

Stochastic dynamics of extra-chromosomal DNA

Yuriy Pichugin¹, Weini Huang^{2,3,*}, Benjamin Werner^{4,*}

¹Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany. ²Department of Mathematics, Queen Mary University of London, United Kingdom. ³Group of Theoretical Biology, The State Key Laboratory of Biocontrol, School of Life Science, Sun Yat-sen University, Guangzhou, 510060 China. ⁴Evolutionary Dynamics Group, Centre for Cancer Genomics and Computational Biology, Barts Cancer Centre, Queen Mary University of London, Charterhouse Square, London, United Kingdom EC1M 6BQ.

* Correspondence should be addressed to weini.huang@qmul.ac.uk and b.werner@qmul.ac.uk

Abstract

Nuclear extra-chromosomal DNA (ecDNA) is highly prevalent in human tumours. ecDNA amplifications can promote accessible chromatin, oncogen over-expression and imply a worse clinical prognosis. Yet, little is known about the evolutionary process of ecDNA in human cancers. Here, we develop the theoretical foundation of the ecDNA somatic evolutionary process combining mathematical, computational and experimental approaches. We show that random ecDNA segregation leads to unintuitive dynamics even if ecDNA is under very strong positive selection. Patterns of inter- and intra-tumour ecDNA and chromosomal heterogeneity differ markedly and standard approaches are not directly applicable to quantify ecDNA evolution. We show that evolutionary informed modelling leads to testable predictions on how to distinguish positively selected from neutral ecDNA dynamics. Our predictions describe the dynamics of circular amplicons in GBM39 cell line experiments, suggesting a 300% fitness increase due to circular extra-chromosomal *EGFRvIII* amplifications. Our work lies the basis for further studies to quantitate ecDNA somatic evolutionary processes.

Introduction

Recent genome profiling studies of both healthy and cancerous human tissues found a high prevalence of nuclear extra-chromosomal DNA (ecDNA) [1, 2, 3, 4]. ecDNA occurs during episodes of chromosomal instability and contributes to cancer initiation and treatment resistance [5, 6, 7, 8, 9]. Double minute oncogene amplifications exist in 31.7% of all neuroblastoma and highly amplified extra-chromosomal KRAS copies appear at relapse of anti-EGFR targeted therapies [1, 4, 9]. Yet, the evolutionary fade

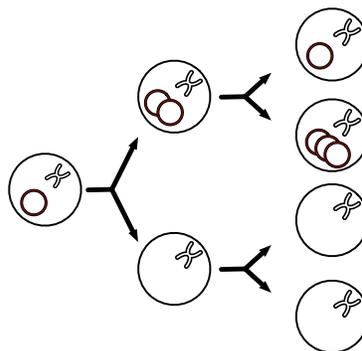


Figure 1. Schematics of extra-chromosomal DNA (ecDNA) population dynamics. ecDNA (circels) is randomly segregated between daughter cells. The resulting dynamics of ecDNA is markedly different compared to chromosomal alterations. For example, a cell can lose all ecDNA by chance, whereas other cells can acquire many copies of ecDNA. A growing cell population naturally becomes a mix of cells with and without ecDNA. Here, we describe how the inter- and intra-tumour ecDNA heterogeneity resembles the underlying evolutionary dynamic process, e.g. how one may quantify fitness consequences of ecDNA.

of specific ecDNA elements within individual tumours remains unknown. It is of great interest and practical need to develop methods to quantify ecDNA driven somatic evolutionary processes.

One major complication is the unique nature of the ecDNA somatic evolutionary process. Circular ecDNA lacks centromeres, causing random segregation of ecDNA into daughter cells after division [5] (Figure 1). Compared to somatic changes of chromosomal DNA, this adds additional stochasticity to the evolutionary dynamics of ecDNA [10, 11, 12]. Consequently, patterns of inter- and intra-patient ecDNA heterogeneity differ compared to chromosomal point mutations [13, 14, 15]. Methods that quantify somatic evolutionary processes based on chromosomal point mutations, such as phylogenetic trees [16, 17, 18, 19, 20, 21], ratios of synonymous and non-synonymous mutations (dN/dS) [22, 23, 24], or site-frequency spectra [25, 26, 27, 10, 28, 29] are not directly applicable to quantify the evolutionary dynamics of ecDNA.

Here, we discuss the stochastic dynamics of ecDNA in a growing tumour population. We show that the random segregation of ecDNA has unintuitive consequences. For example, the frequency of cells with ecDNA can intermittently drop, even if ecDNA conveys a high selective advantage. We present a mathematical and computational analysis of the expected inter- and intra-tumour ecDNA heterogeneity under different fitness scenarios and discuss strategies to infer these scenarios from cancer genomic data. We show that a single time point measurement at single cell resolution can be sufficient to disentangle different somatic evolutionary scenarios. Furthermore, we compare the theoretically predicted ecDNA heterogeneity patterns with experimental observations. Growth experiments with the glioblastoma cell line GBM39 are consistent with very strong positive selection and an approximate fitness increase of 300% due to circular extra-chromosomal *EGFRvIII* amplifications. In summary, our work describes the stochastic population dynamics of ecDNA in a growing tumour and provides a framework to quantify ecDNA fitness effects.

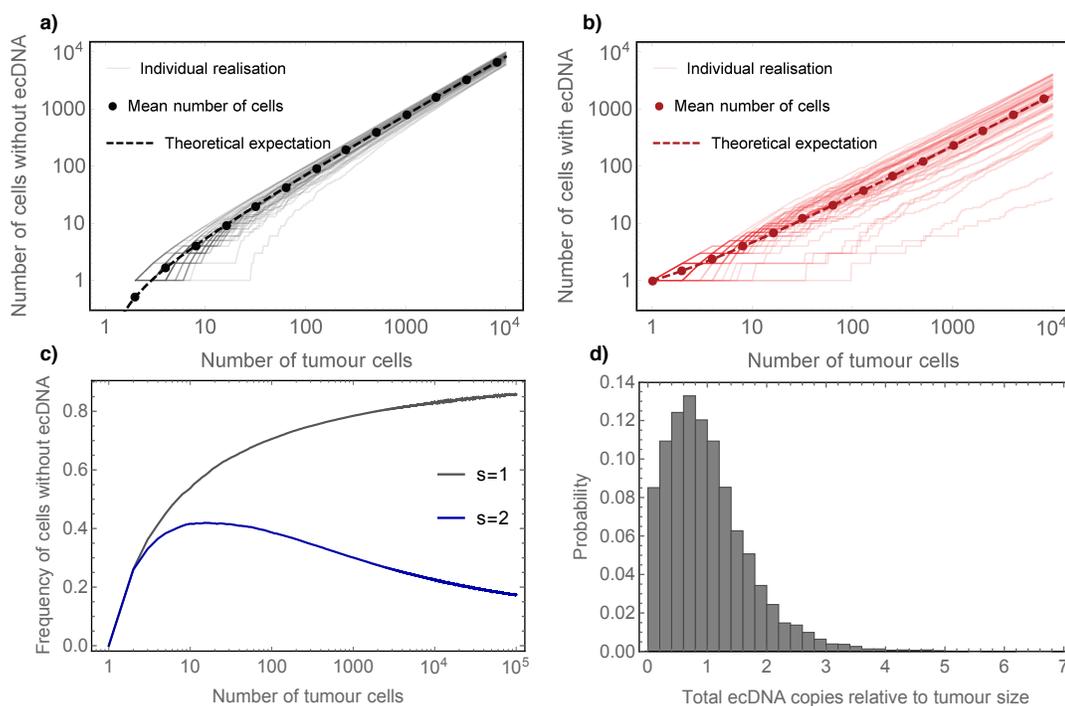


Figure 2. ecDNA population dynamics. Comparison of average deterministic dynamics of cells **a)** without and **b)** with copies of ecDNA. Dots show the average dynamics of neutral stochastic simulations, lines are individual realisation of the same neutral stochastic process and dashed lines show analytical predictions. Between tumour variation is considerable, especially for small tumour populations. **c)** Fraction of cells without ecDNA over time. In the neutral case $s = 1$ the tumour will be dominated by cells without ecDNA, also the fitness of cells with and without ecDNA is the same. Under strong positive selection, where cells with ecDNA have a selection advantage $s = 2$, the frequency of cells without ecDNA approaches 0. Even for strong positive selection we observe a transient increase of cells without ecDNA. **d)** Ratio of total number of ecDNA copies and total number of tumour cells for 10000 independent repeats of the same neutral evolutionary process. In many realisations there are as many or more ecDNA copies than tumour cells.

Results

In the following, we integrate a mathematical analysis of the ecDNA somatic evolutionary process, individual based stochastic computer simulations and *in vitro* ecDNA heterogeneity patterns from growth experiments of the glioblastoma cell line GBM39 with the known circular extra-chromosomal *EGFRvIII* amplification [1, 9]. For a better grasp on some of the peculiar properties of ecDNA evolution, we first discuss deterministic ecDNA dynamics. One important distinct feature of the ecDNA evolutionary process is the random inheritance of ecDNA copies into daughter cells (Figure 1). The probability for a daughter cell to carry k ecDNA copies upon proliferation is given by a Binomial distribution with success probability $1/2$ (Figure 1 and Methods). This implies that one daughter cell might not inherit any ecDNA by chance, which is distinctively different to the inheritance of somatic chromosomal point mutations. Assuming a cell carries k copies of ecDNA, a daughter cell would not inherit ecDNA with probability 2^{-k} . In other words, if a cell carries many copies of ecDNA, a daughter cell without ecDNA is unlikely. However, if a cell carries a single ecDNA copy, the subsequent

probability for a daughter cell without ecDNA is 1/4. We therefore should expect a tumour to be a mix of cells with and without ecDNA. Our results show that the exact tumour composition of cells with and without ecDNA is stochastic especially at the early stages of tumour evolution and critically depends on the underlying ecDNA fitness distribution.

Frequencies of cells with and without ecDNA

We now turn to the expected frequencies of cells with and without ecDNA in a growing tumour population. We start with an analysis of neutral ecDNA dynamics. In this case the presence or absence of ecDNA does not affect cell fitness. In a tumour with N cells, the expected frequency of cells with ecDNA is

$$f_{N^+} = \frac{2}{2 + \ln N}. \quad (1)$$

The frequency of cells with ecDNA f_{N^+} decreases with an increasing tumour population N . This might be surprising at first, given both cell populations have same fitness (Figure 2). However, while cells can lose ecDNA $N^+ \rightarrow N^-$, the reverse is not possible $N^- \rightarrow N^+$. This asymmetry is sufficient to suppress cells with neutral ecDNA in large and growing cell populations. This has important consequences on the interpretation of *in vitro* ecDNA experiments, as the frequency of cells with ecDNA continuously declines under neutral dynamics. For fitness inferences it will be necessary to show significant deviations from the dynamics as expected by equation (1), e.g. the frequency of negatively selected ecDNA should drop faster.

This does not imply that neutral ecDNA elements are not detectable in tumours at diagnosis. Based on equation (1) a tumour of 1 billion cells (10^9) would be composed to 91% of cells without and 9% of cells with ecDNA. The ratio only changes marginally to 93% and 7% for a tumour with 10^{12} cells, which is within the detection threshold of current sequencing approaches. In Figure 2 we show examples of the ecDNA dynamics. The average dynamics of cells with or without ecDNA in a tumour population is captured exactly by equation (1), but stochastic differences between individuals are considerable. Importantly, although the number of cells carrying ecDNA is relatively small in a tumour of detectable size, the total number of ecDNA copies is of the same order of all tumour cells (Figure 2 c&d). This is of particular importance for genes involved in resistance evolution. Highly amplified copies of EGFR are a common resistance mechanism in targeted therapies [30, 31, 32]. Neutral ecDNA dynamics allows for single cells to carry many copies of ecDNA, enabling crowding effects of resistance mechanisms in small cell populations, e.g. single cells have a higher chance to acquire multiple resistance mechanisms. Consequently, the probability of pre-existing multi-drug resistance for genes amplified on ecDNA may be dramatically increased compared to chromosomal resistance mechanisms.

The deterministic dynamics changes once cells with ecDNA are under positive selection. Large tumours are ultimately dominated by cells with ecDNA (Figure 2c). However, even if ecDNA conveys a very strong positive selection advantage, the dynamics remains unintuitive. The frequency of positively

selected cells with ecDNA changes approximately as

$$f_{N^+}(s, t) \approx \frac{2 + t \left(e^{\zeta(s)t} - 1 \right)}{2 + t e^{\zeta(s)t}}, \quad (2)$$

where the exponent $\zeta(s)$ is a function of the fitness coefficient s . Although large tumours are ultimately dominated by cells with ecDNA, cells without ecDNA initially rise in frequency and are only later outcompeted (Figure 2c). This transient effect is probably unobservable in large tumour populations, but may be important for cell line, organoid or xenograft experiments, urging a cautious interpretation of ecDNA fitness assays. For example, an initial decline of the frequency of cells with ecDNA can indicate neutral fitness, but also can be consistent with a fitness advantage. A robust experimental inference of ecDNA fitness will require growth to sufficiently large cell populations. Furthermore, even under very strong positive selection our model predicts the frequency of cells without ecDNA to always be above zero and thus one expects a mixed tumour population at any stage of the evolutionary process. As a consequence, the notion of clonal versus sub-clonal ecDNA copies becomes ambiguous [33, 34, 35], making it more difficult to determine if ecDNA was a founder event or occurred later during tumour evolution.

Inter- and intra-tumour heterogeneity of ecDNA

We now turn to the stochastic properties of the ecDNA dynamics. Primarily, we want to quantify the expected inter- and intra-tumour heterogeneity. We therefore change from a deterministic two population model to a demographic description of ecDNA dynamics, which tracks the number of ecDNA copies in each cell at any given time (see Methods for details). This results in an infinite set of coupled differential equations, for which a fully closed analytical solution is challenging. However, we can derive important approximate properties of the time dependent probability distribution. For neutral ecDNA copies, we find that the moments of the ecDNA copy number distribution in tumour cells change in time according to

$$M^{(l)}(t) \propto t^{l-1}. \quad (3)$$

In particular, this implies that the mean number \bar{k} of neutral ecDNA copies remains constant $\bar{k} = M^{(1)} = 1$, whereas the variance σ^2 of the ecDNA distribution increases linearly in time $\sigma^2 = M^{(2)} - 1 = t$. This also explains why under neutral dynamics the expected number of ecDNA copies in a growing tumour population is approximately the same as the number of tumour cells, despite the declining frequency of cells with ecDNA. In addition, this implies that on average neutral ecDNA amplifications would appear copy number neutral in a standard bioinformatics Copy Number Variation (CNV) pipeline for sufficiently large tumour bulks samples. Thus, neutral ecDNA amplifications may be more common than generally appreciated [3]. Novel bioinformatic tools to discover circular DNA structures help to solve some of these problems [3, 36]. Moreover, the polynomial scaling of all higher moments shows that inter-tumour ecDNA heterogeneity is expected to increase with time.

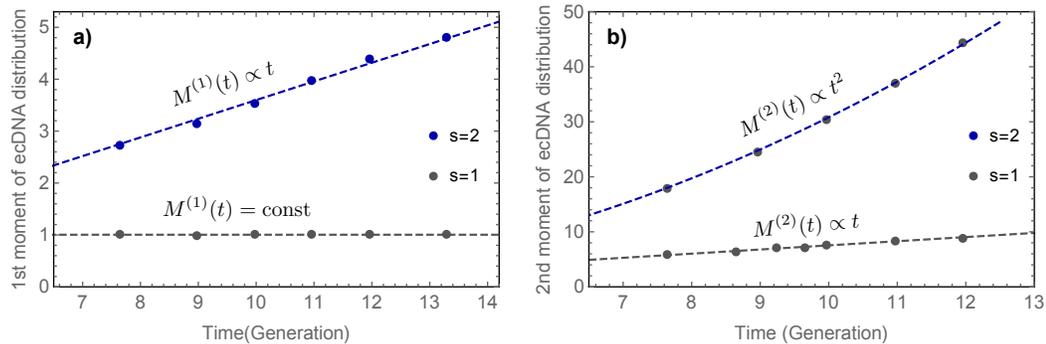


Figure 3. a) First and b) second moment of the ecDNA copy number distribution. In the neutral case ($s = 1$, grey) the mean number of ecDNA copies remains constant and the variance increases linearly in time. Under positive selection ($s = 2$, blue) the mean number of ecDNA copies increases in time. Stochastic simulations (points) are in very good agreement to theoretical predictions of polynomial increasing moments with time (dashed lines), see Eqs. (3) and (4).

If cells with ecDNA have a fitness advantage, the moments also have a polynomial time dependence

$$M^{(l)}(s, t) \propto t^{l-1} M^{(1)}(s, t). \quad (4)$$

However, now the first moment is not constant but increases in time t depending on the selection coefficient s , see Methods. This difference compared to the expected constant mean number of neutral ecDNA copies is important for fitness inferences. Correspondingly, measuring the change of the first moments (e.g. mean and variance) of the ecDNA distribution with time can in principal inform about the (non)-neutral nature of ecDNA amplifications. However, this would require high resolution temporally resolved single cell information of ecDNA copies, which is sparse in current clinical practice, but is available in *in vitro* experiments. With the fast development of single cell sequencing technology, a moment based method for ecDNA fitness inference may become more suitable for clinical settings in near future.

The lack of time resolved information on tumour evolution necessitates alternative evolutionary measures. One possibility is to quantify the expected distribution of ecDNA copies among (many) single cells of tumours, which is inferable from single time data and thus is more suitable to test ecDNA evolutionary dynamics in clinical practice. This is motivated by mutational site-frequency based methods in species and tumour evolution [10, 29]. There is an important distinction. The mutational site frequency spectrum quantifies cell fractions of point mutations, whereas ecDNA copy number distributions correspond to the number of cells with a certain number of ecDNA amplifications. We find that the number of cells N_k with k ecDNA copies is expected to scale as

$$N_k(t) \propto \frac{1}{k}, \quad (5)$$

i.e. inversely with the number of ecDNA copies. Surprisingly, for a constant fitness coefficient s (e.g. no dosing effects of ecDNA), the dynamics among cells with ecDNA is again neutral and thus the scaling

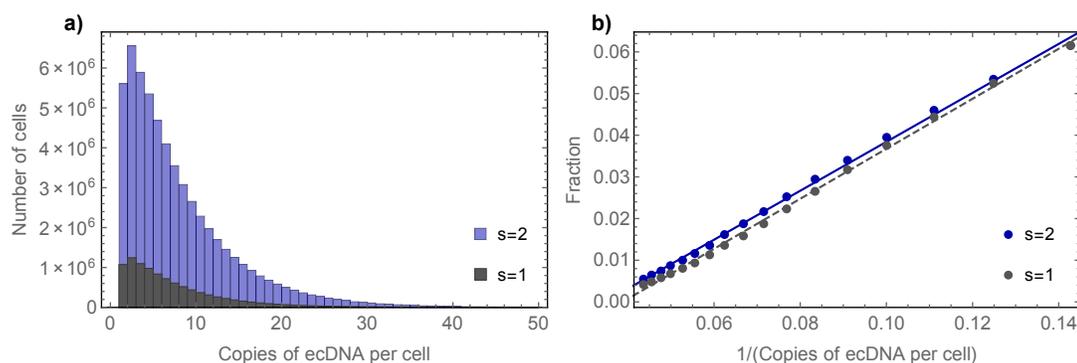


Figure 4. Distribution and scaling of the ecDNA copy number distribution. **a)** Distribution of the ecDNA copy number distribution for neutral (grey) and positively selected (blue) ecDNA evolution for 1000 repeats of stochastic simulations for tumours of 10^4 cells. Overall more cells carry copies of ecDNA if positively selected compared to the neutral case. **b)** The scaling of the ecDNA distribution follows the predicted $1/k$ scaling (dots = stochastic simulations, lines = theoretical expectation). However, the scaling law is the same for neutral evolution and constant positive selection and cannot discriminate neutrality from constant positive selection.

of the ecDNA copy distribution is again given by (5). Based on equation (5) a tumour should contain many cells with few ecDNA copies and fewer cells with many ecDNA copies (Figure 4). Stochastic simulations confirm the theoretical prediction even for small tumours (e.g. $N = 10^4$ in Figure 4).

The difference to the neutral power law scaling of chromosomal point mutations is noteworthy. In an exponentially growing tumour population, point mutations scale as f^{-2} with mutation frequency f [13, 14, 10], whereas the ecDNA copy distribution scales as k^{-1} with the number of ecDNA copies k . In addition, the scaling of the ecDNA copy number distribution alone cannot distinguish neutrality and selection. This is different for point mutations, where the site frequency spectrum can in principal distinguish neutrality from sufficiently strong selection [29, 37]. The scaling of the ecDNA copy number distribution depends solely on the within dynamics of cells with ecDNA. While the scaling of the ecDNA copy number distribution is insufficient to distinguish neutral and constant positive selection, it can distinguish those cases from a scenario where cells with different copy numbers of ecDNA have different fitnesses.

Shannon diversity of ecDNA distributions

A third option for an ecDNA fitness inference is to combine properties of the intra-tumour ecDNA heterogeneity and time series information. One promising candidate is the Shannon diversity index H of a tumour population with regard to the ecDNA distribution. We find that the Shannon diversity depends on both, the strength of selection (fitness coefficient s) and the total time selection shapes the composition of the tumour population. The Shannon diversity index H changes approximately as

$$H(\bar{k}) \propto \ln^2 [\bar{k}]. \quad (6)$$

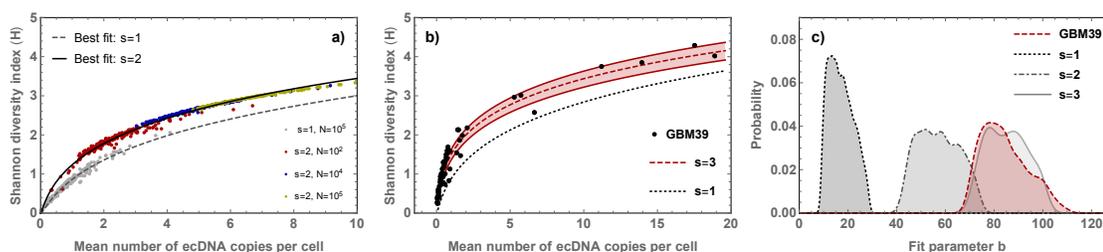


Figure 5. Circular extra-chromosomal *EGFRvIII* amplification in GBM39 is under strong positive selection. **a)** Shown is the Shannon diversity index H for stochastic simulations (each dot is a single realisation of the stochastic process) for varying degrees of selection and final sizes of the tumour population. For different fitness coefficients s , the Shannon diversity index falls on a fixed contour (e.g. dashed line: $s = 1$, black line: $s = 2$). Furthermore, the Shannon diversity increases towards higher values of H and \bar{k} if ecDNA is positively selected but remains clustered around 1 in the neutral scenario. **b)** Shannon diversity index for 72 repeats of *in vitro* growth experiment using the glioblastoma cell line GBM39 (black dots), data from [1]. The Shannon diversity follows a quadratic logarithmic contour as predicted by theory. The dashed line shows the best fit for a fitness advantage of $s = 3$, shaded area corresponds to 95% credibility intervals. Neutral ecDNA dynamics (black dashed line) does not describe the observed patterns of Shannon diversity in the GBM39 cell line. **c)** Posterior distributions of approximate bayesian fits of the parametric curve $H(\bar{k}) = a \ln [1 + b\bar{k}]$ to stochastic simulations under varying selection strength s (gray lines) and GBM39 cell line experiments (red). The ABC suggests very strong positive selection for cells with ecDNA in the GBM39 cell line ($s = 3$). Neutral ecDNA dynamics ($s = 1$, dotted line) does not describe the observed diversity.

This predicts a quadratic logarithmic dependence of the Shannon diversity index H and the mean number of ecDNA elements \bar{k} per cell. This dependence is recapitulated in stochastic individual based simulations. For numerical agreements, as shown in Figure 5a, we fit a parametrised curve $H(\bar{k}) = a \ln [1 + b\bar{k}]^2$ to stochastic simulations, using approximate bayesian computation (see Methods for details).

For neutral ecDNA copies, simulations cluster at small Shannon diversity and small mean number of copies of ecDNA per cell. Interestingly, if cells with ecDNA have a fitness advantage s , although the logarithmic dependence is maintained, there are two major differences from the neutral case. Firstly, the curve shifts to higher diversity values compared to the neutral case. Secondly, the diversity within a tumour of positively selected ecDNA increases in time and travels along sharply defined contours for constant fitness coefficients s (Figure 5a). This is a direct consequence of the linearly increasing mean number of ecDNA copies (Figure 3a). These two properties of the Shannon diversity index suggest a principal strategy to distinguish neutral from positive selection from single time point measurements. For neutral evolution, individual realisations of the stochastic process fall onto a uniquely parametrised curve $H(\bar{k}) = a \ln [1 + b\bar{k}]^2$ that clusters around a mean ecDNA copy number of $\bar{k} = 1$. In contrast, positive selection of ecDNA leads to significant deviations from the neutral contour, shifting H to both higher Shannon diversity and higher average ecDNA copy numbers (Figure 5).

We test this theoretical prediction of the fitness effect of ecDNA on the Shannon diversity in repeated growth experiments of a glioblastoma cell line. In a recent publication Kirsten Turner and colleagues quantified the ecDNA copy number distribution in an *in vitro* growth experiment using the glioblastoma cell line GBM39 with circular extra-chromosomal *EGFRvIII* amplification [1]. In total, 72 repeats of the same cell line experiment were done using low pass (median coverage: 1.19x) whole genome

sequencing data. This allowed us to quantify the contour of the Shannon diversity H for these experimental repeats (Figure 5b). We find a striking consistency of theoretical expectations and experimental observations. The cell line experiment shows a quadratic logarithmic dependence between the Shannon diversity and the average copy number of *EGFRvIII* (ecDNA element) for all experimental repeats. Even more striking, the Shannon diversity index of the circular extra-chromosomal *EGFRvIII* amplification changes along a single contour, suggesting a maintained fitness effect of the *EGFRvIII* amplification across experimental repeats. The curve is shifted towards higher Shannon diversity compared to the neutral expectation, suggesting that the extra chromosomal *EGFRvIII* amplification conveys a considerable fitness advantage in the GBM39 glioblastoma cell line. Figure 5c shows approximate bayesian fits (ABC) for different fitness scenarios in stochastic computer simulations and best inferences for the GBM39 cell line. The ABC inference suggests an almost 300% fitness increase ($s = 3$) for cells with compared to cells without the *EGFRvIII* amplification. The resulting parametric curve for $s = 3$ is shown in Figure 5b (dashed red line). It recovers the experimental Shannon diversity extremely well. In contrast, neutral ecDNA dynamics fails to describe the diversity patterns of ecDNA (black dotted line).

Discussion

Here we developed and analysed theoretical models of ecDNA population dynamics in a growing tumour. The random segregation of ecDNA causes surprisingly unintuitive dynamics that can be markedly different compared to the dynamics of chromosomal point mutations. Consequently, common methods to quantify somatic evolutionary processes such as phylogenetic trees, dN/dS, or mutational site frequency spectra are not directly applicable to quantify ecDNA evolutionary dynamics. However, as we have shown here, evolutionary models of ecDNA dynamics can be constructed and analysed. Our analysis points to a very high fitness advantage (300%) of the *EGFRvIII* amplification in the GBM39 cell line. This is consistent with recent functional evidence of the *EGFRvIII* amplification [1, 9], as well as clinical evidence of poor survival of patients with ecDNA amplifications [38]. It will be highly interesting to quantify evolutionary histories of ecDNA in individual patients in future studies.

Our theoretical model of ecDNA population dynamics is of course build on assumptions. We assume that ecDNA segregation is random and ecDNA copies are amplified (doubled) prior cell divisions. Most importantly, we assume constant fitness. Dosed ecDNA fitness effects and their impact on ecDNA diversity requires a more detailed analysis in follow up studies. Yet, current experimental observations seem to support such a minimal evolutionary model. Larger experimental and/or clinical studies may unearth observations that will require us to consider more involved evolutionary scenarios.

Acknowledgments

We are very grateful to Vineet Bafna and Paul S. Mischel for sharing data on the distribution of ecDNA copy numbers in the GBM39 cell line growth experiments. B.W. is supported by the Barts Cancer Charity.

Methods

Mathematical description of ecDNA dynamics

A deterministic two population model without selection

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

$$\frac{\partial N^-(t)}{\partial t} = N^-(t) + \nu(N^+(t))N^+(t) \quad (7)$$

$$\frac{\partial N^+(t)}{\partial t} = N^+(t) - \nu(N^+(t))N^+(t) \quad (8)$$

Cells with ecDNA N^+ can lose all copies of ecDNA, but the reverse is not possible. ecDNA can not gained back once lost. The function $\nu(N^+(t))$ corresponds to the at the moment undetermined loss rate of ecDNA due to random complete asymmetric segregation, e.g. one daughter cell with k and the other daughter cell with 0 ecDNA copies. We now look at the fraction of cells that do not carry ecDNA $f = \frac{N^-}{N^- + N^+}$. It follows that

$$\frac{\partial}{\partial t} \left(\frac{N^-}{N^- + N^+} \right) = \frac{\partial f}{\partial t} = (1 - f)\nu(N^+). \quad (9)$$

Again, for now both f and ν are unknown functions. However, from the equation above we can also write

$$\nu(N^+) = 1 - \frac{1}{N^+} \frac{\partial N^+}{\partial t} \quad (10)$$

and thus it follows

$$\frac{1}{1-f} \frac{\partial f}{\partial t} + \frac{1}{N^+} \frac{\partial N^+}{\partial t} = 1. \quad (11)$$

This equation can be integrated and after rearranging terms, we find the formal solution for $N^+(t)$,

$$N^+(t) = (1 - f(t))e^t, \quad (12)$$

with initial conditions $N^+(0) = 1$ and $f(0) = 0$. Assuming the tumour grows exponentially over time, e.g. $N^-(t) + N^+(t) = e^t$, we get an expression for the number of cells without ecDNA

$$N^-(t) = fe^t. \quad (13)$$

Stochastic simulations show that $\frac{N^-(t)}{N^+(t)} = \frac{1}{2}t$. It follows that the fraction of cells without ecDNA elements in the neutral case changes as

$$f(t) = \frac{t}{2+t}. \quad (14)$$

in time t (time measured in generations). This also allows us to find an exact expression for the expected loss of ecDNA copies $\nu(N^+)$ within the subpopulation of cells with ecDNA N^+ . Returning to equation (10) we find

$$\nu(N^+) = \frac{1}{2+t} = \frac{1}{2+\ln N}. \quad (15)$$

The rate of novel production of cells without ecDNA ν reduces with tumour size. However, the reduction is logarithmic in tumour size N and thus approaches 0 only slowly. Even in very large tumours, the rate is considerable, e.g. $\nu(N = 10^{12}) = 0.034$. Given 10^{12} cells this would translate to $N^+ \rightarrow N^- \approx 6 \times 10^{10}$ novel cells without ecDNA emerging from cells carrying ecDNA by chance due to uneven ecDNA inheritance.

A deterministic two population model with selection

How will the dynamics change if cells with ecDNA are under positive selection? We therefore introduce a selection coefficient s , where $s = 1$ corresponds to the neutral case and $s > 1$ implies a fitness advantage. This leads to a modification of above equations and they take the form:

$$\frac{\partial N^-(s,t)}{\partial t} = N^-(s,t) + s\nu(N^+(s,t))N^+(s,t) \quad (16)$$

$$\frac{\partial N^+(s,t)}{\partial t} = sN^+(s,t) - s\nu(N^+(s,t))N^+(s,t), \quad (17)$$

With the same procedure as above we can derive a general expression that is independent of $\nu(N^+)$, which becomes

$$(s-1)f + \frac{1}{1-f} \frac{\partial f}{\partial t} + \frac{1}{N^+} \frac{\partial N^+}{\partial t} = s. \quad (18)$$

Note, in the neutral case $s = 1$ this reduces to equation (11). This expression can be solved and we can write

$$N^+(t) = (1-f)e^{st-(1-s)\int_0^t f(\tau)d\tau}. \quad (19)$$

Stochastic simulations suggest a power law dependence of the form

$$f(s,t) \approx \frac{t}{2+te^{\tilde{c}(s)t}}. \quad (20)$$

Hence, the fraction of cell with copies of ecDNA is

$$f_{N^+}(s,t) = 1 - f(s,t) \approx \frac{2+t(e^{\tilde{c}(s)t} - 1)}{2+te^{\tilde{c}(s)t}}. \quad (21)$$

Stochastic dynamics of ecDNA elements

We now discuss the stochastic dynamics of ecDNA elements without selection. We therefore implement a demographic model of ecDNA, where we follow the number of cells $N_i(t)$ that carry i copies of ecDNA. Upon proliferation these copies are amplified and $2i$ copies are randomly distributed between both daughter cells. The number of copies k in a daughter cell then follows a Binomial distribution with success rate $1/2$

$$B(k|n = 2i, p = 1/2) = \binom{2i}{k} \frac{1}{2^{2i}}. \quad (22)$$

This allows us to write for the demographic model of the number of cells N_k with k copies of ecDNA

$$\frac{dN_k(t)}{dt} = -N_k(t) + 2 \sum_{i=\lceil \frac{k}{2} \rceil}^{\infty} N_i(t) \binom{2i}{k} \frac{1}{2^{2i}}. \quad (23)$$

We can decouple population growth and demographic changes via the substitution $N_i(t) = M(t)\rho_i(t)$ such that ρ_i is the density with $\sum_{i=0}^{\infty} \rho_i = 1$ and $M(t)$ corresponds to the population growth. In our case as a sanity check we can write

$$\begin{aligned} \frac{dM(t)}{dt} &= -M + 2M \sum_{k=0}^{\infty} \sum_{i=\lceil \frac{k}{2} \rceil}^{\infty} \rho_i \binom{2i}{k} \frac{1}{2^{2i}} \\ &= -M + 2M \sum_{i=0}^{\infty} \rho_i \frac{1}{2^{2i}} \sum_{k=0}^{2i} \binom{2i}{k} \\ &= -M + 2M \sum_{i=0}^{\infty} \rho_i \frac{1}{2^{2i}} 2^{2i} = M, \end{aligned} \quad (24)$$

and the total cell population does growth exponentially $M(t) = M(0)e^t$. This allows us to rewrite equation (23) for the densities ρ_k , which gives

$$\frac{d\rho_k(t)}{dt} = -2\rho_k(t) + 2 \sum_{i=\lceil \frac{k}{2} \rceil}^{\infty} \rho_i(t) \binom{2i}{k} \frac{1}{2^{2i}}. \quad (25)$$

Moments dynamics

Equation (25) enables us to discuss the Moments dynamics of the ecDNA copy number distribution. In general, the l -th moment of the distribution is defined as

$$M^{(l)}(t) = \sum_{i=0}^{\infty} i^l \rho_i(t). \quad (26)$$

We find that the first three moments of the distribution can be calculated exactly. We furthermore give approximate scaling expressions for all higher moments of the distribution. $M^{(0)}$ is just the sum over the density and by definition constant. The first moment corresponds to the average number of ecDNA copies per cell and we can write

$$\begin{aligned} \frac{dM^{(1)}(t)}{dt} &= -2M^{(1)}(t) + \sum_{k=0}^{\infty} \sum_{i=\lceil \frac{k}{2} \rceil}^{\infty} k \rho_i \binom{2i}{k} \frac{1}{2^{2i}} \\ &= -2M^{(1)}(t) + \sum_{i=0}^{\infty} \rho_i \frac{1}{2^{2i}} \sum_{k=0}^{2i} k \binom{2i}{k} \\ &= -2M^{(1)}(t) + \sum_{i=0}^{\infty} \rho_i \frac{1}{2^{2i}} (2i) 2^{2i-1} = 0 \end{aligned} \quad (27)$$

We therefore find $M^{(1)}(t) = \text{const}$ and the constant is given by $M^{(1)}(0)$, which in most cases discussed here is 1. For the second moment we find

$$\begin{aligned} \frac{dM^{(2)}(t)}{dt} &= -2M^{(2)}(t) + \sum_{k=0}^{\infty} \sum_{i=\lceil \frac{k}{2} \rceil}^{\infty} k^2 \rho_i \binom{2i}{k} \frac{1}{2^{2i}} \\ &= -2M^{(2)}(t) + \sum_{i=0}^{\infty} \rho_i \frac{1}{2^{2i}} \sum_{k=0}^{2i} k^2 \binom{2i}{k} \\ &= -2M^{(2)}(t) + \sum_{i=0}^{\infty} \rho_i \frac{1}{2^{2i}} (2i + (2i)^2) 2^{2i-2} = M^{(1)}(t) \end{aligned} \quad (28)$$

As we have $M^{(1)} = 1$, it follows that the second moment $M^{(2)} = t$ increases linear in time. The general case for the l -th moment leads to a differential equation of the form

$$\frac{dM^{(l)}(t)}{dt} = -2M^{(l)}(t) + \sum_{j=0}^{\infty} \rho_j M_B^{(l)}(2j, 1/2), \quad (29)$$

where $M_B^{(l)}$ is the l -th moment of the Binomial distribution. There are no general closed expressions for the higher moments of the Binomial distribution. Approximate expressions can be obtained using the moment generating function

$$G(z|2i, 1/2) = \left(\frac{1 + e^z}{2} \right)^{2i}, \quad (30)$$

and the moments then become

$$M_B^{(l)} = \frac{\partial^l}{\partial z^l} G(z|2i, 1/2) |_{t=0} = 2^{-2i} \sum_{m=0}^l c_{m,l} e^{mz} (1 + e^z)^{2i-m} \quad (31)$$

and therefore $M_B^{(l)} = \sum_{m=0}^l c_{m,l} 2^{-m}$. To get expressions for the coefficients $c_{m,l}$ we can differentiate the moments equation one more time and get the recursion relation

$$c_{m,l} = mc_{m,l} + (2i - m + 1)c_{m-1,l-1}. \quad (32)$$

For example, $M_B^{(1)} = i$ and $M_B^{(2)} = i^2 + i/2$. The moments of $M_B^{(l)}$ are polynomials over i and once plugged back into equation (29) they become a linear combination of moments $M^{(i)}$ of the ecDNA copy number distribution ρ . The maximal possible rank of the l -th moment emerging in equation (29) is given by the maximal degree of the polynomial in $M_B^{(l)}$ which is given by the coefficient $c_{l,l}$ in above recursion. In this case, the recursion takes the form

$$c_{l,l} = (2i - l + 1)c_{l-1,l-1} \quad (33)$$

which has the solution $c_{l,l} = \frac{(2i)!}{(2i-l)!}$. Keeping only the coefficient with the highest power, we get $M_B^{(l)} = i^l(1 + O(i^{-1}))$. Returning to equation (29) and using the approximate expression for $M_B^{(l)}$ we get

$$\begin{aligned} \frac{dM^{(l)}(t)}{dt} &= -2M^{(l)}(t) + \sum_{i=0}^{\infty} \rho_i i^l (1 + O(i^{-1})) \\ &= -2M^{(l)}(t) + 2M^{(l)}(t) + 2 \sum_{i=0}^{\infty} \rho_i [A_{l-1} i^{l-1} + A_{l-2} i^{l-2} \dots] \\ &= 2A_{l-1} M^{(l-1)}(t) + 2A_{l-2} M^{(l-2)}(t) + \dots \end{aligned} \quad (34)$$

The derivative of the l -th moment depends linearly on all lower moments. Unfortunately, finding closed solutions for all coefficients A_{l-k} appears challenging. However, the general scaling behaviour becomes clear, and we have

$$M^{(l)}(t) \propto t^{l-1} \quad (35)$$

In the neutral case, the first moment remains constant and all higher moments diverge polynomial in time. This is in agreement with individual based stochastic simulations of the ecDNA dynamics Figure 3.

Continues limit and scaling wave solution

In the following we are interested in the scaling behaviour of the ecDNA copy number distribution. The general equation (25) considers discrete copy number states. To make further analytical progress, we consider continues states in the following calculations. Such an approximation likely describes the case of many ecDNA copies well, but might be inaccurate for cells with very few ecDNA copies. Under

the continues assumption, the change of the ecDNA distribution becomes

$$\frac{dN(k, t)}{dt} = -N(k, t) + \frac{2}{\sqrt{\pi}} \int_{\frac{k}{2}}^{\infty} dy \frac{N(y, t)}{\sqrt{y}} e^{\frac{(k-y)^2}{y}}. \quad (36)$$

Where $N(k, t)$ is the number of cells with ecDNA copy number k at time t and the Binomial distribution in equation (25) was replaced by a Normal distribution assuming a limit of large k . Similarly, we can write for the density

$$\frac{d\rho(k, t)}{dt} = -2\rho(k, t) + \frac{2}{\sqrt{\pi}} \int_{\frac{k}{2}}^{\infty} dy \frac{\rho(y, t)}{\sqrt{y}} e^{\frac{(k-y)^2}{y}}. \quad (37)$$

Given the exponential character of the ecDNA distribution, we proceed with an ansatz in the form of a scaling wave solution

$$\rho(k, t) = e^{-vt} \Omega(ke^{-vt}). \quad (38)$$

Plugging Ansatz (38) back into equation (37) and setting $k/2 \rightarrow 0$, we get

$$\frac{\partial}{\partial t} e^{-vt} \Omega(ke^{-vt}) = -2e^{-vt} \Omega(ke^{-vt}) + \int_0^{\infty} dy e^{-vt} \Omega(ke^{-vt}) \frac{1}{\sqrt{y}} e^{\frac{(k-y)^2}{y}}. \quad (39)$$

The left side of equation (39) then becomes

$$\frac{\partial}{\partial t} e^{-vt} \Omega(ke^{-vt}) = -ve^{-vt} \Omega(ke^{-vt}) - kve^{-2vt} \frac{\partial}{\partial z} \Omega(z), \quad (40)$$

with $z = ke^{-vt}$. For $v = 2$ the first terms on the right hand side of equation (39) and (40) cancel and equation (39) becomes

$$-z \frac{\partial}{\partial z} \Omega(z) = \frac{1}{\sqrt{\pi}} \int_0^{\infty} dw \frac{\Omega(w)}{\sqrt{w}} e^{t - \frac{(w-z)^2}{2}} e^{2t}, \quad (41)$$

where as above $z = ke^{-2t}$ and $w = ye^{-2t}$. With a final substitution $m = (w - z)e^t$, substituting m for w and taking the limit $t \rightarrow \infty$ we get

$$-z \frac{\partial}{\partial z} \Omega(z) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dm \Omega(z) \frac{1}{\sqrt{z}} e^{\frac{-m^2}{z}} = \Omega(z), \quad (42)$$

with the solution

$$\Omega(z) = \frac{c}{z}, \quad (43)$$

and an undetermined integration constant c . Finally this gives for the scaling behaviour of the ecDNA copy number distribution

$$\rho(k, t) = e^{-2t} \Omega(z) = \frac{c}{k}. \quad (44)$$

This predicts that the number of cells with k ecDNA copies scales inversely with k . We find the same scaling behaviour in individual based stochastic simulations of the ecDNA copy number distribution, Figure 4.

Dynamics of ecDNA under constant positive selection

Here, we consider the situation, where cells containing ecDNA grow with rate $s > 1$. Then, Eq. (23) becomes

$$\begin{aligned} \left. \frac{dN_k(t)}{dt} \right|_{k>0} &= -sN_k(t) + 2s \sum_{i=\lceil \frac{k}{2} \rceil}^{\infty} N_i(t) B(k|2i, 1/2) \\ &= s \left[\left. \frac{dN_k(t)}{dt} \right|_{s=1} \right], \\ \frac{dN_0(t)}{dt} &= -N_0(t) + 2N_0(t) + 2s \sum_{i=1}^{\infty} N_i(t) B(0|2i, 1/2) \\ &= s \left[\left. \frac{dN_0(t)}{dt} \right|_{s=1} \right] - (s-1)N_0(t), \end{aligned} \quad (45)$$

Again, we decouple population growth and demographic changes via the substitution $N_i(t) = M(t)\rho_i(t)$ such that ρ_i is the density with $\sum_{i=0}^{\infty} \rho_i = 1$ and $M(t)$ corresponds to the population growth. Similar to Eqs. (24) and (25), we get

$$\begin{aligned} \frac{dM(t)}{dt} &= sM - (s-1)M\rho_0 \\ &= s \left[\left. \frac{dM(t)}{dt} \right|_{s=1} \right] - (s-1)M\rho_0, \\ \left. \frac{d\rho_k(t)}{dt} \right|_{k>0} &= -2s\rho_k(t) + 2s \sum_{i=\lceil \frac{k}{2} \rceil}^{\infty} \rho_i(t) B(k|2i, 1/2) + (s-1)\rho_k\rho_0 \\ &= s \left[\left. \frac{d\rho_k(t)}{dt} \right|_{s=1} \right] + (s-1)\rho_k\rho_0, \\ \frac{d\rho_0(t)}{dt} &= -2s\rho_0(t) + 2s \sum_{i=0}^{\infty} \rho_i(t) B(0|2i, 1/2) - (s-1)\rho_0(1-\rho_0) \\ &= s \left[\left. \frac{d\rho_0(t)}{dt} \right|_{s=1} \right] - (s-1)\rho_0(1-\rho_0). \end{aligned} \quad (46)$$

In the case of constant positive selection, the demographic dynamics gains additional non-linear terms compared to the neutral case.

Taking into account these terms, we can find new moments dynamics:

$$\begin{aligned} \frac{dM^{(l)}(t)}{dt} &= -2sM^{(l)}(t) + s \sum_{j=0}^{\infty} \rho_j M_B^{(l)}(2j, 1/2) + (s-1)\rho_0 M^{(l)} \\ &= s \left[\left. \frac{dM^{(l)}(t)}{dt} \right|_{s=1} \right] + (s-1)\rho_0 M^{(l)}. \end{aligned} \quad (47)$$

Hence,

$$\frac{dM^{(1)}(t)}{dt} = (s-1)\rho_0 M^{(1)}, \quad (48)$$

and as a result,

$$M^{(1)}(t) = e^{(s-1)\int_0^t \rho_0(\tau) d\tau}. \quad (49)$$

Similarly,

$$\begin{aligned} \frac{dM^{(2)}(t)}{dt} &= M^{(1)} + (s-1)\rho_0 M^{(2)}, \\ M^{(2)}(t) &= t e^{(s-1)\int_0^t \rho_0(\tau) d\tau} = t M^{(1)}. \end{aligned} \quad (50)$$

Similar to Eq. (34), we can show that

$$M^{(l)}(t) = P_l(t) e^{(s-1)\int_0^t \rho_0(\tau) d\tau} \sim t^{l-1} M^{(1)}, \quad (51)$$

where $P_l(t)$ is some polynomial over t of degree $l-1$. Therefore, the power law of the moment dynamics with incrementing powers observed for neutral selection is also preserved in the constant selection case. However, closed expressions are more challenging, as the equation for $M^{(1)}$ includes additional terms.

Shannon diversity index

For the Shannon diversity index H we can write

$$H = -\sum_i p_i \ln p_i \approx -\sum_i \frac{1}{iN_i} \ln \frac{1}{iN_i} \approx \int d\bar{k} \frac{\ln \bar{k}}{\bar{k}} = \frac{1}{2} \ln^2(\bar{k}) \quad (52)$$

The first approximation considers the scaling of the ecDNA copy distribution followed by approximating the sum by integration. We therefore find

$$H(\bar{k}) \propto \ln^2(\bar{k}). \quad (53)$$

Stochastic computer simulations recapitulate the quadratic logarithmic scaling of the Shannon diversity index (Figure 5a).

Stochastic individual based simulations of ecDNA population dynamics

We implemented individual based stochastic computer simulations of the ecDNA population dynamics in C++. For each cell, the exact number of ecDNA copies is recorded through the simulation. Cells are

chosen randomly but proportional to fitness for proliferation. The proliferation rate of cells without ecDNA is set to $r^- = 1$ (time is measured in generations). A fitness effect for cells with ecDNA then corresponds to a proliferation rate $r^+ = s$. Here $s > 1$ models a fitness advantage, $0 < s < 1$ a fitness disadvantage and $s = 1$ corresponds to no fitness difference (neutral dynamics, $r^+ = r^-$). If a cell is chosen for proliferation, the number of ecDNA copies in that cell are doubled and randomly distributed into both daughter cells according to a Binomial trial with success rate 1/2. If a cell carries no ecDNA, no daughter cell inherits ecDNA. We terminate simulations at a specified cell population size. We output the copy number of ecDNA for each cell at the end of each simulation, which allows us to construct other quantities of interest, such as the ecDNA copy distribution, the time dynamics of the moments, or the Shannon diversity index H .

Approximate Bayesian Computation

We implement Approximate Bayesian computation (ABC) to fit the Shannon diversity index H to stochastic simulations of ecDNA dynamics with different fitness strength s . We run 10000 simulations for a fixed fitness coefficient s to generate a distribution of Shannon diversity indices H in dependence of the mean copy number distribution \bar{k} for each simulation. We then fit the parametric curve $H(\bar{k}) = a \ln^2 [1 + b\bar{k}]$ to each evolutionary scenario independently. Here a and b are free parameters fitted by the ABC algorithm. We therefore choose a and b randomly from uninformed uniform distributions ($a \in [0, 0.3]$ and $b \in [0, 200]$) and reject values for which the coefficient of determination $R^2(H(\bar{k}))$ of the parametric curve with random parameters a and b and the simulated data set is $R^2 < 0.995$. This gives a posterior distribution for the parameters a and b . We do the same ABC procedure for the Shannon diversity distribution of the GBM39 cell line experiment, with a threshold of $R^2(H) < 0.93$. We use the posterior parameter distributions to generate credibility intervals via bootstrapping.

References

1. Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature*. 2017 Mar;543(7643):122–125.
2. deCarvalho AC, Kim H, Poisson LM, Winn ME, Mueller C, Cherba D, et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nature Genetics*. 2018 May;50(5):708–717.
3. Møller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nature Communications*. 2018;9(1):1069.

4. Deshpande V, Luebeck J, Nguyen NPD, Bakhtiari M, Turner KM, Schwab R, et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nature Communications*. 2019;10(392):s41467.
5. Verhaak RG, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nature Reviews Cancer*. 2019;19:283–288.
6. Paulsen T, Kumar P, Koseoglu MM, Dutta A. Discoveries of extrachromosomal circles of DNA in normal and tumor cells. *Trends in Genetics*. 2018;34(4):270–278.
7. Bakhroum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell*. 2018;174(6):1347–1360.
8. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011;144(1):27–40.
9. Wu S, Turner KM, Nguyen N, Raviram R, Erb M, Santini J, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature*. 2019;p. 1–5.
10. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. *Nature Genetics*. 2016;48(3):238.
11. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*. 2017;168(4):613–628.
12. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics*. 2019;20(7):404–416.
13. Durrett R. Population genetics of neutral mutations in exponentially growing cancer cell populations. *The Annals of Applied Probability*. 2013;23(1):230–250.
14. Kessler DA, Levine H. Large population solution of the stochastic Luria–Delbrück evolution model. *Proceedings of the National Academy of Sciences*. 2013;110(29):11682–11687.
15. Fusco D, Gralka M, Kayser J, Anderson A, Hallatschek O. Excess of mutational jackpot events in expanding populations revealed by spatial Luria–Delbrück experiments. *Nature Communications*. 2016;7:12760.
16. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine*. 2012;366(10):883–892.
17. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545(7655):446–451.

18. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JH, O'Brien T, et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. *Cell*. 2018;173(3):611–623.
19. Lee-Six H, Øbro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature*. 2018;561(7724):473–478.
20. Silvestro D, Warnock RC, Gavryushkina A, Stadler T. Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nature Communications*. 2018;9(1):5237.
21. Werner B, Case J, Williams MJ, Chkhaidze K, Temko D, Fernandez-Mateos J, et al. Measuring single cell divisions in human cancers from multi-region sequencing data. *BioRxiv*. 2019;.
22. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. *Cell*. 2017;171(5):1029–1041.
23. Martincorena I, Fowler JC, Wabik A, Lawson AR, Abascal F, Hall MW, et al. Somatic mutant clones colonize the human esophagus with age. *Science*. 2018;362(6417):911–917.
24. Zapata L, Pich O, Serrano L, Kondrashov FA, Ossowski S, Schaefer MH. Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. *Genome Biology*. 2018;19(1):67.
25. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell*. 2012;149(5):994–1007.
26. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. *Nature Genetics*. 2015;47(12):1402–1407.
27. Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P, et al. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proceedings of the National Academy of Sciences*. 2015;112(47):6496–6505.
28. Sun R, Hu Z, Sottoriva A, Graham TA, Harpak A, Ma Z, et al. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nature Genetics*. 2017;49(7):1015–1024.
29. Williams MJ, Werner B, Heide T, Curtis C, Barnes CP, Sottoriva A, et al. Quantification of subclonal selection in cancer from bulk sequencing data. *Nature Genetics*. 2018;50(6):895–903.
30. Diaz Jr LA, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*. 2012;486(7404):537.
31. Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife*. 2013;2:e00747.

32. Khan KH, Cunningham D, Werner B, Vlachogiannis G, Spiteri I, Heide T, et al. Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the PROSPECT-C phase II colorectal cancer clinical trial. *Cancer Discovery*. 2018;8(10):1270–1285.
33. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013;501(7467):338–345.
34. Werner B, Traulsen A, Sottoriva A, Dingli D. Detecting truly clonal alterations from multi-region profiling of tumours. *Scientific Reports*. 2017;7:44991.
35. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JH, O'Brien T, et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. *Cell*. 2018;173(3):611–623.
36. Kumar P, Kiran S, Saha S, Su Z, Paulsen T, Chatrath A, et al. ATAC-seq identifies thousands of extrachromosomal circular DNA in cancers and cell lines. *bioRxiv*. 2019;p. 845040.
37. Chkhaidze K, Heide T, Werner B, Williams MJ, Huang W, Caravagna G, et al. Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. *PLoS Computational Biology*. 2019;15(7):e1007243.
38. Hoon K, Nam N, Kristen T, Sihan W, Jihe L, Viraj D, et al. Frequent extrachromosomal oncogene amplification drives aggressive tumors. *bioRxiv*. 2019;10.1101/859306v1.