A unified simulation model for understanding the diversity of cancer evolution

Atsushi Niida[1*], Takanori Hasegawa[2], Tatsuhiro Shibata[1], Koshi Mimori[3], Satoru Miyano[4]

[1]Laboratory of Molecular Medicine, Human Genome Center,
the Institute of Medical Science, the University of Tokyo, Tokyo, Japan
[2]Division of Health Medical Data Science, Health Intelligence Center,
the Institute of Medical Science, the University of Tokyo, Tokyo, Japan
[3]Department of Surgery, Kyushu University Beppu Hospital, Beppu, Japan,
[4]Laboratory of DNA Information Analysis, Human Genome Center,
the Institute of Medical Science, the University of Tokyo, Tokyo, Japan

* aniida@ims.u-tokyo.ac.jp

Abstract

Because cancer evolution underlies the therapeutic difficulties of cancer, it is clinically important to understand cancer's evolutionary dynamics Thus far, four different evolutionary modes have been proposed for cancer: linear, branching, neutral, and punctuated. However, no simulation model exists that can describe the different cancer evolutionary modes in a unified manner. In this study, we constructed a unified simulation model for describing the four cancer evolutionary modes and performed sensitivity analysis on the model to determine the conditions in which cancer growth is driven by each of the different evolutionary modes. our sensitivity analysis showed that. under the assumption of sufficiently high driver mutation rates, linear and branching evolutions occur with driver mutations of relatively strong and weak driver genes, respectively. Furthermore, we found that, although a high neutral mutation rate is necessary for neutral evolution, a stem cell hierarchy can also prompt neutral evolution by apparently increasing the mutation rate. Finally, we demonstrated the possibility that punctuated evolution underlies the evolutionary shift from branching to neutral evolution, which is observed during colorectal tumorigenesis. Collectively, this study provides a novel foundation for understanding the diversity of cancer evolution.

Author summary

Cancer is regarded as a disease of evolution; during tumorigenesis, a normal cell evolves to a malignant population by mutation accumulation and natural selection. Evolution enables cancer cells to adapt to a new environment and acquire malignant phenotypes such as metastasis and therapeutic resistance. Therefore, understanding of cancer evolutionary dynamics is a clinically important problem. To date, it has been proposed that four different modes exist in cancer. In this paper we introduce a unified simulation model, which can reproduce each of the different evolutionary modes. Moreover, we performed sensitivity analysis of the simulation model; that is, we checked parameter dependency of evolutionary dynamics by performing a large

number of simulations with various parameter settings. This analysis successfully unveiled the conditions in which each type of the evolutionary modes works. In summary, this work deepens our understanding of cancer evolution, which is necessary for developing new therapeutic strategies.

Introduction

Cancer is regarded as a disease of evolution; during tumorigenesis, a normal cell evolves to a malignant population by means of mutation accumulation and natural selection. Evolution allows cancer cells to adapt to a new environment and acquire malignant phenotypes such as metastasis and therapeutic resistance. Therefore, it is clinically important to understand cancer evolutionary dynamics. To date, it has been proposed that several different modes exist in cancer evolution [1]. Traditionally, cancer evolution has been described as "linear evolution," where mutations are acquired linearly in a step-wise process, generating a malignant clonal population.

However, this simple view of cancer evolution has been challenged since the advent of the next generation sequencing technology [2]. Deep sequencing demonstrated that subclonality prevails in both blood and solid tumors, and multiregion sequencing of various types of solid tumor more dramatically unveiled intratumor heterogeneity (ITH), which results from branching evolution. These genomic studies also found that subclones often harbor mutations in known driver genes, suggesting that at least a part of ITH is generated by Darwinian evolution. In some types of cancer, such as renal cell carcinoma [3] and low-grade glioma [4], this Darwinian evolution-driven branching evolution (hereafter, simply referred to as "branching evolution") is especially prominent; we observed convergent evolution in which different subclonal mutations are acquired in the same driver gene or pathway.

Other types of tumors, however, show no clear enrichment of driver mutations in subclonal mutations. Consistently with this observation, several studies employing mathematical modeling have suggested that mainly the accumulation of neutral mutations, which do not affect the growth and survival of cancer cells, shapes ITH; that is, "neutral evolution" is a major origin of ITH in multiple types of cancers [5–7]. The evolutionary principles shaping ITH differ not only among cancer types but also between stages of colorectal tumorigenesis. We and another group have recently reported that ITH is shaped by branching and neutral evolution in the early and late stages of colorectal tumorigenesis, respectively [8, 9].

In the linear, branching and neutral evolution modes, it is commonly assumed that microevolution events such as single nucleotide mutations are accumulated sequentially and gradually over time. However, recent studies have demonstrated that, in multiple types of cancers, chromosome or genome-wide evolutionary events producing copy number alterations and chromosomal rearrangements may occur in a short time at the early stage of evolution [10, 11]. Such rare macroevolution events could confer a large fitness increase on one or a few cells, after which one or a few clones originating from these cells expand to constitute the tumor mass uniformly. This type of evolution is referred to as "punctuated evolution."

Collectively, four types of cancer evolutionary modes were proposed thus far: linear, branching, neutral, and punctuated evolution. Although many mathematical models of cancer evolution have been employed to gain an understanding of cancer evolutionary dynamics [2, 12, 13], no simulation model that can describe the four evolutionary modes in a unified manner exists. In this paper, we propose a unified simulation model that describes the four cancer evolutionary modes. While constructing the simulation model, we also investigated the manner in which the different modes are realized in cancer evolution.

This paper is composed of four parts. In the first part, we introduce the driver model, which contains only driver mutations, and examine the conditions leading to linear or branching evolution. In the second part, the neutral model, which contains only neutral mutations, is introduced to address the conditions leading to neutral mutations. In the third part, we present a combination of these two models as a composite model to reproduce linear, branching, and neutral evolution. In the final part, furthermore we describe the incorporation of punctuated evolution in the composite model, which enables us to reproduce the evolutionary shift from branching to neutral evolution during colorectal tumorigenesis.

Results

Driver model

First, we constructed the "driver" model, which contains only driver genes, aiming to study the two Darwinian evolution modes: linear and branching evolution. We employed an agent-based model, where each agent corresponds to each cell in a tumor. The simulation started from one cell without mutations. In a unit time, a cell divides into two daughter cells with a probability g. This model assumes that immortalized cell, which just divides without dying. In each cell division, each of the two daughter cells acquires k_d driver mutations. Here, k_d is sampled from a Poisson distribution with the parameter $m_d/2$, i.e., $k_d \sim \mathrm{Pois}(m_d/2)$, which means that one cell division generates m_d mutations on average. We assumed that driver mutations acquired by different division events occur at different genomic positions and each cell can accumulate N_d mutations at maximum. In this study, we assumed that all mutations are driver mutations, which increase the cell division rate. When the cell acquires mutations, the cell division rate increases f fold per mutation; that is, when a cell has n_d $(= \sum k_d)$ mutations in total, the cell division probability g is defined as $g = g_0 f^{n_d}$, where g_0 is a base division probability. In each time step, every cell is subject to a cell division trial, which is repeated until population size p reaches P or the number of time steps t reaches T.

To examine the manner in which each parameter affects the evolutionary dynamics of the simulation model, we performed a sensitivity analysis utilizing MASSIVE [14], for which we employed a supercomputer. MASSIVE first performs a very large number of agent-based simulations with a broad range of parameter settings. The results are then intuitively evaluated by the MASSIVE viewer, which interactively displays heat maps of summary statistics and single-cell mutation profiles from the simulations with each parameter setting. In Fig 1, the heat maps of three representative summary statistics, the proportion of clonal mutations (clonal mutation proportion), a measure for ITH (Shannon index 0.05), and an indicator for the occurrence of branching evolution (branching evolution 0.05), are presented for only a part of the parameter space examined. All the results can be interactively explored in the MASSIVE viewer on our website (https://www.hgc.jp/~niiyan/canevosim/driver).

The results of the MASSIVE sensitivity analysis demonstrated that the strength of driver mutations f is the most prominent determinant of the Darwinian evolution modes. A smaller value of f, which indicates weaker driver mutations, is ingeneral associated with branching evolution, which is characterized by large branching evolution 0.05, corresponding to parameter setting i in Fig 1. However, in the case of a low mutation rate, i.e., small m_d, a small f value is insufficient to cause expansions of multiple clones, corresponding to parameter settingiii in Fig 1. When the value of f is large, branching evolution 0.05 is small, but the clonal mutation proportion is large, which suggests that linear evolution occurs, corresponding to parameter setting ii in

Fig 1. By considering these results with time-course snapshots of the simulations, mechanisms driving linear and branching evolution were intuitively interpreted (Fig 2). Under the assumption of weak driver mutations, before a clone that has acquired the first driver mutation becomes dominant, other clones that have acquired different mutations expand, leading to branching evolution (Fig 2 A and B). In contrast, under the assumption of strong driver mutations, a clone that has acquired the first driver mutation rapidly expands to obtain more driver mutations serially, leading to linear evolution (Fig 2 C and D).

We then modified the driver model to create the "driver-d" model where each cell divides with a constant probability g_0 and dies with a probability d. Moreover, we assumed that cell death occurs only in the case of $p > 1$, to prevent the simulation from halting before clonal expansion. In the driver-d model, each driver mutation decreases the cell division probability by f fold: $d = d_0 e^{-n_d}$, where d_0 is the base death probability. The MASSIVE analysis of the modified driver model suggested that, when driver mutations decrease the death rate, branching evolution is pervasive, irrespective of the strength of the driver mutations (https://www.hgc.jp/~niiyan/canevosim/driver_d). This observed behavior is supposed to occur because a driver mutation that decreases the death rate cannot provide a cell with the strong growth advantage necessary for linear evolution. By simulating tumor growth on a one-dimensional lattice, we also previously examined the effects of the spatial bias of a resource necessary for cell divisions: the MASSIVE analysis of the spatial model suggested that spatial resource bias could prompt branching evolution [14].

Neutral model

Next, we examined the neutral evolution mode by analyzing the "neutral" model, where we considered only neutral mutations that do not affect cell division and death. In a unit time, a cell divides into two daughter cells with a constant probability g_0 without dying. Similarly to driver mutations in the driver model, in each cell division, each of the two daughter cells acquires $k_n \sim \mathrm{Pois}(m_n/2)$ neutral mutations. We assumed that neutral mutations acquired by different division events occur at different genomic positions and each cell can accumulate N_n mutations at maximum. In this study, we set $N_n = 1000$, which is sufficiently large that no cell reaches the upper limit, except in a few exceptional cases. The simulation started from one cell without mutations and ended when population size p reached P or time t reached T.

The MASSIVE analysis of the neutral model demonstrated that, as expected, the mutation rate is the most important factor for neutral evolution (Fig 3; https://www.hgc.jp/~niiyan/canevosim/neutral_s; note that the neutral model is included by the neutral-s model, which is described below). When the mean number of mutations generated by per cell division, m_n, was less than 1, the neutral models just generated the sparse mutation profiles with relatively small values of the ITH score, Shannon index 0.05, whereas when m_n exceeded 1, the mutation profiles presented extensive ITH, characterized by fractal-like patterns and large values of the ITH score (hereafter, this type of ITH is referred to as "neutral ITH"). According to these results, it is intuitively supposed that neutral ITH is shaped by neutral mutations that trace the cell lineages in the simulated tumors. Note that the mutation profiles were visualized after filtering out low-frequency mutations. Under the assumption of a high mutation rate, more numerous subclones having different mutations should be found to exist if we count the mutations having lower frequencies.

To verify this speculation, we counted the number of subclones generated from a simulated tumor, while varying the frequency cutoffs for filtering out mutations. Fig S1 shows the plot of the relationship between the number of subclones and the

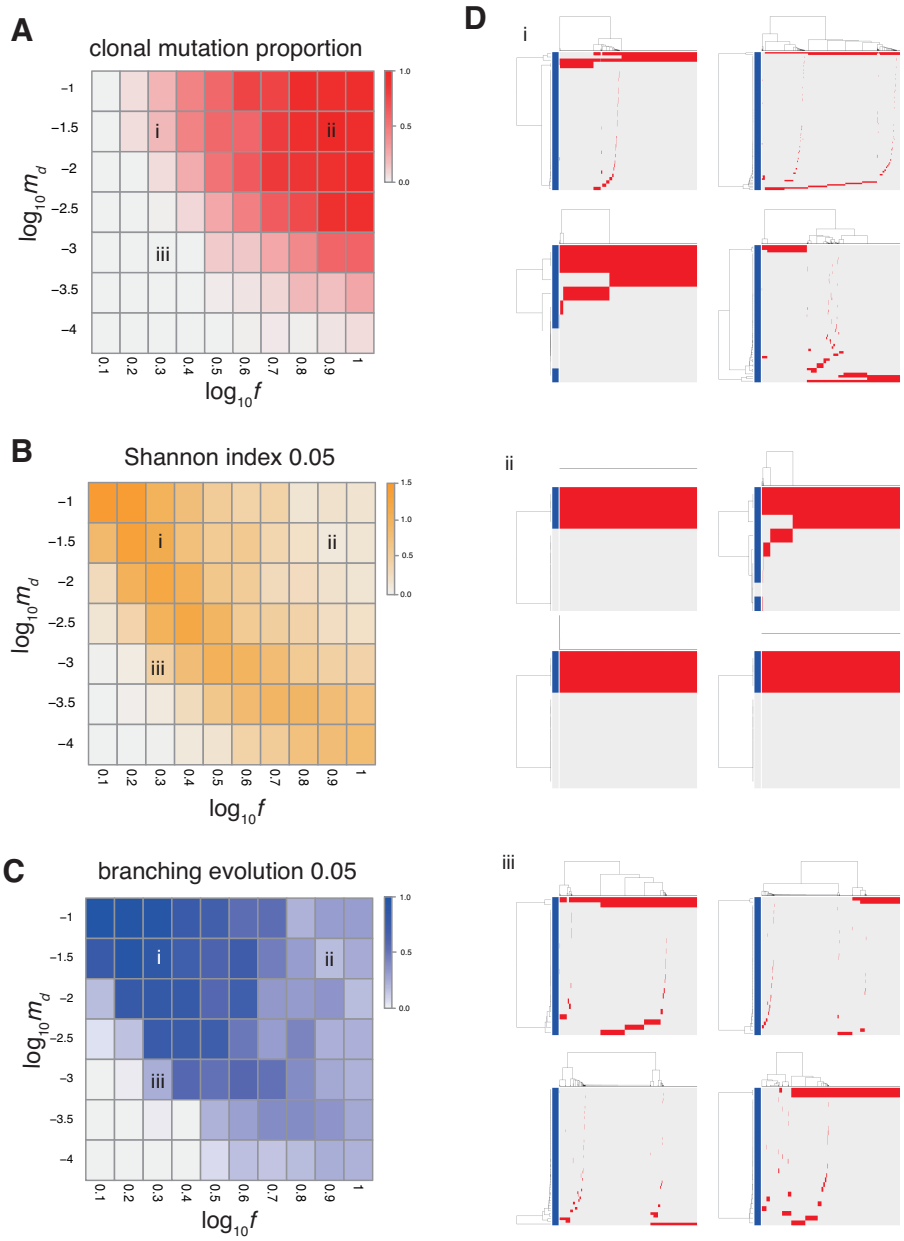

Fig 1. Sensitivity analysis of the driver model. While changing the driver mutation rate m_d and the strength of driver mutations f, heat maps of summary statistics were prepared for the proportion of clonal mutations, clonal mutation proportion (A), a measure for ITH, Shannon index 0.05 (B) and an indicator for the occurrence of branching evolution, branching evolution 0.05 (C). N_d and P were set to 3 and 10^5, respectively. (D) Single-cell mutations profiles obtained for three parameter settings, which are indicated on the heat maps presented in A, B, and C. Rows and columns of the clustered single-cell mutations profile matrices denote mutations and cells, respectively. Blue side bars indicate driver mutations.

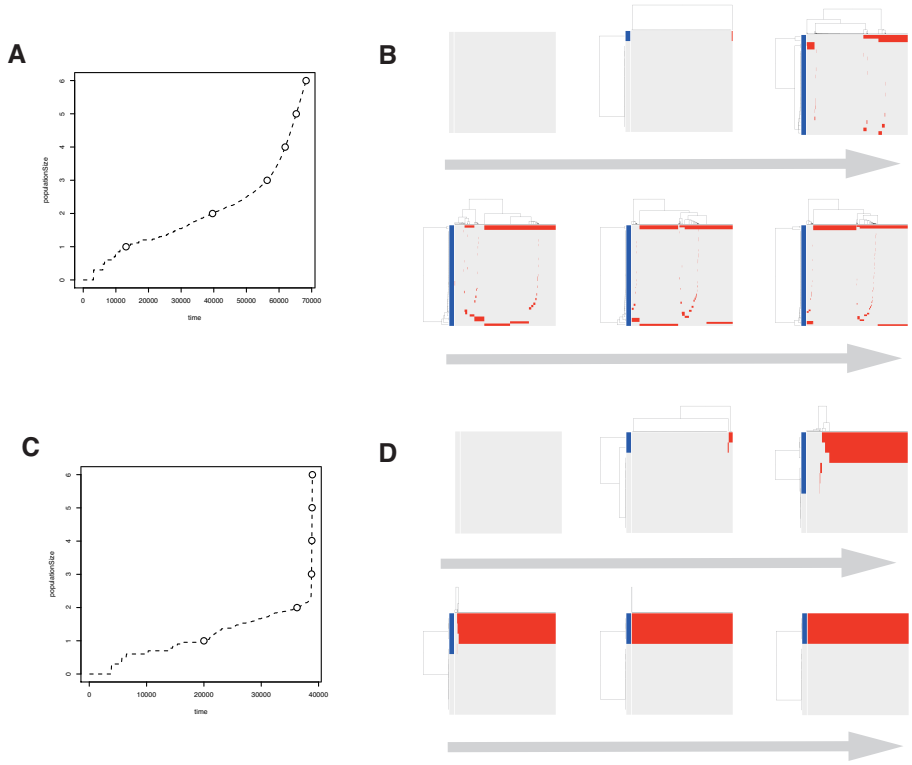

Fig 2. Time-course snapshots of simulations based on the driver model. Growth curve (A) and time-course snapshots of mutation profiles (B) simulated from the driver model with $N_d = 3$, $P = 10^6$, $f = 10^{0.3}$ and $m_d = 10^{-1.5}$ (corresponding to parameter setting i in Fig 1). Growth curve (C) and time-course snapshots of mutation profiles (D) simulated from the driver model with $N_d = 3$, $P = 10^6$, $f = 10^{0.9}$, and $m_d = 10^{-1.5}$ (corresponding to parameter setting ii in Fig 1). The time points when snapshots were obtained are indicated by empty circles on the growth curves.

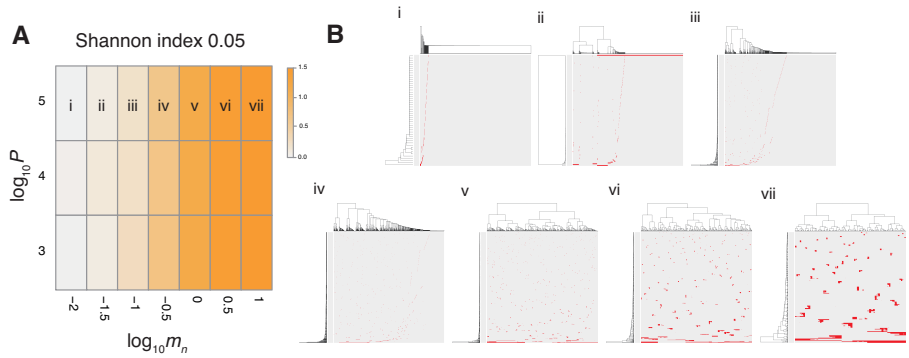

Fig 3. Sensitivity analysis of the neutral model. (A) Heap map obtained by calculating Shannon index 0.05 while changing the neutral mutation rate m_n and maximum population size P. (B) Single-cell mutations profiles obtained for seven parameter settings, which are indicated on the heat map in (A).

frequency cutoffs. As expected, the results indicate that the simulated tumor presents an increasing number of subclones as the frequency cutoff is lowered. The linearity of the log-log plot demonstrates that the power law is hidden in the mutation profile, consistently with its fractal-like pattern [15]. Note that, although the ITH score does not depend on population size P and fractal-like patterns shaped in the earliest stage appears to be subsequently unchanged in the time-course snapshots (Fig 4), these are also because low-frequency mutations were filtered out before visualization; the simulated tumor in fact expands neutral ITH by accumulating numerous low-frequency mutations as it grows.

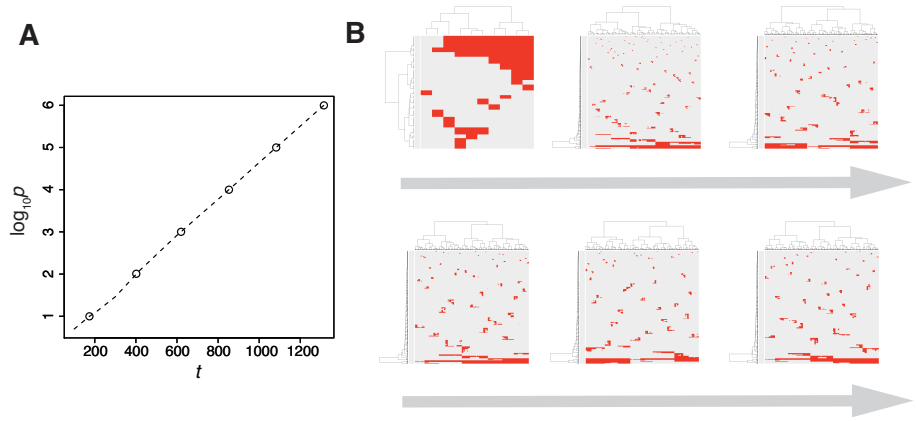

Fig 4. Time-course snapshots of simulations based on the neutral model. Growth curve (A) and time-course snapshots of mutation profiles (B) simulated from the driver model with $P = 10^6$ and $m_n = 10$ (corresponding to parameter setting vii in Fig 3). The time points when snapshots were obtained are indicated by empty circles on the growth curves.

Thus far, several theoretical and computational studies have shown that a stem cell hierarchy can boost neutral evolution [16, 17], which prompted us to extend the neutral model to the "neutral-s" model such that it contains a stem cell hierarchy (Fig S2). The neutral-s model assumes that two types of cell exist: stem and differentiated. Stem cells divide with a probability g_o without dying. For each cell division of stem cells, a symmetrical division generating two stem cells occurs with a probability s, while an asymmetrical division generating one stem cell and one differentiated cell occurs with a probability $1 - s$. A differentiated cell symmetrically divides to generate two differentiated cells with a probability g_0 but dies with a probability d_0^d. The means of accumulating neutral mutations in the two types of cell is the same as that in the original neutral model, which means that the neutral-s model is equal to the original neutral model when $s = 0$ or $d_0^d = 0$. For convenience, we defined $\delta = \log_{10}(d_0^d/g_0)$ and hereafter use δ instead of d_0^d.

The MASSIVE analysis of the neutral-s model confirmed that the incorporation of the stem cell hierarchy boosts neutral evolution (https://www.hgc.jp/~niiyan/canevosim/neutral_s). To obtain the heat map in Fig 5, the ITH score was measured while d_0^d and δ were changed, but $m_n = 0.1$ and $P = 1000$ were constantly set. In the heat map, a decrease of s leads to an increase in the ITH score when $\delta \geq 0$ (i.e., $d_0^d \geq g_0$). A smaller value of s means that more differentiated cells are generated per stem cell division, and $\delta \geq 0$ means that the population of the differentiated cells cannot grow in total, which is a valid assumption for typical stem cell hierarchy models. That is, this observation indicates that the stem cell hierarchy can induce neutral ITH even with a relatively low mutation rate setting

(i.e., $m_n = 0.1$), with which the original neutral model cannot generate neutral ITH.

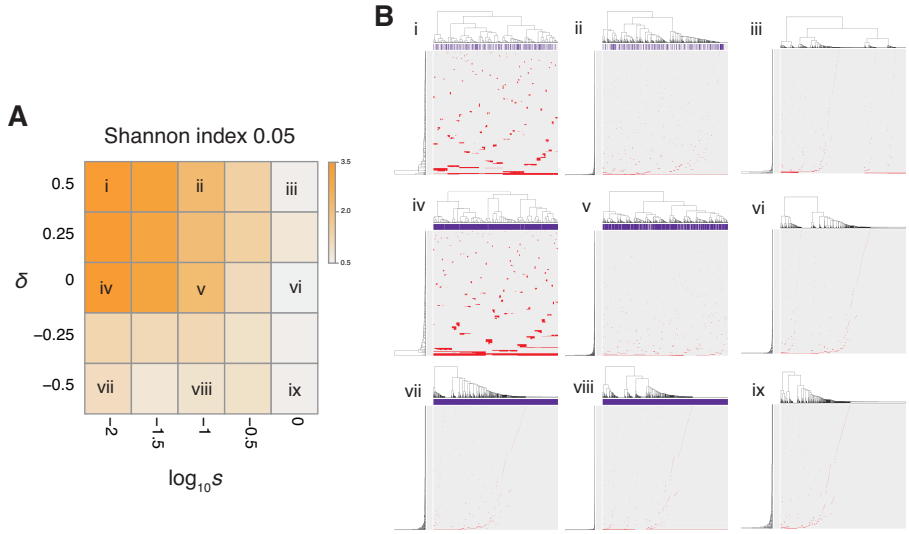

Fig 5. Sensitivity analysis of the neutral-s model. (A) Heat map obtained by calculating Shannon index 0.05 while changing the relative death rate of differentiated cells $\delta = \log_{10}(d_0^d/g_0)$ and the asymmetrical division rate s. The neutral mutation rate m_n and maximum population size P set to 10^{-1} and 10^5, respectively. (D) Single-cell mutation profiles obtained for nine parameter settings, indicated on the heat map presented in (A).

The underlying mechanism boosting neutral evolution can be explained as follows. We here consider only stem cells for an approximation, because differentiated cells do not contribute to tumor growth with $\delta \geq 0$. While one cell grows to a population of P cells, let cell divisions synchronously occur across x generations during the clonal expansion. Then, $(1+s)^x = P$ holds, because the mean number of stem cells generated per cell division is estimated as $1+s$. Solving the equation for x gives $x = \log P/\log(1+s)$; That is, it can be estimated that, during the clonal expansion, each of the P cells experiences $\log P/\log(1+s)$ cell divisions and accumulates $m_n \log P/2\log(1+s)$ mutations on average. We confirmed that the expected mutation count based on this formula fits the values observed in our simulation well, except in the exceptional cases where the mutation counts reached the upper limit, $N_n = 1000$ (Fig S3). These arguments mean that a tumor with a stem cell hierarchy accumulates more mutations until reaching a fixed population size than does a tumor without a stem cell hierarchy; that is, a stem cell hierarchy increases the apparent mutation rate by $\log 2/\log(1+s)$ folds, which induces neutral evolution neutral ITH even with relatively low mutation rate settings.

Combining the driver and neutral model

We now present the "composite" model that was constructed by combining the driver and neutral model to reproduce ITH more similar to those in real tumors. In a unit time, a cell divides into two daughter cells with a constant probability g without dying. In each cell division, each of the two daughter cells acquires $k_d \sim \text{Pois}(m_d/2)$ driver mutations and $k_n \sim \text{Pois}(m_n/2)$ neutral mutations. For each type of mutation, N_d and N_n mutations can be accumulated at maximum. For a cell that has n_d $(= \sum k_d)$ mutations, cell division probability g is defined as $g = g_0 f^{n_d}$, where g_0 is a

base division probability. The simulation started from one cell without mutations and ended when the population size p reached P or time t reached T. As expected from the MASSIVE analyses of the driver and neutral model that were performed separately, our MASSIVE analysis of the composite model confirmed that, depending on the parameter setting, the behaviors of the composite model are roughly categorized into the following six modes (Fig 6; https://www.hgc.jp/~niiyan/canevosim/composite) :

- With small m_d and small m_n, i.e., with low driver and neutral mutation rates, no evolution involving driver and neutral mutations occurs.
- With large m_d, small m_n, and small f (i.e., with high driver and low neutral mutation rates, and weak driver mutations), branching evolution occurs while no neutral evolution occurs.
- With large m_d, small m_n, and large f (i.e., with high driver and low neutral mutation rates, and strong driver mutations), linear evolution occurs while no neutral evolution occurs.
- With small m_d and large m_n (i.e., with low driver and high neutral mutation rates), neutral evolution occurs while no evolution involving driver mutations occurs.
- With large m_d, large m_n, and small f (i.e., with low driver and high neutral mutation rates, and weak driver mutations), branching and neutral evolution mixedly occur.
- With large m_d, large m_n, and large f (i.e., with high driver and high neutral mutation rates, and strong driver mutations), linear and neutral evolution mixedly occur.

Note that, because tumors having high driver mutation rates must have high neutral mutation rates also, linear and branching mutations must in general be accompanied by neutral evolution. Therefore, the last three modes are supposed to constitute the process that can actually occur in real tumors.

Adding punctuated evolution

Previously, we analyzed multiregion sequencing data of advanced colorectal cancer and precancerous lesions jointly to determine the evolutionary principles generating ITH shift from branching to neutral evolution during colorectal tumorigenesis [8]. We also determined that the number of copy number alterations drastically increases during the progression from colorectal precancerous lesions to advanced colorectal cancer, which prompted us to suspect that punctuate evolution triggers the evolutionary shift from branching to neutral evolution. To examine this possibility, we additionally incorporated punctuated evolution in the composite model to build the "punctuated" model.

For the models considered thus far, we assumed that cells can infinitely grow without a decrease in their growth speed. However, it is more natural to assume that there exists a limit of population size because of the resource limitation and that the growth speed gradually slows down as the population size approaches the limit. The limit of population sizes is called the carrying capacity and employed in the logistic equation [18]. By mimicking the logistic equation, we modified the division probability as $g = g_0 f^{n_d}(1 - p/p_c)$, where p_c is the carrying capacity. To reproduce punctuated evolution, we additionally employ an "explosive" driver mutation, which negates the

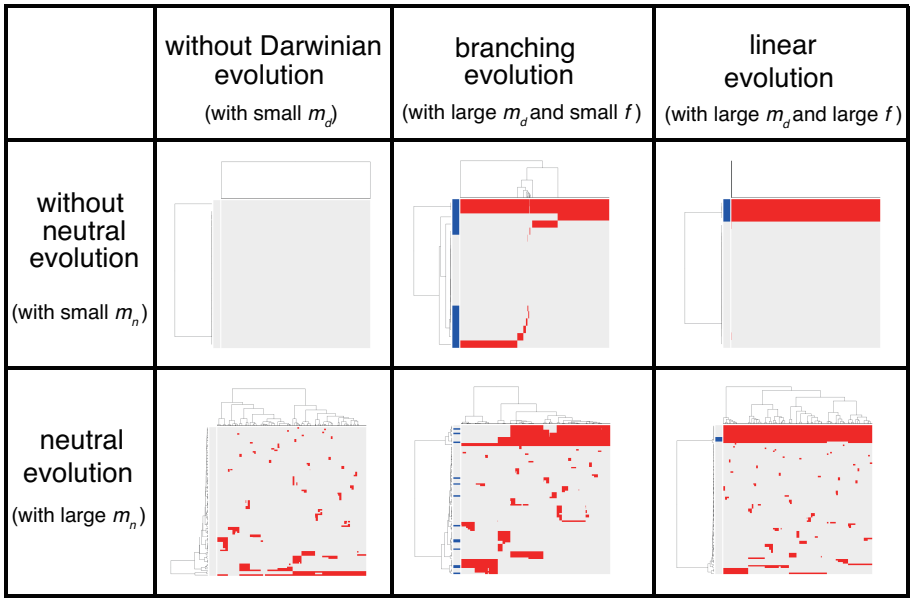

Fig 6. Six evolutionary modes simulated by the composite model. Our sensitivity analysis demonstrated that, depending on the parameter setting, the behaviors of the composite model are roughly categorized into the six modes. Representative mutation profiles of the six modes are presented.

effect of the carrying capacity. After a cell accumulates driver mutations up to the maximum N_d, the explosive driver mutation is introduced at a probability m_e. For a cell that has the explosive driver mutation, the carrying capacity p_c is set to infinite; That is, it is assumed that the explosive driver mutation rapidly evolve the cell so that it can conquer the growth limit and attain infinite proliferation ability.

Next, we searched for parameter settings that lead the punctuated model to reproduce punctuated evolution. The MASSIVE analysis confirmed that, with sufficiently large m_e, punctuated evolution is reproducible in the punctuated model (https://www.hgc.jp/~niiyan/canevosim/punctuated; note that, for simplicity, we omitted neutral mutations by setting $m_n = 0$ in the MASSIVE analysis). We also examined time-course snapshots of simulations conducted with these parameter settings. In the example shown in Fig 7 A and B, we observed that multiple subclones having different driver genes exist; that is, branching evolution, which mixedly occurs with neutral evolution, is prominent during the early phase of the simulation. Note that a growth curve plot indicates that, as the population size approaches the carrying capacity, the growth speed slows down; however, the tumor regrows after the appearance of a clone that has acquired an explosive driver mutation. The clone with the explosive driver mutation is then subjected to a selective sweep, which causes subclonal driver mutations in the clone to shift to clonal mutations. Then, only neutral mutations are accumulated as subclonal mutations; That is, ITH is finally generated by neutral evolution.

We also found that two subclones having different subclonal driver mutations sometimes appear by obtaining two independent explosive driver mutations (Fig 7 C and D). This observation recalls to mind the multiverse model, which was proposed for glioblastoma evolution [19]. The multiverse model is derived from the Big-Bang model, a model for jointly describing punctuated and neutral evolution during colorectal tumorigenesis [6]. The Big-Bang model assumes that a single clone

explosively expands from a precancerous lesion while generating neutral ITH, consistently with our evolutionary shift model. However, in the multiverse model, it is assumed that multiple subclones are subject to explosive expansion. Collectively, our simulation based on the punctuated model not only supports our hypothesis that punctuated evolution underlies the evolutionary shift during colorectal tumorigenesis, but also can reproduce multiple types of punctuated models proposed thus far.

Discussion

In this paper, we introduce a series of simulation models that reproduce the four types of cancer evolutionary modes: linear, branching, neutral, and punctuated. Our sensitivity analysis of these models provided a number of insights into cancer evolution. For example, under the assumption of sufficiently high mutation rates, branching evolution occurs with strong driver mutations, whereas linear evolution occurs with weak driver mutations. However, a major concern about our sensitivity analysis is whether the ranges of parameter values examined is realistic. Although dependent on tumor types, the number of driver mutations were previously estimated as in the low single digits for most tumor types, consistently with our settings for d. As the increase in the cell division probability per driver mutation f, which is interpreted as the strength of driver mutations, we examined values ranging from $10^{0.1}$ to $10^{1.0}$. Although the value of f has not been the subject to extensive experimental determination, it has been reported that induction of K-ras$^{\mathrm{G12D}}$ in murine small intestine increases growth rate from one cell cycle per 24 hr to one cell cycle per 15 hr, from which f is estimated as $10^{0.204}$ [20].

The driver mutation rate m_d and population size P appear to be problematic. Although the driver mutation rate was previously estimated as $\sim 3.4 \times 10^{-5}$ per cell division [21], our sensitivity analysis examined values from 10^{-4} to 10^{-1}, which are below the estimated value by orders of magnitude. It should also be noted that, in our simulation, it was assumed that a tumor contains 10^6 cells at maximum, whereas the number of cancer cells in one gram of tumor tissue is reportedly 10^9 or one order less [22]. Clearly, for m_d and P, the parameter space we examined does not cover those in a real tumor. However, the results of the MASSIVE analysis allow the behaviors of the driver model to be envisioned in a realistic parameter space. When P is small, neither linear nor branching evolution occurs. As P increases, we observe linear or branching evolution with smaller m_d, although the range of f that leads to branching evolution shifts to larger values. Moreover, as shown by the sensitivity analysis of the neutral-s model, the presence of a stem cell hierarchy increases the apparent mutation rate. Therefore, a real tumor having a a stem cell hierarchy apparently should have a higher m_d value. Collectively, it is natural to assume that a real tumor having large P and small m_d can be similarly generated by linear or branching evolution, although, in such cases, the actual value of f might be larger than those that we examined.

The sensitivity analysis of the neutral model showed that neutral ITH is generated if the expected number of driver mutations generated per cell division, m_n, exceeds 1. In a recent report, the estimated somatic mutation rate was given as 2.66×10^{-9} mutations per base pair per mitosis. Given that most mutations are neutral on the human genome comprised of 3×10^9 bases, even a cell division of normal cells generates more than 1 mutation. As cancer cells should have higher mutation rates, which can be further accelerated by stem cell hierarchies, it is reasonable to assume that a tumor in general satisfies the conditions to generate neutral ITH. However, not every tumor necessarily has neutral ITH; neutral ITH is distorted by natural selection if the tumor additionally satisfies the conditions for branching evolution, as shown by the analysis employing the composite model.

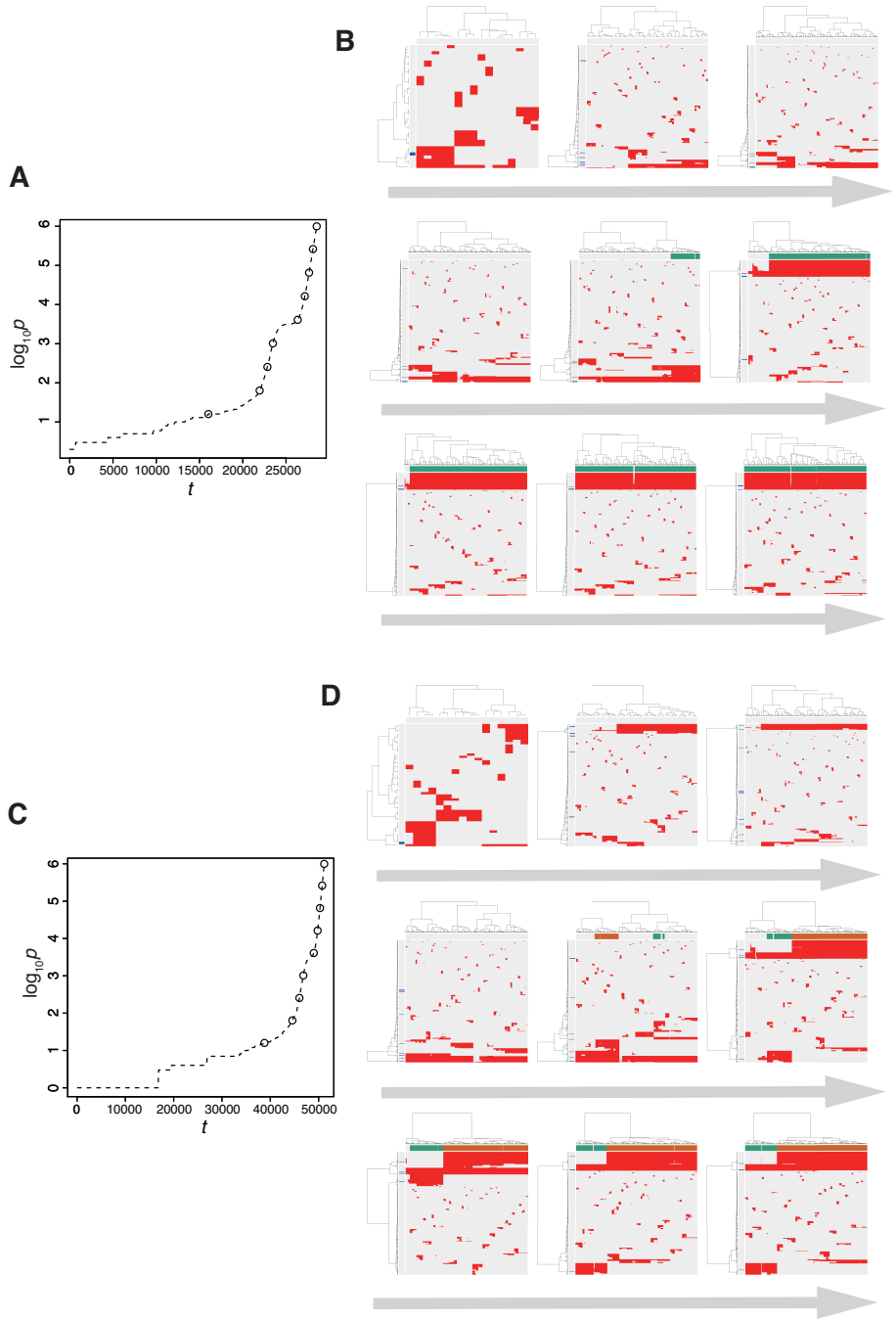

Fig 7. Time-course snapshots of simulations based on the punctuated model. Growth curve (A) and time-course snapshots of mutation profiles (B) simulated from the punctuated model with $P = 10^6$, $P_c = 10^{3.5}$, $m_d = 10^{-1}$, $m_p = 10^{0.5}$, and $m_e = 10^{-4}$. Growth curve (C) and time-course snapshots of mutation profiles (D) simulated from the punctuated model with $P = 10^6$, $P_c = 10^{3.5}$, $m_d = 10^{-1}$, $m_p = 10^{0.5}$, and $m_e = 10^{-3}$. The time points when snapshots were obtained are indicated by empty circles on the growth curves.

A highlight of this work is that the punctuated model demonstrated that punctuated evolution triggers the evolutionary shift from branching to neutral evolution. For carrying capacity p_c and the probability of acquiring an explosive mutation m_e in the punctuated model, the parameter values that we examined are clearly outside realistic ranges. Similarly to P, p_c should take a larger value. Although it cannot easily to be experimentally determined, m_e also appears to be overestimated; although the human body in fact potentially harbors numerous precancerous lesions, which are assumed not to have acquired explosive driver mutations yet, only a tiny fraction of cases progresses to advanced stages by acquiring explosive driver mutations. However, it is intuitively understandable that the behaviors of the punctuated model, as well as of the driver model, are not dependent on precise values of these parameters, and in our opinion our analysis is sufficient to provide a semi-quantitative understanding of cancer evolution.

The models we introduced in this paper can be described collectively as the unified model, a formal description of which is provided in the method section. The unified model is very simple but sufficient to reproduce linear, branching, neutral, and punctuated evolution. Of course, the unified model harbors many limitations, which should be addressed in future studies. Our current version of the model completely ignores spatial information, which potentially influences evolutionary dynamics. Recently reported studies have shown that spatial structures regulate evolutionary modes in tumors [23, 24]. We also determined that resource bias prompts branching evolution, by simulating tumor growth on a one-dimensional lattice [14]. Although our model assumed that driver mutations independently have effects of equal strength, different driver mutations should have different strengths and might work synergistically [25]. Although we assumed that punctuated evolution occurs only once in the course of cancer evolution, it is possible that a tumor is confronted with different types of resource limitations during progressions and undergoes punctuated evolution multiple times to conquer them [26]. However, in spite of these imitations, the MASSIVE analyses of the models encompassed by the unified model have successfully provided a number of insights into cancer evolutionary dynamics. In our opinion the unified model serves as a starting point for constructing more realistic simulation models to understand in greater depth the diversity of cancer evolution, which is being unveiled by the evergrowing amount of cancer genomics data.

Methods

Simulation model

Although we described a series of simulation models in the main text, we here reformulate the unified model, which encompasses these models. Starting from a stem cell without mutations, the following time steps are repeated until the number of population size p reaches P or the number of time steps t reaches T. For each time step, each cell is subject to cell division with a probability g and cell death with a probability d. g depends on a base division rate g_0, the increase in the cell division probability per driver mutation f, the number of driver mutations accumulated in the cell n_d, population size p, and the carrying capacity p_c: $g = g_0 f^{n_d}(1 - p/p_c)$. d depends on the base death rate d_0, the decrease in the cell death probability per driver mutation, and the number of driver mutations accumulated in the cell n_d: $d = d_0 e^{-n_d}$. When the cell is a differentiated cell, d_0 is replaced by d_0^d, which is the base death rate for differentiated cells: $d = d_0^d e^{-n_d}$. The order of the trials of cell division and death is flipped with probability 0.5. We also assumed that cell death occurs only in the case where $p > 1$, to prevent the simulation from halting before clonal expansion.

In a cell division, the cell is replicated into two daughter cells. If the parent cell is a stem cell, one of the two daughter cells is differentiated with a probability $1 - s$; that is, s expresses the probability of symmetric division. For each of the two daughter cells, we introduce k_d driver and k_n neutral mutations. k_d and k_n are sampled from Poison distributions, the parameters of which are $m_d/2$ and $m_n/2$, respectively: $k_d \sim \text{Pois}(m_d/2)$ and $k_n \sim \text{Pois}(m_n/2)$. Note that this means that each cell division generates m_d driver and m_n neutral mutations on average. We assumed each mutation acquired by different division events occurs at different genomic positions and each cell can accumulate N_d driver and N_n neutral mutations at maximum. When each of the two daughter cells has N_d driver mutations, we further attempted to introduce an explosive driver mutation; the explosive driver mutation is introduced with a probability m_e and sets the carrying capacity p_c of the cell to infinite. The pseudocode for the unified model is provided as Algorithm 1. The variables and parameters employed in the unified model are listed in Tables 1 and 2.

Table 1. Variables

Symbol	Description
k_d	Number of driver mutations obtained in a cell division
n_d	Number of driver mutations accumulated in a cell
k_n	Number of neutral mutations obtained in a cell division
p	Population size
t	Number of time steps
g	Cell division probability
d	Cell death probability

Table 2. Parameters

Symbol	Description
m_d	Expected number of driver mutations generated per cell division
m_n	Expected number of neutral mutations generated per cell division
m_e	Probability of acquiring an explosive mutation
N_d	Maximum number of driver mutations accumulated in a cell
N_n	Maximum number of neutral mutations accumulated in a cell
f	Increase of the cell division probability per driver mutation
e	Decrease of the cell death probability per driver mutation
g_0	Base cell division probability
d_0	Base cell death probability for stem cells
d_0^d	Base cell death probability for differentiated cells
s	Symmetric division probability
p_c	Population capacity
P	Maximum population size
T	Maximum number of time steps

Post-processing of simulation results

To evaluate the simulation results quantitatively, we calculated summary statistics based on 1000 cells randomly sampled from each simulated tumor. these summary statistics are listed in Table 3. time and population size indicate the numbers of time steps and cells, respectively, when the simulation is complete. mutation count per cell represents the mean number of mutations accumulated in each of the randomly

Algorithm 1 Unified model

```
prepare a stem cell without mutations
while p < P or t < T do
    for each cell do
        g = g_0 f^{n_d} (1 - p/p_c)
        d = d_0 e^{-n_d}
        if the cell is a differentiated cell then
            d = d_0^d e^{-n_d}
        if rand < 0.5 then
            if rand() < g then
                divide(the cell)
            else if rand() < d then
                if p > 1 then
                    kill the cell (accordingly, p = p - 1)
        else
            if rand() < d then
                if p > 1 then
                    kill the cell (accordingly, p = p - 1)
            else if rand() < g then
                divide(the cell)
    t = t + 1


function rand()
    return a random number ranging from 0 to 1

function divide(a cell)
    replicate the cell into two daughter cells (accordingly, p = p + 1)
    if the parent cell is a stem cell then
        if rand() > s then
            differentiate one of the two daughter cells
    for each of the two daughter cells do
        introduce k_d ~ Pois(m_d/2) driver mutations
        introduce k_n ~ Pois(m_n/2) neutral mutations
        if n_d = Σ k_d reaches the upper limit N_d then
            if rand() < m_e then
                set p_c of the cell to infinite
```

sampled 1000 cells. By combining the mutations of the 1000 cells, we defined the mutations that occur in 0.95% or more of the 1000 cells as clonal mutations, and the others as subclonal mutations. The numbers of clonal, subclonal, and both types of mutations were then defined as clonal mutation count, subclonal mutation count, and total mutation count, from which clonal mutation proportion and subclonal mutation proportion were further calculated. The degree of ITH was also measured by Shannon and Simpson indices, which were calculated based on the proportions of different subclones (i.e., cell subpopulations with different mutations) after removing mutations having a frequency less of than 5% or 10%: Shannon index 0.05, Shannon index 0.1, Simpson index 0.05, and Simpson index 0.1. Similarly, after removing mutations having a frequency of less than 5% or 10%, we also checked whether multiple subclones harboring different driver mutations coexist, which is represented as binary statistics, branching evolution 0.05, and branching evolution 0.1. When the simulated tumor had differentiated cells or subclones with explosive driver mutations, the proportion of the subpopulation was calculated as subpopulation proportion .

The single-cell mutation profiles of the 1000 cells are represented as a binary matrix, the row and column indices of which are mutations and samples, respectively. To interpret the simulation results intuitively, we also visualized the binary matrix by utilizing the heatmap function in R after the following pre-processing, if necessary. When the number of rows was less than 10, empty rows were inserted in the matrix so that the number of rows was 10. When the number of rows was more than 300, we extracted the 300 rows with the highest mutation occurrence so that the number of rows was 300. In the neutral and neutral-s models, we exceptionally set the maximum row number to 1000 in order to visualize low-frequency mutations. The visualized matrix is accompanied by a left-side blue bar indicating the driver mutations. When the simulated tumor had differentiated cells or subclones with explosive driver mutations, the subpopulation is indicated by the purple bar on the top of the visualized matrix.

Sensitivity analysis based on MASSIVE

To cover a sufficiently large parameter space in the sensitivity analysis, we employed a supercomputer, SHIROKANE4 (at Human Genome Center, The Institute of Medical Science, The University of Tokyo). The simulation and post-processing steps for different parameter settings were parallelized on Univa Grid Engine. For each model, we employed a full factorial design involving four parameters (i.e, we tested every combination of candidate values of the four parameters) while other parameters were fixed. The parameter values used for our analysis are listed in Table 2. For each parameter setting, 50 Monte Carlo trials were performed and the summary statistics were averaged over the 50 trials. The averaged summary statistics calculated for each parameter setting were visualized by interactive heat maps on a web-based visualization tool, the MASSIVE viewer. The MASSIVE viewer also displays single-cell mutation profiles from 5 of the 50 trial with the same parameter setting. For details, please refer to our methodological report [14]. All the results in this study can be interactively explored in the MASSIVE viewer on our website (https://www.hgc.jp/~niiyan/canevosim).

Table 3. Summary statistics

Name	Description
time	Number of time steps when simulation is finished
population size	Number of cells when simulation is finished
mutation count per cell	Mean number of mutations accumulated in each cell
clonal mutation count	Number of clonal mutations
subclonal mutation count	Number of subclonal mutations
total mutation count	clonal mutation count + subclonal mutation count
clonal mutation proportion	clonal mutation count / total mutation count
subclonal mutation proportion	subclonal mutation count / total mutation count
Shannon index 0.1	Shannon index calculated with a mutation frequency cutoff of 0.1
Shannon index 0.05	Shannon index calculated with a mutation frequency cutoff of 0.05
Simpson index 0.1	Simpson index calculated with a mutation frequency cutoff of 0.1
Simpson index 0.05	Simpson index calculated with a mutation frequency cutoff of 0.05
branching evolution 0.05	Binary statistic indicating that multiple subclones harboring different driver mutations coexist, calculated with a mutation frequency cutoff of 0.05
branching evolution 0.1	Binary statistic indicating that multiple subclones harboring different driver mutations coexist, calculated with a mutation frequency cutoff of 0.1
subpopulation proportion	proportion of differentiated cells or subclones with explosive driver mutations

Supporting information

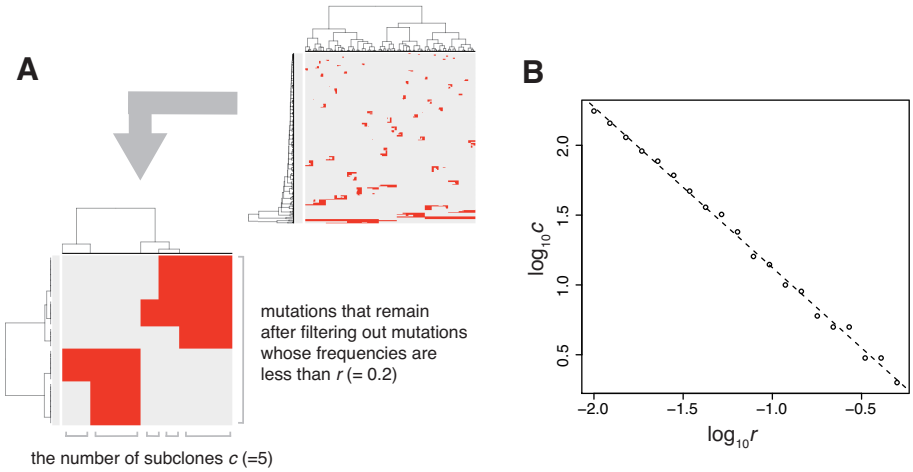

Fig S1. Self-similarity of neutral ITH. (A) Illustrative explanation of the preparation of the log-log plot presented in (B). After mutations having frequencies less than r are filtered out, the number of subclones c is counted based on the mutation profiles. (B) Log-log plot for r and c obtained from a simulation with $P = 10^5$ and $m_n = 10$. Similar linearity holds when $m_n \geq 1$.

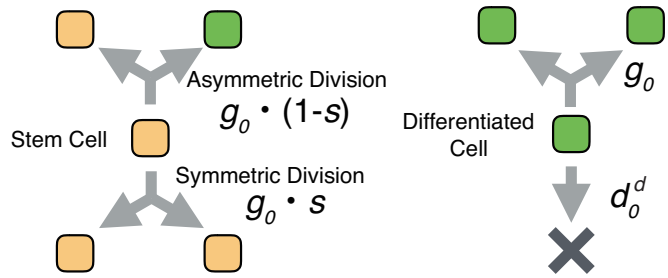

Fig S2. Schema of the neutral-s model. Stem cells divide with a probability g_o without dying. For each cell division of stem cells, a symmetrical division generating two stem cells occurs with probability s, while an asymmetrical division generating one stem cell and one differentiated cell occurs with probability $1 - s$. A differentiated cell symmetrically divides to generate two differentiated cells with probability g_0 but dies with probability d_0^d.

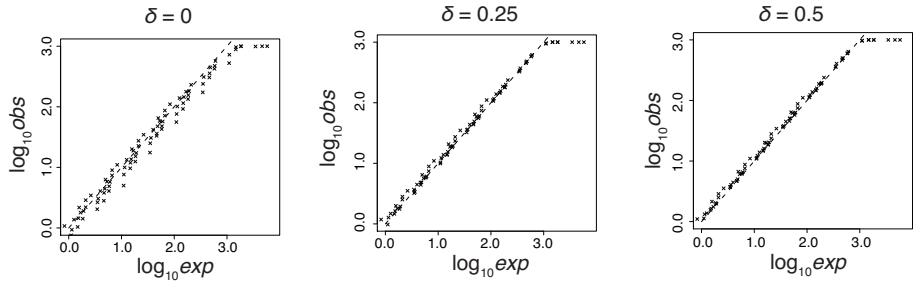

Fig S3. Observed and expected mutation counts from the neutral-s model. The observed mutation counts (obs) were prepared from values of mutation count per cell in the MASSIVE analysis, while the expected mutation counts (exp) were analytically estimated as $m_n \log P/2 \log(1 + s)$ under the assumption that $\delta \geq 0$. Each cross representing each parameter setting was plotted in log10 scale for different values of δ. Positioning on the dashed line indicates the equality of the observed and expected mutation counts.

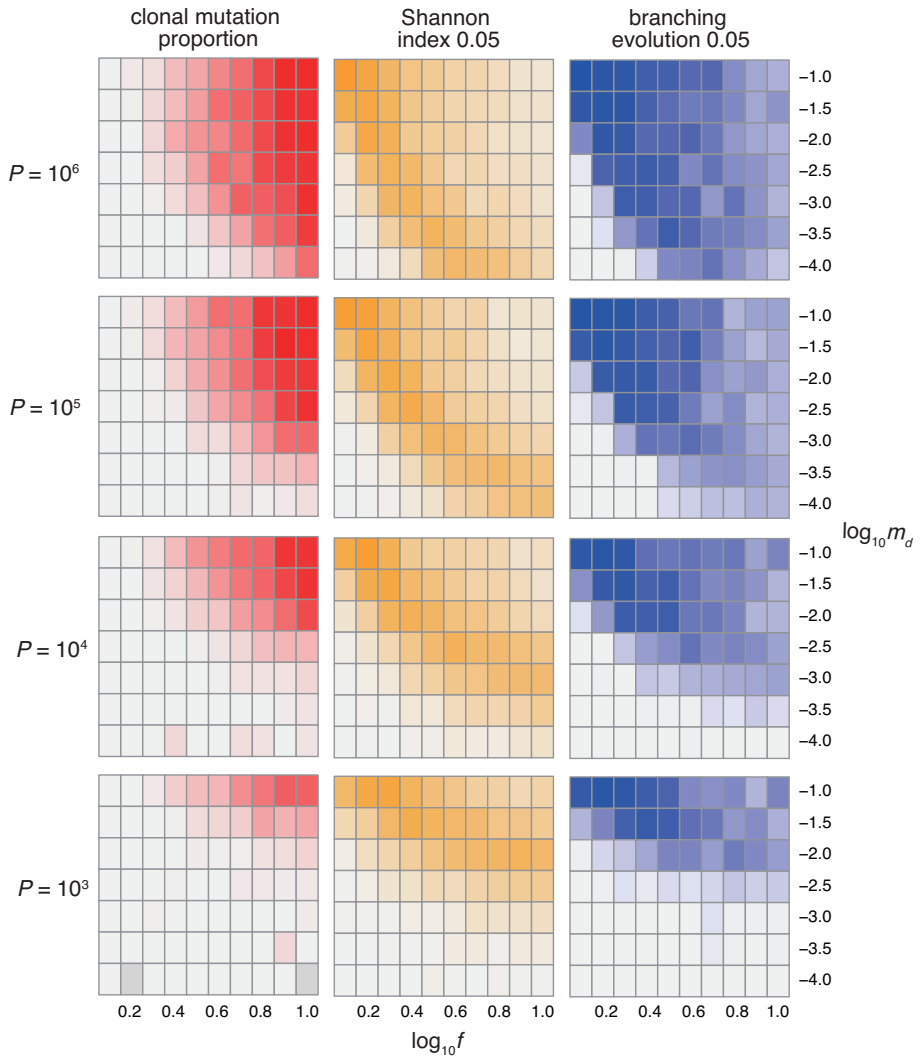

Fig S4. Sensitivity analysis of the driver model. While changing the driver mutation rate m_d, the strength of driver mutations f, and maximum population size P, heat maps of summary statistics were prepared for the proportion of clonal mutations, clonal mutation proportion (A), a measure for ITH, Shannon index 0.05 (B) and an indicator for the occurrence of branching evolution, branching evolution 0.05 (C). N_d was set to 3.

Acknowledgment

This research was supported by AMED under Grant Number JP17cm0106504h0002; MEXT as Priority Issue on Post-K computer (Supercomputer Fugaku) (Project ID: hp170227); and JSPS as Grant-in-Aid for Scientific Research on Innovative Areas (15H0912). We thank Hiroshi Haeno, Hideki innan, and Watal M. Iwasaki for helpful discussions.

References

1. Davis A, Gao R, Navin N. Tumor evolution: Linear, branching, neutral or punctuated? Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 2017;1867(2):151–161.

2. Niida A, Nagayama S, Miyano S, Mimori K. Understanding intratumor heterogeneity by combining genome analysis and mathematical modeling. Cancer science. 2018;109(4):884–892.

3. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. Cell. 2018;173(3):595–610.

4. Suzuki H, Aoki K, Chiba K, Sato Y, Shiozawa Y, Shiraishi Y, et al. Mutational landscape and clonal architecture in grade II and III gliomas. Nature genetics. 2015;47(5):458.

5. Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, et al. Integrated multiregional analysis proposing a new model of colorectal cancer evolution. PLoS genetics. 2016;12(2):e1005778.

6. Sun R, Hu Z, Curtis C. Big Bang tumor growth and clonal evolution. Cold Spring Harbor perspectives in medicine. 2018;8(5):a028381.

7. 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):E6496–E6505.

8. Saito T, Niida A, Uchi R, Hirata H, Komatsu H, Sakimura S, et al. A temporal shift of the evolutionary principle shaping intratumor heterogeneity in colorectal cancer. Nature communications. 2018;9(1):2884.

9. Cross W, Kovac M, Mustonen V, Temko D, Davis H, Baker AM, et al. The evolutionary landscape of colorectal tumorigenesis. Nature ecology & evolution. 2018;2(10):1661.

10. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. Nature genetics. 2016;48(10):1119.

11. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013;153(3):666–677.

12. Beerenwinkel N, Schwarz RF, Gerstung M, Markowetz F. Cancer evolution: mathematical models and computational inference. Systematic biology. 2014;64(1):e1–e25.

13. Altrock PM, Liu LL, Michor F. The mathematics of cancer: integrating quantitative models. Nature Reviews Cancer. 2015;15(12):730.

14. Niida A, Hasegawa T, Miyano S. Sensitivity analysis of agent-based simulation utilizing massively parallel computation and interactive data visualization. PloS one. 2019;14(3):e0210678.

15. Brown JH, Gupta VK, Li BL, Milne BT, Restrepo C, West GB. The fractal nature of nature: power laws, ecological complexity and biodiversity. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2002;357(1421):619–626.

16. Sottoriva A, Verhoeff JJ, Borovski T, McWeeney SK, Naumov L, Medema JP, et al. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. Cancer research. 2010;70(1):46–56.

17. Solé RV, Rodríguez-Caso C, Deisboeck TS, Saldaña J. Cancer stem cells as the engine of unstable tumor progression. Journal of theoretical biology. 2008;253(4):629–637.

18. Verhulst PF. Notice sur la loi que la population suit dans son accroissement. Corresp Math Phys. 1838;10:113–126.

19. Lee JK, Wang J, Sa JK, Ladewig E, Lee HO, Lee IH, et al. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. Nature genetics. 2017;49(4):594.

20. Snippert HJ, Schepers AG, van Es JH, Simons BD, Clevers H. Biased competition between Lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion. EMBO reports. 2014;15(1):62–69.

21. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. Proceedings of the National Academy of Sciences. 2010;107(43):18545–18550.

22. Del Monte U. Does the cell number 109 still really fit one gram of tumor tissue? Cell Cycle. 2009;8(3):505–506.

23. Noble R, Burri D, Kather JN, Beerenwinkel N. Spatial structure governs the mode of tumour evolution. bioRxiv. 2019; p. 586735.

24. West J, Schenck R, Gatenbee C, Robertson-Tessi M, Anderson AR. Tissue structure accelerates evolution: premalignant sweeps precede neutral expansion. bioRxiv. 2019; p. 542019.

25. Castro-Giner F, Ratcliffe P, Tomlinson I. The mini-driver model of polygenic cancer evolution. Nature Reviews Cancer. 2015;15(11):680.

26. Aktipis CA, Boddy AM, Gatenby RA, Brown JS, Maley CC. Life history trade-offs in cancer evolution. Nature Reviews Cancer. 2013;13(12):883.