Currently available bulk sequencing data do not necessarily support a model of neutral tumor evolution

To the Editor — Williams et al.[1] analyzed next-generation sequencing data from bulk tumor samples, supporting the hypothesis that selection is limited to times before malignant transformation while tumor cell populations afterward evolve exclusively by neutral evolution. The authors arrived at their conclusions by showing that the expected number of mutations (M) in an exponentially increasing population undergoing only neutral evolution grows linearly with the inverse allele frequency ($1/f$) of mutant alleles. The allele frequency of a variant is defined as the proportion of cells containing that particular variant. Their reasoning was based on the fact that, in a neutrally evolving cell population, all subclones grow at the same rate, so the allele frequency is fixed as the inverse of the number of cells present at the time of appearance of a new variant. The authors then concluded that a tumor displaying a linear relationship between the number of mutations and the inverse allele frequency should imply that the population evolves without selection. They also performed simulations of a branching-process model with selection to show that selection cannot explain a linear relationship between M and $1/f$. Furthermore, they observed a high correlation between M and $1/f$ in about one-third of the patient samples investigated as further indication of neutral tumor evolution. Their findings corroborate previous results[2] based on the 'Big Bang' model that suggested a similar conclusion in colorectal cancer using single-time-point bulk sequencing data.

We believe that the authors arrived at an erroneous conclusion, which also stands in contrast to other recent findings in this field[3–5], based on flawed logic known as the 'fallacy of the converse'. The fact that a model of neutral evolution leads to a linear relationship between M and $1/f$ does not imply that a linear relationship proves the presence of neutral evolution. In more abstract terms, A implying B does not necessarily mean that B implies A. Here we demonstrate that models with selection can also lead to a linear relationship between M and $1/f$ and that, therefore, linearity is a test statistic that cannot be used to distinguish between populations evolving with and without selection. Our results indicate that the claims made by Williams et al.[1] have little merit.

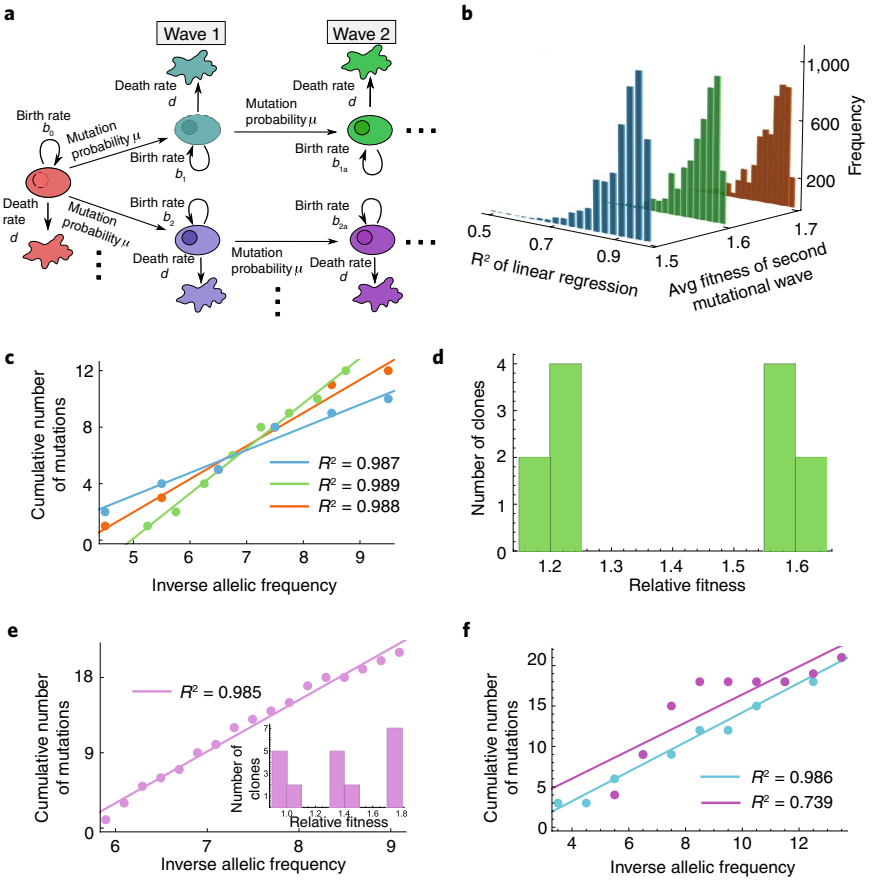

Fig. 1 | A simple branching-process model of tumor evolution. a, Schematic representation for the accumulation of mutations in our model. **b**, Histogram of R^2 values for the model with two mutational waves. Fitness values are as follows: type 0, 0.9; type i, $1 + N(0.2, 0.012)$; type ia, $1 + N(\mu, 0.012)$, with $\mu = 0.5$ (blue), 0.6 (green) or 0.7 (orange). Histograms were generated from 5,000 draws from $N(\mu, \sigma)$. **c**, One example draw with $R^2 > 0.985$ from each of the three cases in **b** is shown with corresponding color codes. **d**, Fitness distribution of the clones corresponding to $\mu = 0.6$ (green in **b** and **c**). **e**,**f**, As in **c** and **d**, but with three waves of mutations. Fitness values in **f** were chosen from a log-normal distribution with the same parameters as in **e**. The total size of the tumor in all cases is allowed to reach anywhere between 6×10^7 to 7×10^{11} cells.

We designed and analyzed two alternative stochastic evolutionary models that both return a linear relationship and show similarly high R^2 values for neutral as well as selection scenarios. To be consistent with the assumptions of the model and results used by Williams et al.[1], both models are based on exponentially growing, non-competing cellular populations with no spatial or microenvironmental effects and thus represent, by design, simplified versions of the tumorigenic process. The first model is a simple birth–death process of mutation accumulation (Fig. 1). In this model, each new mutation event gives rise to a single variant allele. This approach allows derivation of exact expressions for

the expected size of all mutant clones, thus providing an easy way of testing the authors' claim that a linear relationship can arise only from neutrality. The second model is a more complex infinite-allele branching-process model (Fig. 2) where multiple mutations may arise and lead to unique clones, making it analogous to the model developed by the authors. In both models, additive fitness effects in new clones are chosen from a fitness distribution such that any new mutant has a different birth rate that can lead to faster (or slower) growth in comparison to the parent clone. Furthermore, the second model incorporates more complex assumptions such as the infinite-allele model (rendering all mutants unique) as well as cell sampling and a Poisson-distributed number of variants (see below for details) to more closely match the model analyzed previously[1]. Simulation results from both models demonstrate that neutral (i.e., drift only) and selective evolution both give rise to linear relationships between M and $1/f$. Using both models, we tested a number of scenarios of selection, including intentionally chosen extreme (although biologically improbable) cases where every clone has a large fitness advantage or disadvantage. Even under such selection scenarios, we demonstrate that linearity with high R^2 values arises, highlighting a serious flaw in the authors' method of determining neutral evolution from curves with high R^2 values.

In the first model (Fig. 1a), clonal expansion begins with a single cell of the original, tumor-initiating type (type 0), which proliferates and dies with rates b_0 and d, respectively, and may accumulate mutations with probability μ during each cell division. The resulting mutant cells of type i ($i > 0$, birth rate b_i and death rate d) constitute the first wave of mutated cells, which in turn can mutate to produce the second wave (type ia, birth rate b_{ia} and death rate d), and so on. By deriving the differential equations governing the time evolution of this model, the expected number of cells of any type can be solved for exactly over time (see ref. [6] for details of the solution). We first explored the case of two mutational waves (Fig. 1c,d). The additive fitness values (additional birth rates) of the cell types are chosen from a normal distribution $N(\mu, \sigma)$ with mean μ and s.d. σ. We allowed μ to become progressively larger with wave number. The cumulative frequencies of cell types (M) were then calculated as a function of inverse frequency ($1/f$), and the R^2 values of linear regression between M and $1/f$ were calculated (examples where $R^2 > 0.985$ are shown in Fig. 1c). This process was performed 5,000

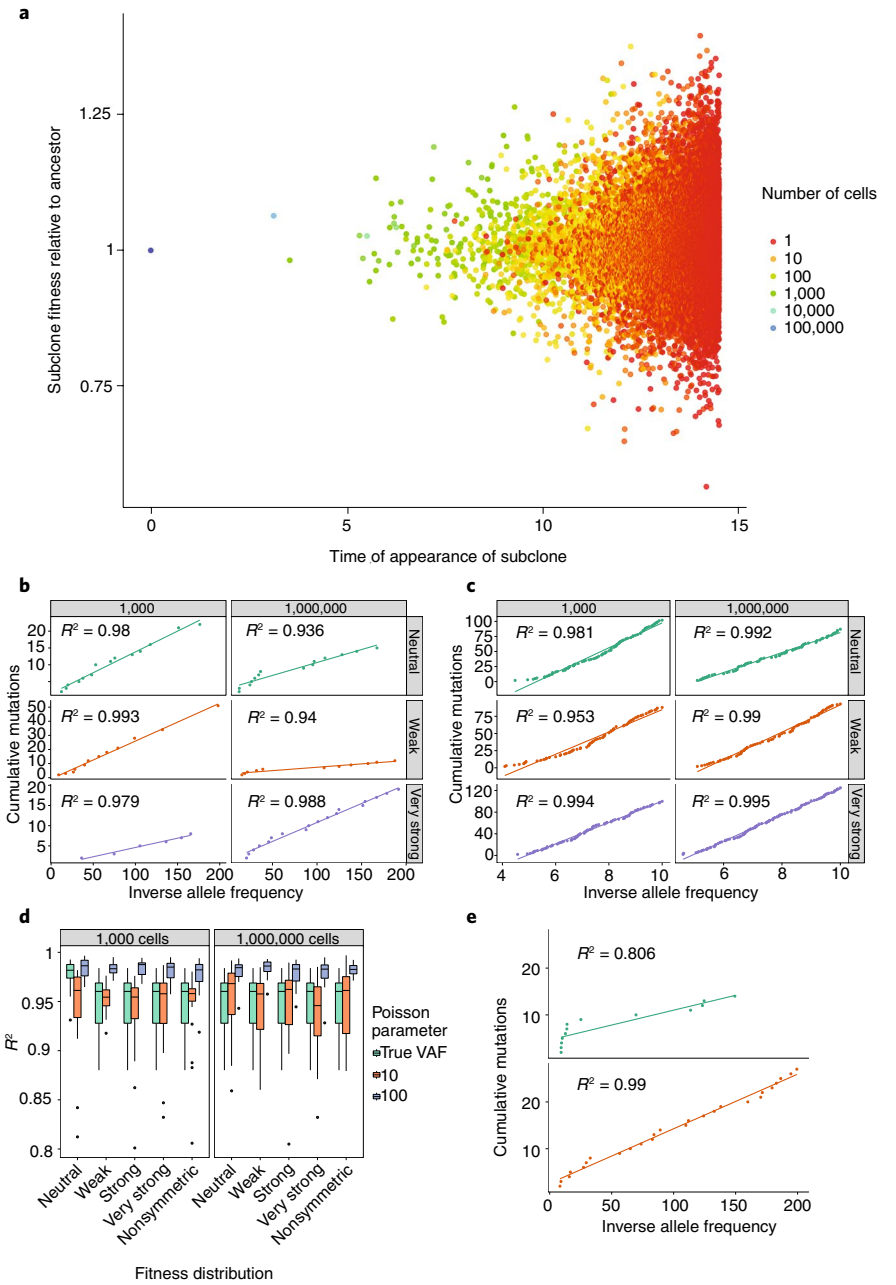

Fig. 2 | An infinite-allele branching-process model of tumor evolution, including sampling as in the original study. We initiate each process with a single ancestor with birth rate of 1, a death rate of 0.1, and a double-exponential fitness distribution with mean fitness change of 0.01 (weak), 0.04 (strong) or 1 (very strong) along with a neutral evolution model where there is no change in fitness and a model with only increasing fitness changes. **a**, The time of a new subclone's appearance with the birth rate colored by the subclone's size at the end of the simulation, showing that subclone size in a simulation with strong selection is associated with age but also with fitness. Allowing the simulation to run longer would result in younger subclones with high fitness outcompeting older ones. **b**, A plot of the cumulative number of mutations (M) and inverse allele frequency ($1/f$) shows linear trends in simulations where a single mutation arises from any mutation event and no additional noise is added to mimic the effect of sequencing. **c**, A linear trend is apparent between M and $1/f$ in the same model where each new mutation event contains Poisson(100) mutations and alleles are sampled to account for sequencing errors to create a result that follows the methods of Williams et al.[1]. **d**, Box plots for 25 simulations in all models for 1,000 and 1,000,000 cells show there is little change in R^2 as selection becomes larger, but allowing multiple mutations to occur at any mutation event has a large effect on linearity. VAF, variant allele frequency. **e**, The model is able to recapitulate nonlinear curves, suggesting the models with selection do not necessarily result in linear curves but can result in both linear and nonlinear curves.

times, and histograms were generated showing the distribution of R^2 values (Fig. 1b). This simple model of mutation accumulation already demonstrates that linear curves with high R^2 values can be easily obtained even when mutant cells are allowed to have large (~50–80%) fitness advantages. This model is limited by the number of distinct clones present, which results in only a few data points for the linearity test (Fig. 1c). To check whether increasing the number of clones affected our results, we allowed for the possibility of a third mutational wave (Fig. 1e,f), which significantly increased the number of data points on the basis of which R^2 is calculated. We also tested the ability of asymmetric fitness distributions to produce high R^2 values by choosing the additive fitness of mutants from a log-normal distribution with the same parameters as the previously used normal distribution (Fig. 1f). In all cases, R^2 values greater than 0.98 were easily obtained, thereby confirming our claim that models with selection can generate linear M versus $1/f$ curves. Finally, we also show a curve where the R^2 metric is less than 0.98 (Fig. 1f). The nonlinearity seen in this example is qualitatively similar to the examples shown by the authors in their Supplementary Fig. 11, thereby showing that this oversimplified model does recapitulate all scenarios demonstrated originally by the authors.

To go beyond this simplest model, we then constructed a continuous-time birth–death mutation process analogous to the model created in ref. [1]. Our process allows cells to live for an exponentially distributed time before dividing or dying, and cells may accumulate mutations during each cell division according to a given probability distribution (for details of the simulation technique, see ref. [7]). To match the approach in ref. [1], a new mutant cell contains a Poisson-distributed number of variants with rate 100, which is the same distribution and rate chosen by the authors. This approach allows multiple variants to arise at each mutation event, and sampling to account for sequencing noise results in multiple alleles with similar, but not identical, frequencies. Under the neutral model, mutant cells have the same birth rate as their parents, but when allowing for selection a mutant cell has a birth rate equal to the sum of the parent's rate and an additional fitness term generated from a double-exponential distribution, allowing mutations to be deleterious or advantageous. We continue the process until 1,000 cells (as in ref. [1]) and 1,000,000 cells accumulate to demonstrate how time, in addition to selection, affects linearity between M and $1/f$. The ancestor individual splits into two new cells with rate $b = 1$ and dies with rate $d = 0.1$. Given a split, one of the daughter cells may become a mutant with probability μ, which is 0.1 for the 1,000-cell scenario and 0.03 for the 1,000,000-cell scenario. A mutation results in a new clone with a birth rate of $b + s$, where s is chosen from the fitness distribution. We consider multiple levels of selection and a model with only advantageous selection. These levels of selection are based on the width of the fitness distribution, parameterized by the rate of the exponential distribution. Weak selection is associated with a rate parameter of 100 for the double-exponential fitness distribution, leading to an average change in the birth rate of 0.01 for a single mutation, while strong selection has a wider fitness distribution with a rate parameter of 25 that changes the fitness by an average of 0.04. Very strong selection is also included where the rate parameter is 1 such that the fitness doubles or halves on average with each new clone, representing a very extreme and significant increase. Finally, the asymmetric distribution is a one-sided exponential distribution with a rate of 25 where fitness only increases in the population. As mutations accumulate, the fitness of subclones increases as well, and the stronger selection scenarios are expected to lead to many more subclones with large fitness values relative to ancestor fitness.

The results of our more complex model (Fig. 2) indicate that the contribution of clones to the final total population size is mainly due to early mutations, but the accumulating fitness suggests that later subclones have the ability to outcompete earlier ones given enough time. Later subclones with large fitness values are still

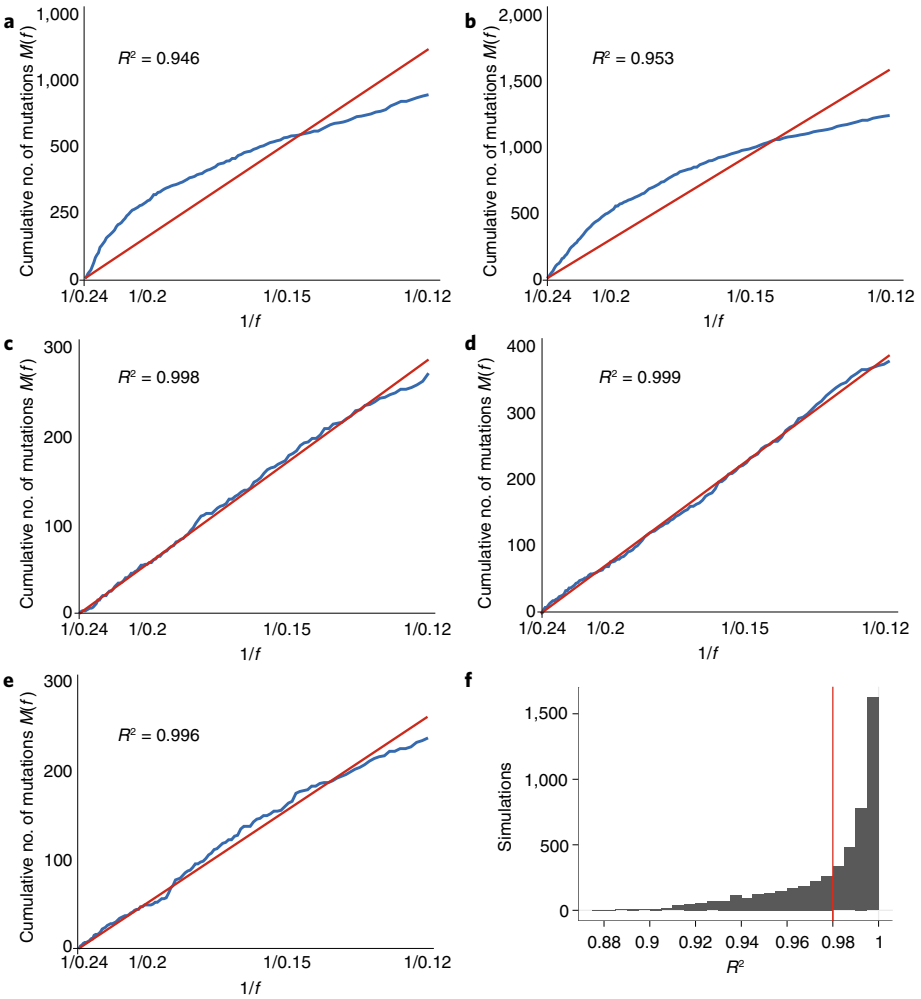

Fig. 3 | The model from the original study for multiple simulations. Using the code provided by Williams et al.[1] for situations with selection, we show that linearity between the cumulative mutation count and inverse allele frequency is widespread. **a**–**e**, We used the code with different seed values than provided by the authors to initiate the random number generator, including 5 (**a**), 7 (**b**), 2 (**c**), 911 (**d**) and 1,234 (**e**) seeds. **f**, A histogram of R^2 values for 5,000 runs of the code.

small owing to their young age but will eventually outcompete older, less fit clones. These subclones have usually accumulated multiple mutations, which allow for larger fitness values. The overlap in sizes among clones (Fig. 2a) also indicates that we cannot use cell counts or allele frequencies as a surrogate for time or age in such a population. Limiting our analysis to allele frequencies of [0.12, 0.24] as in ref. [1], the true allele frequency without accounting for multiple mutations or sequencing error is noticeable (Fig. 2b), but this effect is much stronger when multiple mutations are allowed to occur in an individual event and alleles are sampled from the population to represent noise in order to obtain results similar to the original analysis (Fig. 2c). Changing the Poisson parameter has the largest effect on linearity, as indicated by the box plots, and there is no apparent effect due to the strength of selection in the model (Fig. 2d). This observation suggests that R^2, or even linearity, is not a proper statistic to distinguish neutral from selection regimes, as both regimes tend toward a linear relationship.

Even more impressive is the drastic increase in R^2 as we increase the final cell count (Fig. 2b,c). The cumulative number of mutations for an individual sample in each scenario increases linearly with respect to inverse allele frequency in all scenarios, and the conformation to linearity becomes much stronger as the population size increases. However, we also show examples of relatively high R^2 at 1,000 cells that have nonlinear relationships. This analysis illustrates the problem in using R^2 as a cutoff, especially at such a high value where minor differences change the conclusion of neutral or selective evolution. Thus, we show relatively broad scenarios of evolution with selection that fit the original model[1]. This observation suggests that convergence to a linear relationship between mutation count and allele frequency is shared among branching-process models under the infinite-allele assumption, as previously shown[8,9], suggesting that linearity may be achieved in even more general scenarios evolving according to branching processes.

However, nonlinearity is not necessarily guaranteed for models with selection. The authors created a model with selection that leads to nonlinear trends between M and $1/f$, as we were also able to do with our models (Figs. 1f and 2e). Using the code provided in ref. [1], we found nonlinear trends for some simulation runs (Fig. 3a,b) but linear trends for others (Fig. 3c–e). In fact, 5,000 simulations using their model with selection generated a distribution where a majority of simulations (66%) had R^2 values greater than 0.98 (Fig. 3f), showing that their own code does not support the authors' conclusion.

Our results demonstrate the difficulty in drawing conclusions about parameters in population kinetics on the basis of data obtained at a single time point per patient. Even in the most simplified scenarios such as the absence of density-dependent interactions among cells and spatial components, the growth rate, mutation rate and tumor/clone age are all unknowns and provide too many degrees of freedom to elucidate estimates from single-time-point data. Our findings demonstrate that it is challenging to differentiate neutral from selective evolution given the data used in ref. [1] without obtaining additional quantitative molecular information about the tumor. In this context, one might be tempted to brandish Occam's razor and choose an apparently simpler neutral model for the 30% of cases where a linear relationship between M and $1/f$ was observed[1]. However, because 70% of the data show evidence of selection, a mixture model would then be required to account for all cases. This observation suggests that choosing a neutral model to describe 30% of the data because of parsimony is inadequate—a more complex model would be needed to describe all data. Considering that we are able to account for linearity and nonlinearity in a single model, our approach could therefore be considered more parsimonious.

Finally, we argue that an arbitrary R^2 value should not be used as a cutoff for linearity, especially when simulating branching processes to a number as low as 1,000 cells. We present simulation results for neutral and selective evolution that are similar, yet within multiple simulation runs we observe a large amount of variability between sampled alleles. Increasing the final population size helps resolve that variability in both scenarios, further demonstrating the problem of using R^2 without other analysis or exploratory work and suggesting a trend toward linearity as the number of cells increases regardless of the type of process. Given the inability to conclude that neutral evolution necessarily underlies the observed tumor mutation frequencies, we believe that estimates of patient-specific in vivo mutation rates, contrary to the authors' claims, can also be seen as scientifically inaccurate.

Reporting Summary

Further information on research design can be found in the Nature Research Reporting Summary linked to this article. ❐

Thomas O. McDonald[1,2,3,4,5], Shaon Chakrabarti[1,2,4,5] and Franziska Michor[1,2,3,4]*

[1]*Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA.* [2]*Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA.* [3]*Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA, USA.* [4]*Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA.* [5]*These authors contributed equally: Thomas O. McDonald, Shaon Chakrabarti.*
**e-mail: michor@jimmy.harvard.edu*

Published online: 29 October 2018
https://doi.org/10.1038/s41588-018-0217-6

References

1. Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. *Nat. Genet.* **48**, 238–244 (2016).
2. Sottoriva, A. et al. *Nat. Genet.* **47**, 209–216 (2015).
3. Gerlinger, M. et al. *N. Engl. J. Med.* **366**, 883–892 (2012).
4. Ding, L. et al. *Nature* **481**, 506–510 (2012).
5. Welch, J. S. et al. *Cell* **150**, 264–278 (2012).
6. Chakrabarti, S. & Michor, F. *Cancer Res.* **77**, 3908–3921 (2017).
7. McDonald, T. O. & Michor, F. *Bioinformatics* **33**, 2221–2223 (2017).
8. McDonald, T. O. & Kimmel, M. *J. Appl. Probab.* **52**, 864–876 (2015).
9. Jagers, P. & Nerman, O. *Adv. Appl. Probab.* **16**, 221–259 (1984).

Acknowledgements

The authors would like to acknowledge discussions with members of the Michor laboratory and with N. Navin, D. Pellman and K. Polyak. This work was supported by the Dana-Farber Cancer Institute Physical Sciences Oncology Center (NCI U54CA193461).

Author contributions

T.O.M. and S.C. developed the models, performed simulations and analyzed results. F.M. conceived the idea. All authors wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0217-6.

nature research

Corresponding author(s): Franziska Michor

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Simulations were ran using custom code for branching processes according to the conditions specified in the manuscript. Additionally, the open source package SIApop was used for simulating as well as the code released by the authors of the original manuscript.
Data analysis	Data analysis was done using R and graphics for analysis were created in R with ggplot2 and Matlab.

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

Data

Policy information about availability of data

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

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

No raw data

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	N/A
Data exclusions	N/A
Replication	Multiple simulations were ran under all scenarios.
Randomization	N/A
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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