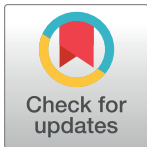


RESEARCH ARTICLE

Mathematical modeling the order of driver gene mutations in colorectal cancer

Lingling Li^{1,2}, Yulu Hu², Yunshan Xu³, Sanyi Tang^{1*}

1 School of Mathematics and Statistics, Shaanxi Normal University, Xi'an, China, **2** School of Science, Xi'an Polytechnic University, Xi'an, China, **3** Mathematics Department, Faculty of Science and Technology, University of Macau, Taipa, Macau, China

* sytang@snnu.edu.cn

OPEN ACCESS

Citation: Li L, Hu Y, Xu Y, Tang S (2023) Mathematical modeling the order of driver gene mutations in colorectal cancer. PLoS Comput Biol 19(6): e1011225. <https://doi.org/10.1371/journal.pcbi.1011225>

Editor: Philip K. Maini, Oxford, UNITED KINGDOM

Received: October 8, 2022

Accepted: May 29, 2023

Published: June 27, 2023

Copyright: © 2023 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data cannot be shared publicly, since public availability would compromise patient privacy. The data can be accessed upon request <https://www.seer.cancer.gov/seertrack/data/request/>, which need complete the application form to obtain the SEER*Stat account.

Funding: The Natural Science Foundation of China supported financially this work: 12001417 to LL and 12031010 to ST. LL also received funding from Shaanxi Science and Technology Association Young Talent Lifting Program (grant number 20220519). The funders had no role in study

Abstract

Tumor heterogeneity is a large obstacle for cancer study and treatment. Different cancer patients may involve different combinations of gene mutations or the distinct regulatory pathways for inducing the progression of tumor. Investigating the pathways of gene mutations which can cause the formation of tumor can provide a basis for the personalized treatment of cancer. Studies suggested that KRAS, APC and TP53 are the most significant driver genes for colorectal cancer. However, it is still an open issue regarding the detailed mutation order of these genes in the development of colorectal cancer. For this purpose, we analyze the mathematical model considering all orders of mutations in oncogene, KRAS and tumor suppressor genes, APC and TP53, and fit it on data describing the incidence rates of colorectal cancer at different age from the Surveillance Epidemiology and End Results registry in the United States for the year 1973–2013. The specific orders that can induce the development of colorectal cancer are identified by the model fitting. The fitting results indicate that the mutation orders with $KRAS \rightarrow APC \rightarrow TP53$, $APC \rightarrow TP53 \rightarrow KRAS$ and $APC \rightarrow KRAS \rightarrow TP53$ explain the age-specific risk of colorectal cancer with very well. Furthermore, eleven pathways of gene mutations can be accepted for the mutation order of genes with $KRAS \rightarrow APC \rightarrow TP53$, $APC \rightarrow TP53 \rightarrow KRAS$ and $APC \rightarrow KRAS \rightarrow TP53$, and the alternation of APC acts as the initiating or promoting event in the colorectal cancer. The estimated mutation rates of cells in the different pathways demonstrate that genetic instability must exist in colorectal cancer with alterations of genes, KRAS, APC and TP53.

Author summary

Cumulative mutations in driver genes are the essential cause of cancer disease. For the colorectal cancer, KRAS, APC and TP53 are the common driver genes, and approximately 15% patients with colorectal cancer carry all mutations of the three genes. Exploring the pathway of mutations in these gene is extremely useful for the diagnosis and treatment of cancer. However, the mutation orders of these genes may vary in different patients due to the heterogeneity of tumor. Hence, we discuss all possible mutation orders in the genes,

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

KRAS, APC and TP53 by using the model with five hits and find out the mutation pathways of genes that can effectively fit the incidence rate of colorectal cancer at different age in this article. In addition, we give the estimated values of mutation rates in each pathway that can explain the procession of colorectal cancer. The results obtained can offer guidelines to the treatment strategy of colorectal cancer.

Introduction

Colorectal cancer is one of the most common malignant tumors of digestive tract, which is the cancer with the third highest incidence and the second leading cause of mortality in the world [1]. Cancer is widely believed to have been caused by the accumulation of genetic and epigenetic alterations that result in the transformation of normal colonic epithelium to malignant cells. Recent studies showed that the progression from normal cells to the first malignant cell was supported by the alternations in three driver genes involving APC, KRAS and TP53, all of which were the most significant driver mutations in the colorectal cancer [2, 3]. However, not all colorectal cancers harbor the alterations of these three genes. Statistical analysis suggested that approximately 15% of colorectal cancers contained all mutations in APC, KRAS and TP53, and approximately 20% of tumor carried the mutations in both APC and KRAS [4, 5]. Among them, APC and TP53 are tumor suppressor genes (TSGs), and KRAS is an oncogene. It is well known that activation of oncogene, KRAS, only needs one hit, and inactivation of the TSGs, APC and TP53, requires two hits [6]. Hence, it takes five hits to turn the normal cells into the malignant cells for the primary colorectal cancer involving three driver genes, APC, KRAS and TP53.

Studying the mechanism of cancer formation by using the mathematical model is of essential importance in the prediction of tumor risk. The mathematical models considering the number of gene mutation were presented to explore the mechanism of cancerogenesis by a large number of scholars. The most revealing cancer models were the multistage model proposed by Armitage and Doll as early as 1950s [7, 8]. Their models were used to match age-specific mortality rate of various cancers, and the results indicated that the logarithm of mortality rate was linear with the logarithm of age. Subsequently, they considered the clonal expansion of cells and presented the two-stage model with clonal expansion of cells, widely applied to study the risk of various cancer [9]. Notably, Knudson utilized this model to fit incidence rate data of retinoblastoma and discovered the RB gene, the earliest TSG [10]. In addition, Moolgavkar et al. developed the method of solving the hazard function and the probability of tumor for the stochastic two-stage model involving selective growth advantage of cells and explained the detailed biological meaning of the model [11, 12]. However, two mutations are not sufficient to pose all cancers. Therefore, the two-stage model with clonal expansion of cells was extended to the model with more than two events where mutations accumulate in a specific order, allowing the intermediate stages to provide selective growth advantage to mutated cells [13–16]. It is showed that the model with more than two mutations better fits the incidence data of colorectal cancer compared to the two-stage model. Besides, Lang et al. used the two-type branching process model to study the dynamic behaviour of adenoma growth and transition into carcinomas in colorectal [17]. Nevertheless, these models did not consider the detailed mechanism of gene mutations.

The detailed mutation mechanism of driver gene is extremely urgent for colorectal cancer patients at different ages, which can provide insights into the strategies for detection and treatment of tumor with a more effective way. Recently, the mathematical model involving the

specific driver genes of colorectal cancer was developed to analyze the procession of malignant transformation in the colorectal [18, 19]. The stochastic model of malignant transformation with the losses of APC and TP53 and gain of KRAS presented by Paterson et al. was used to study the specific order of these genes and indicated that inactivation of APC initiated the progression of tumor in more than half of colorectal cancer cases [18]. Subsequently, Zhang et al. studied the sizes of cells with different gene mutation and the waiting time distribution of driver gene mutations in colorectal cancer by using the five-hit branching process model based on the result of Paterson et al [20]. These work proposed the approximate analytic solution with only net growth of cells and analyzed the procession of malignant transformation in the colorectal cancer by using the solution and fixing the values of some parameters from references and some mutational data. However, they did not consider the effect of the death rate of cells on the risk of colorectal cancer. In addition, the rates of mutation and loss of heterozygosity (LOH) are not easy decided, since the rates of mutation and LOH may vary as the type of genetic instability and the mutation order of gene mutation in the colorectal cancer [6, 21]. Therefore, our work considers the effect of the growth rate and death rate of cells on the risk of colorectal cancer and does not employ the value of gene mutation rate gave by reference [18].

We use the mathematical model with five hits considering the selective advantage provided by the KRAS, APC and TP53 genes. In the model, any order of mutations in genes KRAS, APC and TP53 is allowed, leading to multiple evolutionary pathways that induce the development of colorectal cancer. Moreover, all possible mutation pathways of each order are considered. We first compare the approximate analytic solution and the exact numerical solution we used. Our result shows that approximate solution will get worse in estimating the risk of cancer at young ages (before 25 ages) when the growth rate or death rate of cells is enough big (e.g. 50 per year). Thus, the exact numerical solution of the model is chosen to explore the mechanism of mutations in genes KRAS, APC and TP53 that are most likely to occur in the development of colorectal cancer. In our fitting, approximate Bayesian computation schemes with the simulated likelihood density is used to estimate the mutation rates of model [22]. Finally, we analyze the models with specific mutation order that can fit well the specific-age incidence data of colorectal cancer and apply them to estimate the expected number of mutated cells with alternations of different genes in the early stages of colorectal carcinogenesis.

Materials and methods

Recent studies show that three driver genes are sufficient to produce the tumor in colorectal cancer, which involve two TSGs and one oncogene [3, 18, 23]. It is familiar that TSG needs two mutations to loss function keeping the growth of cells, and the oncogene is activated by a

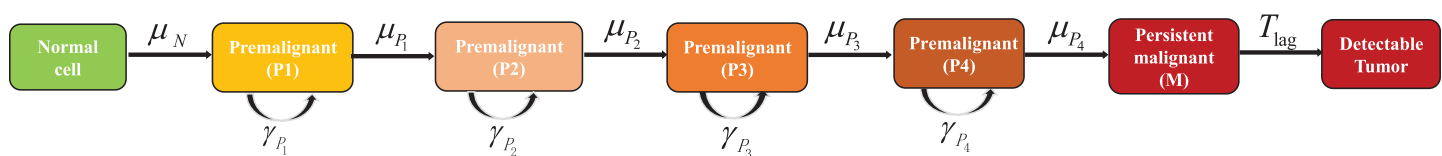


Fig 1. The model with five mutations involving two TSGs (APC and TP53) and one oncogene (KRAS). P_i ($i = 1, 2, 3, 4$) denotes the premalignant cells with i mutations, and M denotes the persistent malignant cells. μ_N denotes the mutation rate per normal stem cell per year, μ_{P_i} denotes the mutation rate per premalignant cell with i mutations per year, and λ_{P_i} denotes the net growth rate per premalignant cell per year, which is equal to be the difference between the rate of growth of cells and the death rate of cells. T_{lag} denotes the lag time from a persistent malignant cell to detectable tumor.

<https://doi.org/10.1371/journal.pcbi.1011225.g001>

single mutation. Hence, the normal cells require five mutations to become tumor for colorectal cancer involving TP53, APC and KRAS. The detailed model is displayed in Fig 1.

In Fig 1, the net growth rate of cells, λ_{p_i} , is equal to the difference between growth rate of cells (α_{p_i}) and death rate of cells (β_{p_i}). We define the following variables,

N —the number of normal cells in all crypts;

$P_i(t)$ —the number of premalignant cells with i mutations at time t ;

$M(t)$ —the number of persistent malignant cells at time t ;

$D(t)$ —whether the tumor is detected clinically at time t , and the value is 1 if tumor detected, otherwise the value is 0.

In the model, μ_{p_4} represents the effective transformation rate from premalignant cell to persistent malignant one that does not die. Let τ be the time of the first persistent malignant cell. The persistent malignant cell is produced at rate $\mu_{p_4} P_4(s)$ at time s , the probability that persistent malignant cell doesn't show up by time t yields [24],

$$P(\tau > t | P_4(s), s \leq t) = \exp\{-\mu_{p_4} \int_0^t P_4(s) ds\}. \quad (1)$$

Then the probability that a persistent malignant cell occurs by time t is given by

$$P(t) = P(t \geq \tau) = 1 - \exp\{-\mu_{p_4} \int_0^t P_4(s) ds\}. \quad (2)$$

Here, we are mainly concerned with the progression from normal cells to a persistent malignant cell, and the progression from a persistent malignant cell to the tumor detected clinically is assumed to be a fixed time, T_{lag} . Thus, the probability that tumor is clinically detected at time t can be written as

$$P(D(t) = 1) = P(t - T_{lag} \geq \tau) = 1 - \exp\{-\mu_{p_4} \int_0^{t-T_{lag}} P_4(s) ds\}. \quad (3)$$

The hazard function, that is, the incidence rate of primary malignant tumors at time t , $h(t)$, follows that

$$h(t) = -\frac{\ln(P(D(t) = 0))}{dt} = -\frac{\ln(P(\tau > t - T_{lag}))}{dt} = \mu_{p_4} P_4(t - T_{lag}) \approx \mu_{p_4} E[P_4(t - T_{lag})]. \quad (4)$$

The expected number of premalignant cells can be decided by the following system of equations

$$\begin{cases} \frac{dE[P_1(t)]}{dt} = \lambda_{p_1} E[P_1(t)] + \mu_N N, \\ \frac{dE[P_2(t)]}{dt} = \lambda_{p_2} E[P_2(t)] + \mu_{p_1} E[P_1(t)], \\ \frac{dE[P_3(t)]}{dt} = \lambda_{p_3} E[P_3(t)] + \mu_{p_2} E[P_2(t)], \\ \frac{dE[P_4(t)]}{dt} = \lambda_{p_4} E[P_4(t)] + \mu_{p_3} E[P_3(t)]. \end{cases} \quad (5)$$

The detailed derivation of the above equation is depicted in S1 Text. By solving Eqs (4) and

(5), we get

$$h(t) \approx N\mu_N\mu_{p_1}\mu_{p_2}\mu_{p_3}\mu_{p_4} \left[\frac{\exp\{\lambda_{p_4}(t - T_{lag})\}}{\lambda_{p_4}(\lambda_{p_4} - \lambda_{p_3})(\lambda_{p_4} - \lambda_{p_2})(\lambda_{p_4} - \lambda_{p_1})} - \frac{\exp\{\lambda_{p_3}(t - T_{lag})\}}{\lambda_{p_3}(\lambda_{p_3} - \lambda_{p_2})(\lambda_{p_3} - \lambda_{p_1})(\lambda_{p_3} - \lambda_{p_4})} + \frac{\exp\{\lambda_{p_2}(t - T_{lag})\}}{\lambda_{p_2}(\lambda_{p_2} - \lambda_{p_1})(\lambda_{p_2} - \lambda_{p_3})(\lambda_{p_2} - \lambda_{p_4})} - \frac{\exp\{\lambda_{p_1}(t - T_{lag})\}}{\lambda_{p_1}(\lambda_{p_1} - \lambda_{p_2})(\lambda_{p_1} - \lambda_{p_3})(\lambda_{p_1} - \lambda_{p_4})} + \frac{1}{\lambda_{p_1}\lambda_{p_2}\lambda_{p_3}\lambda_{p_4}} \right]. \quad (6)$$

The above approximate solution is supported when the probability of persistent malignant cell, $P(M(t) > 0)$, is close to zero [11]. However, it is not applied to cancers with high incidence rate [11]. Moreover, this approximate solution depends on the net growth rates and the mutation rates of cells, which doesn't reflect the effects of the growth or death rate of cells on the risk of cancer. Therefore, we analyze another solution of hazard function. Here we consider the growth rate and death rate of cells in the development of cancer. Let α_{p_i} and β_{p_i} denote the growth rate and death rate of premalignant cells with i mutation(s), respectively.

To solve the hazard function, we define the following probability generating functions

$$\left\{ \begin{array}{l} \Psi(x_1, x_2, x_3, x_4, y; 0, t) = \sum_{i_1, i_2, i_3, i_4, j} \text{prob}\{P_1(t) = i_1, P_2(t) = i_2, P_3(t) = i_3, \\ P_4(t) = i_4, M(t) = j | P_1(0) = 0, P_2(0) = 0, P_3(0) = 0, P_4(0) = 0, M(0) = 0\} x_1^{i_1} x_2^{i_2} x_3^{i_3} x_4^{i_4} y^j, \\ \Phi_1(x_1, x_2, x_3, x_4, y; 0, t) = \sum_{i_1, i_2, i_3, i_4, j} \text{prob}\{P_1(t) = i_1, P_2(t) = i_2, P_3(t) = i_3, \\ P_4(t) = i_4, M(t) = j | P_1(0) = 1, P_2(0) = 0, P_3(0) = 0, P_4(0) = 0, M(0) = 0\} x_1^{i_1} x_2^{i_2} x_3^{i_3} x_4^{i_4} y^j, \\ \Phi_2(x_2, x_3, x_4, y; 0, t) = \sum_{i_2, i_3, i_4, j} \text{prob}\{P_2(t) = i_2, P_3(t) = i_3, P_4(t) = i_4, M(t) = j | P_2(0) = 1, P_3(0) = 0, P_4(0) = 0, M(0) = 0\} \\ x_2^{i_2} x_3^{i_3} x_4^{i_4} y^j, \\ \Phi_3(x_3, x_4, y; 0, t) = \sum_{i_3, i_4, j} \text{prob}\{P_3(t) = i_3, P_4(t) = i_4, M(t) = j | P_3(0) = 0, P_4(0) = 0, M(0) = 0\} x_3^{i_3} x_4^{i_4} y^j, \\ \Phi_4(x_4, y; 0, t) = \sum_{i_4, j} \text{prob}\{P_4(t) = i_4, M(t) = j | P_4(0) = 0, M(0) = 0\} x_4^{i_4} y^j. \end{array} \right. \quad (7)$$

It is easy to obtain the following survival function, that is, probability of no tumor at time t ,

$$S(t) = \Psi(1, 1, 1, 1, 0; 0, t - T_{lag}) \quad (8)$$

and then

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{\Psi'(1, 1, 1, 1, 0; 0, t - T_{lag})}{\Psi(1, 1, 1, 1, 0; 0, t - T_{lag})} \quad (9)$$

The probability generating functions in formulas (7) satisfy the following Kolmogorov backward equations [25, 26],

$$\begin{cases} \frac{d\Psi(1, 1, 1, 1, 0; 0, t)}{dt} = \mu_N N[\Phi_1(1, 1, 1, 1, 0; 0, t) - 1], \\ \frac{d\Phi_1(1, 1, 1, 1, 0; 0, t)}{dt} = \alpha_{p_1} \Phi_1^2(1, 1, 1, 1, 0; 0, t) - (\alpha_{p_1} - \beta_{p_1} - \mu_{p_1}) \Phi_1(1, 1, 1, 1, 0; 0, t) + \mu_{p_1} \Phi_1(1, 1, 1, 1, 0; 0, t) \Phi_2(1, 1, 1, 0; 0, t) + \beta_{p_1}, \\ \frac{d\Phi_2(1, 1, 1, 0; 0, t)}{dt} = \alpha_{p_2} \Phi_2^2(1, 1, 1, 0; 0, t) - (\alpha_{p_2} - \beta_{p_2} - \mu_{p_2}) \Phi_2(1, 1, 1, 0; 0, t) + \mu_{p_2} \Phi_2(1, 1, 1, 0; 0, t) \Phi_3(1, 1, 0; 0, t) + \beta_{p_2}, \\ \frac{d\Phi_3(1, 1, 0; 0, t)}{dt} = \alpha_{p_3} \Phi_3^2(1, 1, 0; 0, t) - (\alpha_{p_3} - \beta_{p_3} - \mu_{p_3}) \Phi_3(1, 1, 0; 0, t) + \mu_{p_3} \Phi_3(1, 1, 0; 0, t) \Phi_4(1, 0; 0, t) + \beta_{p_3}, \\ \frac{d\Phi_4(1, 0; 0, t)}{dt} = \alpha_{p_4} \Phi_4^2(1, 0; 0, t) - (\alpha_{p_4} - \beta_{p_4} - \mu_{p_4}) \Phi_4(1, 0; 0, t) + \beta_{p_4}. \end{cases} \quad (10)$$

The detailed derivation of above equation can be seen in S1 Text. By the definitions of the probability generating functions, the initial values of equations (10) are as follows

$$\begin{cases} \Psi(1, 1, 1, 1, 0; 0, 0) = 1, \\ \Phi_1(1, 1, 1, 1, 0; 0, 0) = 1, \\ \Phi_2(1, 1, 1, 0; 0, 0) = 1, \\ \Phi_3(1, 1, 0; 0, 0) = 1, \\ \Phi_4(1, 0; 0, 0) = 1. \end{cases} \quad (11)$$

Then, hazard function can be obtained by solving equations (10) with the initial conditions (11).

Results

Comparison of two solutions

To analyze the effect of the growth rate of cells on the risk of cancer, we fix the following parameter values

$$N = 10^7, \quad \mu_{p_i} = 10^{-5}, \quad \lambda_{p_i} = 0.1 + (i - 1) * 0.05,$$

and let the growth rate of cells, α_{p_i} , takes different values, for example, $\alpha_{p_i} = 0.5$, $\alpha_{p_i} = 1$, $\alpha_{p_i} = 10$ and $\alpha_{p_i} = 50$. The logarithm of hazard functions from formula (6) and equations (10) with the initial conditions (11) are illustrated in Fig 2(A), and the corresponding absolute errors are seen in Fig 2(C). From Fig 2(A) and 2(C), we find that the difference between the approximate solution and the exact numerical solution of the model goes up with the growth rate of cells increases, especially for older patients. The higher the growth rate α_{p_i} , the larger the error between the approximate solution and the exact numerical solution of the model at an earlier time (younger ages). In like manner, the result holds true by varying death rate of cells (see Fig 2(B) and 2(D)). For the pure birth process ($\beta_{p_i} = 0$), the difference between the approximate solution and the exact numerical solution of the model is still significant. Then, Fig 2 shows that the approximate solution is far greater than the exact numerical solution when the division rate or death rate of cells is quite large. In addition, the approximate solution only reflects the effect of net growth rate on hazard function. Nevertheless, the exact numerical solution involves the growth rate and death rate of cells in addition to net growth rate of cells. In fact, the approximate solution is given assuming that the probability of malignant cells is zero [11]. The approximate solution will get worse with an increase in the growth rate or death rate of

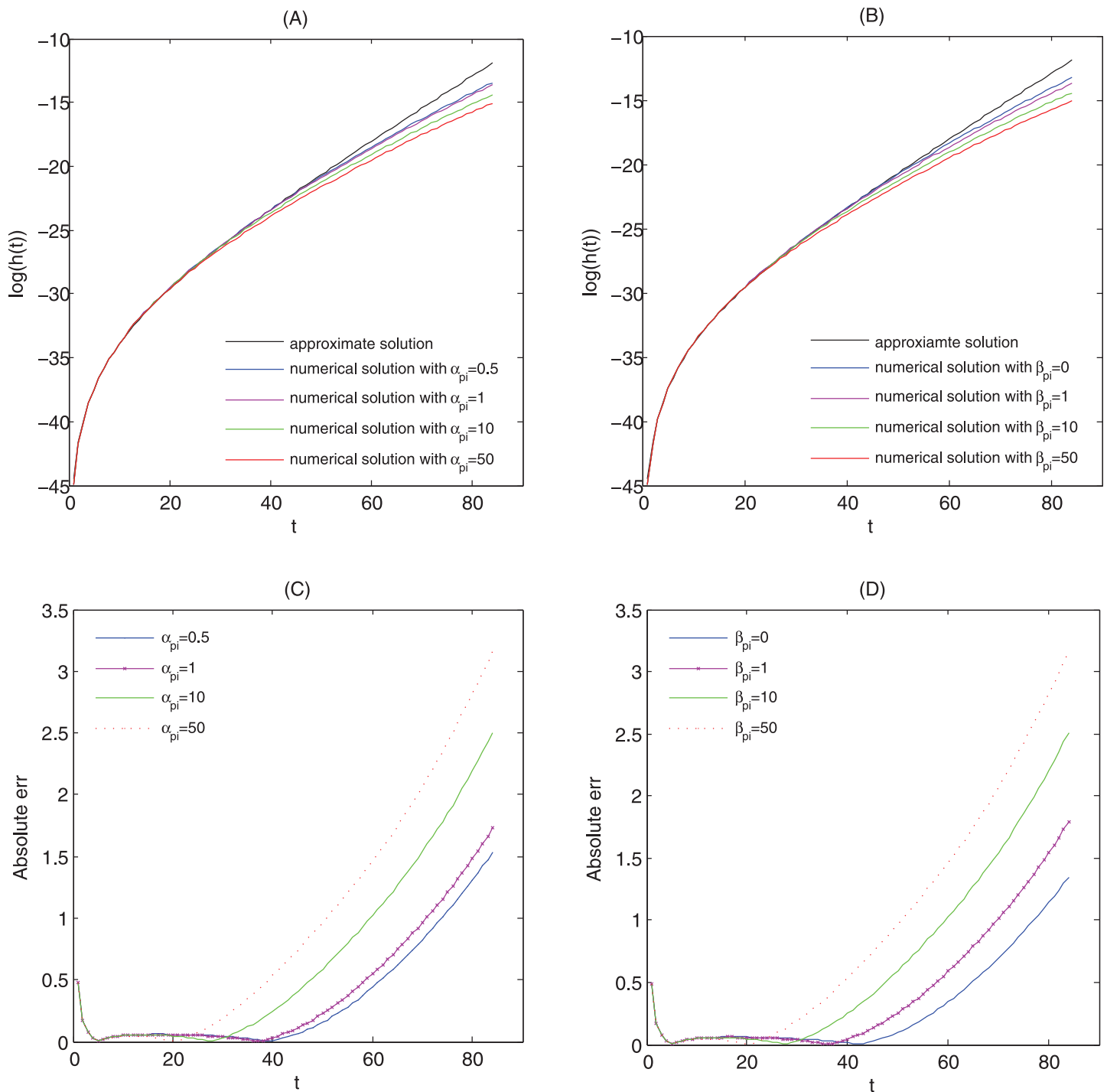


Fig 2. Comparison of the approximate hazard function given by (6) with the exact numerical solution of hazard function given by (9). All parameters of the model are set as follows: $N = 10^7$, $\mu_N = \mu_{pi} = 10^{-5}$, $(i = 1, 2, 3, 4)$, $\lambda_i = 0.1 + (i - 1) * 0.05$.

<https://doi.org/10.1371/journal.pcbi.1011225.g002>

cells. Therefore, the approximate solution may not be an excellent choice to simulate the risk of cancer, especially for a cancer with a high incidence rate.

For stem cell in colon, it divides on average every five days [27]. It is shown that a stem cell produces two identical daughter cells by symmetry division and one mutated cell and one

equivalent daughter cell by asymmetric division. In the model, the mutation rate, μ_{p_i} , signifies an asymmetric division rate of per cell per year. That is, the sum of the mutation rate (μ_{p_i}) and the growth rate (α_{p_i}) of the model is equal to the division rate of per cell per year. However, the mutation rate, μ_{p_i} , is far less than the growth rate, α_{p_i} . Therefore, it is easy to estimate the growth rate of stem cells, approximately 73 per cell per year. By the Fig 2, using the approximate solution may produce a large difference in estimating the risk of colorectal cancer compare to the exact numerical solution. As a consequence, we simulate the incidence rate of colorectal cancer at specific age by using the numerical solution obtained by equations (9), (10) and (11).

The order of gene alterations

Gene mutation is the leading cause of drug resistance whose emergence is a great obstacle for tumor treatment [28, 29]. For example, the mutation of KRAS leads to the resistant to gefitinib, erlotinib and cetuximab which induce the apoptosis of cancer cells in the therapy of cancer [30, 31]. Hence, knowing the order of gene alterations in a tumor is extremely crucial for the early treatment of cancer, which provides the guidance to the selection of medicines. Here, we investigate all possible mutation orders of three driver gene in the colorectal cancer by the mathematical modelling. Tumor, including all alterations of the three drive genes, involves six cases with the order of gene alterations. The detailed descriptions are seen in Fig 3. All possible pathways of alterations in the genes, KRAS, APC and TP53, are displayed in Figs A–F in S1 Text.

TSG can inhibit tumor formation by inducing the death of abnormal cell, whose function is blocked only when both alleles are inactivated ($TSG^{-/-}$). In other words, the cells do not grow abnormally when only one allele is inactivated ($TSG^{+/-}$). However, the activation of oncogene just takes one hit (oncogene⁺), which promotes the proliferation of cells. Thus, by the clonal

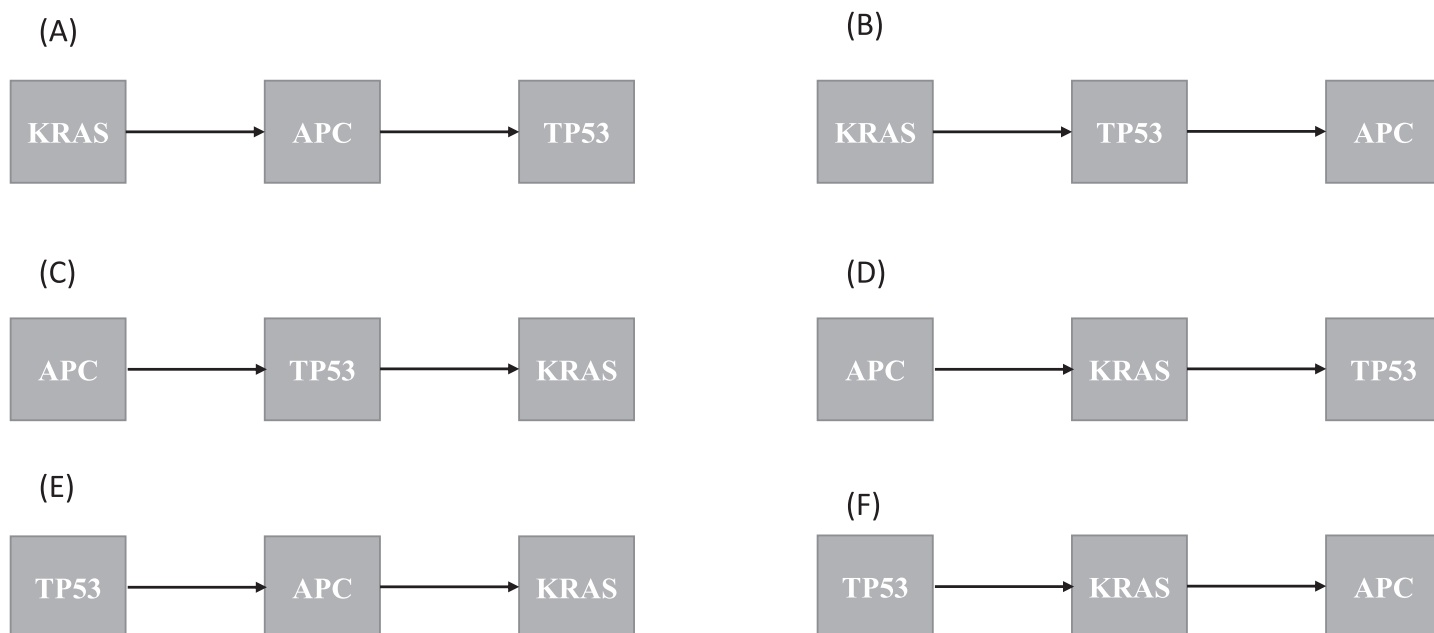


Fig 3. All orders of gene mutations for three driver genes, KRAS, APC and TP53, in colorectal cancer.

<https://doi.org/10.1371/journal.pcbi.1011225.g003>

expansion of premalignant cells, all possible pathways for the alterations of three genes can be summarized as follows:

1. $\lambda_{p_2} = \lambda_{p_1}$ and $\lambda_{p_4} = \lambda_{p_3}$, which involves the following sequences of mutations

- $KRAS^+ \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$
- $KRAS^+ \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-}$

2. $\lambda_{p_3} = \lambda_{p_2} = \lambda_{p_1}$, which involves the following sequences of mutations

- $KRAS^+ \rightarrow APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$
- $KRAS^+ \rightarrow TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$
- $KRAS^+ \rightarrow APC^{+/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{-/-}$
- $KRAS^+ \rightarrow TP53^{+/-} \rightarrow APC^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{-/-}$

3. $\lambda_{p_1} = 0$ and $\lambda_{p_3} = \lambda_{p_2}$, which involves the following sequences of mutations

- $APC^{+/-} \rightarrow KRAS^+ \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$
- $TP53^{+/-} \rightarrow KRAS^+ \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$
- $APC^{+/-} \rightarrow KRAS^+ \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{-/-}$
- $TP53^{+/-} \rightarrow KRAS^+ \rightarrow APC^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{-/-}$
- $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^+$
- $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-}$
- $TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^+$
- $TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{+/-} \rightarrow KRAS^+ \rightarrow APC^{-/-}$

4. $\lambda_{p_1} = 0$ and $\lambda_{p_4} = \lambda_{p_3}$, which involves the following sequences of mutations

- $APC^{+/-} \rightarrow KRAS^+ \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$
- $TP53^{+/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-}$
- $APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^+ \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$
- $TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^+ \rightarrow APC^{+/-} \rightarrow APC^{-/-}$

5. $\lambda_{p_1} = 0$ and $\lambda_{p_2} = 0$, which involves the following sequences of mutations

- $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow KRAS^+ \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$
- $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow KRAS^+ \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$
- $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-} \rightarrow APC^{-/-}$
- $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-} \rightarrow APC^{-/-}$
- $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^+$
- $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^+$
- $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-}$
- $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-}$

- $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+}$
- $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow TP53^{-/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+}$
- $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+} \rightarrow APC^{-/-}$
- $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+} \rightarrow APC^{-/-}$

There are five cases for the model with five hits by the above analyses, involving 30 pathways of gene mutations. To determine the specific order of three drive genes alterations, we choose the numerical solution of the model to fit the incidence rate data of colorectal cancer from the Surveillance Epidemiology and End Results (SEER) registry during the period 1973–2013. However, not all parameters of the model can be estimated by the data alone. Here, we assumed the growth rate of normal cells to be 73 per year, and the lag time from a persistent malignant cell to tumor detected clinically to be 5 years [27, 32]. In addition, all mutation rates of cells (μ_{p_i}) are limited in the range $(0, 10^{-2})$, since the probability of a gene mutation (μ_{p_i}) is far less than one.

The activation of oncogene promotes the proliferation of cells, and the differentiation or apoptosis mechanism of cells will disarrange if the inactivation of TSG happens [6]. Studies showed that crypts carrying one of APC and KRAS alterations (that is the inactivation of APC and activation of KRAS) confer a selective growth advantage to cells [33, 34]. It has been shown that the net growth rate of the cells including only $APC^{-/-}$ is 0.2 per year, and 0.07 per year for the cells with only $KRAS^{+}$ [35, 36]. Furthermore, the inactivation of TP53 does not provide any growth advantage to cells in normal conditions [37]. We assume that mutations between genes have no interaction in cell growth. Therefore, the growth rate and death rate of premalignant cells with different gene alterations in the model are set to be as follows:

- for the premalignant cells with $TP53^{-/-}$, the growth rate and death rate are 73 per year and 73 per year, respectively;
- for the premalignant cells with $KRAS^{+}$ or both $TP53^{-/-}$ and $KRAS^{+}$, the growth rate and death rate are 73.07 per year and 73 per year, respectively;
- for the premalignant cells with $APC^{-/-}$ or both $TP53^{-/-}$ and $APC^{-/-}$, the growth rate and death rate are 73 per year and 72.8 per year, respectively;
- for the premalignant cells with both $APC^{-/-}$ and $KRAS^{+}$, the growth rate and death rate are 73.07 per year and 72.8 per year, respectively.

In our simulations, the fourth-order Runge-Kutta is utilized to solve Equations (10) with the initial values (formulas (11)), and the model parameters, $\Theta = (v, \mu_{p_1}, \mu_{p_2}, \mu_{p_3}, \mu_{p_4})$ where $v = N\mu_N$, are estimated by using approximate Bayesian computation schemes involving the simulated likelihood density [22].

We do twenty fits of model with the different proliferation rate of cells (λ_{p_i}), since mutation in first allele of TP53 and APC ($TP53^{+/-}$ and $APC^{+/-}$) does not bring about the change in proliferation rate of cell. These fits correspond to the pathways in Figs A–F in S1 Text. Our fitting results suggest that only the orders with $KRAS-APC-TP53$, $APC-TP53-KRAS$ and $APC-KRAS-TP53$ can accepted to explain the incidence rate of colorectal cancer at different ages, involving eleven pathways (see Fig 4). There are four pathway of mutations for the orders with $KRAS-APC-TP53$ and $APC-KRAS-TP53$, and three sequences of mutations for the order with $APC-TP53-KRAS$. As a consequence, the first event is just activation of KRAS or inactivation of APC, and the inactivation of TP53 is usually a late event in the development of colorectal cancer.

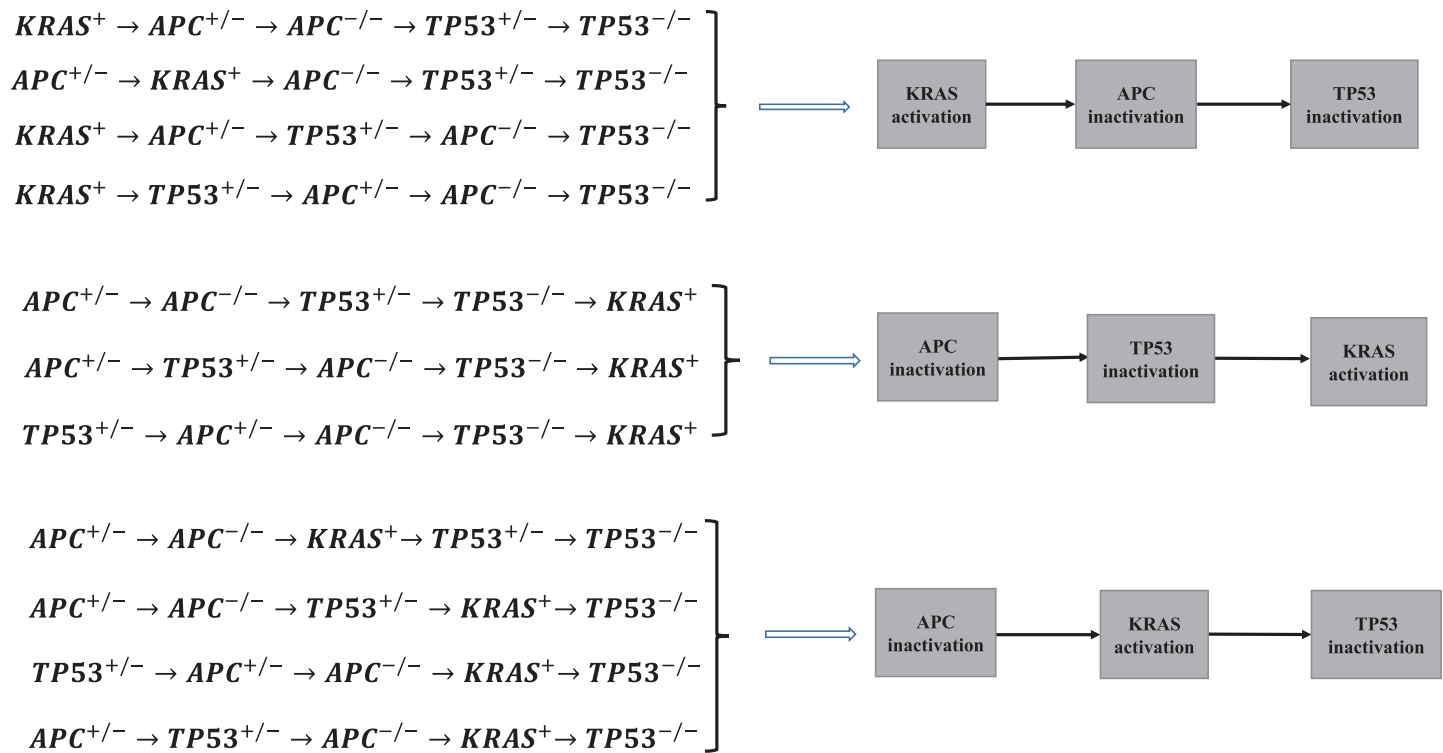


Fig 4. All pathways are accepted to explain the development of colorectal cancer at different ages.

<https://doi.org/10.1371/journal.pcbi.1011225.g004>

By the previous analyses, the five-stage models with $KRAS^+ \rightarrow APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$ and $KRAS^+ \rightarrow TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$ (correspond to Fig A3 of Fig A in S1 Text) have the same parameter values for the order with $KRAS - APC - TP53$. It is the same for the order with $APC - TP53 - KRAS$ involving $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^+$ and $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^+$ (correspond to Fig C2 of Fig C in S1 Text), and the order with $APC - KRAS - TP53$ involving $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-}$ and $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-}$ (correspond to Fig D3 of Fig D in S1 Text). The fitting results of the pathways in Fig 4 are displayed in Figs 5–12, and the detailed values of parameters are shown in Tables 1–8. They show that the first mutation rate is extremely low for all pathways in Fig 4, and the mutation rate of the same gene is different in different mutation pathways. These parameter estimates can provide some evidences for inferring the type of gene alteration.

For the pathway with $KRAS^+ \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$, Table 1 shows the rate of mutation on first APC allele is larger than that of mutation on second APC allele, approximately 2 times, which infers that inactivation of APC is caused by mutation in both alleles, or by LOH in first allele and mutation in second allele. In addition, the inactivation rates of APC and TP53 are much greater than the activation rate of KRAS and the point mutation rate [38, 39]. The total number of driver positions in APC is 604 from reference [18], and then the mutation rate of APC is 7.55×10^{-6} if the base pair mutation rate takes 1.25×10^{-8} from [39]. Our estimated value of alteration in APC (4.70×10^{-3}) is much larger than 7.55×10^{-6} . These imply that the inactivations of APC or TP53 may be accompanied by genetic instability (microsatellite instability or chromosomal instability) that increases the mutation rate of gene. Evidences manifested that chromosomal instability might result from mutations in APC or TP53 [40, 41].

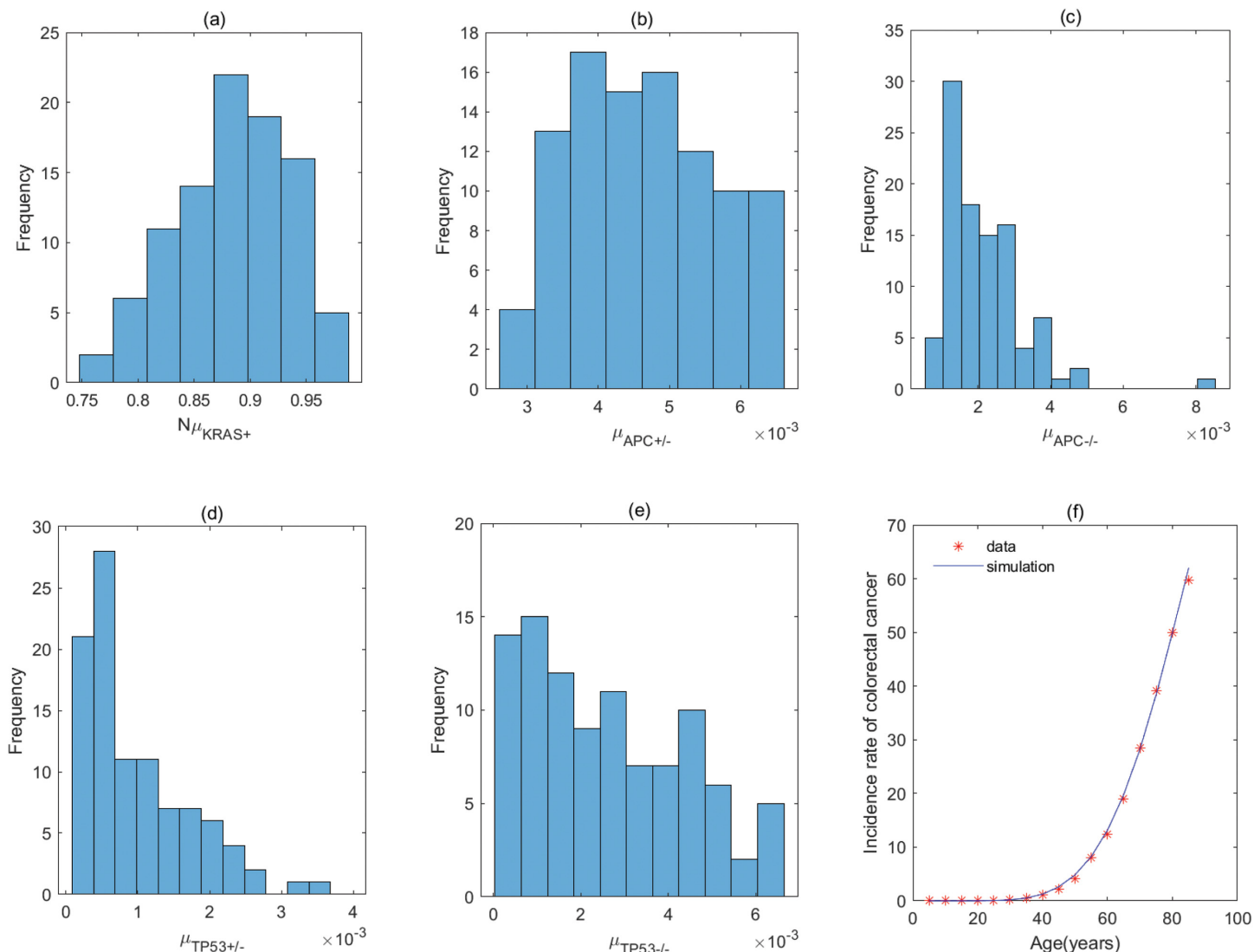


Fig 5. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $KRAS^+ \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$. (a) $N\mu_{KRAS^+}$ ($N\mu_N$ of model), (b) $\mu_{APC^{+/-}}$ (μ_{p_1} of model), (c) $\mu_{APC^{-/-}}$ (μ_{p_2} of model), (d) $\mu_{TP53^{+/-}}$ (μ_{p_3} of model), (e) $\mu_{TP53^{-/-}}$ (μ_{p_4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g005>

By comparing the mutation rate of the first event in the orders with $KRAS - APC - TP53$, $APC - TP53 - KRAS$ and $APC - KRAS - TP53$, we find that the initiated rate in the order with $KRAS - APC - TP53$ is lower than that in other orders with high probability in initiation of colorectal cancer. Tables 1–8 indicate that the transform rate (μ_{p_4}) of malignant cells is an extraordinarily high value for all pathways. It is shown that genetic instability must exist in the development of colorectal cancer. In addition, inactivation rate of APC ($\mu_{APC^{-/-}}$) in the pathways with $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^+$, $APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^+ \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$ and $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow KRAS^+ \rightarrow TP53^{-/-}$ is much less than that in other pathways, approximately $\frac{1}{10}$. Evidences suggested that the point mutation rate of genes with microsatellite instability is about ten times as much as that without

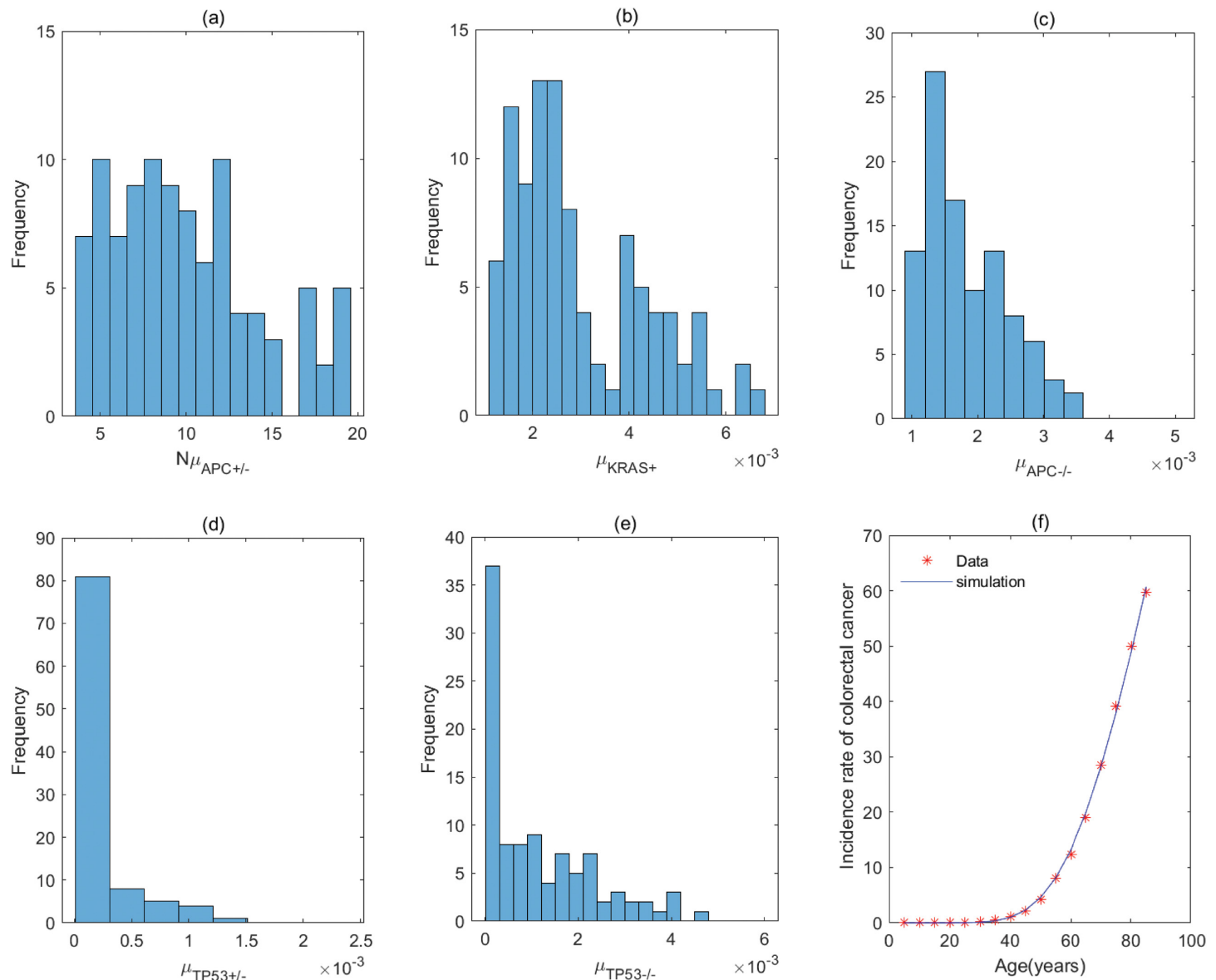


Fig 6. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $APC^{+/-} \rightarrow KRAS^{+} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$. (a) $N\mu_{APC+/-}$ ($N\mu_N$ of model), (b) μ_{KRAS+} (μ_{p_1} of model), (c) $\mu_{APC-/-}$ (μ_{p_2} of model), (d) $\mu_{TP53+/-}$ (μ_{p_3} of model), (e) $\mu_{TP53-/-}$ (μ_{p_4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g006>

microsatellite instability for colorectal cancer [42, 43]. It can be used as a mark in identifying the detailed mutation pathway for the orders with $APC - TP53 - KRAS$ and $APC - KRAS - TP53$.

The number of cells with different genetic alterations

We next use the five-stage models with $KRAS - APC - TP53$, $APC - TP53 - KRAS$ and $APC - KRAS - TP53$ to detect the changes of the mutated cells with different genetic mutations over time. By Fig 4, the cells with single genetic mutation have two types: the cells with activation of

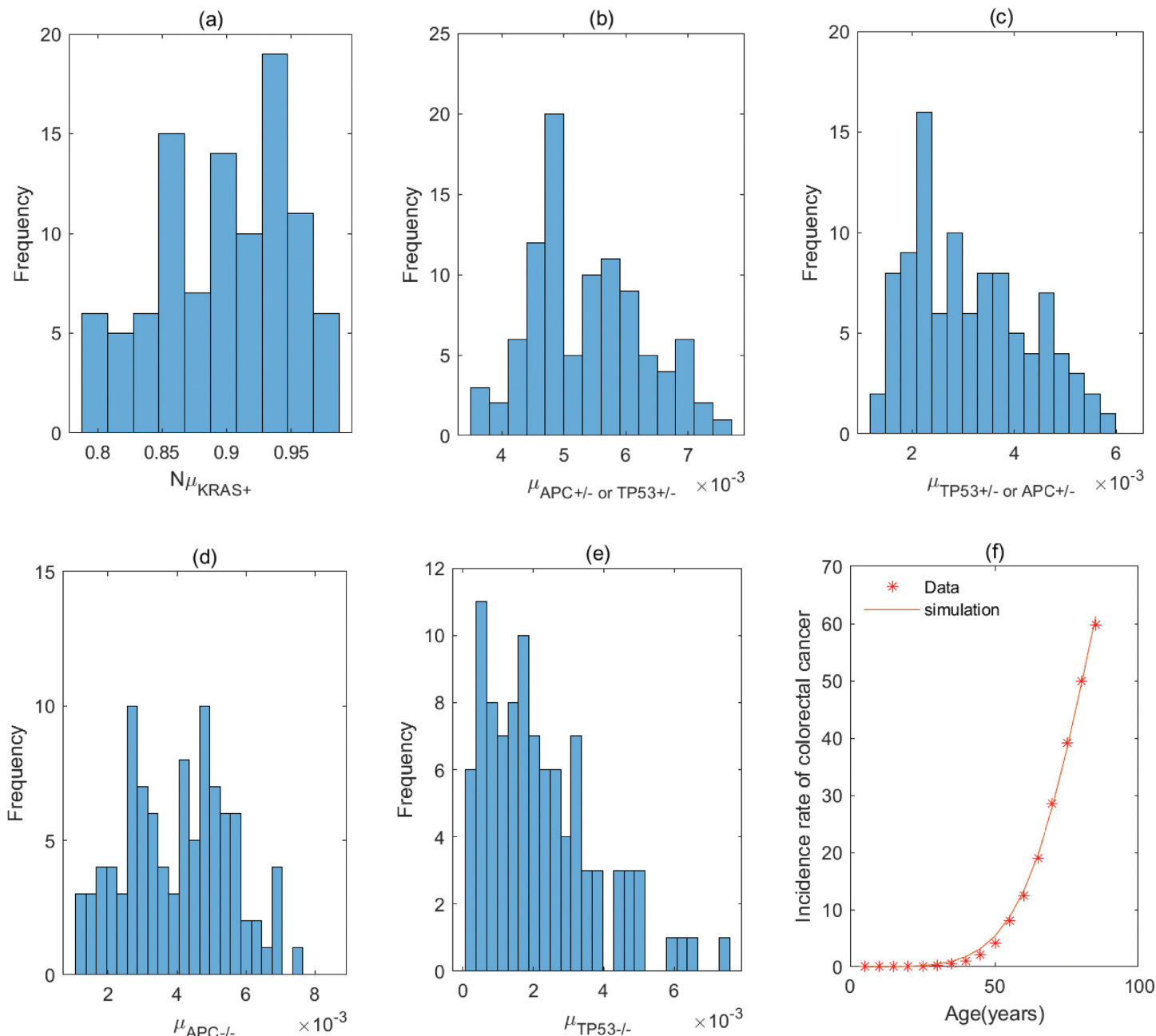


Fig 7. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $KRAS^+ \rightarrow APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$ or $KRAS^+ \rightarrow TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$. (a) $N\mu_{KRAS^+}$ ($N\mu_N$ of model), (b) $\mu_{APC^{+/-}} \text{ or } \mu_{TP53^{+/-}}$ (μ_{p_1} of model), (c) $\mu_{TP53^{+/-}} \text{ or } \mu_{APC^{+/-}}$ (μ_{p_2} of model), (d) $\mu_{APC^{-/-}}$ (μ_{p_3} of model), (e) $\mu_{TP53^{-/-}}$ (μ_{p_4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g007>

KRAS and those with inactivation of APC. In addition, there are three types of mutated cells with the combination of two genetic mutations. They are mutated cells with activation of KRAS followed by inactivation of APC, those with inactivation of APC followed by activation of KRAS, and those with inactivation of APC followed by inactivation of TP53. The detailed formulas about the expected numbers of mutated cells with different genetic alterations are

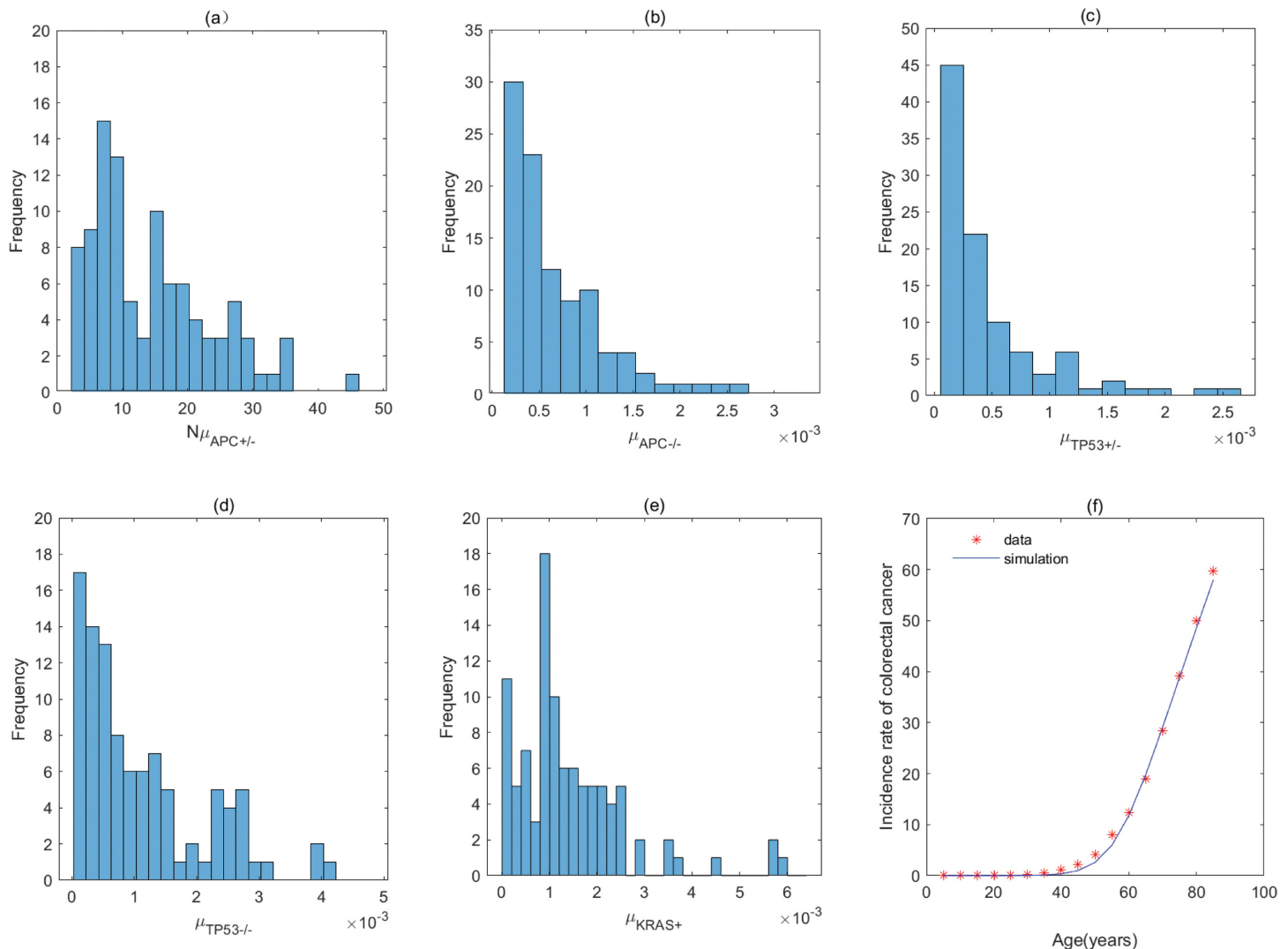


Fig 8. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+}$. (a) $N\mu_{APC+/-}$ ($N\mu_N$ of model), (b) $\mu_{APC-/-}$ (μ_{p_1} of model), (c) $\mu_{TP53+/-}$ (μ_{p_2} of model), (d) $\mu_{TP53-/-}$ (μ_{p_3} of model), (e) μ_{KRAS+} (μ_{p_4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g008>

given in S1 Text, and the changes in the number of mutated cells over time in different order of gene alterations are plotted in Fig 13.

Fig 13 shows that the number of KRAS-mutated cells is much higher than that of cells with both activated KRAS and inactivated APC at the beginning, and then the number of cells with both KRAS-mutated and APC-mutated surpasses that of KRAS-mutated cells by middle age in the order with $KRAS - APC - TP53$. For the orders with $APC - TP53 - KRAS$ and $APC - KRAS - TP53$, the number of cells with single-mutant outnumber those with double-mutant, throughout the human lifetime. Additionally, the number of cells with double-mutant is close to those with a single gene mutation over time for the order with $APC - KRAS - TP53$. These changes in the number of cells with different gene alterations are able to offer some clues in identifying the detailed order of gene alterations for patients with colorectal cancer.

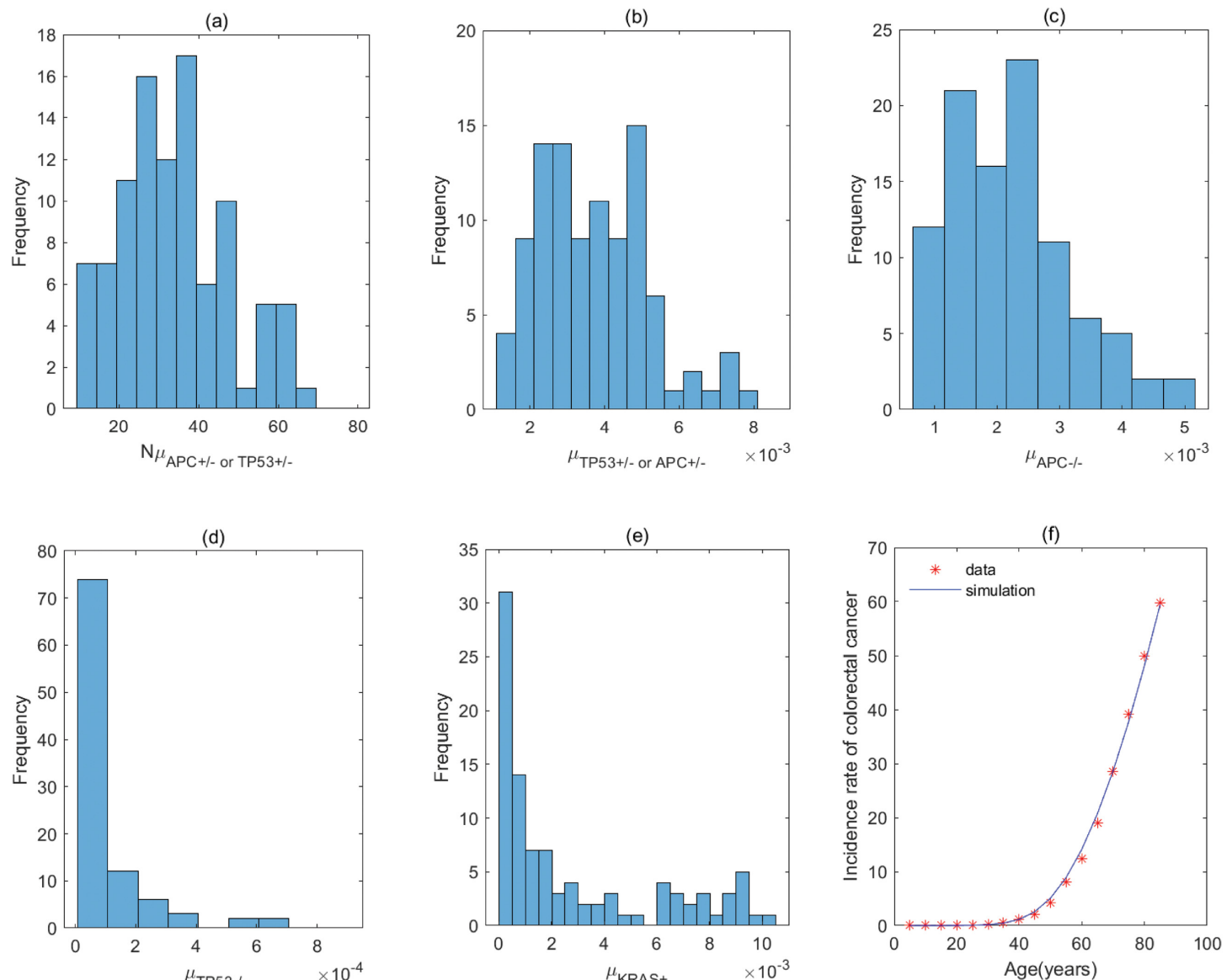


Fig 9. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+}$ or $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+}$. (a) $N\mu_{APC+/-}$ or $N\mu_{TP53+/-}$ ($N\mu_N$ of model), (b) $\mu_{TP53+/-}$ or $\mu_{APC+/-}$ (μ_{p_1} of model), (c) $\mu_{APC+/-}$ (μ_{p_2} of model), (d) $\mu_{TP53+/-}$ (μ_{p_3} of model), (e) μ_{KRAS+} (μ_{p_4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g009>

Discussion

The order of driver gene mutations is very vital for the clinical treatment and even prevention in the course of cancer. In this article, we construct the mathematical model considering all possible orders of mutations in APC, TP53 and KRAS that are the most significant driver genes in colorectal cancer. There are 30 pathways involving the alterations of all three driver genes, which need five hits from a normal cell to a malignant cell with all mutations of these driver genes. These pathways are classified as twenty cases based on the different proliferation rates of mutated cells, which corresponds with the five-stage model with different net growth

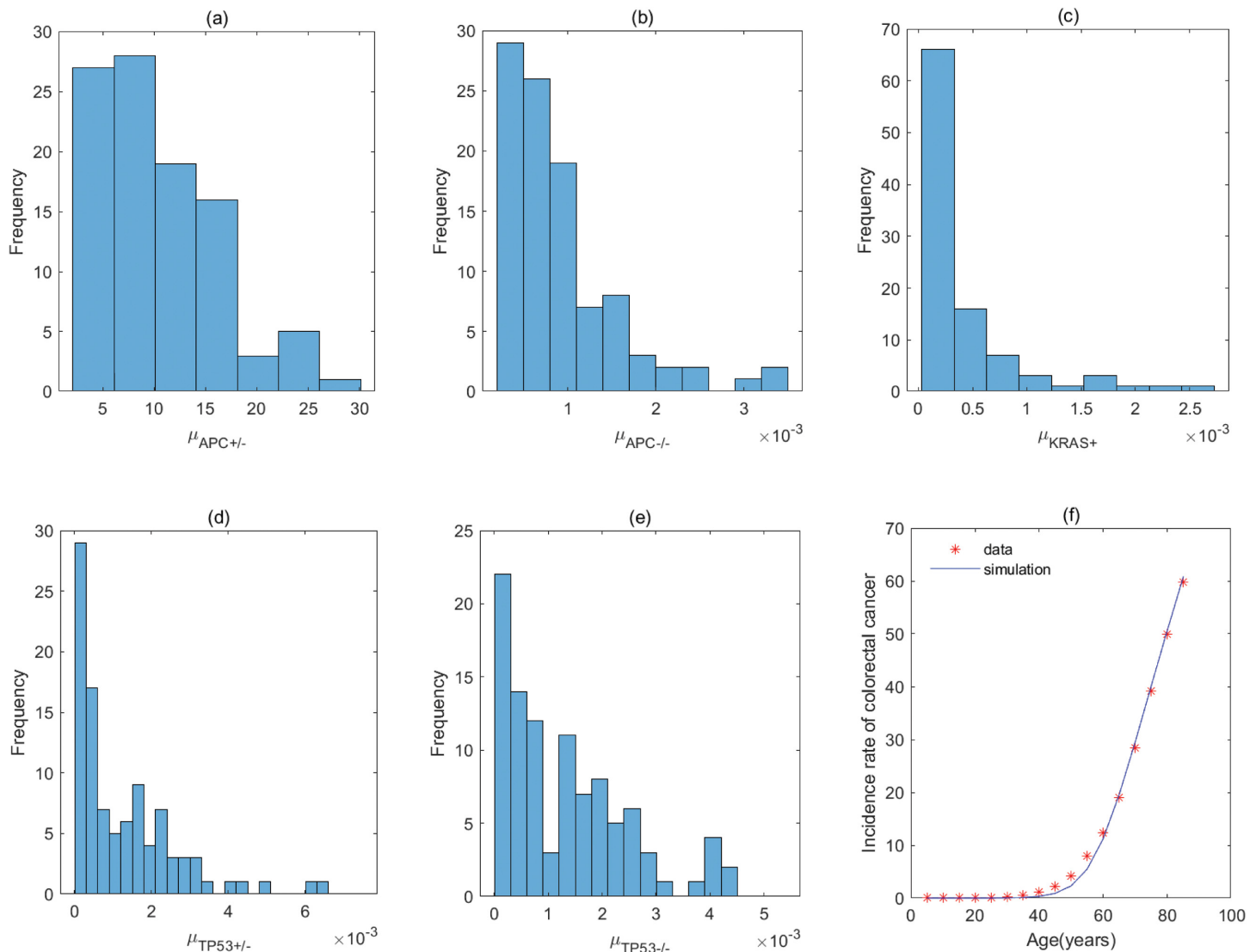


Fig 10. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$. (a) $N\mu_{APC+/-}$ ($N\mu_N$ of model), (b) $\mu_{APC+/-}$ (μ_{p1} of model), (c) μ_{KRAS+} (μ_{p2} of model), (d) $\mu_{TP53+/-}$ (μ_{p3} of model), (e) $\mu_{TP53-/-}$ (μ_{p4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g010>

rate of mutated cells (see Figs A–F in S1 Text). Studies showed that approximately 15% colorectal cancer harbored alterations in all of the three genes [4, 5]. Therefore, we use the constructed model based on the net proliferation rates of mutated cells to match 15% incidence rate of colorectal cancer for male and female in different ages instead of those of a fixed age [18]. Our fitting results indicate that three mutation orders with $KRAS \rightarrow APC \rightarrow TP53$, $APC \rightarrow TP53 \rightarrow KRAS$, $APC \rightarrow KRAS \rightarrow TP53$, are supported to lead to tumor, involving eleven pathways. Among them, there are seven pathways that inactivation of APC acts as the first event in the colorectal cancer, and the inactivation of TP53 is the last event when the first event is activation of KRAS in the development of colorectal cancer. This is line with the results that the alternations of APC is an early driver event that accounts for 80% of colorectal cancers [44, 45]. Inactivation of TP53 is critical in the transformation from early adenoma to advanced

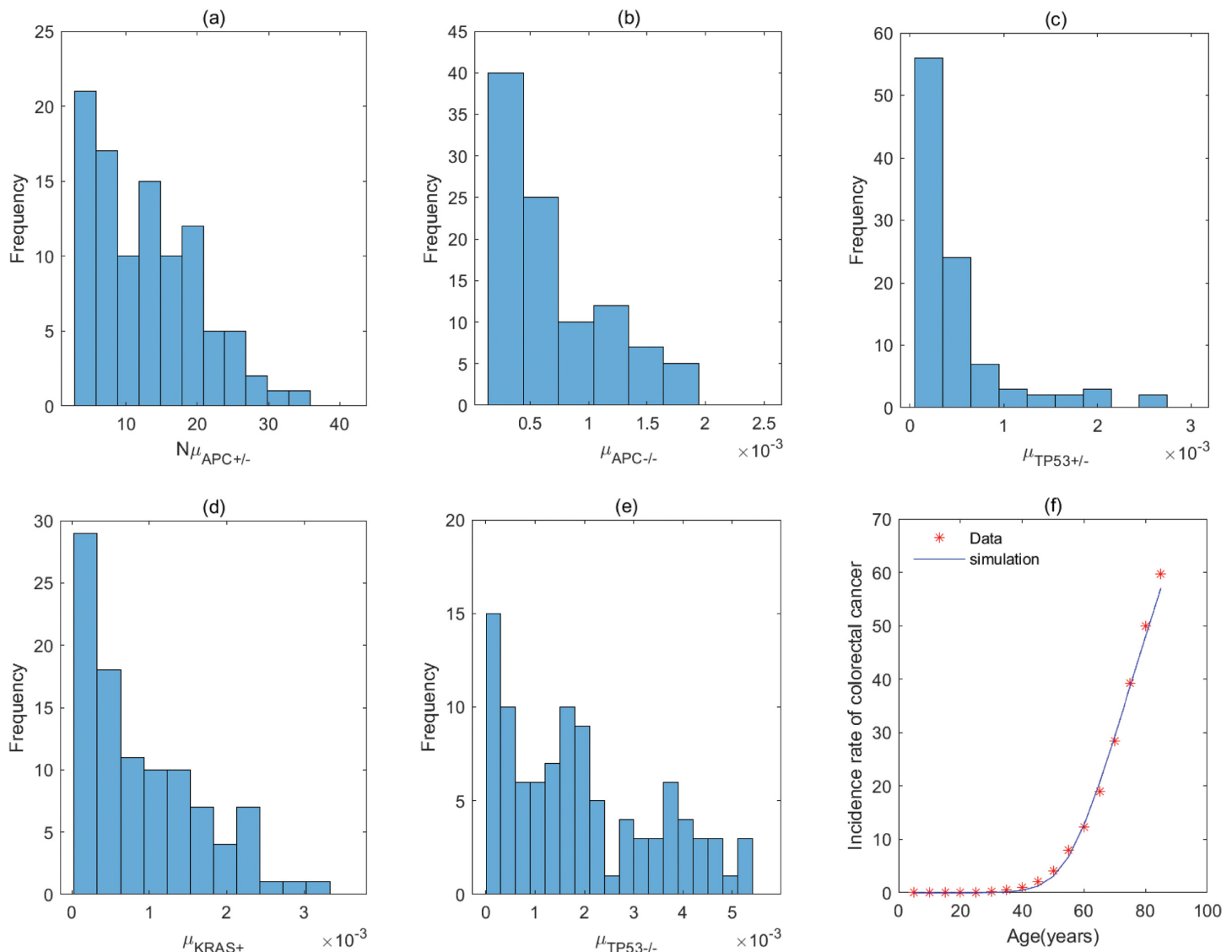


Fig 11. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow KRAS^{+} \rightarrow TP53^{-/-}$. (a) $N\mu_{APC+/-}$ ($N\mu_N$ of model), (b) $\mu_{APC+/-}$ (μ_{p_1} of model), (c) $\mu_{TP53+/-}$ (μ_{p_2} of model), (d) μ_{KRAS+} (μ_{p_3} of model), (e) $\mu_{TP53-/-}$ (μ_{p_4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g011>

tumor, which regulates G1 cycle and apoptosis of cells. As a result, the inactivation of TP53 is usually a late event in the development of colorectal cancer. drugs that do not develop the resistance caused by the mutation of KRAS and APC should be adopted preferentially in the therapy of colorectal cancer.

The multi-type branching process is a useful tool to solve the risk function of the multistage model. The approximate solution of hazard function is used to predict the risk of cancer by some work [18, 46, 47]. Here, we consider the effect of the growth rate and death rate of cells on risk cancer and make a comparison for the approximate solution without the death rate of cells and the exact numerical solution including the growth rate and death rate of cells. It is shown that the approximate solution will overestimate the risk of cancer compare to the exact

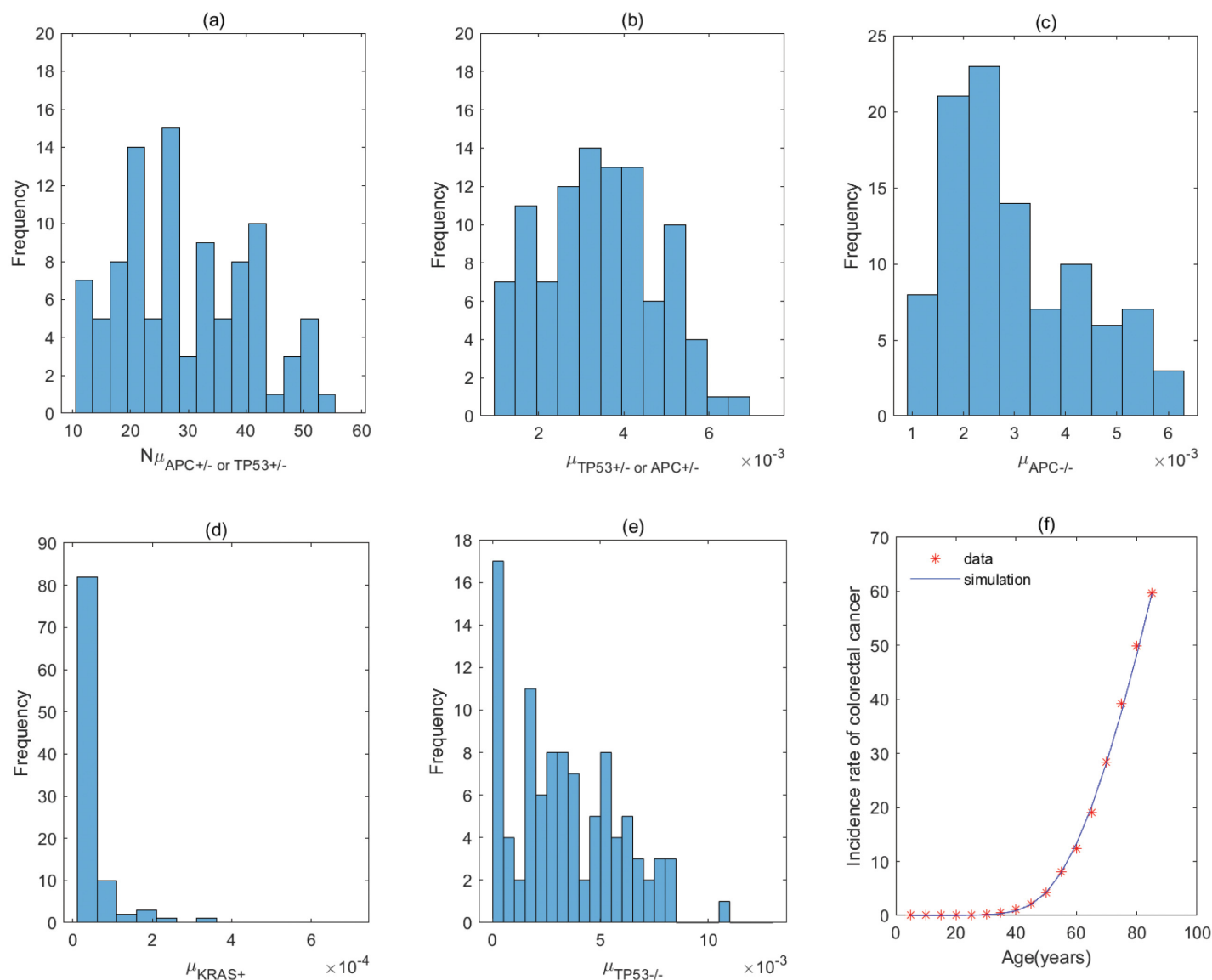


Fig 12. The fitting result and Probabilistic distributions of the estimated parameters using approximate Bayesian computation schemes involving the simulated likelihood density for the pathway with $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+} \rightarrow TP53^{-/-}$ or $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+} \rightarrow TP53^{-/-}$. (a) $N\mu_{APC^{+/-}}$ or $N\mu_{TP53^{+/-}}$ ($N\mu_N$ of model), (b) $\mu_{TP53^{+/-}}$ or $\mu_{APC^{+/-}}$ (μ_{p_1} of model), (c) $\mu_{APC^{-/-}}$ (μ_{p_2} of model), (d) $\mu_{KRAS^{+}}$ (μ_{p_3} of model), (e) $\mu_{TP53^{-/-}}$ (μ_{p_4} of model), (f) The colorectal cancer incidence rate per 100,000 patients from SEER registry and rates predicted by the model.

<https://doi.org/10.1371/journal.pcbi.1011225.g012>

Table 1. Estimate values of parameters in the pathway with $KRAS^{+} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$.

Mutation rate	$\mu_{KRAS^{+}}$ (μ_N)	$\mu_{APC^{+/-}}$ (μ_{p_1})	$\mu_{APC^{-/-}}$ (μ_{p_2})	$\mu_{TP53^{+/-}}$ (μ_{p_3})	$\mu_{TP53^{-/-}}$ (μ_{p_4})
Value (per year)	8.93×10^{-9}	4.70×10^{-3}	1.99×10^{-3}	6.99×10^{-4}	2.52×10^{-4}
95% CI	$[8.93, 9.31] \times 10^{-9}$	$[3.92, 5.58] \times 10^{-3}$	$[1.43, 2.76] \times 10^{-3}$	$[4.30, 14.09] \times 10^{-4}$	$[1.04, 4.23] \times 10^{-3}$

The estimated values are obtained by calculating quartile of Fig 5(a)–5(e), and $\mu_{KRAS^{+}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t001>

Table 2. Estimate values of parameters in the pathway with $APC^{+/-} \rightarrow KRAS^{+} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$.

Mutation rate	$\mu_{APC^{+/-}} (\mu_N)$	$\mu_{KRAS^{+}} (\mu_{P_1})$	$\mu_{APC^{-/-}} (\mu_{P_2})$	$\mu_{TP53^{+/-}} (\mu_{P_3})$	$\mu_{TP53^{-/-}} (\mu_{P_4})$
Value (per year)	9.30×10^{-8}	2.60×10^{-3}	1.60×10^{-3}	1.02×10^{-4}	8.12×10^{-4}
95% CI	$[6.66, 12.49] \times 10^{-8}$	$[1.90, 4.10] \times 10^{-3}$	$[1.30, 2.30] \times 10^{-3}$	$[0.46, 2.47] \times 10^{-4}$	$[0.16, 2.00] \times 10^{-3}$

The estimated values are obtained by calculating quartile of Fig 6(a)–6(e), and $\mu_{APC^{+/-}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t002>

Table 3. Estimate values of parameters in the pathway with $KRAS^{+} \rightarrow APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$ or $KRAS^{+} \rightarrow TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-}$.

Mutation rate	$\mu_{KRAS^{+}} (\mu_N)$	$\mu_{APC^{+/-}} (\mu_{TP53^{+/-}}) (\mu_{P_1})$	$\mu_{TP53^{+/-}} (\mu_{APC^{+/-}}) (\mu_{P_2})$	$\mu_{APC^{-/-}} (\mu_{P_3})$	$\mu_{TP53^{-/-}} (\mu_{P_4})$
Value (per year)	9.03×10^{-9}	5.30×10^{-3}	3.00×10^{-3}	4.20×10^{-3}	1.90×10^{-3}
95% CI	$[8.63, 9.41] \times 10^{-9}$	$[4.70, 6.10] \times 10^{-3}$	$[2.20, 3.90] \times 10^{-3}$	$[2.80, 5.20] \times 10^{-3}$	$[0.97, 3.10] \times 10^{-3}$

The estimated values are obtained by calculating quartile of Fig 7(a)–7(e), and $\mu_{KRAS^{+}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t003>

Table 4. Estimate values of parameters in the pathway with $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+}$.

Mutation rate	$\mu_{APC^{+/-}} (\mu_N)$	$\mu_{APC^{-/-}} (\mu_{P_1})$	$\mu_{TP53^{+/-}} (\mu_{P_2})$	$\mu_{TP53^{-/-}} (\mu_{P_3})$	$\mu_{KRAS^{+}} (\mu_{P_4})$
Value (per year)	1.23×10^{-7}	5.03×10^{-4}	2.94×10^{-4}	7.92×10^{-4}	1.15×10^{-3}
95% CI	$[0.71, 2.01] \times 10^{-7}$	$[3.08, 9.83] \times 10^{-4}$	$[1.37, 6.21] \times 10^{-4}$	$[3.34, 16.12] \times 10^{-4}$	$[0.73, 1.97] \times 10^{-3}$

The estimated values are obtained by calculating quartile of Fig 8(a)–8(e), and $\mu_{APC^{+/-}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t004>

Table 5. Estimate values of parameters in the pathway with $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+}$ or $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+}$.

Mutation rate	$\mu_{APC^{+/-}} (\mu_{TP53^{+/-}}) (\mu_N)$	$\mu_{TP53^{+/-}} (\mu_{APC^{+/-}}) (\mu_{P_1})$	$\mu_{APC^{-/-}} (\mu_{P_2})$	$\mu_{TP53^{-/-}} (\mu_{P_3})$	$\mu_{KRAS^{+}} (\mu_{P_4})$
Value (per year)	3.27×10^{-7}	3.61×10^{-3}	2.18×10^{-3}	5.08×10^{-5}	1.45×10^{-3}
95% CI	$[2.44, 4.33] \times 10^{-7}$	$[2.55, 4.75] \times 10^{-3}$	$[1.49, 2.78] \times 10^{-3}$	$[2.87, 11.18] \times 10^{-5}$	$[0.43, 5.84] \times 10^{-3}$

The estimated values are obtained by calculating quartile of Fig 9(a)–9(e), and $\mu_{APC^{+/-}}$ or $\mu_{TP53^{+/-}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t005>

Table 6. Estimate values of parameters in the pathway with $APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$.

Mutation rate	$\mu_{APC^{+/-}} (\mu_N)$	$\mu_{APC^{-/-}} (\mu_{P_1})$	$\mu_{KRAS^{+}} (\mu_{P_2})$	$\mu_{TP53^{+/-}} (\mu_{P_3})$	$\mu_{TP53^{-/-}} (\mu_{P_4})$
Value (per year)	8.89×10^{-8}	7.23×10^{-4}	1.74×10^{-4}	8.64×10^{-4}	9.83×10^{-4}
95% CI	$[5.73, 14.26] \times 10^{-8}$	$[4.42, 11.75] \times 10^{-4}$	$[0.69, 4.20] \times 10^{-4}$	$[2.59, 19.22] \times 10^{-4}$	$[3.54, 19.85] \times 10^{-4}$

The estimated values are obtained by calculating quartile of Fig 10(a)–10(e), and $\mu_{APC^{+/-}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t006>

Table 7. Estimate values of parameters in the pathway with $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow KRAS^{+} \rightarrow TP53^{-/-}$.

Mutation rate	$\mu_{APC^{+/-}} (\mu_N)$	$\mu_{APC^{-/-}} (\mu_{P_1})$	$\mu_{KRAS^{+}} (\mu_{P_2})$	$\mu_{TP53^{+/-}} (\mu_{P_3})$	$\mu_{TP53^{-/-}} (\mu_{P_4})$
Value (per year)	1.21×10^{-7}	5.26×10^{-4}	2.91×10^{-4}	7.49×10^{-4}	1.71×10^{-3}
95% CI	$[0.68, 1.83] \times 10^{-7}$	$[3.36, 10.33] \times 10^{-4}$	$[1.47, 5.67] \times 10^{-4}$	$[2.85, 14.53] \times 10^{-4}$	$[0.63, 3.22] \times 10^{-3}$

The estimated values are obtained by calculating quartile of Fig 11(a)–11(e), and $\mu_{APC^{+/-}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t007>

Table 8. Estimate values of parameters in the pathway with $APC^{+/-} \rightarrow TP53^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+} \rightarrow TP53^{-/-}$ or $TP53^{+/-} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+} \rightarrow TP53^{-/-}$.

Mutation rate	$\mu_{APC^{+/-}} (\mu_{TP53^{+/-}}) (\mu_N)$	$\mu_{TP53^{+/-}} (\mu_{APC^{+/-}}) (\mu_{P_1})$	$\mu_{APC^{-/-}} (\mu_{P_2})$	$\mu_{KRAS^{+}} (\mu_{P_3})$	$\mu_{TP53^{-/-}} (\mu_{P_4})$
Value (per year)	2.76×10^{-7}	3.42×10^{-3}	2.64×10^{-3}	2.78×10^{-5}	3.14×10^{-3}
95% CI	$[2.03, 3.89] \times 10^{-7}$	$[2.46, 4.29] \times 10^{-3}$	$[1.95, 4.00] \times 10^{-3}$	$[2.02, 4.19] \times 10^{-5}$	$[1.61, 5.22] \times 10^{-3}$

The estimated values are obtained by calculating quartile of Fig 12(a)–12(e), and $\mu_{APC^{+/-}}$ or $\mu_{TP53^{+/-}}$ is derived by assuming $N = 10^8$ [20].

<https://doi.org/10.1371/journal.pcbi.1011225.t008>

numerical solution when all parameters of model are fixed, especially for high growth rate or death rate of cells. As a consequence, we choose the exact numerical solution of hazard function to fit the data with incidence rate of colorectal cancer from SEER register during the year 1973–2013. In our work, we do not fix all parameters of model like references [18], and only growth rate and death rate of cells are decided from published references. The mutation rates of gene are not accurately inferred only based on the mutational data due to genetic instability and the type of gene mutation. For this purpose, the mutation rates of model are estimated by fitting incidence rate data of colorectal cancer based on approximate Bayesian computation

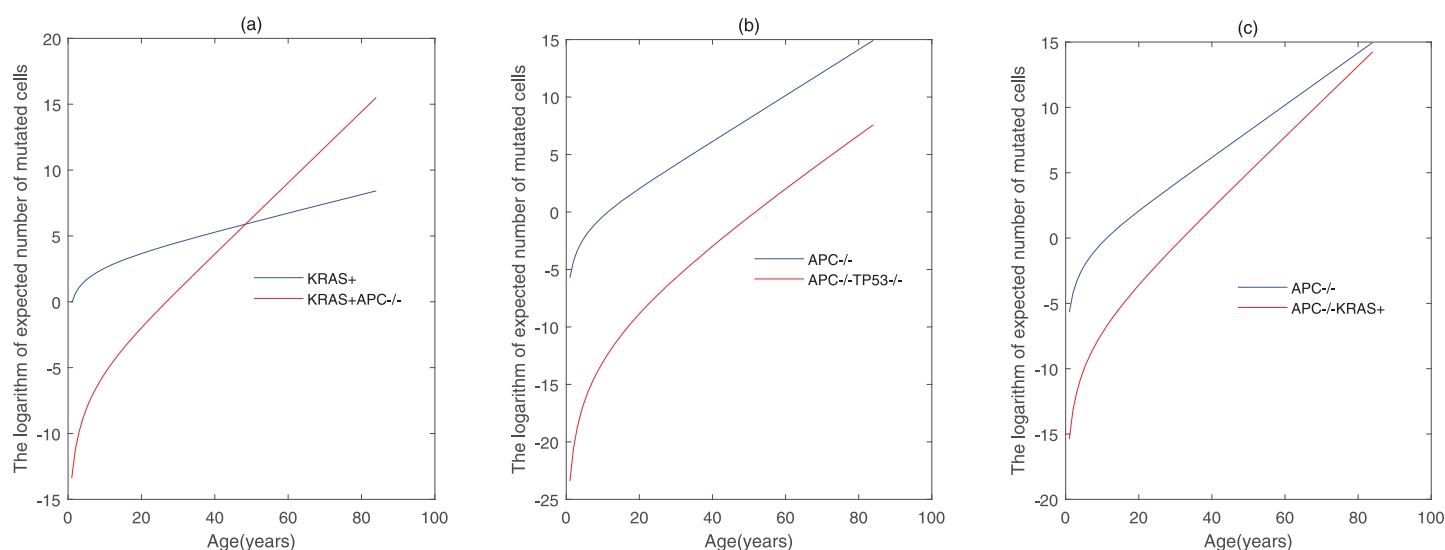


Fig 13. The logarithm of expected number of mutated cells with the combination of different genetic mutations for the pathways involving $KRAS^{+} \rightarrow APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$, $APC^{+/-} \rightarrow APC^{-/-} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-} \rightarrow KRAS^{+}$ and $APC^{+/-} \rightarrow APC^{-/-} \rightarrow KRAS^{+} \rightarrow TP53^{+/-} \rightarrow TP53^{-/-}$. The values of parameters are decided by Tables 1–8.

<https://doi.org/10.1371/journal.pcbi.1011225.g013>

schemes with simulated likelihood density. We consider the simulation of incidence rate data of colorectal cancer from 0 to 84 years old instead of that at 80 years old in reference [18]. The estimated mutation rates are extremely valuable to predict the risk of colorectal cancer. Moreover, approximate Bayesian computation schemes is a good method to estimate the biological parameters. Among them, approximate Bayesian computation with Markov chain Monte Carlo and sequential Monte Carlo was applied to biological systems with success [48, 49]. Here, we use approximate Bayesian computation schemes with simulated likelihood density to estimate our model parameters, which is effective for inferring parameters in high-dimensional stochastic model, and can reduce the probability of getting stuck the low probability regions compared to approximate Bayesian computation with Markov chain Monte Carlo [22].

To further identify the feature of different gene mutation orders, the sizes of premalignant cells in the evolutionary pathways with $KRAS \rightarrow APC \rightarrow TP53$, $APC \rightarrow TP53 \rightarrow KRAS$, $APC \rightarrow KRAS \rightarrow TP53$ are analyzed. We find that the sizes of cells with single-mutant and double-mutant are significant difference in different gene mutation orders. The number of KRAS-mutated cells is higher than that of cells with both activated KRAS and inactivated APC at the beginning, and the number of cells with both KRAS-activated and APC-inactivated will surpass that of KRAS-mutated cells over time for the order with $KRAS \rightarrow APC \rightarrow TP53$. For the orders with $APC \rightarrow TP53 \rightarrow KRAS$ and $APC \rightarrow KRAS \rightarrow TP53$, the cells with $APC^{-/-}$ far outnumber those with $APC^{-/-}TP53^{-/-}$ and $APC^{-/-}KRAS^{+}$. These results can be used as the mark of colorectal cancer diagnosis and inferring the sequence of gene mutation.

In this paper, we not only give the mutation order of three driver genes, but also provide the detailed pathway of gene mutations. However, there are still some limits in our work. Our results obtained do not consider the interaction effect between mutations in two genes, since no evidence suggests that epistatic interactions occur in mutations of two genes [4]. As is well known, inactivation of each allele of TSG has two ways involving point mutation and the LOH. We do not discuss the detailed pattern of alteration in TSGs, APC and TP53 in our model. The reason is that the model involving the pattern of inactivation in TSGs includes too many parameters, which results in non-identifiability issue of model parameters. Moreover, the type of genomic instability is unclear because of heterogeneity of tumor. However, we discuss the mutation rate of gene in different pathway obtained by fitting the incidence rate data of colorectal cancer at different ages. These estimated values of mutation rates can provide some information for inferring the pattern of alteration in TSGs and the existence of genetic instability. With the development of DNA sequencing, sequenced cancer genomes and gene expression data are important data source for inferring the evolutionary pathway in cancer progression, and some methods are developed to predict tumor evolution [50–52]. Although these data and methods do not give the rate of gene mutation, they can provide new ideas for constructing specific mathematical model of cancer. Thus, more diverse data deserve to be considered in further studying evolutionary pathways of cancer, which will be the direction for future research.

Supporting information

S1 Text. Mathematical derivation and supplementary figures. Fig A: All pathways of gene mutations for the case (A) in colorectal cancer. Fig B: All pathways of gene mutations for the case (B) in colorectal cancer. Fig C: All pathways of gene mutations for the case (C) in colorectal cancer. Fig D: All pathways of gene mutations for the case (D) in colorectal cancer. Fig E: All pathways of gene mutations for the case (E) in colorectal cancer. Fig F: All pathways of gene mutations for the case (F) in colorectal cancer.
(PDF)

Acknowledgments

We thank Dr. Tianhai Tian for the help of code in the manuscript.

Author Contributions

Conceptualization: Yulu Hu, Yunshan Xu, Sanyi Tang.

Data curation: Yulu Hu, Yunshan Xu.

Formal analysis: Lingling Li.

Funding acquisition: Lingling Li, Sanyi Tang.

Investigation: Yulu Hu, Yunshan Xu.

Methodology: Lingling Li, Sanyi Tang.

Project administration: Lingling Li, Sanyi Tang.

Resources: Lingling Li, Sanyi Tang.

Software: Lingling Li.

Supervision: Sanyi Tang.

Validation: Lingling Li, Sanyi Tang.

Visualization: Lingling Li.

Writing – original draft: Lingling Li, Yulu Hu, Sanyi Tang.

Writing – review & editing: Lingling Li, Sanyi Tang.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68(6): 394–424. <https://doi.org/10.3322/caac.21492> PMID: 30207593
2. Fearon ER. Molecular genetics of colorectal cancer. *Annu Rev Pathol.* 2011; 6: 479–507. <https://doi.org/10.1146/annurev-pathol-011110-130235> PMID: 21090969
3. Tomasetti C, Marchionni L, Nowak MA, Parmigiani G, Vogelstein B. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc Natl Acad Sci U S A.* 2015; 112(1): 118–123. <https://doi.org/10.1073/pnas.1421839112> PMID: 25535351
4. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014; 505(7484): 495–501. <https://doi.org/10.1038/nature12912> PMID: 24390350
5. Bos JL, Fearon ER, Hamilton SR, Vries MV, van Boom JH, van der Eb AJ, et al. Prevalence of ras gene-mutations in human colorectal cancers. *Nature.* 1987; 327(6120): 293–297. <https://doi.org/10.1038/327293a0> PMID: 3587348
6. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med.* 2004; 10(8): 789–799. <https://doi.org/10.1038/nm1087> PMID: 15286780
7. Nordling CO. A new theory of cancer inducing mechanism. *Br J Cancer.* 1953; 7(1): 68–72. <https://doi.org/10.1038/bjc.1953.8> PMID: 13051507
8. Armitage P, Doll R. The age distribution of cancer and multi-stage theory of carcinogenesis. *Br J Cancer.* 1954; 8(1): 1–12. <https://doi.org/10.1038/bjc.1954.1> PMID: 13172380
9. Armitage P, Doll R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer.* 1957; 9(12): 161–169. <https://doi.org/10.1038/bjc.1957.22>
10. Knudson AG. Mutation and cancer: Statistical study of retinoblastoma. *Proc Natl Acad Sci U S A.* 1971; 68(4): 820–823. <https://doi.org/10.1073/pnas.68.4.820> PMID: 5279523
11. Moolgavkar SH, Dewanji A, Venzon DJ. A stochastic two-stage model for cancer risk assessment I: The hazard function and the probability of tumor. *Risk Anal.* 1988; 8(3): 383–392. <https://doi.org/10.1111/j.1539-6924.1988.tb00502.x> PMID: 3201016

12. Meza R, Jeon J, Moolgavkar SH, Luebeck EG. Age-specific incidence of cancer: phases, transitions, and biological implications. *Proc Natl Acad Sci U S A*. 2008; 105(42): 16284–16289. <https://doi.org/10.1073/pnas.0801151105> PMID: 18936480
13. Moolgavkar SH, Luebeck EG. Multistage carcinogenesis: population-based model for colon cancer. *J Nat Cancer Inst*. 1992; 84(8): 610–618. <https://doi.org/10.1093/jnci/84.8.610> PMID: 1313509
14. Luebeck EG, Moolgavkar SH. Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci U S A*. 2002; 99(23): 15095–15100. <https://doi.org/10.1073/pnas.222118199> PMID: 12415112
15. Little MP, Wright EG. A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. *Math Biosci*. 2003; 183(2): 111–134. [https://doi.org/10.1016/S0025-5564\(03\)00040-3](https://doi.org/10.1016/S0025-5564(03)00040-3) PMID: 12711407
16. Little MP, Vineis P, Li G. A stochastic carcinogenesis model incorporating multiple types of genomic instability fitted to colon cancer data. *J Theor Biol*. 2017; 254(2): 229–238. <https://doi.org/10.1016/j.jtbi.2008.05.027>
17. Lang BM, Kuipers J, Misselwitz B, Beerenwinkel N. Predicting colorectal cancer risk from adenoma detection via a two-type branching process model. *PLoS Comput Biol*. 2020; 16(2): e1007552. <https://doi.org/10.1371/journal.pcbi.1007552> PMID: 32023238
18. Paterson C, Clevers H, Bozic I. Mathematical model of colorectal cancer initiation. *Proc Natl Acad Sci U S A*. 2020; 117(34): 20681–20688. <https://doi.org/10.1073/pnas.2003771117> PMID: 32788368
19. Haupt S, Zeilmann A, Ahadova A, Biker H, Heuveline V. Mathematical modeling of multiple pathways in colorectal carcinogenesis using dynamical systems with kronecker structure. *PLoS Comput Biol*. 2021; 17(5): e1008970. <https://doi.org/10.1371/journal.pcbi.1008970> PMID: 34003820
20. Zhang R, Ukogu OA, Bozic I. Waiting times in a branching process model of colorectal cancer initiation. *Theor Popul Biol*. 2023; 151: 44–63. <https://doi.org/10.1016/j.tpb.2023.04.001> PMID: 37100121
21. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998; 396(6712): 643–649. <https://doi.org/10.1038/25292> PMID: 9872311
22. Wu Q, Smith-Miles K, Tian T. Approximate Bayesian computation schemes for parameter inference of discrete stochastic models using simulated likelihood density. *BMC Bioinform*. 2014; 15(Suppl 12): S3. <https://doi.org/10.1186/1471-2105-15-S12-S3> PMID: 25473744
23. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science*. 2013; 339(6127): 1546–1558. <https://doi.org/10.1126/science.1235122> PMID: 23539594
24. Durrett R. Branching Process Models of Cancer. In: *Branching Process Models of Cancer*. Mathematical Biosciences Institute Lecture Series, vol 1.1. Springer, Cham; 2015.
25. Harris TE. The theory of branching processes. 1st ed. Springer-Verlag Berlin Heidelberg; 1963.
26. Portier CJ, Sherman CKA. Calculating tumor incidence rates in stochastic models of carcinogenesis. *Math Biosci*. 1996; 135(2): 129–146. [https://doi.org/10.1016/0025-5564\(96\)00011-9](https://doi.org/10.1016/0025-5564(96)00011-9) PMID: 8768218
27. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*. 2015; 347(6217): 78–81. <https://doi.org/10.1126/science.1260825> PMID: 25554788
28. Komarova NL, Wodarz D. Drug resistance in cancer: principles of emergence and prevention. *Proc Natl Acad Sci U S A*. 2005; 102(27): 9714–9719. <https://doi.org/10.1073/pnas.0501870102> PMID: 15980154
29. Shamieh SE, Saleh F, Assaad S, Farhat F. Next-generation sequencing reveals mutations in RB1, CDK4 and TP53 that may promote chemo-resistance to palbociclib in ovarian cancer. *Drug Metabol Therapy*. 2019; 34(2): 20180027. PMID: 31145688
30. Lièvre A, Bachet JB, Corre DL, Boige V, Landi B, Emile JF, et al. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res*. 2012; 66(8): 3992–3995.
31. Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, et al. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Medicine*. 2005; 2(1): e17. <https://doi.org/10.1371/journal.pmed.0020017> PMID: 15696205
32. Luebeck EG, Curtius K. Impact of tumor progression on cancer incidence curves. *Cancer Res*. 2013; 73(3): 1086–1096. <https://doi.org/10.1158/0008-5472.CAN-12-2198> PMID: 23054397
33. Lamlum H, Papadopoulou A, Ilyas M, Rowan A, Gillet C, Hanby A, et al. APC mutations are sufficient for the growth of early colorectal adenomas. *Proc Natl Acad Sci U S A*. 2000; 97(5): 2225–2228. <https://doi.org/10.1073/pnas.040564697> PMID: 10681434
34. 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 Rep*. 2014; 15(1): 62–69. <https://doi.org/10.1002/embr.201337799> PMID: 24355609

35. Baker AM, Cereser B, Melton S, Fletcher AG, Rodriguez-Justo M, Tadrous PJ, et al. Quantification of crypt and stem cell evolution in the normal and neoplastic human colon. *Cell Rep.* 2014; 8(4): 940–947. <https://doi.org/10.1016/j.celrep.2014.07.019> PMID: 25127143
36. Nicholson AM, Olpe C, Hoyle A, Thorsen AS, Rus T, Colombé M, et al. Fixation and spread of somatic mutations in adult human colonic epithelium. *Cell Stem Cell.* 2018; 22(6): 909–918. <https://doi.org/10.1016/j.stem.2018.04.020> PMID: 29779891
37. Vermeulen L, Morrissey E, van der Heijden D, Nicholson AM, Sottoriva A, Buczacki S, et al. Defining stem cell dynamics in models of intestinal tumor initiation. *Science.* 2013; 342(6161): 995–998. <https://doi.org/10.1126/science.1243148> PMID: 24264992
38. Tomasetti C, Bozic I. The (not so) immortal strand hypothesis. *Stem Cell Res. (Amst.)* 2015; 14(2): 238–241. <https://doi.org/10.1016/j.scr.2015.01.005> PMID: 25700960
39. Blokzijl F, Ligt JD, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature.* 2016; 538(7624): 260–264. <https://doi.org/10.1038/nature19768> PMID: 27698416
40. Fukasawa K, Choi T, Kuriyama R, Rulong S, Vande Woude GF. Abnormal centrosome amplification in the absence of p53. *Science.* 1996; 271(5256): 1744–1747. <https://doi.org/10.1126/science.271.5256.1744> PMID: 8596939
41. Fodde R, Kuipers J, Rosenberg C, Smits R, Kielman M, Gaspar C, et al. Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nature Cell Biol.* 2001; 3(4): 433–438. <https://doi.org/10.1038/35070129> PMID: 11283620
42. Timmermann B, Kerick M, Roehr C, Fischer A, Isau M, Boerno ST, et al. Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLoS One.* 2010; 5(12): e15661. <https://doi.org/10.1371/journal.pone.0015661> PMID: 21203531
43. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012; 487(7407): 330–337. <https://doi.org/10.1038/nature11252>
44. Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN, et al. APC mutations occur early during colorectal tumorigenesis. *Nature.* 1992; 359(6392): 235–237. <https://doi.org/10.1038/359235a0> PMID: 1528264
45. Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, et al. Evaluation of tumor microsatellite instability using five quasimonomorphic repeats and pentaplex PCR. *Gastroenterology.* 2002; 123(6): 1804–1811. <https://doi.org/10.1053/gast.2002.37070>
46. Zhang X, Fang Y, Zhao YD, Zheng WM. Mathematical modeling the pathway of human breast cancer. *Math Biosci.* 2014; 253: 25–29. <https://doi.org/10.1016/j.mbs.2014.03.011> PMID: 24680645
47. Paterson C, Bozic I, Smith MJ, Hoad X, Evans DGR. A mechanistic mathematical model of initiation and malignant transformation in sporadic vestibular schwannoma. *Br J Cancer.* 2022; 127: 1843–1857. <https://doi.org/10.1038/s41416-022-01955-8> PMID: 36097176
48. Golightly A, Wilkinson DJ. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus.* 2011; 1(6): 807–820. <https://doi.org/10.1098/rsfs.2011.0047> PMID: 23226583
49. Bollen Y, Stelloo E, van Leenen P, van den Bos M, Ponsioen B, Lu BX, et al. Reconstructing single-cell karyotype alterations in colorectal cancer identifies punctuated and gradual diversification patterns. *Nat Genet.* 2021; 53(8): 1187–1195. <https://doi.org/10.1038/s41588-021-00891-2> PMID: 34211178
50. Caravagna G, Graudenzi A, Ramazzotti D, Sanz-Pamplona R, Sano LD, Mauri G, et al. Algorithmic methods to infer the evolutionary trajectories in cancer progression. *Proc Natl Acad Sci U S A.* 2016; 113(28): E4025–E4034 <https://doi.org/10.1073/pnas.1520213113> PMID: 27357673
51. Fleck JL, Pavel AB, Cassandras CG. Integrating mutation and gene expression cross-sectional data to infer cancer progression. *BMC Syst Biol.* 2016; 10(1): 1–12. <https://doi.org/10.1186/s12918-016-0255-6> PMID: 26810975
52. Diaz-Colunga J, Diaz-Uriarte R. Conditional prediction of consecutive tumor evolution using cancer progression models: What genotype comes next? *PLoS Comput Biol.* 2021; 17(12): e1009055–e1009055. <https://doi.org/10.1371/journal.pcbi.1009055> PMID: 34932572