We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Agent-Based Modeling and Analysis of Cancer Evolution

Atsushi Niida and Watal M. Iwasaki

Abstract

Before the development of the next-generation sequencing (NGS) technology, carcinogenesis was regarded as a linear evolutionary process, driven by repeated acquisition of multiple driver mutations and Darwinian selection. However, recent cancer genome analyses employing NGS revealed the heterogeneity of mutations in the tumor, which is known as intratumor heterogeneity (ITH) and generated by branching evolution of cancer cells. In this chapter, we introduce a simulation modeling approach useful for understanding cancer evolution and ITH. We first describe agent-based modeling for simulating branching evolution of cancer cells. We next demonstrate how to fit an agent-based model to observational data from cancer genome analyses, employing approximate Bayesian computation (ABC). Finally, we explain how to characterize the dynamics of the simulation model through sensitivity analysis. We not only explain the methodologies, but also introduce exemplifying applications. For example, simulation modeling of cancer evolution demonstrated that ITH in colorectal cancer is generated by neutral evolution, which is caused by a high mutation rate and stem cell hierarchy. For cancer genome analyses, new experimental technologies are actively being developed; these will unveil various aspects of cancer evolution when combined with the simulation modeling approach.

Keywords: cancer, evolution, agent-based model, approximate Bayesian computation, sensitivity analysis

1. Introduction

Cancer is a clump of abnormal cells that originates from normal cells. Normal cells proliferate or stop proliferating depending on their surrounding environment. For example, when skin cells are injured, they proliferate to cover the wound; however, when the wound heals, they stop proliferating. In contrast, cancer cells continue proliferating by ignoring the surrounding environment. Moreover, cancer cells invade surrounding tissues, metastasize to distant organs, and impair functions in the human body.

Malignant transformation from normal to cancer cells generally results from the accumulation of somatic mutations, which are induced by various causes such as aging, ultraviolet rays, cigarette, alcohol, chemical carcinogens, etc. Mutations that contribute to malignant transformation are known as “driver mutations”, whereas genes whose function are impaired by driver mutations are named as “driver genes”. There are two types of driver genes are categorized: “oncogenes” and

"tumor suppressor genes". Oncogenes act as gas pedals for cell proliferation, which are constitutively turned on by driver mutations. Tumor suppressor genes act as brakes to stop cell proliferation, and inhibiting the function of the brakes is necessary for malignant transformation.

Normal cells are transformed into cancer cell when two to 10 driver mutations are acquired. Because these mutations are not induced simultaneously, but rather gradually over a long period of time, this process is known as "multi-stage carcinogenesis" [1]. This process is also regarded as a linear evolutionary process, driven by repeated acquisition of multiple driver mutations and Darwinian selection. Understanding cancer from an evolutionary perspective is important, as therapeutic difficulties against cancer originate from the high evolutionary capacity, which easily endows cancer cells with therapeutic resistance.

Mutations in cancer cells are experimentally detected by DNA sequencing. Next-generation sequencing (NGS) technology, which raised around 2010, enabled cancer genome analysis to comprehensively detect mutations in cancer cells. During the last decade, cancer genome analysis has revolutionized our understanding of cancer [2]. Cancer genome analysis showed that cancer cells harbor a large number of mutations, only a small fraction of which is driver mutations; namely, most mutations in cancer cells are "neutral mutations", which have no selective advantages (also referred to as "passenger mutations" in paired with driver mutations). By sequencing hundreds of tumor samples from different patients with the same cancer type, the repertories of driver genes were also determined across various types of cancer. Moreover, cancer genome analysis has revealed heterogeneity of mutations within one tumor, which is termed intratumor heterogeneity (ITH) [3]. As described above, carcinogenesis was regarded as a linear evolutionary process until the arrival of NGS; however, ITH is actually generated by branching evolution of cancer cells.

However, cancer genome analysis is not sufficient to explain the origin of ITH. To understand the evolutionary principles underlying the generation of ITH, a simulation modeling approach is useful and increasingly employed in the field of cancer research. In this chapter, we introduce such simulation modeling approaches. We first describe agent-based modeling for simulating branching evolution of cancer cells. We next demonstrate how to fit an agent-based model to observational data obtained by cancer genome analyses, employing approximate Bayesian computation (ABC). Finally, we explain how to characterize the dynamics of the simulation models through sensitivity analysis.

2. Agent-based modeling of cancer evolution

To simulate heterogenous cancer evolution, agent-based modeling is widely employed. An agent-based model assumes a set of system constituents, known as independent agents, and specifies rules for the independent behavior of the agents themselves, as well as for the interactions between agents and the agent environment [4]. The agent-based model is a flexible representation of the model, and given the initial conditions and parameters of the system, the behavior of the system can be easily analyzed by computational simulation. For modeling of cancer evolution, if each cell is assumed to be an agent, ITH can be easily represented by the differences in the internal states of each agent. As an example, we explain an agent-based model named as the branching evolutionary process (BEP) model, which was originally introduced by Uchi *et al.* [5] to studying ITH of colorectal cancer (**Figure 1A**). In the BEP model, a cell assumed to be an agent has a genome containing n genes, each of which is represented as a binary value, 0 (wild-type) or

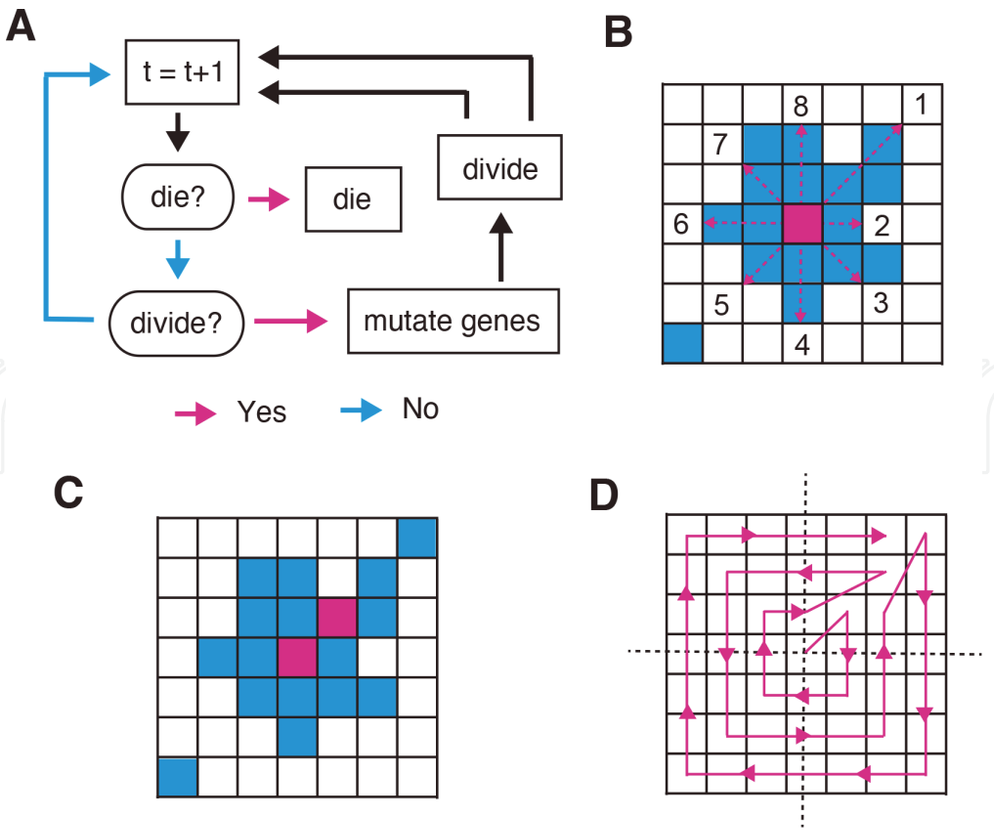


Figure 1.
Illustration of the BEP model. (A) A flowchart of a simulation based on the BEP model. First, a cell is tested for survival and killed with a provability q. Next, the cell is tested for cell division and replicated with a provability p. Before cell division, each gene in the cell is mutated with a provability r. After this process is applied to each cell, the simulation proceeds to the next time step. (B–D) illustration of division operation (see the main text for details). This image originally appeared in [5].

1 (mutated). Thus, the genome is represented as a binary vector **g** with length n. In a unit time step, a cell replicates with a probability p and dies with a probability q. When the cell replicates, a wild-type gene is mutated with a probability r. The first d genes in **g** are considered as driver genes, whose mutations accelerate replication. A normal cell without mutations has a replication probability p_0, and each driver mutation increases p by 10^f-fold; i.e., $p = p_0 \cdot 10^{fk}$, where $k = \sum_{i=1}^{d} g_i$, the number of mutated driver genes. The death probability is fixed as $q = q_0$. Let c and t denote the size of the simulated cell population and number of the time steps, respectively. A simulation is started with c_0 normal cells and the unit time step is repeated while the population size $c \leq c_{max}$ and time step $t \leq t_{max}$.

The BEP model assumes that a simulated tumor grows in a two-dimensional square lattice where each cell occupies one lattice point. Initially, c_0 cells are initialized as close as possible to the center of the lattice. In a unit time step, along an outward spiral starting from the center, we replicate and kill each cell with probabilities p and q, respectively. When cells replicate, the BEP model places the daughter cell in the neighborhood of the parent cell, assuming a Moore neighborhood (i.e., eight points surrounding a central point). If empty neighbor points exist, we randomly select one of these points. Otherwise, we create an empty point in any of the eight neighboring points as follows. First, for each of the eight directions, we count the number of consecutive occupied points that range from each neighboring point to immediately before the nearest empty cell as indicated in **Figure 1B**. Next, any of the 8 directions is randomly selected proportionally with $1/l_i$, where $l_i (1 \leq i \leq 8)$ is the count of the consecutive occupied points for each direction. The consecutive occupied points in the selected direction are then shifted by one point

so that an empty neighboring point appears as shown in **Figure 1C**. Note that simulation results depend on the order of the division operation in the two-dimensional square lattice. The BEP model first marks cells to be divided and then applies the division operation to the marked cells along an outward spiral starting from the center. In each round on the spiral, the direction is randomly flipped in order to maintain spatial symmetry. An example of such spirals is shown in **Figure 1D**.

Given that a cell without mutations divides according to this rule, after a normal cell acquires its first driver mutation, which accelerates cell division, the proportion of the clone originating from the cell increases in the whole cell population. By repeating these steps, each cell gradually accumulates driver mutations and accompanying passenger mutations, which do not affect the cell division rate, finally forming a tumor with many mutations. Depending on the parameter values during the course of cancer evolution, each cancer cell can accumulate different combinations of mutations to generate different types of ITH. **Figure 2** show an example of snapshots of two-dimensional tumor growth simulated based on the BEP model with an appropriate parameter setting. In this example, driver mutations gradually accumulated in the cells, and a clone with four mutations was selected through Darwinian selection and finally became dominant in the tumor.

The BEP model is a very simple model and has many limitations. Although this BEP model assumes that driver mutations increase the replication probability, it is considered that driver mutations decrease the death probability. The BEP model also assume that each diver mutation has the same effect on the replication probability; however, actual tumors contain different driver mutations of different strengths. Although actual tumors grow in a three-dimensional space, the BEP model assumes tumor growth on the a two-dimensional square lattice; extension to a three-dimensional lattice should be considered as a future improvement. For on-lattice models, various other simulators has been developed for studying tumor growth (off-lattice models which do not assume that tumor growth on the lattice reflects the actual situations more accurately, but are computationally intensive and not commonly used [7]). For example, the pioneering works of agent-based

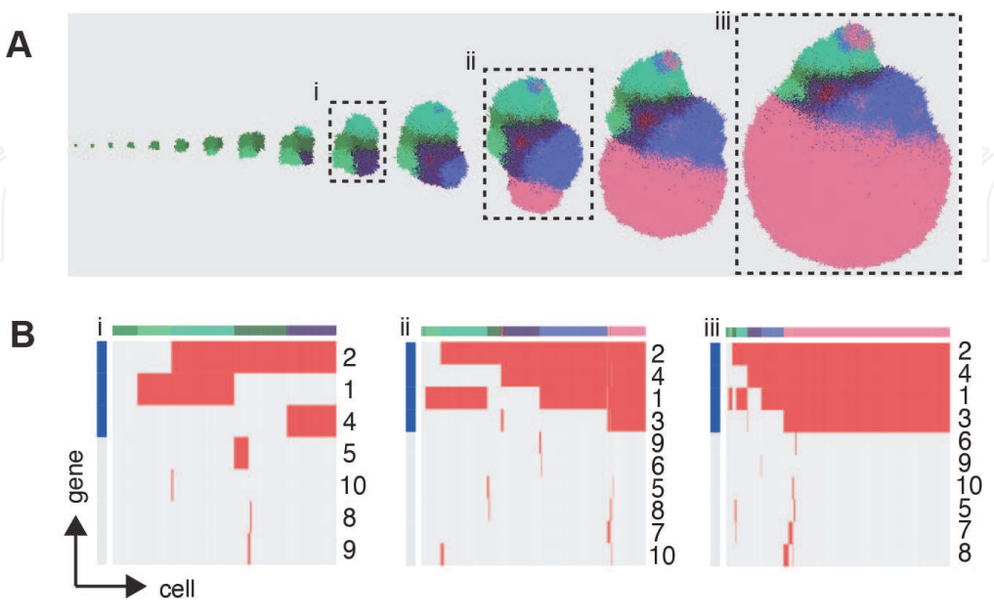


Figure 2.
Visualization of a simulation based on the BEP model. (A) Evolutionary snapshots obtained by simulating two-dimensional tumor growth based on the BEP model with an appropriate parameter setting. The region with the same color represents a clone with the same set of mutated genes. (B) Single-cell mutation profiles at three time points in the simulated tumor growth. Top colored bands represent clones, whereas the blue bands on the left represent driver genes. This image was obtained by modifying a figure that originally appeared in [6].

modeling were performed by Anderson and colleagues [8, 9]. Enderling and colleagues extended the model to incorporate cell differentiation from cancer stem cells where differentiated cells have a limited potential for cell division [10, 11]. Sottoriva *et al.* [12] found that hierarchical organization of cell differentiation affects tumor heterogeneity, which leads to an invasive morphology with finger-like front. Waclaw *et al.* [13] predicted that dispersal and cell turnover limit intratumor heterogeneity.

Each group developed a model different from the others, and thus only limited conditions were considered in each study. To address this issue, Iwasaki and Innan [14] developed a flexible and comprehensive simulation framework named as *tumopp* so that all previous models were included in a single program. This enables researchers to explore the effects of various model settings with simple command-line options. For example, the combined effect of local density on the cell division rate and method of placing new cells had a major impact on ITH. Under the condition with random push and no local competition, all cells undergo cell division at a constant rate regardless of the local density, and new cells are placed while randomly pushing out pre-existing neighbor cells. This behavior creates shuffled patterns of ITH with weak isolation by distance. In contrast, under the conditions of strong resource competition, the division rate of a cell is higher when it has more space (empty sites) in the neighborhood, which generally applies to cells near the surface of the simulated tumor; new cells are placed to fill the empty space without pushing existing cells. This setting tends to create a biased complex shape with clusters of genetically closely related cells, resulting in strong isolation by distance. Thus, it has been demonstrated that various patterns in the shape and heterogeneity of tumors arose depending on the model setting even without Darwinian selection. This suggests a caveat in analyzing ITH data with simulations using limited settings because another setting may predict a different ITH pattern, which could result in a different conclusion.

Moreover, *tumopp* introduced several other factors to relax various assumptions. First, it adopted a gamma function for the waiting time involved in cell division, whereas all previous studies assumed a simple decreasing (i.e., exponential) function implicitly or explicitly. The shape of the gamma distribution can be specified by parameter k, which affects the growth curve and inequality of clones in a tumor. An exponential distribution, which is the most widely used, is included as a special case with $k = 1$, whereas punctual and deterministic cell divisions can be achieved with $k = \infty$. It is reasonable to expect that the true value of k lies somewhere in-between, such as at ~ 10, because cell division is neither a memoryless Poisson process ($k = 1$; equivalent to an exponential distribution) nor perfectly synchronized ($k = \infty$; equivalent to Dirac delta distribution) [15, 16]. Second, a hexagonal lattice has been implemented, which is thought to be biologically more reasonable because the distance to all neighboring cells is identical so that there is only one definition of the neighborhood. A hexagonal lattice should be used for future work when the spatial pattern is of interest to the study. These factors will contribute to simulating the evolutionary processes of cellular populations under more realistic conditions.

3. Fitting the simulation model to observational data

As described in the "Introduction" section, cancer genome analysis demonstrated intratumor heterogeneity and branching evolution of cancer; paticulally, an approach known as multiregion sequencing has been popularly employed for analyzing solid tumors. Here, we introduce a concrete example of a multiregion

sequencing study and explain the utility of cancer evolution simulation when combined with multiregion sequencing data.

In multiregion sequencing, multiple samples obtained from physically separate regions within the tumor of a single patient are analyzed (**Figure 3A**), with two categories of somatic single-nucleotide mutations identified: "founder" and "progressor" mutations (**Figure 3B**). Founder mutations are defined as present in all regions, whereas progressor mutations are defined as present in some regions (note that they are also referred to using different terms in different studies, e.g., public/private or trunk/branch mutations). Founder mutations are thought to accumulate during the early phases of cancer evolution. The common ancestor clone acquires all founder mutations, and then branches into subclones, which accumulate progressor mutations and contribute to forming ITH. Through these multiregion mutational profiles, we can infer an evolutionary history of the cancer by constructing a phylogenetic tree (**Figure 3C**).

As a pioneering study, Gerlinger et al. [17] performed multiregion sequencing, revealing extensive ITH and clonal branching evolution in renal cancer. Ther also identified not only founder mutations in some known driver genes such as *VHL*, but also progressor mutations in other known driver genes such as *SETD2* and *BAP1*. Interestingly, in some cases, different mutations in the same driver gene or genes with the same function were acquired independently. This phenomenon known as parallel evolution also indicates that part of the ITH was generated by Darwinian selection.

Uchi et al. [5] also investigated ITH in nine cases of surgically resected late-stage colorectal tumors by multiregion sequencing to identify founder and progressor mutations in each case. **Figure 4** shows the results obtained from one of the nine cases, which contains 20 samples from the primary lesion and one sample from the metastatic lesion. Note that the progressor mutations showed a mutational pattern that was geographically correlated with the sampling locations. Moreover, they found that mutation allele frequencies, which can be approximately regarded as the proportion of cells with mutations in each region, tended to be lower for progressor mutations than for founder mutations. This observation suggests that the founder mutations existed in all the cancer cells while the progressor mutations existed in only a fraction of the cancer cells in each region. Thus, even in each region, extensive ITH may have existed, which was not captured by the resolution of multiregion sequencing. In addition, most mutations in known driver genes such as *APC* and *KRAS* were identified as founder mutations. However, progressor mutations contain few driver mutations and parallel evolution was not confirmed, which contrasts to the findings obtained in renal cancer. These observations suggest that apart from

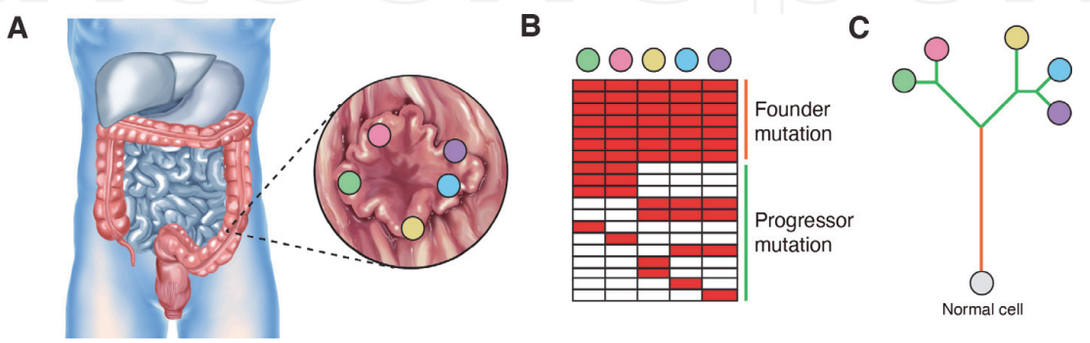


Figure 3.
Multiregion sequencing (A) DNA samples from multiple regions of a single tumor are analyzed by next-generation sequencing. (B) Through multiregion mutation profiling, founder and progressor mutations are identified as common mutations in all regions tested and only restricted regions, respectively. (C) In a phylogenetic tree constructed from the multiregion mutation profile, the trunk and branches correspond to the founder and progressor mutations, respectively. This image originally appeared in [6].

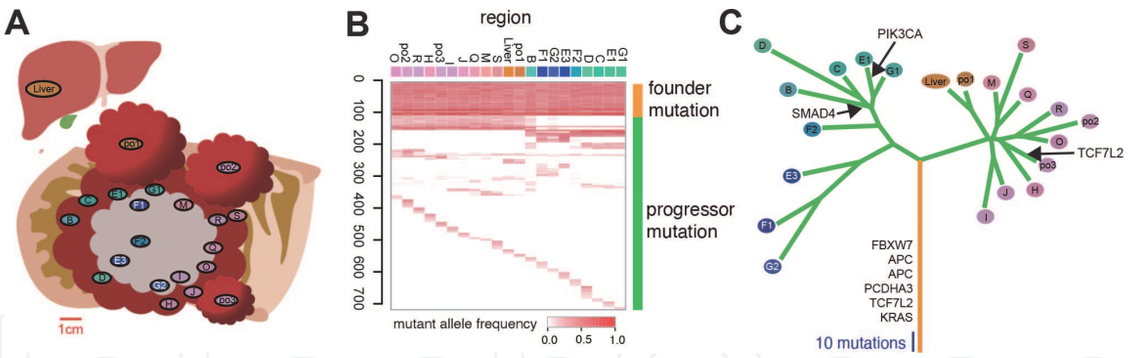


Figure 4.
Multiregion sequencing of colorectal cancer. (A) Schema of the tumor subjected to multiregion sequencing. (B) Multiregion mutation profile. The depth of red represents the mutant allele frequency, whereas the colors of the sample labels were prepared so that the similarities of colors represent those of mutation patterns. (C) Phylogenetic tree constructed from the multiregion mutation profile. The time when mutations in known driver genes of colorectal cancer were acquired is indicated along the tree. This image was obtained by modifying a figure that originally appeared in [5].

Darwinian selection, there are other evolutionary principles generating ITH. To identify these principles, they developed the BEP model as described above; by fitting the BEP model to the multiregion sequencing data, they evaluated the evolutionary principles generating ITH in colorectal cancer.

To fit the simulation model to the observational data, we can employ ABC [18], which constitutes a class of computational methods rooted in Bayesian statistics that can be used to estimate the posterior distributions of model parameters. A common incarnation of Bayes' theorem relates the conditional probability of a specific parameter value θ given data D to the probability of D given θ by the rule, $p(\theta|D) \propto p(D|\theta)p(\theta)$, where $p(\theta|D)$ denotes the posterior, $p(D|\theta)$ the likelihood, and $p(\theta)$ the prior. The prior represents beliefs or knowledge about θ before D is available. To obtain the the posterior, the likelihood function is required. For simple models, an analytical formula for the likelihood function can typically be derived. However, for more complex models, an analytical formula may be elusive or the likelihood function may be computationally very costly to evaluate. Agent-based models also fall into the latter case. ABC methods bypass evaluation of the likelihood function by using summary statistics and simulations, which widen the realm of models for which statistical inference can be considered. ABC has rapidly gained popularity over the last few years, for analizing complex problems arising in biological sciences, e.g., in population genetics, ecology, epidemiology, and systems biology.

In the basic form of ABC, which is known as rejection sampling, we first sample a parameter value (or a combination of parameter values, if there is more than one parameter) from a prescribed prior distribution of the parameter value. Simulated data are then generated from the sampled parameter value. The similarity between the simulated and observational data is evaluated using summary statistics (typically multiple), which is designed to represent the maximum amount of information in the simplest possible form. If the distance of the summary statistics between the simulated and observational data is below a tolerance parameter, the parameter value is accepted and pooled into the posterior probability of the parameter value. Repeating these steps many times, we can approximate the probability distribution. A conceptual overview of the ABC rejection sampling algorithm is presented in **Figure 5**.

In the study of colorectal cancer study by Uchi et al. [5], as summary statistics, they adopted the proportions of founder mutations and unique mutations, which is uniquely observed in each sample, in a multiregion mutation profile.

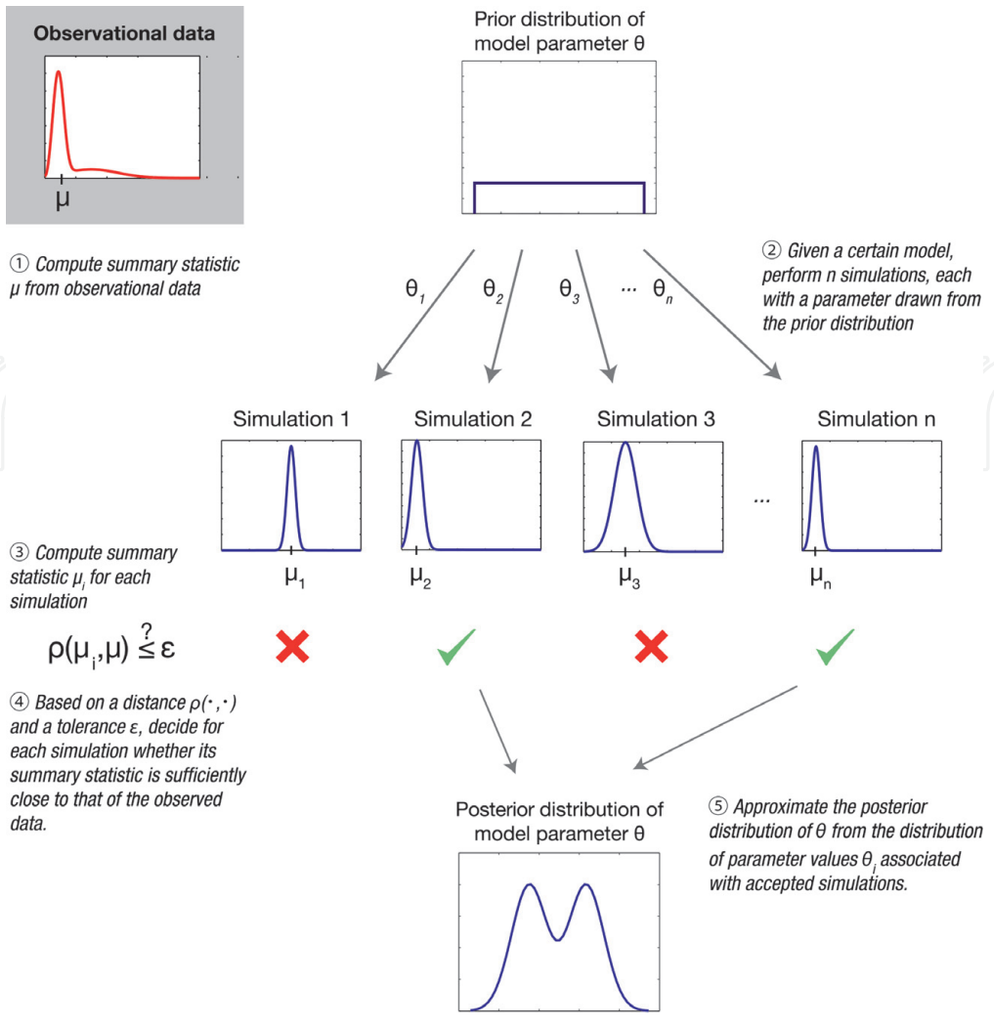


Figure 5.
Conceptual overview of the ABC rejection sampling algorithm. This image originally appeared in [18].

They obtained multiregion mutation profiles for 9 cases with different sample numbers. As the proportions of founder mutations and unique mutations depend on the number of samples, they set the sample number to 5, which is the minimum sample number of the 9 cases, by downsampling the samples in cases containing more than 5 samples. They then estimated the mean of the proportions of founder mutations and unique mutations and used these values as summary statistics values of the observational data (**Figure 6**; note that although we should apply ABC to each of the 9 case separately, they targeted the population mean for simplicity).

For ABC, they generated simulation data while varying 3 parameters, m (the mutation rate), d (the number of driver genes), and f (the effect of driver mutations), which appear to be critical for simulation results (for strategies used to find such parameters, read the next section). In each simulation trial, we simulated multiregion sequencing from a tumor simulated by the BEP model; a multiregion mutation profile was obtained by digging 5 squares out from a simulated tumor and averaging the mutation status of cells in the squares. From the multiregion mutation profile, the proportions of founder mutations and unique mutations were obtained as summery statistics. They performed 50 simulations for each grid point in a three-dimensional rectangular parameter space; namely, they assumed a uniform prior for each of the three parameters. For each grid point in the parameter space, they calculate the proportion of the simulation instances whose statistics fall within 1 standard deviation from the mean of the values observed in the real multiregion

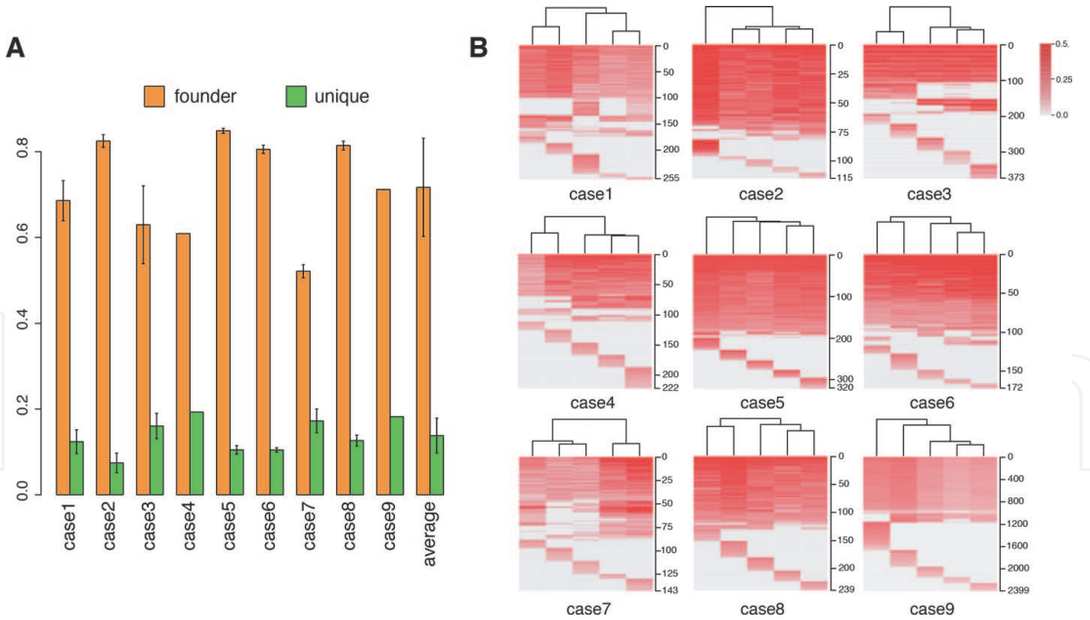


Figure 6.
Fitting the BEP model to multiregion sequencing data by ABC. (A) Observed values of summary statistics in the multiregion sequencing data. After downsampling the samples in cases containing more than 5 samples, the proportions of founder and unique mutations were estimated for 9 cases (case 1–9) and an "average" over the 9 cases was obtained as summary statistics values of the observational data. The error bars at case 1–3 and 5–8 indicate standard deviations for 10 downsampling trials while the error bar at average indicates standard deviations over the 9 cases. (B) Multiregion mutation profiles of 9 colorectal tumors. For the cases except case 4 and 9, representative samples from the downsampling trials were presented as in ***Figure 4B****. This image originally appeared in [5].*

mutation profiles. The distribution of the proportions can be regarded as the posterior and visualized in heat maps (**Figure 7**).

As a result, when cancer evolution was simulated with the assumption of a high mutation rate, we reproduced mutation profiles similar to those obtained by our multiregion sequencing of colorectal cancers (compare **Figure 8A** and **B** with **Figure 4A** and **B**). That is, irrespective of the presence of founder mutations, progressor mutations contributed to the formation of a heterogeneous mutation profile, which was geographically correlated with the sampling locations. Moreover, we also reconstructed local heterogeneity, as illustrated by the finding that progressor mutations existed as mutations with lower allele frequencies in each region. Interestingly, although driver mutations were acquired as founder mutations, progressor mutations contained few driver mutations, and most comprised neutral mutations that did not affect the cell division rate. This suggests that, after the appearance of the common ancestor clone with accumulated driver mutations, extensive ITH was generated by neutral evolution. Moreover, the single-cell mutation profiles of the simulated tumor suggest that the tumor comprises a large number of minute clones with numerous neutral mutations accumulated (**Figure 8C**).

By employing a agent-based model and ABC, Sottoriva et al. [19] also proposed a Big Bang model of human colorectal tumor growth; in their model, tumors grow predominantly as a single expansion producing numerous intermixed subclones that are not subject to stringent selection, which is consistent with the model developed by Uchi *et al.* [5], and both public (clonal) and most detectable private (subclonal) alterations arise early during growth. Hu et al. [20] also employed an agent-based model and ABC to examine the timing of metastasis in colorectal cancer. Multiregion sequencing data containing both primary and metastatic samples were prepared from patients with metastases to the liver or brain. Simultaneously, a spatial agent-based model was developed to simulate tumor growth,

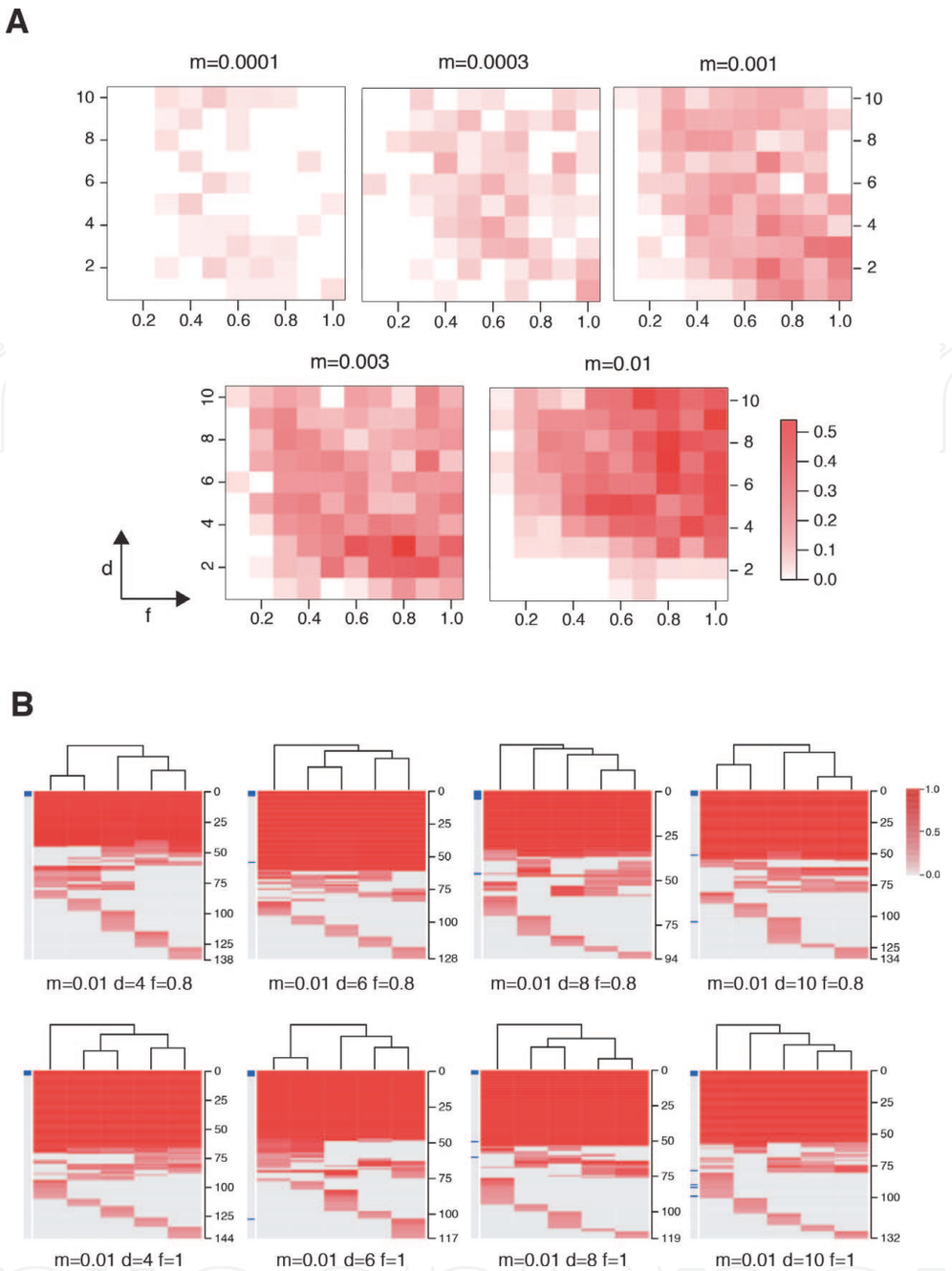


Figure 7.
Fitting the BEP model to multiregion sequencing data by ABC (continued). (A) The proportion of simulation instances fitted to the real data. Multiregion mutation profiles were simulated while varying 3 parameters and, for each parameter settings, the proportion of simulation instances that were judged to be similar to the real data based on summary statistics are visualized as heat maps. (B) Multiregion mutation profiles from the simulations. Representative instances from simulation with indicated parameter settings were presented as in ***Figure 4B****. Left blue bars indicate driver genes. This image originally appeared in [5].*

mutation accumulation, and metastatic dissemination. From multiregion sequencing data of each patient, the time of dissemination, which is a parameter in the agent-based model, was estimated by ABC. The results demonstrated that early disseminated cells commonly (81%, 17 of 21 patients) showed metastases, whereas the carcinoma was clinically undetectable (typically, less than 0.01 cm^3). Collectively, these examples demonstrated that ABC successfully fitted the simulation model of cancer evolution to cancer genome data, providing insight into the mechanisms of cancer evolution.

Although the problem of computational cost generally accompanies ABC, new sampling approaches utilizing Markov chain Monte Carlo and its derivatives [21]

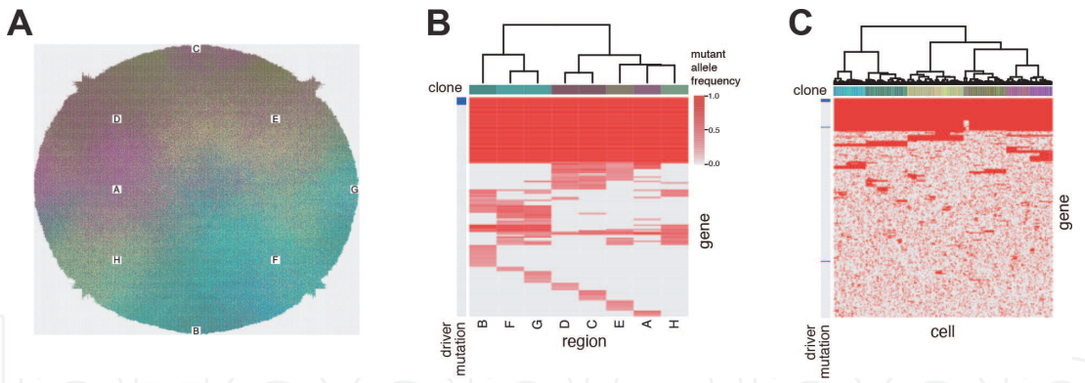


Figure 8.
Computer-simulated tumor with extensive ITH generated by neutral evolution. (A) Tumor simulated based on the BEP model with an assumption of a high mutation rate. (B) Simulated multiregion mutation profile of the simulated tumor. Cell populations in the regions labeled with A–H were extracted from the simulated tumor and their averaged mutation profiles were obtained. (C) Simulated single-cell mutation profile of the simulated tumor. This image was obtained by modifying a figure that originally appeared in [5].

have been developed to overcome this limitation. Moreover, considering the increasing computing power, this problem will potentially be less important. Notably, ABC has many potential pitfalls [18]. For example, setting the tolerance parameter to zero will give accurate results, but typically at a very high computational cost. In practice, therefore, values of greater than zero are used, but this introduces bias. Similarly, sufficient statistics are sometimes not available and other summary statistics are used instead, but this introduces additional bias because of the loss of information. Additionally, prior distributions and choices of parameter ranges are often subject to criticisms, although they are not unique to ABC and apply to all Bayesian methods. Model complexity (i.e., the number of model parameters) is also an important point. If a model is too simple, it can lack predictive power. In contrast, if the model is too complex, there is a risk of overfitting. Moreover, the complex model faces a problem known as the curse of dimensionality, in which the computational cost is severely increased and may, in the worst case, render the computational analysis intractable. When constructing a simulation model, we should follow the Occam's razor principle: i.e., achieve the lowest model complexity that is sufficient to explain the observational data. To determine the optimal model complexity, we can also employ the model selection scheme based on Bayes factor if a choice of summary statistics is appropriate [22].

4. Characterizing the dynamics of the simulation model

In the previous section, we explained how to fit a simulation model to observational data. Another direction for studying a simulation is by characterizing the dynamics of the simulation model without observational data. Namely, we can examine parameter dependance by performing a large number of simulations while varying the parameter values. This approach is known as sensitivity analysis and can provide insights into the modeled system as well as identify parameters that are critical for the system dynamics. In sensitivity analysis, as in ABC, we define a summary statistic Y. A simulation model is then regarded as a function: $Y = \mathrm{F}(\mathbf{X})$ where $\mathbf{X} = \{X_1, X_2, \ldots X_k\}$ are model parameters. The aim of sensitivity analysis can also be considered as characterizing the function "F".

So far, a number of approaches have been proposed for sensitivity analysis. For example, one-factor-at-a-time (OFAT) sensitivity analysis is one of the simplest and most common approaches that changes one parameter at a time to determine

the effects on a summery statistic [23]. In OFAT sensitivity analysis, we move one parameter, while leaving the other parameters at their baseline (nominal) values, and then return the parameter to its nominal, which is repeated for each of the other parameters. We then plot the relationship between each parameter and a summary statistic to examine the dependency of the summary statistic on the parameter, or the relationship can be measured by partial derivatives or linear regression. In exchange for its simplicity, this approach does not fully explore the input space, as it does not consider the simultaneous variation of multiple parameters. This means that the OFAT approach cannot detect interactions between parameters.

Global sensitivity analysis aims to address this point by sampling a summary statistic over a wide parameter space involving multiple parameters. Sobol's method is a popular approach for estimating the contributions of different combinations of parameters to the variance of the summary statistic while assuming that all parameters are independent [24]. The sensitivity of the summary statistic Y to a parameter X_i is measured by the amount of variance in Y caused by the parameter X_i and can be expressed as a conditional expectation, $\mathrm{Var}(\mathrm{E}_{\mathbf{X}_{\sim i}}(Y|X_i))$, where "V*ar*" and "E" denote the variance and expected value operators, respectively, and $\mathbf{X}_{\sim i}$ denotes the set of all input variables except for X_i. This expression essentially measures the contribution X_i alone to uncertainty (variance) in Y (averaged over variations in other variables), and is known as the first-order sensitivity index or main effect index. Importantly, it does not measure the uncertainty caused by interactions with other variables. A further measure, known as the total effect index, gives the total variance in Y caused by X_i and its interactions with any of the other input variables. Both quantities are typically standardized by dividing by $\mathrm{Var}(Y)$. In Sobol's method, we typically attempt full exploration of the parameter space based on a Monte Carlo method to grasp parameter interactions and nonlinear responses.

However, such approaches appears to be insufficient to comprehensively grasp how the parameters judged to be influential control the behaviors of agent-based models. To overcome this point, Niida *et al.* [25] recently developed a new approach to sensitivity analysis for agent-based simulations, named as MASSIVE (Massively parallel Agent-based Simulations and Subsequent Interactive Visualization-based Exploration). MASSIVE overcomes the limitations of existing methods by taking advantage of two currently rising technologies: massively parallel computation and interactive data visualization. MASSIVE employs a full factorial design involving a multiple number of parameters (i.e., test every combination of candidate values of the multiple parameters), which can broadly cover a target parameter space. In addition, when analyzing a stochastic simulation model such as an agent-based model, multiple simulation trials with the same parameter setting are required to examine stochastic effects. To cope with the computational cost problem caused by these features, MASSIVE utilizes a supercomputer, in which agent-based simulations with different parameter settings and the following post-processing step of simulation results are performed in parallel. The massively parallel simulations generate massive results, which then pose a problem for interpretation. This problem was solved by developing a web-based tool that interactively visualizes not only the values of multiple summary statistics, but also output images (e.g., mutation profiles) from simulations with each parameter setting.

Below I explain an example of sensitivity analysis, which was performed by Niida et al. [26] to understand the precise mechanisms underlying neutral evolution induced by a high mutation rate. First, they built an agent-based model, referred to as the "neutral" model, for simulating neutral evolution in cancer. Although the neutral model is similar to the BEP model, the neutral model assumes only neutral mutations and omits spatial information. They also improved the approach used for mutation accumulation in the BEP model. Namely, in the neutral model, they

considered only neutral mutations that did 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. In each cell division, each of the two daughter cells acquires $k_n \sim \text{Pois}(m_n/2)$ neutral mutations. They assumed that neutral mutations acquired by different division events occur at different genomic positions.The simulation started from one cell without mutations and ended when the population size p reached P or time t reached T.

Through sensitivity analysis based on the MASSIVE method, they confirmed that the mutation rate is the most important factor affecting neutral evolution (**Figure 9**). As a summary statistic for evaluating ITH, they calculated Shannon index 0.05 by the following procedure. After removing mutations with frequencies of less than 0.05, the proportions of different subclones (cell subpopulations with different mutations) were obtained and the Shannon index H was calculated using the following formula: $H = -\sum_{i=1}^{n} p_i \log(p_i)$, where n is the total number of different subclones and p_i is the proportion of each subclone. Based on this definition, a larger Shannon index 0.05 value indicates more extensive ITH. Together with a heat map of the Shannon index 0.05 values, we visualized single-cell mutations profiles obtained for different parameter settings. The mutation profile matrix was obtained by sampling 1,000 cells from a simulated tumor, and visualized after filtering out lower-frequency mutations, such that the maximum number of rows was 300. The rows and columns are reordered by hierarchical clustering and index mutations and samples, respectively. They found that when the mean number of mutations generated by per cell division, m_n, was less than 1, the neutral model just generated sparse mutation profiles with relatively small values of Shannon index 0.05. In contrast, when m_n exceeded 1, the mutation profiles presented extensive ITH, which are characterized by a fractal-like pattern and large values of the ITH score (hereinafter, this type of ITH is referred to as "neutral ITH"). These results suggest 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. Assuming a high mutation rate, more numerous subclones with different mutations should be observed if mutations existing at lower frequencies are counted. However, the ITH score does not depend on the population size P because low-frequency mutations were filtered out before calculation.

Thus far, several theoretical and computational studies have shown that a stem cell hierarchy can boost neutral evolution in a population of cancer cells [12, 27];

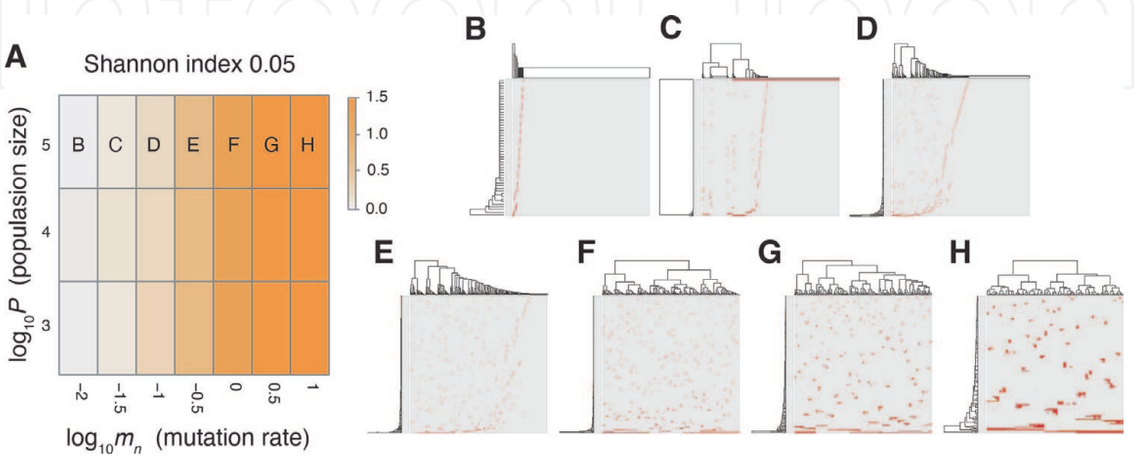


Figure 9.
Sensitivity analysis of the neutral model. (A) Heap map obtained by calculating Shannon index 0.05 *while changing the neutral mutation rate m_n and maximum population size P. (B–H) Single-cell mutations profiles obtained for seven parameter settings, which are indicated on the heat map in A. This image originally appeared in [26].*

based on this, they extended the neutral model to the "neutral-s" model such that it contains a stem cell hierarchy (**Figure 10**). The neutral-s model assumes that two types of cell exist: stem and differentiated. Stem cells divide with a probability g_0 without dying. For each cell division of stem cells, a symmetrical division generating two stem cells occurs with a probability s, whereas 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, they define $\delta = \log_{10}\left(d_0^d/g_0\right)$ and hereinafter use δ rather than d_0^d.

MASSIVE analysis of the neutral-s model confirmed that incorporation of the stem cell hierarchy boosts neutral evolution (**Figure 11**). To obtain the heat map in **Figure 11A**, the ITH score was measured while d_0^d and δ were changed, whereas $m_n=0.1$ and $P=1000$ were maintained as constant. In the heat map, a decrease in 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 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.

The underlying mechanism boosting neutral evolution can be explained as follows. Only stem cells were considered for an approximation, as 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 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. They confirmed that the expected mutation count based on this formula fit well with the values observed in their simulation (data not shown). These arguments mean that a tumor with a stem cell hierarchy accumulates more mutations until reaching a fixed population size than

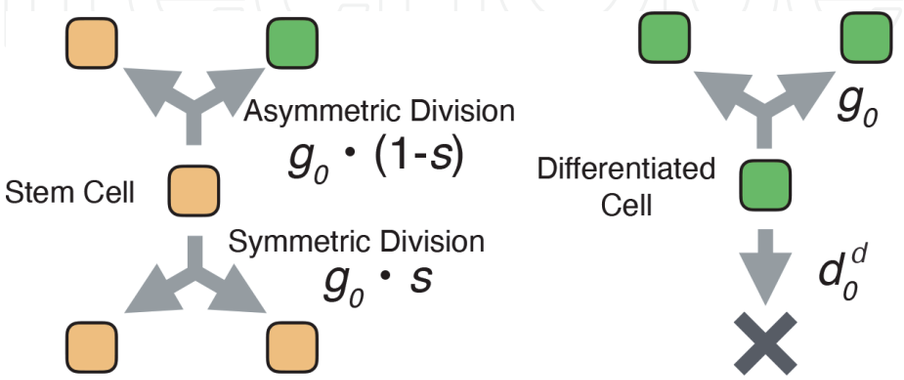


Figure 10.
Schema of the neutral-s model. Stem cells divide with a probability g_o without dying. For each cell division of stem cells, a symmetrical division generating two stem cells occurs with probability s, whereas an asymmetrical division generating one stem cell and one differentiated cell occurs with probability $1-s$. A differentiated cell symmetrically divides to generate two differentiated cells with probability g_o but dies with probability d_o^d. This image originally appeared in [26].

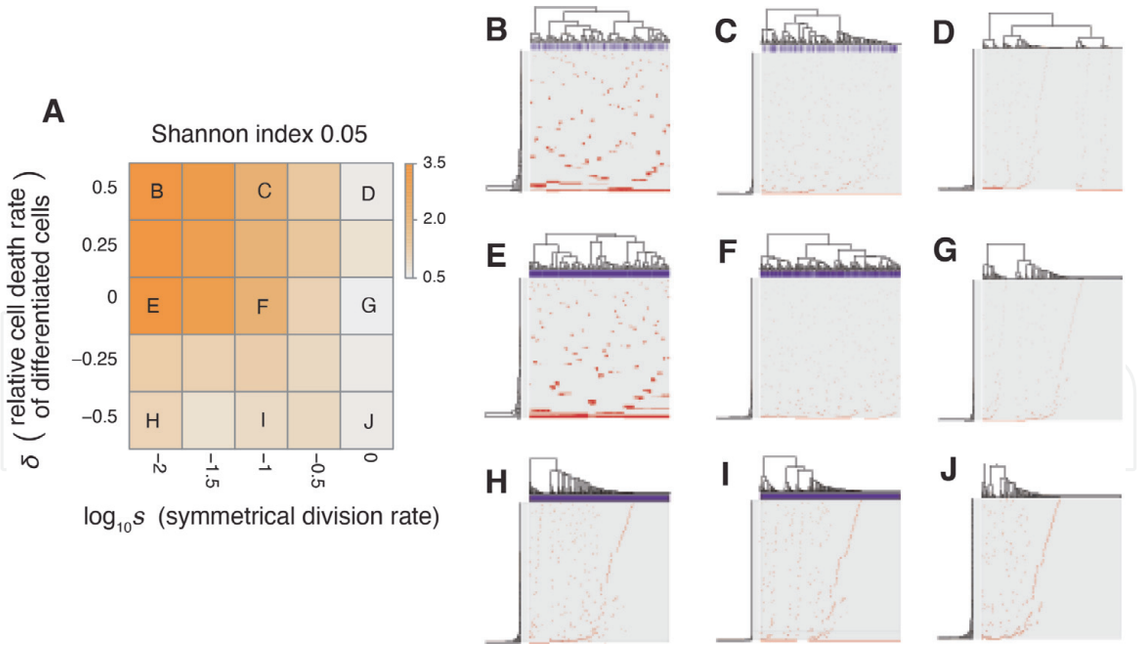


Figure 11.
Sensitivity analysis of the neutral-s model. (A) Heat map obtained by calculating Shannon index 0.05 while changing the relative death rate of differentiated cells $\delta = \log_{10}\left(d_o^d/g_o\right)$ *and symmetrical division rate s. The neutral mutation rate* m_n *and maximum population size P set to* 10^{-1} *and* 10^5*, respectively. (B–J) Single-cell mutation profiles obtained for nine parameter settings, indicated on the heat map presented in A. This image originally appeared in [26].*

does a tumor without a stem cell hierarchy. That is, a stem cell hierarchy increases the apparent mutation rate by $\log 2/\log(1+s)$-fold, which induces neutral evolution even with relatively low mutation rate settings.

Recent genomic analysis demonstrated that multiple evolutionary modes exists in cancer systems. For example, as described above, ITH in renal cancer is generated by Darwinian selection, which is in contrast to neutral evolution in colorectal cancer. Moreover, by multiregion sequencing of early-stage colorectal tumors, Saito *et al.* [26] showed that ITH is shaped by Darwinian selection in the early phase of colorectal cancer evolution, which means that a temporal shift of the evolutionary principle shaping ITH occurs during colorectal tumorigenesis. Employing agent-based modeling and MASSIVE analysis, Niida *et al.* [26] also constructed a model that explain this evolutionary shift. Darwinian ITH in an early-stage tumor is reproduced by the assumption of multiple driver mutations of relatively weak strength. At some point, growth of the early colorectal tumor slows because resource limitations, which is reproduced by introducing the carrying capacity into the simulation model. When they assumed that an explosive mutation that negates the carrying capacity was obtained with a small probability, a clone acquiring the explosive mutation overcame the resource limitation and expanded as late-stage tumors, in which ITH was generated neutral evolution. Another simulation study by West *et al.* [28] proposed that spatial constraints and limited cellular mixing play important roles in a similar Darwinian-neutral shift.

Sensitivity analysis also provides insight into metastatic tumor progression, which is poorly understood despite its clinical importance. Evaluation of genomic divergence between paired metastatic and primary tumors (M-P divergence) from multiregion sequencing is a good starting point for addressing this problem. Sun and Nikolakopoulos [29] extended *tumopp* [14] to simulate paired primary and metastatic tumors, and explored factors affecting M-P divergence by sensitivity analysis. As a result, they found that M-P divergence depends not only on the metastatic dissemination time, but also on the evolutionary dynamics and

detectability of seedling cell lineages in a primary tumor. It was concluded that investigating tumor growth dynamics in detail is important, particularly when researchers interpret heterogeneity among longitudinal samples to infer the evolutionary timeline of cancer progression. Collectively, these examples demonstrated that agent-based modeling combined with sensitivity analysis is a useful tool for studying cancer evolutionary dynamics.

5. Conclusion

In this chapter, we introduced agent-based modeling of cancer evolution along with methodologies for data fitting and sensitivity analysis. Although there is a long history of theoretical science in the field of cancer research, this approach has been overshadowed by experimental science until recently. However, with a recent explosive increase in cancer genome data, there is now an increasing need to integrate experimental and theoretical science. As an example, this chapter introduced methods for modeling and analyzing the evolutionary processes generating ITH, which is experimentally observed by multiregion sequencing. We also presented exemplifying applications: e.g., agent-based simulation modeling and analysis successfully demonstrated that ITH in colorectal cancer is generated by neutral evolution, which is caused by a high mutation rate and stem cell hierarchy. For cancer genome analyses, new experimental technologies are actively being developed. For example, single-cell sequencing technologies can profile IHT at the ultimate resolution [30] while liquid biopsy technologies, such as the sequencing of circulating tumor DNA, enables us to non-invasively track cancer evolution during treatment [31]. These technologies will unveil more various aspects of cancer evolution when combined with the approach introduced in this chapter. This chapter also exemplified how simulation modeling helps to solve scientific problems raised by new experimental technologies. We hope that this chapter will provides readers with some hints to solve their own problems using simulation modeling.

Acknowledgements

This work was supported by the JSPS KAKENHI (19K12214) and AMED (JP21cm0106504).

DOI: http://dx.doi.org/10.5772/intechopen.100140

IntechOpen

IntechOpen

Author details

Atsushi Niida[1*] and Watal M. Iwasaki[2]

1 The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

2 Graduate School of Life Sciences, Tohoku University, Sendai, Japan

*Address all correspondence to: aniida@ims.u-tokyo.ac.jp

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

References

[1] Eric R Fearon and Bert Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61(5):759–767, 1990.

[2] Bert Vogelstein, Nickolas Papadopoulos, Victor E Velculescu, Shibin Zhou, Luis A Diaz, and Kenneth W Kinzler. Cancer genome landscapes. *Science*, 339 (6127):1546–1558, 2013.

[3] Nicholas McGranahan and Charles Swanton. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*, 27(1):15–26, 2015.

[4] Charles M Macal and Michael J North. Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.*, pages 14–pp. IEEE, 2005.

[5] Ryutaro Uchi, Yusuke Takahashi, Atsushi Niida, Teppei Shimamura, Hidenari Hirata, Keishi Sugimachi, Genta Sawada, Takeshi Iwaya, Junji Kurashige, Yoshiaki Shinden, et al. Integrated multiregional analysis proposing a new model of colorectal cancer evolution. *PLOS Genetics*, 12(2): e1005778, 2016.

[6] Atsushi Niida, Satoshi Nagayama, Satoru Miyano, and Koshi Mimori. Understanding intratumor heterogeneity by combining genome analysis and mathematical modeling. *Cancer Science*, 109(4):884–892, 2018.

[7] PADM Van Liedekerke, A Buttenschön, and D Drasdo. Off-lattice agent-based models for cell and tumor growth: numerical methods, implementation, and applications. In *Numerical methods and advanced simulation in biomechanics and biological processes*, pages 245–267. Elsevier, 2018.

[8] Alexander RA Anderson, Alissa M Weaver, Peter T Cummings, and Vito Quaranta. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell*, 127(5):905–915, 2006.

[9] Alexander RA Anderson, Katarzyna A Rejniak, Philip Gerlee, and Vito Quaranta. Microenvironment driven invasion: a multiscale multimodel investigation. *Journal of Mathematical Biology*, 58(4):579–624, 2009.

[10] Heiko Enderling, Lynn Hlatky, and Philip Hahnfeldt. Migration rules: tumors are conglomerates of self-metastases. *British Journal of Cancer*, 100 (12):1917–1925, 2009.

[11] Jan Poleszczuk, Philip Hahnfeldt, and Heiko Enderling. Evolution and phenotypic selection of cancer stem cells. *PLOS Computational Biology*, 11 (3):e1004025, 2015.

[12] Andrea Sottoriva, Joost JC Verhoeff, Tijana Borovski, Shannon K McWeeney, Lev Naumov, Jan Paul Medema, Peter MA Sloot, and Louis Vermeulen. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer Research*, 70(1): 46–56, 2010.

[13] Bartlomiej Waclaw, Ivana Bozic, Meredith E Pittman, Ralph H Hruban, Bert Vogelstein, and Martin A Nowak. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525 (7568):261–264, 2015.

[14] Watal M Iwasaki and Hideki Innan. Simulation framework for generating intratumor heterogeneity patterns in a cancer cell population. *PLOS One*, 12(9): e0184229, 2017.

[15] Camilla Hurwitz and LJ Tolmach. Time-lapse cinemicrographic studies of x-irradiated hela s3 cells: I. cell progression and cell disintegration. *Biophysical Journal*, 9(4):607–633, 1969.

[16] Darren R Tyson, Shawn P Garbett, Peter L Frick, and Vito Quaranta. Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data. *Nature Methods*, 9 (9):923–928, 2012.

[17] Marco Gerlinger, Andrew J Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine*, 366:883–892, 2012.

[18] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate bayesian computation. *PLOS Computational Biology*, 9(1):e1002803, 2013.

[19] Andrea Sottoriva, Haeyoun Kang, Zhicheng Ma, Trevor A Graham, Matthew P Salomon, Junsong Zhao, Paul Marjoram, Kimberly Siegmund, Michael F Press, Darryl Shibata, et al. A big bang model of human colorectal tumor growth. *Nature Genetics*, 47(3): 209–216, 2015.

[20] Zheng Hu, Jie Ding, Zhicheng Ma, Ruping Sun, Jose A Seoane, J Scott Shaffer, Carlos J Suarez, Anna S Berghoff, Chiara Cremolini, Alfredo Falcone, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nature Genetics*, 51(7):1113–1122, 2019.

[21] SA Sisson and Y Fan. Abc samplers. *Handbook of Approximate Bayesian Computation*, pages 87–123, 2018.

[22] Jean-Michel Marin, Natesh S Pillai, Christian P Robert, and Judith Rousseau. Relevant statistics for bayesian model choice. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 833–859, 2014.

[23] Veronica Czitrom. One-factor-at-a-time versus designed experiments. *The American Statistician*, 53(2):126–131, 1999.

[24] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1–3):271–280, 2001.

[25] Atsushi Niida, Takanori Hasegawa, and Satoru Miyano. Sensitivity analysis of agent-based simulation utilizing massively parallel computation and interactive data visualization. *PLOS one*, 14(3):e0210678, 2019.

[26] Atsushi Niida, Takanori Hasegawa, Hideki Innan, Tatsuhiro Shibata, Koshi Mimori, and Satoru Miyano. A unified simulation model for understanding the diversity of cancer evolution. *PeerJ*, 8: e8842, 2020.

[27] Ricard V Solé, Carlos Rodrguez-Caso, Thomas S Deisboeck, and Joan Saldaña. Cancer stem cells as the engine of unstable tumor progression. *Journal of Theoretical Biology*, 253(4):629–637, 2008.

[28] Jeffrey West, Ryan O Schenck, Chandler Gatenbee, Mark Robertson-Tessi, and Alexander RA Anderson. Normal tissue architecture determines the evolutionary course of cancer. *Nature Communications*, 12(1):1–9, 2021.

[29] Ruping Sun and Athanasios N Nikolakopoulos. Elements and evolutionary determinants of genomic divergence between paired primary and metastatic tumors. *PLOS Computational Biology*, 17(3):e1008838, 2021.

[30] Bora Lim, Yiyun Lin, and Nicholas Navin. Advancing cancer research and medicine with single-cell genomics. *Cancer Cell*, 37(4):456–470, 2020.

[31] David W Cescon, Scott V Bratman, Steven M Chan, and Lillian L Siu. Circulating tumor dna and liquid biopsy in oncology. *Nature Cancer*, 1(3):276–290, 2020.