segregation of MYCN-carrying double minutes in
347 neuroblastoma illustrates the role of randomness in oncogene amplification. *PLoS One* **3**, e3099,
348 doi:10.1371/journal.pone.0003099 (2008).
- 349 19 Shoshani, O. *et al.* Chromothripsis drives the evolution of gene amplification in cancer. *Nature*
350 **591**, 137-141, doi:10.1038/s41586-020-03064-z (2021).
- 351 20 Fuller, B. G. *et al.* Midzone activation of aurora B in anaphase produces an intracellular
352 phosphorylation gradient. *Nature* **453**, 1132-1136, doi:10.1038/nature06923 (2008).
- 353 21 Tasan, I. *et al.* CRISPR/Cas9-mediated knock-in of an optimized TetO repeat for live cell imaging
354 of endogenous loci. *Nucleic Acids Res* **46**, e100, doi:10.1093/nar/gky501 (2018).
- 355 22 Grimm, J. B. *et al.* A general method to optimize and functionalize red-shifted rhodamine dyes.
356 *Nat Methods* **17**, 815-821, doi:10.1038/s41592-020-0909-6 (2020).
- 357 23 Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion
358 sequencing. *N Engl J Med* **366**, 883-892, doi:10.1056/NEJMoa1113205 (2012).
- 359 24 Qazi, M. A. *et al.* Intratumoral heterogeneity: pathways to treatment resistance and relapse in
360 human glioblastoma. *Ann Oncol* **28**, 1448-1456, doi:10.1093/annonc/mdx169 (2017).
- 361 25 Paffhausen, T., Schwab, M. & Westermann, F. Targeted MYCN expression affects cytotoxic
362 potential of chemotherapeutic drugs in neuroblastoma cells. *Cancer Lett* **250**, 17-24,
363 doi:10.1016/j.canlet.2006.09.010 (2007).
- 364 26 Hung, K. L. *et al.* EcDNA hubs drive cooperative intermolecular oncogene expression. *bioRxiv*,
365 2020.2011.2019.390278, doi:10.1101/2020.11.19.390278 (2020)
- 366
367

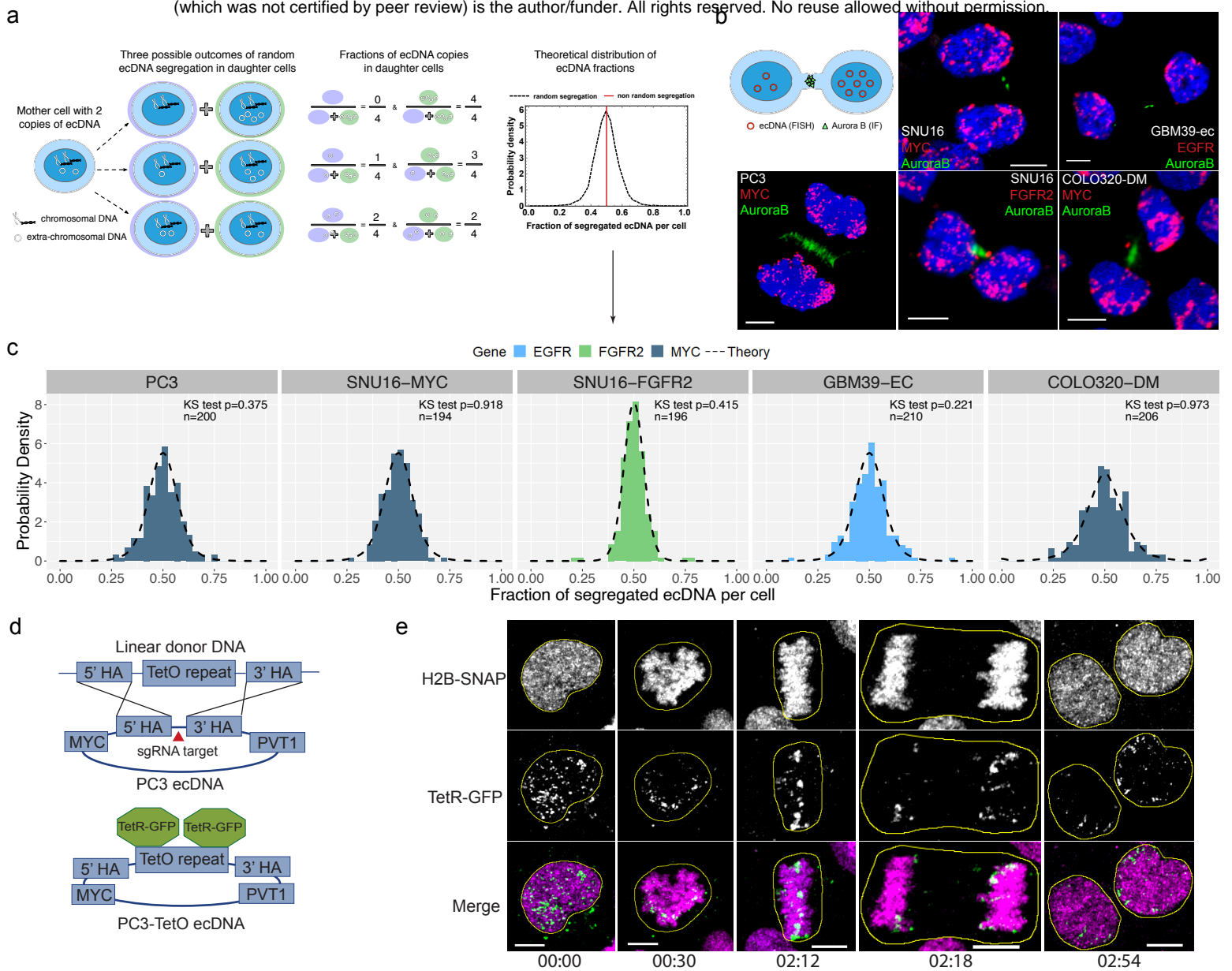


Figure 1. ecDNA is randomly segregated to daughter cells. **a**, Schematic of ecDNA segregation and predicted distribution of ecDNA fractions. **b**, Representative images of ecDNA distribution to daughter cells, identified by Aurora B midbody staining, in multiple cancer cell lines in late-stage mitosis. **c**, Distribution of ecDNA fractions in cancer cell lines analyzed in **b**, showing agreement between theoretical prediction (dashed lines) and observation (histograms) (KS test p values >0.05). **d**, Schematic of CRISPR-based genetic approach used for live-cell imaging of ecDNA in prostate cancer cells. **e**, Live-cell time lapse imaging reveals unequal distribution of ecDNA between daughter cells. Time stamps hh:mm. All scale bars 5µm.

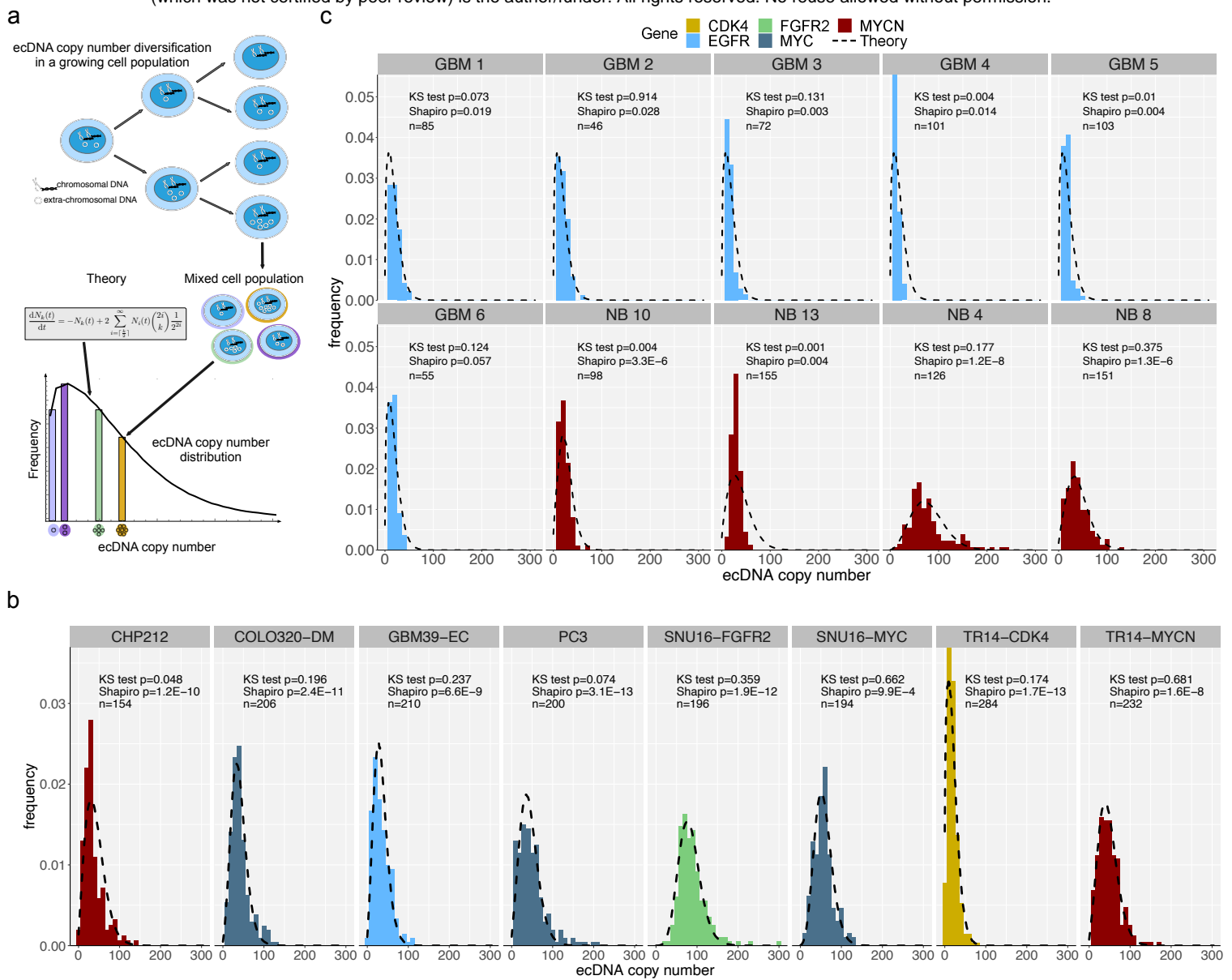


Figure 2. Random segregation of ecDNA promotes intratumoral heterogeneity of oncogenes in cancer cell lines and patient tumor samples. **a**, Schematic showing predicted impact of random ecDNA segregation on single cell oncogene copy number distribution. **b**, Distribution of ecDNA oncogene copy number, assessed by interphase FISH, in cancer cell lines. Agreement between theoretical prediction (dashed lines) and observation (histograms), reveals that oncogene copy number largely follows the prediction distribution (KS test $p > 0.05$). Shapiro-Wilk $p < 0.05$ suggests ecDNA number does not resemble a normal distribution. **c**, Distribution of ecDNA oncogene copy number, assessed by interphase FISH, in glioblastoma (GBM) and neuroblastoma (NB) patient tumor samples.

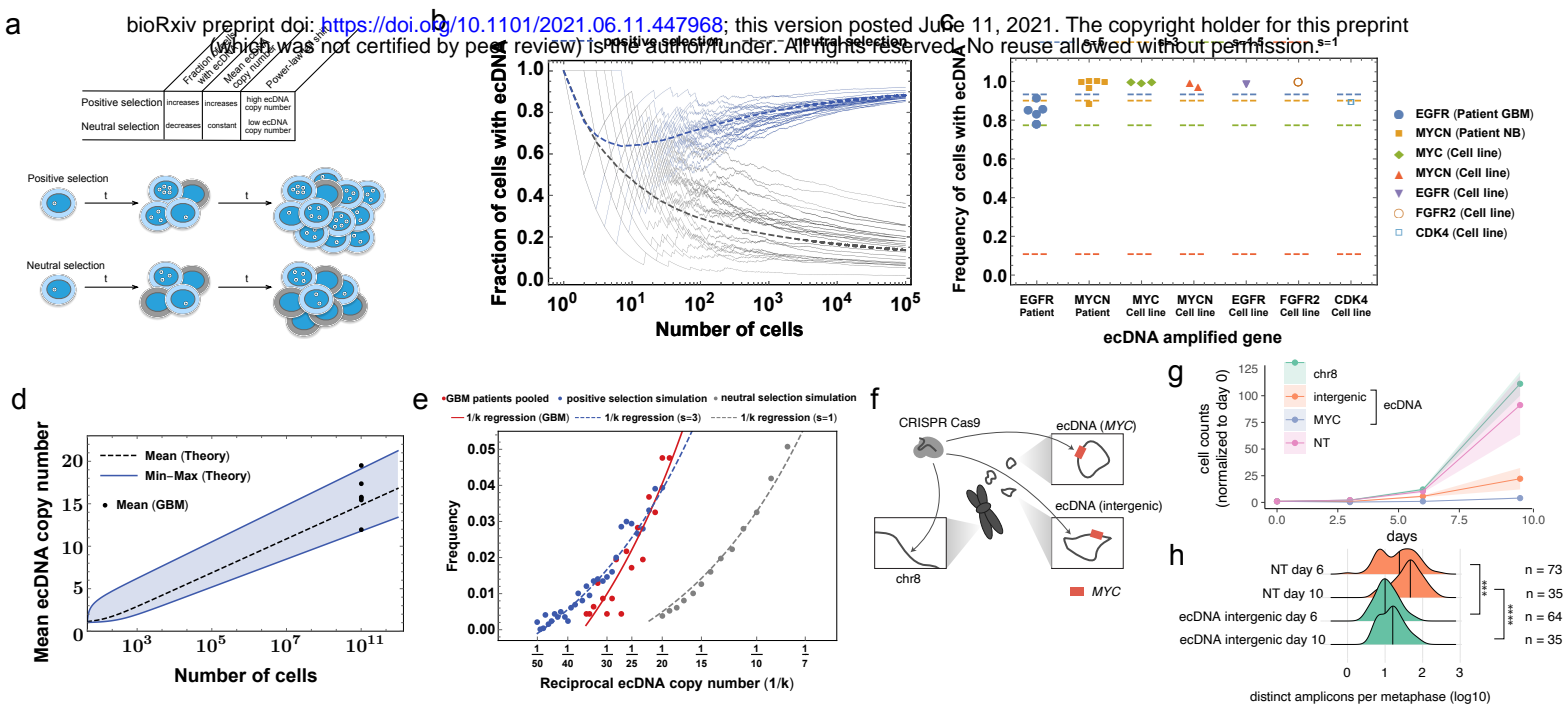


Figure 3. Strong selection for ecDNA in cancer. **a**, Schematic comparing predicted impact of ecDNA under positive or neutral selection. **b**, Simulations showing ecDNA prevalence in populations derived from a single ecDNA+ cell with ecDNA under positive or neutral selection. Positive selection $s=3$; neutral $s=1$. **c**, Empirical data comparing the frequency of ecDNA+ cells in cancer cell line and patient samples (colored shapes) to various selection strengths (dashed lines), shows strong evidence for positive ecDNA selection across oncogenes and cancer types. **d**, Comparison between the minimum, maximum and mean copy number predicted by strong ecDNA selection pressure (blue cone) and the ecDNA copy number reached in modeled GBM tumors (blue cone and lines) and the observed mean ecDNA copy number detected in 6 GBM patient samples (dots). **e**, Power law tail shift of ecDNA copy number in GBM patients indicative of strong positive selection. **f**, Depiction of CRISPR-based strategy to test selective advantage given to COLO320DM cells by MYC ecDNA. Arrows indicate regions targeted by sgRNA. **g**, Genome editing of MYC encoded on ecDNA causes massive decrease in cell number that exceeds the impact of intergenic editing, indicative of strong selection for oncogenes on ecDNA. **h**, Quantification of ecDNA numbers per metaphase at 6 and 10 days post CRISPR transfection. P values calculated using Mann-Whitney tests. *** $p \leq 0.0005$; **** $p \leq 0.00005$.

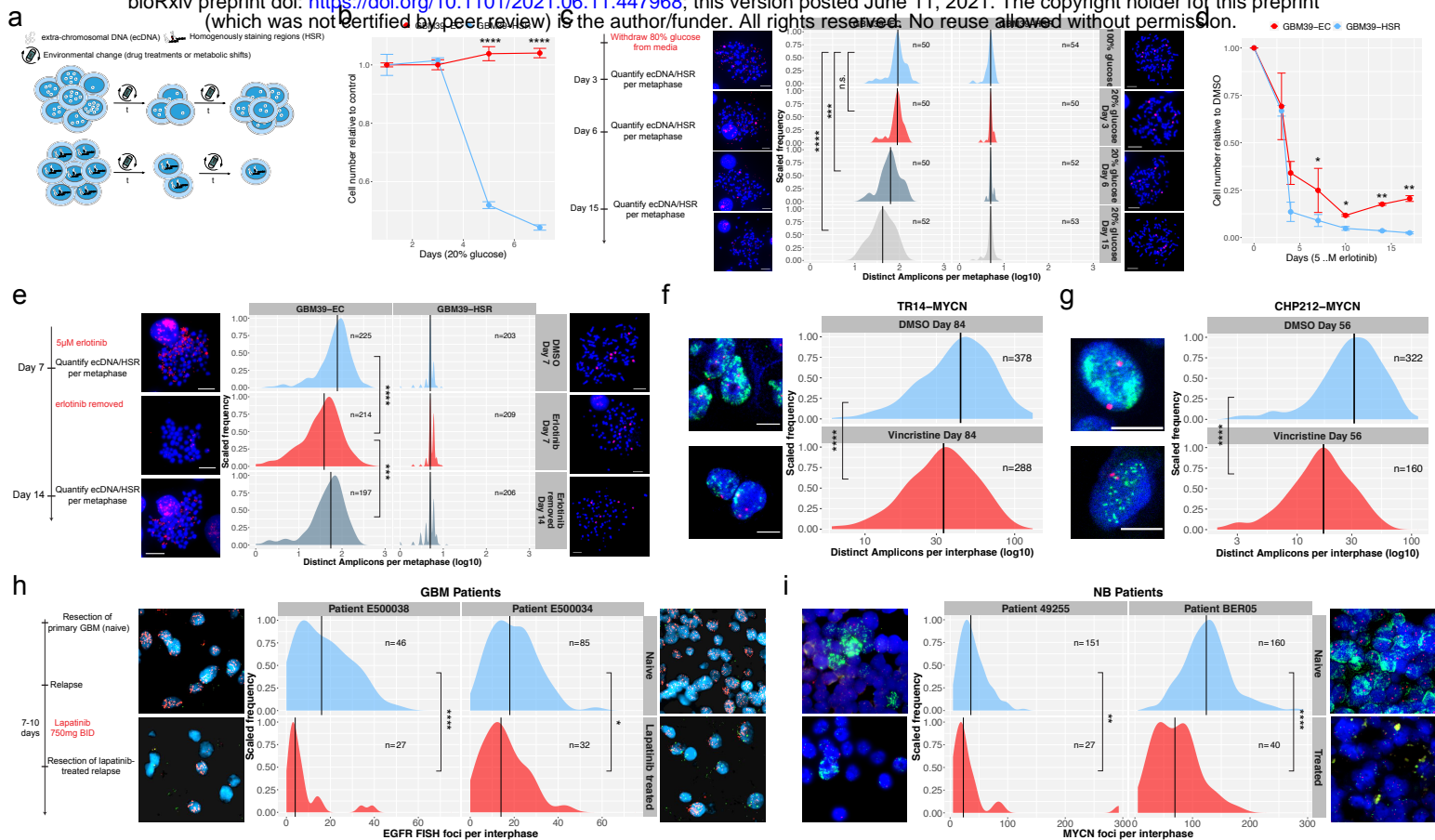
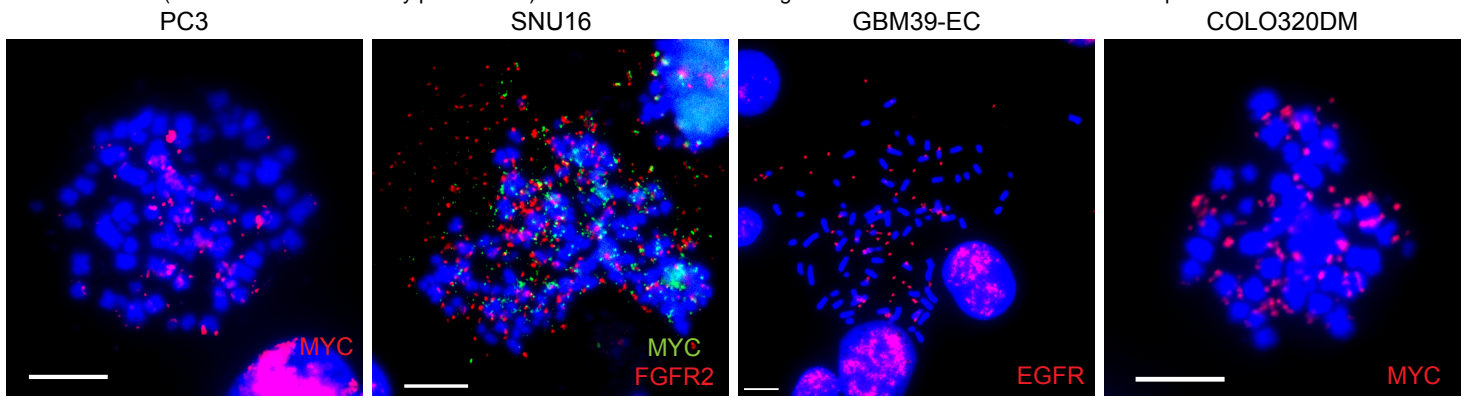
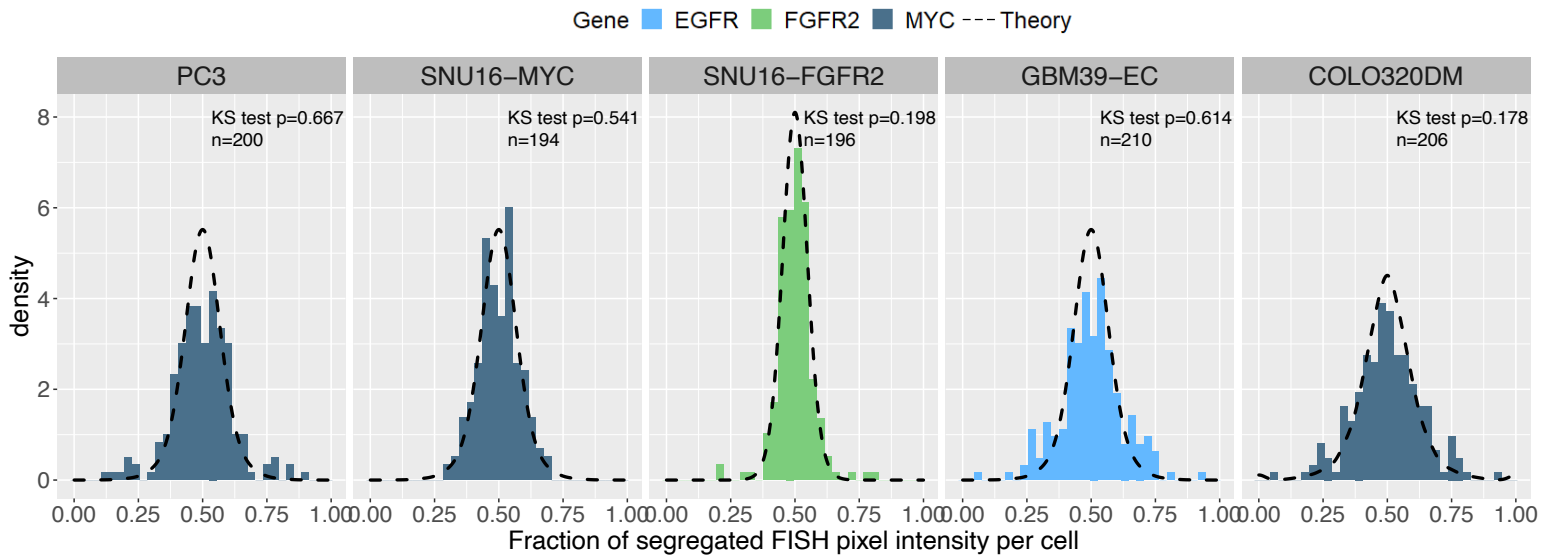


Figure 4. Non-chromosomal inheritance of ecDNA promotes rapid adaptation and resistance to glucose withdrawal and targeted drug treatment. **a**, Schematic depicting how the random segregation of ecDNA and ensuing heterogeneity can drive rapid adaptation and resistance. **b**, ecDNA-containing GBM cells are relatively resistant to glucose withdrawal, whereas GBM cells in which the same oncogene has lodged onto chromosomal loci at near identical copy number (GBM39-HSRs) cannot tolerate glucose withdrawal. Significance determined by two-sided t-tests. Error bars indicate standard deviation. Each data point indicates mean of 3 biological replicates. **c**, Adaptation of ecDNA containing cells to glucose withdrawal is linked to a rapid shift in the distribution of amplicons per cell, unlike the highly sensitive HSR containing cells that cannot modulate amplicon copy number. Timeline of experiment is depicted at the left of the panel. Red FISH signal is from EGFR FISH probe. **d**, GBM cells with EGFRvIII amplified on ecDNA, after an initial response, rapidly become resistant to the EGFR tyrosine kinase inhibitor erlotinib, whereas the GBM39-HSR cells remain highly sensitive. Significance determined by two-sided t-tests. Error bars indicate standard deviation. Each data point indicates mean of 2 biological replicates (4 for day 7). **e**, GBM cells with EGFRvIII amplified on ecDNA rapidly shift the distribution of EGFRvIII amplicons per cell, measured at 7 days, which can also be rapidly reversed within one week by drug withdrawal. Timeline of experiment is depicted at the left on the panel. Red signal is EGFR FISH probe. **f**, NB cell line TR14 shift the copy number distribution of MYCN ecDNA when treated with 43nM vincristine for 12 weeks. Green signal is MYCN FISH probe. **g**, NB cell line CHP-212 shift the copy number distribution of MYCN ecDNA when treated with 5.3nM vincristine for 8 weeks. Green signal is MYCN FISH probe. **h**, Comparison of the distribution of EGFR amplification per cell in two GBM patients before therapy (naive) and after 7-10 days of lapatinib treatment. Red FISH signal is from EGFR FISH probe. Green FISH signal is from Chr. 7 control probe. **i**, Comparison of MYCN ecDNA copy numbers assessed by MYCN (green) FISH in two NB patients before and after receiving chemotherapy including Vincristine. Red signal from Chr. 2 control FISH probe. Scale bars represent 5µm in all images. P values calculated using Mann-Whitney tests except where indicated. N.s. not significant; * p ≤ 0.05; ** p ≤ 0.005; *** p ≤ 0.0005; **** p ≤ 0.0005.

a

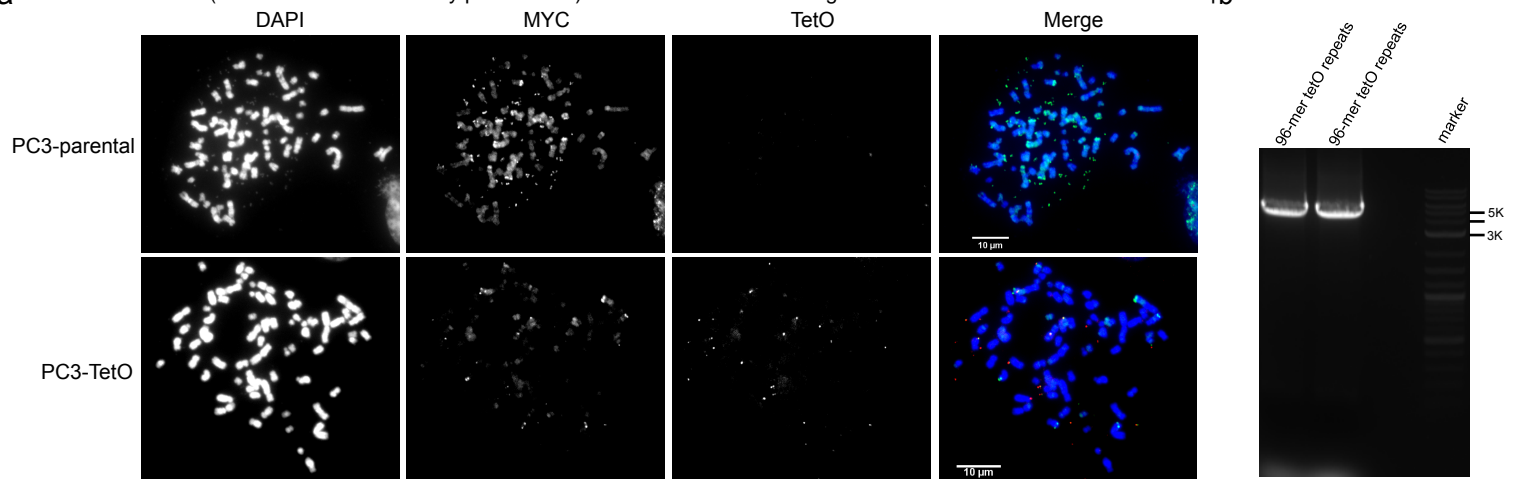


b

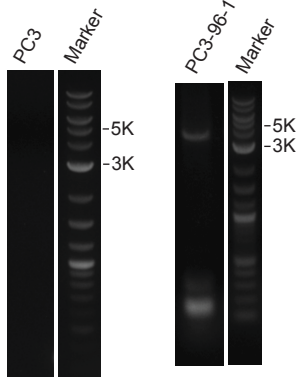


Extended Data Fig. 1 | Quantification of ecDNA pixel intensity shows uneven random segregation to daughter cells. **a**, Representative metaphase FISH images for cell lines used to quantify segregation dynamics in **Fig. 1**. **b**, The same daughter cells analyzed in **Fig. 1c** were analyzed by quantifying the pixel intensity of FISH signal in each daughter cell, as a proxy for ecDNA number. Analysis was unbiased and useful for cases in which ecDNA were packed together making counting distinct foci difficult. Agreement between theoretical predictions (dashed lines) and observation (histograms) shown by KS test p value > 0.05. Scale bars 10 μ m.

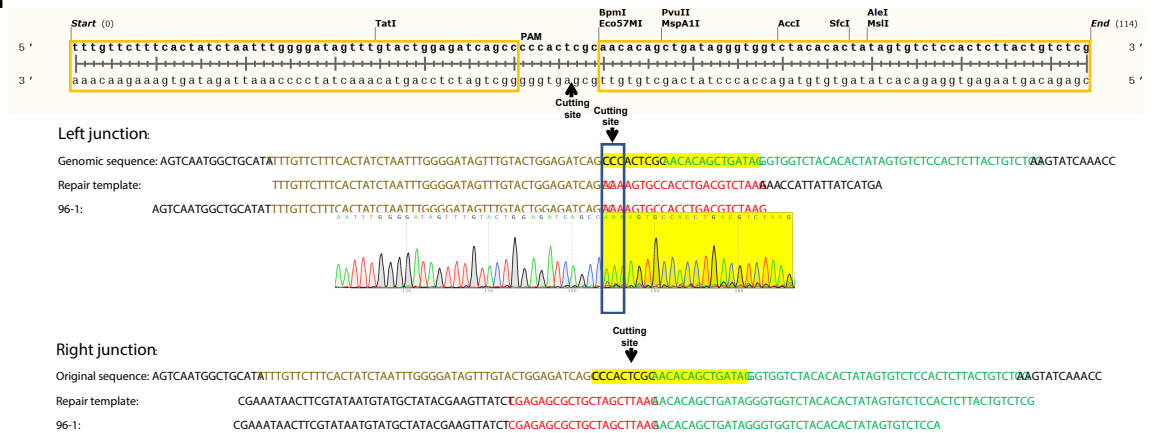
a



c

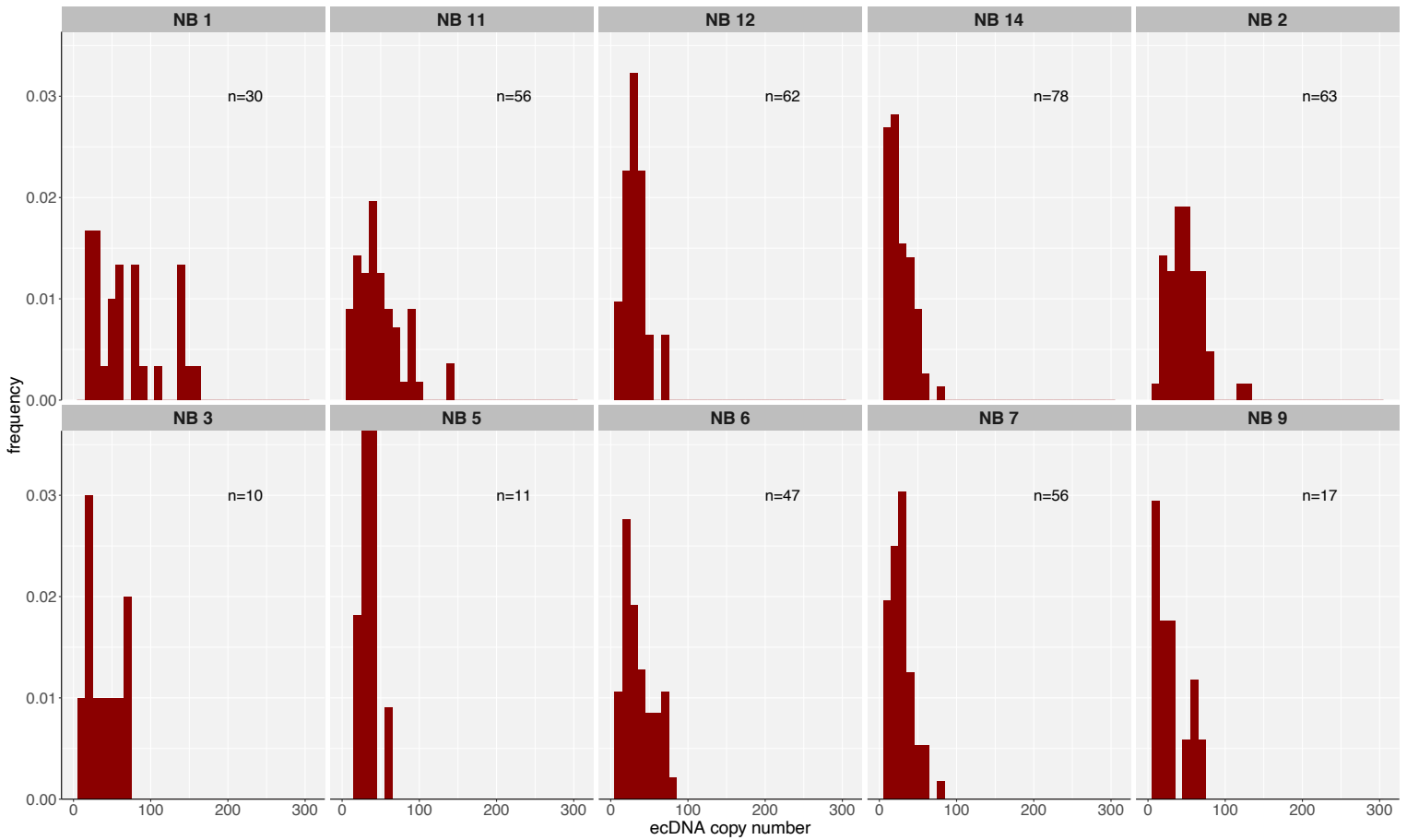


d

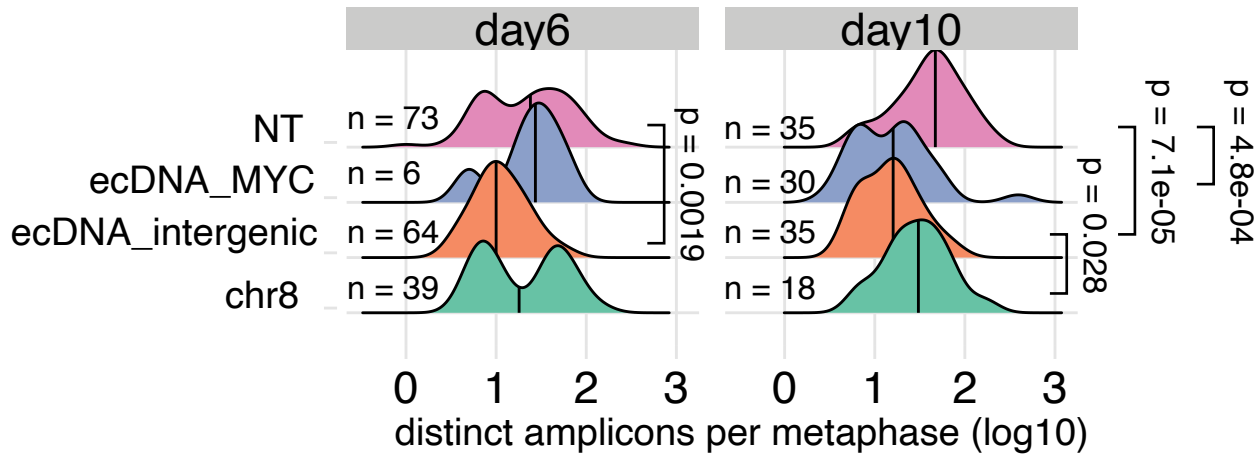


Extended Data Fig. 2 | Live cell tracking of ecDNA through insertion of Tet-O array into the ecDNA of PC3 cells. a, Representative images of PC3 parental and PC3-TetO cell lines showing extensive MYC amplification on both. PC3-TetO shows significant TetO FISH signal on multiple ecDNA bodies as well. **b,** PCR amplification of 96-mer TetO repeats from DNA isolated from PC3-TetO cells confirming insertion. **c,** PCR amplification of 96-mer TetO repeats from DNA isolated from PC3-TetO cells confirming insertion. **d,** Sanger sequencing of PCR amplification product from PC3-TetO cells. Both left and right junctions were repaired by homologous recombination at the insertion site.

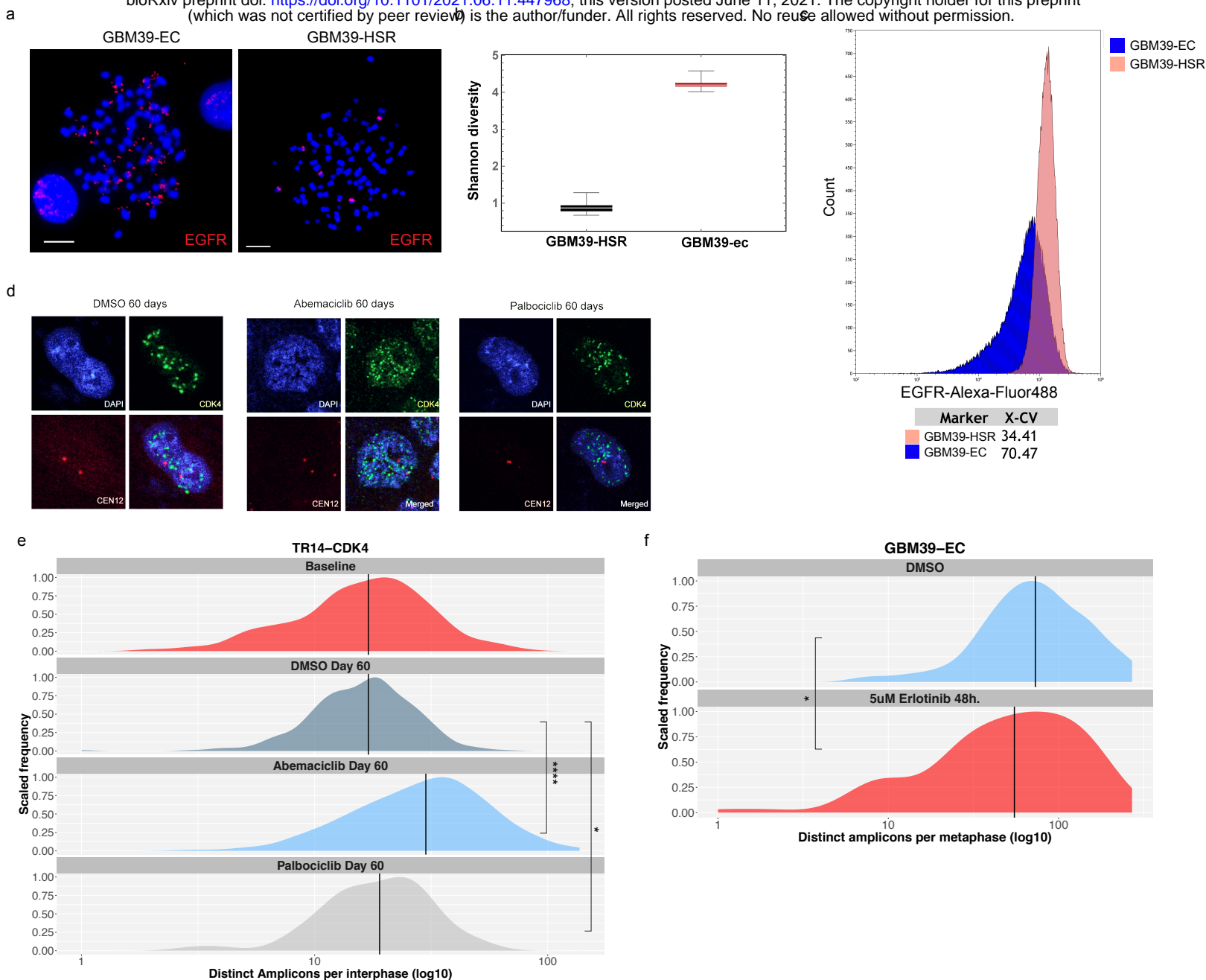
a



b



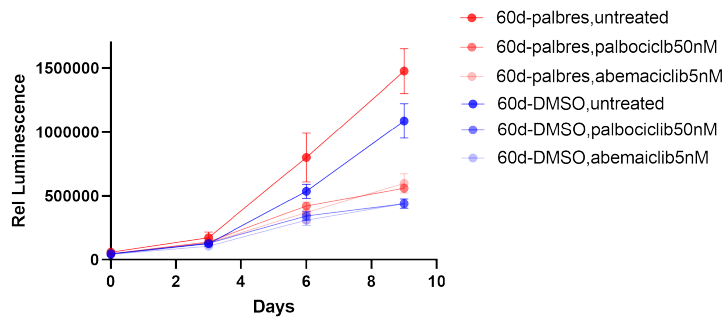
Extended Data Fig. 3 | ecDNA heterogeneity and selection. a, Histograms of ecDNA copy number assessed by interphase FISH on patient tumor tissue from neuroblastoma (NB) patients. b, Quantification of ecDNA numbers at Day 6 and Day 10 after CRISPR cutting of regions of the COLO320-DM genome, either on or off of ecDNA. Shows clear evidence for selection of ecDNA both by the severe drop in copy number when targeted and the indication that the copy number begins to return to initial levels. Note ecDNA_MYC at day 6 is severely limited in its growth and only 6 metaphases were able to be identified and imaged.



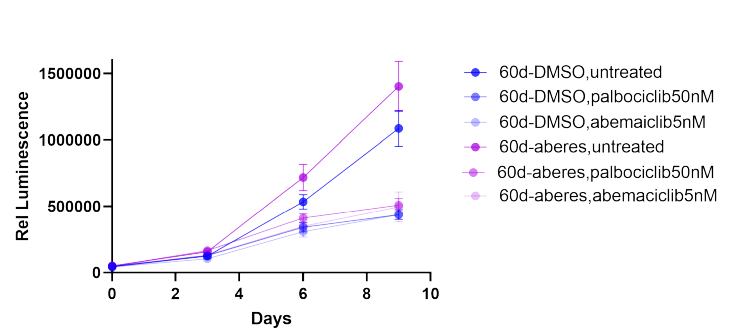
Extended Data Fig. 4 | ecDNA dynamically responds to therapeutics. **a**, Representative images of metaphase spread FISH from isogenic GBM39 cell line. **b**, Quantification of the Shannon diversity index between isogenic GBM39-HSR and GBM39-EC cell lines based on counts of ecDNA amplicons per cell. **c**, Flow cytometry analysis of EGFR protein expression in isogenic GBM39-EC and GBM39-HSR cell lines shows a pattern of heterogeneity similar to that seen in copy number. X-CV quantifies the % coefficient of variation for the two samples. **d**, Representative images of TR14 cells treated with Abemaciclib or Palbociclib for 60 days. CDK4 FISH signal is shown in green, CEN12 control FISH probe is shown in red. **e**, Quantification of the experiment described in **c** shows a significant shift in CDK4 ecDNA copy number distribution under both drug conditions. **f**, Quantification of EGFR ecDNA in GBM39-EC cells after short-term treatment with erlotinib shows a rapid change in ecDNA copy number distribution. Lines indicate medians. P values calculated using Mann-Whitney tests. * $p \leq 0.05$; **** $p \leq 0.0001$. Scale bars 10 μm .

a

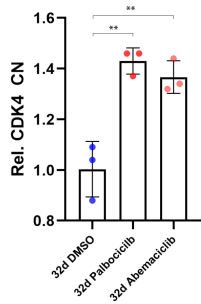
Cell growth (CellTiterGlo)
60 day palbociclib-res vs. 60 day
DMSO-treated



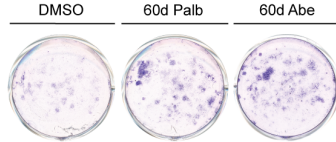
Cell growth (CellTiterGlo)
60 day abemaciclib-res vs.
60 day DMSO-treated



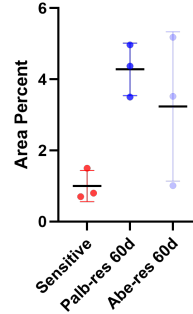
c



d



e



Extended Data Fig. 5 | ecDNA dynamics correlate with formation of resistance. **a**, Treatment of long term palbociclib resistant populations of TR14 cells with palbociclib or abemaciclib, showing resistance to treatment. **b**, Treatment of long term abemaciclib resistant populations of TR14 cells with palbociclib or abemaciclib showing resistance to treatment. **c**, Validation of increased ecDNA copy number by qPCR for CDK4. **d**, Crystal violet staining of TR14 cells re-challenged with palbociclib or abemaciclib after development of resistance, or not (DMSO). **e**, Quantification of **d** showing resistance in populations treated with CDK4 inhibitors for 60 days.

Supplemental Movie 1. Live-cell imaging of PC3-TetO cells. Live-cell imaging of PC3-TetO cells with chromatin labelled by H2B-SNAP (purple) and ecDNA labelled in green (GFP).

368 **METHODS**

369 **Cell culture**

370 Cell lines were purchased from ATCC or DSMZ-German Collection of Microorganisms
371 and Cell Cultures (Leibniz Institute) or were a kind gift from J.H. Schulte. GBM39-HSR
372 and GBM39-EC were derived from a patient GBM as previously described (Nathanson
373 cite).

374
375 PC3 cells were cultured in DMEM with 10% fetal bovine serum (FBS). COLO320-HSR
376 and COLO320-DM were cultured in DMEM/F12 50%:50% with 10% FBS. SNU16 were
377 grown in RPMI-1640 with 10% FBS. GBM39-HSR and GBM39-EC neurospheres were
378 grown in DMEM/F12 with B27, Glutamax, Heparin (5 μ g/ml), EGF (20ng/ml), and FGF
379 (20ng/ml). TR-14 cells were grown in RPMI-1640 with 20% FCS. TR-14 cells were
380 cultured in RPMI-1640 with 20% FCS. Cell numbers were counted with a TC20
381 automated cell counter (Bio-Rad). For drug treatments, drug was replaced every 3-4
382 days.

383

384 **Metaphase chromosome spreads**

385 Cells were concentrated in metaphase by treatment with KaryoMAX colcemid (Gibco) at
386 100ng/ml for between 3 hours and overnight (depending on cell cycle speed). Cells were
387 washed once with PBS and a single cell suspension was incubated in 75mM KCl for 15
388 minutes at 37°C. Cells were then fixed with Carnoy's fixative (3:1 methanol:glacial acetic
389 acid) and spun down. Cells were washed with fixative 3 additional times. Cells were then
390 dropped onto humidified glass slides.

391

392 **Fluorescence *in situ* hybridization (FISH)**

393 Fixed samples on coverslips or slides were equilibrated briefly in 2x SSC buffer. They
394 were then dehydrated in ascending ethanol concentrations of 70%, 85%, and 100% for
395 approximately 2 minutes each. FISH probes were diluted in hybridization buffer (Empire
396 Genomics) and added to the sample with addition of a coverslip or slide. Samples were
397 denatured at 72°C for 2 minutes and then hybridized at 37°C overnight in a humid and
398 dark chamber. Samples were then washed with 0.4x SSC then 2x SSC 0.1% Tween-20
399 (all washes approximately 2 minutes). DAPI (100ng/ml) was applied to samples for 10
400 minutes. Samples were then washed again with 2x SSC 0.1% Tween-20 then 2x SSC.
401 Samples were briefly washed in ddH₂O and mounted with Prolong Gold. Slides were
402 sealed with nail polish.

403

404 **Dual immunofluorescence – fluorescence *in situ* hybridization (IF-FISH)**

405 Asynchronous cells were grown on poly-l-lysine coated coverslips (laminin, for GBM39-
406 EC). Cells were washed once with PBS and fixed with cold 4% paraformaldehyde (PFA)
407 at room temperature for 10-15 minutes. Samples were permeabilized with 0.5% Triton-X
408 in PBS for 10 minutes at room temperature and then washed with PBS. Samples were
409 then blocked with 3% BSA in PBS-0.05% Triton-X for 30 minutes at room temperature.
410 Samples were incubated in primary antibody, diluted in blocking buffer, for either 1 hour
411 at room temperature or overnight at 4°C. Samples were washed thrice in PBS-0.05%
412 Triton-X. Samples were incubated in secondary antibody, diluted in blocking buffer, for 1
413 hour at room temperature (all subsequent steps in the dark) and then washed thrice in

414 PBS-0.05% Triton-X. Cells were washed once with PBS and re-fixed with cold 4% PFA
415 for 20 minutes at room temperature. Cells were washed once with PBS then once with 2x
416 SSC buffer. FISH proceeded as described above with the following difference:
417 denaturation was performed at 80°C for 20 minutes.

418

419 **Microscopy**

420 Conventional fluorescence microscopy was performed using an Olympus BX43
421 microscope; images were acquired with a QI-Click cooled camera. Confocal microscopy
422 was performed using a Leica SP8 microscope with lightning deconvolution (UCSD School
423 of Medicine Microscopy Core). Neuroblastoma cell lines were imaged with a Leica TCS
424 SP5 microscope, HCC PL APO lambda blue 63x 1.4 oil lens.

425

426 **Neuroblastoma patient tissue FISH**

427 FISH analysis was performed on 4 µm sections of FFPE blocks. Slides were
428 deparaffinized, dehydrated and incubated in pre-treatment solution (Dako, Denmark) for
429 10 min at 95–99°C. Samples were treated with pepsin solution for 2 min at 37°C. For
430 hybridization, the *ZytoLight*® SPEC MYCN/2q11 Dual Color Probe (ZytoVision,
431 Bremerhaven, Germany) was used. Incubation took place overnight at 37°C, followed by
432 counterstaining with 4,6-diamidino-2-phenylindole (DAPI). For each case, signals were
433 counted in 50 non-overlapping tumour cells using a fluorescence microscope (BX63
434 Automated Fluorescence Microscope, Olympus Corporation, Tokyo, Japan). Computer-
435 based documentation and image analysis was performed with the SoloWeb imaging
436 system (BioView Ltd, Israel) MYCN amplification (MYCN FISH+) was defined as
437 MYCN/2q11.2 ratio > 4.0, as described in the INRG report²⁷.

438

439 **Quantification of FISH foci**

440 Quantification of FISH foci was performed using the ImageJ-Find maxima function in a
441 supervised fashion. For quantification of pixel intensity, the ImageJ-Pixel intensity function
442 was used. These two GBM patient tissue FISH images were obtained as part of a phase
443 II lapatinib GBM clinical trial described previously. In brief, patients were administered
444 750 mg of lapatinib orally twice a day (BID) for 7 to 10 days (depending on whether
445 treatment interval fell over a weekend) before surgery, the time to steady state. Blood and
446 tissue samples were obtained at the time of resection².

447

448 **Construction of PC3-TetO cell line**

449 The insertion of tetO repeats was conducted through CRISPR/cas9 mediated
450 approaches. And the plasmids: pSP2-96-mer TetO-EFS-BlaR and F9-TetR-EGFP-IRES-
451 PuroR used in this section were kind gifts from Dr. Huimin Zhao²¹. Briefly, the intergenic
452 region between MYC and PVT1 was selected as the insertion region on the basis that it
453 is amplified in PC3 cells on ecDNA with high frequency. DNA sequences were retrieved
454 from UCSC Genome Browser, repetitive and low complexity DNA sequences were
455 annotated and masked by RepeatMasker in the UCSC Genome Browser. The guide
456 sequences of sgRNAs were designed by CRISPRdirect web tool²⁸, and their amplification
457 was confirmed with WGS data. The guide sequence selected was constructed into
458 pSpCas9(BB)-2A-Puro (PX459) [pSpCas9(BB)-2A-Puro(PX459) was a gift from Feng

459 Zhang (Addgene plasmid #62988; <http://n2t.net/addgene:62988>; RRID:
 460 Addgene_62988)]. Repair donor was obtained through PCR amplification, using pSP2-
 461 96-merTetO-EFS-BlaR plasmid as template, as well as primers containing the 50nt
 462 homology arm upstream and downstream of the predicted cutting site.

463
 464 The transfection of CRISPR/Cas9 plasmid and 96-mer TetO EGFP-BlastR donor into
 465 PC3 cells was conducted with X-tremeGENE HP transfection reagent according to
 466 manufactory instruction with CRISPR/Cas9 plasmid only or 96-mer TetO EGFP-BlastR
 467 only using as negative control. 2 days after transfection, Blastidicin was added to the
 468 culture medium for 3 days, at a time point that the majority of the cells in the negative
 469 control groups have died while more cells survived in the group with transfection of
 470 CRISPR/Cas9 plasmid and donor. The surviving cells were subjected to limited dilution
 471 in 96-well plate, with Blastidicin being added all the time. Surviving clones were expanded
 472 and their genomic DNA were extracted and subjected to genotyping with a pair of primer
 473 flanking the inserted region. PCR product of genotyping results were subjected to sanger
 474 sequencing to confirm the insertion at predicted cutting site. Clones with positive
 475 genotyping band will be expanded and metaphase cells were collected. Double FISH with
 476 FISH probe against Tet operator and against MYC FISH probe was performed on
 477 metaphase spread. PC3 cells with TetO repeats were infected with lentivirus containing
 478 the F9-TetR-EGFP-IRES-PuroR, and 2 days after infection puromycin was added into
 479 culture medium to establish a stable cell line that is able to image ecDNA with the aid of
 480 EGFP visualization.
 481

Primer Name	Sequence
crispr-MYC-P-4-F	CACCGCTATCAGCTGTGTTGCGAGT
crispr-MYC-P-4-R	AAACACTCGCAACACAGCTGATAGC
donor-4-for	T*T*T*GTTCTTTCACTATCTAATTTGGGGATAGTTTGT ACTGGAGATCAGCCAAAAGTGCCACCTGACGTCTA AG
donor-4-rev	C*A*GTAAGAGTGGAGACACTATAGTGTGTAGACCA CCCTATCAGCTGTGTTCTTAAGCTAGCAGCGCTCTC G
genotyping In-Forward	CACGAGGCCCTTTCGTCTTC
genotyping 4-rev	CGAGACAGTAAGAGTGGAGACAC
1 st primer for tetO-pBEST	CACAGGAAACAGCTATGACCatgcatDDDDDDDDDDT CCCTATCAGTGATAGAGADDDDDDDDDTCCCTATC AGTGATAGA
2 nd primer for tetO-pBEST	GADDDDDDDDDTCCCTATCAGTGATAGAGADDDDD DDDDDDctgcagTAGGATGAAGctgcagGTTGTAAAACG ACGGCCAGT

482

483

484 Live cell imaging of ecDNA

485 PC3 TetO TetR-GFP cell line was transfected with PiggyBac vector expressing H2B-
 486 SNAPf and the super PiggyBac transposase (2:1 ratio) as previously described²⁹. Stable
 487 transfectants were selected by 500µg/ml G418 and sorted by flow cytometry. To facilitate

488 long-term time lapse imaging, 10 μ g/ml human fibronectin was coated in each well of 8-
489 well lab-tek chambered cover glass. Prior to imaging, cells were stained with 25nM SNAP
490 tag ligand JF₆₆₉²² at 37°C for 30 minutes followed by 3 washed with regular medium for
491 30 minutes total. Cells were then transferred to an imaging buffer containing 20% serum
492 in 1x Opti-Klear live cell imaging buffer at 37°C. Cells were imaged on a Zeiss LSM880
493 microscope pre-stabilized at 37°C for 2 hours. We illuminated the sample with 1.5%
494 488nm laser and 0.75% 633nm laser with the EC Plan-Neofluar 40x/1.30 oil lens, beam
495 splitter MBS 488/561/633 and filters BP 495-550 + LP 570. Z-stacks were acquired with
496 0.3 μ m z step size with 4-minute intervals between each volumetric imaging for a total of
497 16 hours.

498

499 **Colony formation assay**

500 TR-14 cells were taken from 60 days of treatment with either DMSO, 50 nM Palbociclib,
501 or 5 nM Abemaciclib, and seeded into a poly-D-lysine coated 24-well plate at 20,000 cells
502 per well. After 24 h, the cells from each condition were treated with either DMSO, 50 nM
503 Palbociclib, or 5 nM Abemaciclib over 20 days, in triplicate. At 20 days, crystal violet
504 staining procedure was performed. Briefly, cell culture media was aspirated, cells were
505 washed gently with PBS, fixed in 4% paraformaldehyde in PBS for 20 min, stained with 2
506 mL of crystal violet solution (50 mg in 50 mL 10% ethanol in MilliQ water), washed 1x with
507 PBS, and dried for 30 min. The area intensity was calculated using the ColonyArea plugin
508 in ImageJ³⁰.

509

510 **CellTiter-Glo**

511 TR-14 cells were taken from 60 days of treatment with either DMSO, 50 nM Palbociclib,
512 or 5 nM Abemaciclib and seeded into white flat-bottom 96 WPs (Corning) in 100 μ l media
513 at a density of 500 cells/well. After 24 h, the cells were treated with either vehicle, 50 nM
514 Palbociclib, or 5 nM Abemaciclib (50 μ L of drug solution/well). Cell viability was
515 determined using CellTiter-Glo Luminescent Cell Viability Assay (Promega) at 3, 6, and
516 9 days after drug was added, following the manufacturer's protocol.

517

518 **Immunoblotting**

519 Whole-cell protein lysates were prepared by lysing cells in Silly lysis buffer. Protein
520 concentrations were determined by bicinchoninic acid assay (BCA, Thermo Fisher). 10
521 μ g of protein were denatured in Laemmli buffer at 95 °C for 5 minutes and 1mM DTT was
522 added. Lysates were loaded onto 10% Tris-Glycin (Thermo Fisher) for gel
523 electrophoresis. Proteins were transferred onto Immobilon-FL PVDF membranes (Sigma
524 Aldrich), blocked Odyssey Blocking Buffer in TBS for 1 hour and incubated with primary
525 antibodies overnight at 4°C, then secondary antibodies for 1 hour at room temperature.
526 Fluorescent signal was detected using the Odyssey CLx imaging system. Quantification
527 was performed with LI-COR Image Studio Software.

528

529 **Flow cytometry**

530 Single cell suspensions were made and passed through a cell filter to ensure single cell
531 suspension. Cells were suspended in flow cytometry buffer (HBSS buffer without calcium
532 and magnesium, 1x Glutamax, 0.5% (v/v) FBS, 10mM HEPES). EGFRvIII mab 806³¹ was
533 added at 1ug per million cells and incubated on ice for one hour. Cells were washed in

534 flow cytometry buffer and resuspended in buffer with anti-mouse alexa-488 antibody
535 (1:1000) for 45 minutes on ice in the dark. Cells were washed again with flow cytometry
536 buffer and resuspended in flow cytometry buffer at approximately 4 million cells per
537 milliliter. Cells were sorted using a Sony SH800 FACS sorter and was calibrated and
538 gating was informed using a secondary only negative control.

539

540 **Quantitative PCR (qPCR)**

541 DNA extraction was performed using the NucleoSpin Tissue kit (Macherey-Nagel),
542 following the manufacturer's protocol. qPCR was performed using 50 ng or 1.5 μ l of
543 template DNA and 0.5 μ M primers with SYBR Green PCR Master Mix (Thermo Fisher
544 Scientific) in FrameStar 96-well PCR plates (4titude). Reactions were run and monitored
545 on a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) and Ct values were
546 calculated with the StepOne Plus software v.2.3 (Thermo Fisher Scientific).

547 CDK4 Fwd: AAAGTTACCACACACCC

548 CDK4 Rev: AGTGCTAAGAAAGCGGCACT

549

550 **Quantification of single cell ecDNA segregation patterns**

551 We generate the theoretically expected distribution of ecDNA copy number fractions after
552 a single cell division under different models of ecDNA segregation by stochastic computer
553 simulations implemented in C++. Briefly a single cell is initiated with a random number of
554 ecDNA copies n , drawn from a uniform distribution $U(20,200)$. EcDNA is amplified and
555 $2n$ ecDNA copies are segregated between two daughter cells following a Binomial trial
556 $B(2n, p)$, with segregation probability p . Here, $p = 1/2$ corresponds to random
557 segregation and $p > 1/2$ to a biased random segregation. This results in two daughter
558 cells with ecDNA copy number $n_1 \sim B(2n, p)$ and $n_2 = n - n_1$. The fraction of segregated
559 ecDNA f is then calculated via

560

$$561 \quad f_1 = \frac{n_1}{n_1+n_2} \quad \text{and} \quad f_2 = \frac{n_2}{n_1+n_2}.$$

562

563 Iterating the process 10^7 times generates the expected distribution of f as shown in Figure
564 1c. Similarly, we can generate an expected distribution of f for chromosomal patterns of
565 inheritance. For perfect chromosomal segregation, we have $f_1 = f_2 = 1/2$. To allow for
566 mis-segregation we introduce a probability $u = 0.05$ such that $n_1 = n \pm 1$ and $n_2 = n -$
567 n_1 . We use Kolmogorov-Smirnov statistics to compare the theoretically expected and
568 experimentally observed distributions of ecDNA copy number fractions under these
569 different scenarios.

570

571 **Stochastic simulations of ecDNA population dynamics**

572 We implemented individual based stochastic computer simulations of the ecDNA
573 population dynamics in C++. For each cell, the exact number of ecDNA copies is recorded
574 through the simulation. Cells are chosen randomly but proportional to fitness for
575 proliferation using a Gillespie algorithm. The simulation is initiated with one cell carrying
576 n_0 copies of ecDNA. The proliferation rate of cells without ecDNA is set to $r^- = 1$ (time is
577 measured in generations). A fitness effect for cells with ecDNA then corresponds to a
578 proliferation rate $r^+ = s$. Here, $s > 1$ models a fitness advantage, $0 < s < 1$ a fitness
579 disadvantage and $s = 1$ corresponds to no fitness difference (neutral dynamics, $r^+ = r^-$).

580 During proliferation, the number of ecDNA copies in that cell are doubled and randomly
581 distributed into both daughter cells according to a Binomial trial $B(n, p)$ with success rate
582 $p = 1/2$. If a cell carries no ecDNA, no daughter cell inherits ecDNA. We terminate
583 simulations at a specified cell population size. We output the copy number of ecDNA for
584 each cell at the end of each simulation, which allows us to construct other quantities of
585 interest, such as the ecDNA copy number distribution, the time dynamics of moments,
586 the power law scaling of tails or the Shannon diversity index. We use Kolmogorov-
587 Smirnov statistics to test similarity between simulated and experimental ecDNA copy
588 number distributions and Shapiro-Wilk statistics to test for deviations from normality.

589

590 **Sampling and resolution limits**

591 We ran an *in-silico* trial to test our ability to reconstruct the true ecDNA copy number
592 distribution from a sampled subset of varying sizes. We constructed a simulated ecDNA
593 copy number distribution from 2×10^6 cells using our stochastic simulations. We then
594 performed 500 random samples of 25, 50, 100 and 500 cells, reconstructed the sampled
595 ecDNA copy number distribution and compared similarity to the true copy number
596 distribution using Kolmogorov-Smirnov statistics. The distribution converges to the true
597 distribution with increasing sampling size and a comparably small sample of 100 to 500
598 cells is sufficient to reconstruct the true underlying ecDNA copy number distribution.

599

600

601 **Mathematical description of ecDNA dynamics**

602 **Deterministic two population model without selection**

603 In the simplest representation of the model, we discriminate cells that do or do not carry
604 copies of ecDNA. We denote cells with copies of ecDNA as $N^+(t)$ and cells without copies
605 of ecDNA with $N^-(t)$. We can write for the change of these cells in time t

$$\begin{aligned} 606 \quad \frac{\partial N^-(t)}{\partial t} &= N^-(t) + v(N^+(t))N^+(t) \\ 607 \quad \frac{\partial N^+(t)}{\partial t} &= N^+(t) - v(N^+(t))N^+(t) \end{aligned}$$

608 where $v(N^+(t))$ corresponds to the loss rate of random complete asymmetric ecDNA
609 segregation. We find for the fraction of cells carrying ecDNA $f^+(t)$ in an exponentially
610 growing population

$$611 \quad f^+(t) = \frac{2}{2+t}$$

612 The fraction of cells carrying ecDNA decreases with $\sim 1/t$ if ecDNA is neutral. Thus,
613 copies of neutral ecDNA are only present in a small subpopulation of tumour cells.

614

615 **Deterministic two population model with selection**

616 Above equations can be modified to allow for a fitness advantage $s > 1$ for cells carrying
617 ecDNA.

$$\begin{aligned} 618 \quad \frac{\partial N^-(t)}{\partial t} &= N^-(t) + sv(N^+(t))N^+(t) \\ 619 \quad \frac{\partial N^+(t)}{\partial t} &= sN^+(t) - sv(N^+(t))N^+(t) \end{aligned}$$

620 The solution to this set of equations is

621
$$N^+(t) = (1 - f^-)e^{st - (1-s)\int_0^t f^-(\tau)d\tau}$$

622 In the case of positive selection, the fraction of cells with ecDNA $f^+ \rightarrow 1$. For sufficiently
623 long times, the tumour will be dominated by cells carrying ecDNA.

624

625 **Stochastic dynamics of neutral ecDNA**

626 We are also interested in the stochastic properties of ecDNA dynamics in a growing
627 population. We therefore move to a more fine-grained picture and consider the number
628 of cells $N_k(t)$ with k copies of ecDNA at time t . The dynamic equation for neutral copies
629 of ecDNA becomes

630
$$\frac{\partial N_k(t)}{\partial t} = -N_k(t) + 2 \sum_{i=\lfloor k/2 \rfloor}^{\infty} N_i(t) \binom{2i}{k} \frac{1}{2^{2i}}$$

631 It is more convenient to work with the cell density ρ instead of cell numbers N .
632 Normalizing above equation, we get for the density ρ_k of cells with k ecDNA copies

633
$$\frac{\partial \rho_k(t)}{\partial t} = -2\rho_k(t) + 2 \sum_{i=\lfloor k/2 \rfloor}^{\infty} \rho_i(t) \binom{2i}{k} \frac{1}{2^{2i}}$$

634

635 **Moment dynamics for neutral ecDNA copies**

636 With above equation for the density of cells with k ecDNA copies, we can calculate the
637 moments of the underlying probability density function. In general, the l -th moment can
638 be calculated via

639
$$M^{(l)}(t) = \sum_{i=0}^{\infty} i^l \rho_i(t)$$

640 It can be shown that all moments scale with $M^{(l)}(t) \sim t^{l-1}$ and we find explicitly for the first
641 two moments

642
$$M^{(1)} = 1 \quad \text{and} \quad M^{(2)}(t) = t$$

643 The mean ecDNA copy number in an exponentially growing population remains constant
644 for neutral ecDNA copies. The variance of the ecDNA copy number increases linearly in
645 time.

646

647 **Stochastic dynamics of ecDNA under positive selection**

648 Above equations can be generalized to accommodate positive selection ($s > 1$) for
649 ecDNA copies. The set of dynamical equations for cell densities becomes

650
$$\left. \frac{\partial \rho_k(t)}{\partial t} \right|_{k>0} = s \left. \frac{\partial \rho_k(t)}{\partial t} \right|_{s=1} + (s-1)\rho_k\rho_0$$

651
$$\left. \frac{\partial \rho_0(t)}{\partial t} \right|_{s=1} = s \left. \frac{\partial \rho_k(t)}{\partial t} \right|_{s=1} + (s-1)(1-\rho_0)\rho_0$$

652 A general solution to these equations is challenging, but still important quantities, e.g.,
653 the moment dynamics and the scaling behavior can be calculated explicitly.

654

655 **Moment dynamics for ecDNA under positive selection**

656 A generalized equation for the dynamics of moments directly follows from above
657 equations. We have

658
$$\frac{\partial M^{(l)}(t)}{\partial t} = s \frac{\partial M^{(l)}(t)}{\partial t} \Big|_{s=1} + (s-1)\rho_0 M^{(l)}(t)$$

659 This implies for the first moment $\frac{\partial M^{(1)}(t)}{\partial t} = (s-1)\rho_0 M^{(1)}(t)$, which then can be solved for
660 the first moment

661
$$M^{(1)}(t) = e^{(s-1)\int_0^t d\tau \rho_0(\tau)}$$

662 Similarly, the dynamic equation for the second moment becomes $\frac{\partial M^{(2)}(t)}{\partial t} = M^{(1)}(t) + (s-1)\rho_0 M^{(2)}(t)$ and we find

664
$$M^{(2)}(t) = tM^{(1)}(t)$$

665 The first moment increases exponentially initially. However, with increasing mean copy
666 number, the rate of cells transition into a state without ecDNA is decreasing and the
667 increase of the mean ecDNA copy number slowly levels off. Note, for $s = 1$ we recover
668 the previous results for the moments of neutral ecDNA amplifications.

669
670 **Scaling wave solution and limiting behavior of the ecDNA copy number distribution**

671 In the following, we are interested in the scaling behavior of the ecDNA copy number
672 distribution. Our general time dynamics considers discrete copy number states. To make
673 further analytical progress, we now consider continuous states in the following calculations.
674 This is an approximation that works well for the case of many ecDNA copies, but might
675 be inaccurate for cells with very few copies of ecDNA. Under this continuous assumption,
676 the change of the ecDNA copy number distribution becomes

677
$$\frac{\partial \rho(k, t)}{\partial t} = -2\rho(k, t) + \frac{2}{\sqrt{\pi}} \int_{k/2}^{\infty} dy \frac{\rho(y, t)}{\sqrt{y}} e^{-\frac{(k-y)^2}{y}}$$

678 Here, we also replaced the Binomial by a Normal distribution. Given the exponential
679 character of the ecDNA distribution, we proceed with an Ansatz in the form of a scaling
680 wave solution

681
$$\rho(k, t) = e^{-vt} \Omega(ke^{-vt})$$

682 Plugging this Ansatz into the dynamical equation and taking all terms into careful
683 consideration, it follows that $\Omega(k, t) = \frac{c}{k} e^{vt}$, where c is an undetermined constant and thus
684 we have

685
$$\rho(k, t) \sim k^{-1}$$

686 For sufficiently large ecDNA copy number. With other words, we expect the right-hand
687 tail of the ecDNA copy number distribution to follow a power law proportional to the
688 inverse of the ecDNA copy number. This prediction is confirmed in stochastic computer
689 simulations and can also be observed in experimentally measured distributions.

690
691 **Genome editing using CRISPR-Cas9 ribonucleoprotein**

692 Genome editing in COLO320-DM cells were performed using Alt-R S.p. Cas9 Nuclease
693 V3 (IDT, Cat# 1081058) complexed with sgRNA (Synthego) following Synthego's RNP
694 transfection protocol using the Neon Transfection System (ThermoFisher, Cat#
695 MPK5000). Briefly, 10 pmol Cas9 protein and 60 pmol sgRNA for each 10 ul reaction
696 were incubated in Neon Buffer R for 10 minutes at room temperature. Cell were washed
697 with 1X PBS, resuspended in Buffer R, and 200,000 cells were mixed with for the pre-

698 incubated ribonucleoprotein complex for each 10 ul reaction. The cell mixture was
699 electroporated following the manufacturer's protocol using the following settings: 1700 V,
700 20 ms, 1 pulse. Cells were cultured for 10 days afterwards; cell counts and ecDNA copy
701 number data were collected at day 3, 6, and 10. To estimated ecDNA copy numbers, we
702 performed metaphase chromosome spreading followed by FISH as described above. All
703 sgRNA sequences are in table below.

704

705

ID	gRNA_sequence	gRNA_target
1	GAACGACUAGUUAGGCGUGUA	Gal4 (non-targeting control)
2	GUGCUGCAAGGCGAUUAAGU	LacZ (non-targeting control)
3	CCAGCAAUCGUUAACCACUG	ecDNA intergenic region
4	GGUGAUAGAUUUUAUGCCCAG	ecDNA intergenic region
5	CUUCGGGGAGACAACGACGG	ecDNA MYC CDS
6	GCCGUUUUCUACUGCGACG	ecDNA MYC CDS
7	GUGAUUUUGAACCGCCCUG	Chr8
8	GAGGAUACAGUACUUCGCA	Chr8

706

707 **FISH probes**

708 ZytoLight SPEC CDK4/CEN 12 Dual Color Probe (ZytoVision)

709 ZytoLight SPEC MYCN/2q11 Dual Color Probe (ZytoVision)

710 Empire Genomics EGFR FISH Probe

711 Empire Genomics MYC FISH Probe

712 Empire Genomics FGFR2 FISH Probe

713

714 **Antibodies**

715 β -Actin (8H10D10) Mouse mAb #3700 (Cell Signaling)

716 CDK4 (D9G3E) Rabbit mAb #12790 (Cell Signaling)

717 IRDye 780RD Secondary Antibody (Licor)

718 IRDye 800CW Secondary Antibody (Licor)

719 Aurora B Polyclonal Antibody #A300-431A (ThermoFisher)

720

721

722 **Methods References**

723 27 Ambros, P. F. *et al.* International consensus for neuroblastoma molecular diagnostics: report
724 from the International Neuroblastoma Risk Group (INRG) Biology Committee. *Br J Cancer* **100**,
725 1471-1482, doi:10.1038/sj.bjc.6605014 (2009).

- 726 28 Naito, Y., Hino, K., Bono, H. & Ui-Tei, K. CRISPRdirect: software for designing CRISPR/Cas guide
727 RNA with reduced off-target sites. *Bioinformatics* **31**, 1120-1123,
728 doi:10.1093/bioinformatics/btu743 (2015).
- 729 29 Xie, L. *et al.* 3D ATAC-PALM: super-resolution imaging of the accessible genome. *Nat Methods*
730 **17**, 430-436, doi:10.1038/s41592-020-0775-2 (2020).
- 731 30 Guzman, C., Bagga, M., Kaur, A., Westermarck, J. & Abankwa, D. ColonyArea: an ImageJ plugin
732 to automatically quantify colony formation in clonogenic assays. *PLoS One* **9**, e92444,
733 doi:10.1371/journal.pone.0092444 (2014).
- 734 31 Jungbluth, A. A. *et al.* A monoclonal antibody recognizing human cancers with
735 amplification/overexpression of the human epidermal growth factor receptor. *Proc Natl Acad*
736 *Sci U S A* **100**, 639-644, doi:10.1073/pnas.232686499 (2003).
- 737