

Mutation Mechanisms of Human Breast Cancer

LINGLING LI,¹ TIANHAI TIAN,² and XINAN ZHANG,¹

ABSTRACT

Cancer is a class of diseases caused by the accumulation of gene mutations. All mutated genes constitute a genetic network for cancer progression. It is very helpful for tumor diagnosis and therapy if we know how many mutated genes are needed for human breast cancer. In this article, we investigate the mutation mechanisms of human breast cancer by modeling the data of surveillance, epidemiology, and end results registry. The data are age-specific incidence rates of breast cancer of females in the United States. We set up stochastic multistage models to estimate the age-specific incidence rates by using several coupled ordinary differential equations derived from the Kolmogorov backward equations. Our results suggest that 2–14 mutations in the genome of breast stem cells are required for normal breast stem cells to become a malignant cell, and 3 gene mutations are most likely to occur in the development of female breast cancer.

Keywords: breast cancer, gene mutation; incidence rate, stochastic multistage model.

1. INTRODUCTION

CANCER IS A GENETIC DISEASE. The tumorigenesis is the result of the accumulated mutations by oncogenes, tumor suppressor genes, and genetic instability genes (Vogelstein and Kinzler, 2004), which is regarded to be developed slowly over many years generally (Leob et al., 2003). It has been proposed that eight biological capabilities should be acquired during the multistep development of human tumors: namely self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, resistance of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis, reprogramming of energy metabolism, and evading immune destruction (Hanahan and Weinberg, 2011). Those hallmarks of human cancer indicate that normal stem cells that become a malignant cell must undergo a series of mutations in the genome. A recent study of 18,191 human genes from breast and colorectal cancers shows that <15 mutations were likely to be responsible for the driver mutations of breast cancer and colorectal cancer (Wood et al., 2007). It will be very helpful for tumor diagnosis and therapy if we know how many driver mutations are needed for human breast cancer.

Armitage and Doll (1954) developed the first multistage model for the development of human cancers by fitting the data of age-specific death rates caused by cancer diseases. However, this model did not give the exact biological meanings of the multistage nor consider the clonal expansion of intermediate cells. These questions were studied later by examining 48 cases of retinoblastoma (Knudson, 1971). The hypothesis was put forward that retinoblastoma is a cancer caused by two mutation events in the genome of stem cells. In

¹School of Mathematics and Statistics, Central China Normal University, Wuhan, P.R. China.

²School of Mathematical Science, Monash University, Melbourne, Australia.

the dominantly inherited form, the first mutation was inherited from the germinal cells and the second occurred in somatic cells. In the nonhereditary form, both mutations occurred in somatic cells. This result led to the discovery of the earliest tumor suppressor gene RB. The earliest two-stage (two-hit) model with clonal expansion of intermediate cells was developed by Moolgavkar et al. (1980). This model can fit the age-specific incidence rates of breast cancer and colorectal cancer. There are many works extending the two-stage (two-hit) model with clonal expansion of intermediate cells (Luebeck and Moolgavkar, 2002; Jeon et al., 2006; Meza et al., 2008; Moolgavkar et al., 2012; Curtius et al., 2015; Hazelton et al., 2015; Brouwer et al., 2016).

Three or more mutations are required for most cancers. For example, colorectal cancer can be described by three or more mutations to the P53, RAS, and APC genes (Knudson, 2001), and three driver gene mutations are needed for lung cancer (Tomasetti et al., 2015). Zhang and Simon (2005) extended the two-stage model to the six-stage model with clonal expansion in each compartment of intermediate cells and found that two-stage to six-stage models can fit the data of age-specific incidence rates of breast cancer and the three-stage model fits the data better than the two-stage model.

However, there are a few shortages for the mentioned multistage models. Some of them have used an approximation formula of the hazard function. Other models have not considered both the growth rates and death rates simultaneously for all compartments of intermediate cells. Mutation and clonal expansion are two important concepts in tumor processes. Loeb et al. had put forward mutator phenotype to account for the large number of mutations observed in cancer cells (Loeb, 1991; Loeb et al., 2003). Tomlinson and Bodmer (1999) had challenged the mutator phenotype hypothesis and claimed that selection for clonal expansion is more important than the increasing mutation rate in the development of a tumor. This is the general debate for the selection of mutation or clonal expansion. Thus, a multistage model should include mutation rate, growth rate, and death rate in each compartment of intermediate cells. This work was motivated by the work of Crump et al. (2005) for a two-stage model with clonal expansion of intermediate cells. We have extended their work to any k -stage or k -hit ($k=2, 3, \dots$) model. By fitting the data set of age-specific incidence rates of breast cancer, we can determine how many mutations are needed for breast cancer processes. Our results are also tested by statistical inference to examine their availability.

2. METHODS

2.1. The surveillance, epidemiology, and end result data

Incidence data for breast cancer were obtained from the surveillance, epidemiology, and end result (SEER) registry for the years 1973–1999 (Ries et al., 2002). We use the reported incidence of breast cancer by gender, race, age, and calendar year in the nine SEER geographic areas, which cover an estimated 10% of the U.S. population. The population bases come from SEER population files (based on the data from the U.S. Census Bureau) by sex and race and are cross-tabulated by calendar years 1973–1999 and 5-year age groups (ages 0–85+). Our analyses combined all races for breast cancer in females during the period 1990–1999. Rates are expressed as cases per 100,000 females.

2.2. The model

A comprehensive model must be postulated for at least two rate-limiting events for most human cancer types (Vogelstein and Kinzler, 2004). From the findings of Hanahan and Weinberg (2011) and Wood et al. (2007), more than six-stage models should be considered, which have not been considered by previous work. A schematic representation of the deterministic-stochastic k -stage ($k=2, 3, \dots$) model is illustrated in Figure 1.

N represents the population of stem cells per breast with mutation rate μ_0 per cell per year. Those stem cells are subject to mutate into a type of intermediate cells that carried an irreversible mutation. Cells in compartment I_i ($i=1, 2, \dots, k-1$) accumulating i mutations with mutation rate μ_i may be divided into two daughter cells at a rate α_i and differentiate or die at a rate β_i . The time delay from a malignant cell to clinical detection is assumed as 5 years (Radmacher and Simon, 2000). If we assume that the time delay is 2 or 3 years, there is no difference for estimation of parameters of the multistage model (Zhang and Simon, 2005).

For a stochastic k -stage model, we use $N(t)$ to denote the number of normal progenitor cells per breast at time t , $Y_i(t)$ ($i=1, \dots, k-1$) the number of intermediate cells in compartment I_i per breast at time t , $Y_k(t)$ the number of fully malignant cells per breast at time t . For $\tau \leq t$, we define

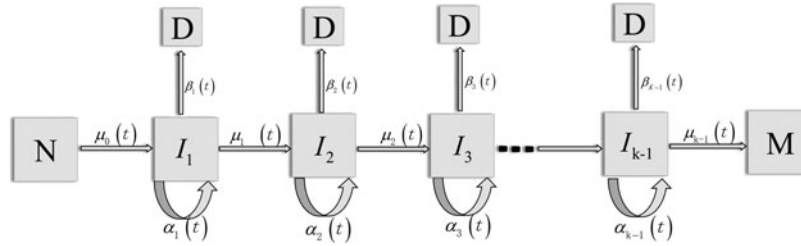


FIG. 1. The schematic representation of stochastic multistage model for carcinogenesis. N denotes the normal cell, I_i ($i = 1, 2, \dots, k - 1$) the compartment of intermediate cells with i mutations, D the dead or differentiated cell, and M the malignant cell. μ_0 and μ_i are the mutation rates per cell per year of normal cells and I_i compartment, respectively. α_i and β_i are the growth rate and death or differentiation rate per cell per year for the I_i compartment, respectively.

$$\Psi(y_1, y_2, \dots, y_k; \tau, t) = \sum_{i_1, \dots, i_k} p\{Y_1(t) = i_1, Y_2(t) = i_2, \dots, Y_k(t) = i_k | Y_1(\tau) = 0, Y_2(\tau) = 0, \dots, Y_k(\tau) = 0\} y_1^{i_1} y_2^{i_2} \dots y_k^{i_k}, \tag{1}$$

and

$$\Phi_i(y_i, y_{i+1}, \dots, y_k; \tau, t) = \sum_{i_1, \dots, i_k} p\{Y_i(t) = i_i, Y_{i+1}(t) = i_{i+1}, \dots, Y_k(t) = i_k | Y_i(\tau) = 1, Y_{i+1}(\tau) = 0, \dots, Y_k(\tau) = 0\} y_i^{i_i} y_{i+1}^{i_{i+1}} \dots y_k^{i_k}, \tag{2}$$

where Ψ and Φ_i are the probability-generating functions for the number of intermediate cells and malignant cells at time t , starting from no intermediate cell in compartment I_i to one intermediate cell in compartment I_i at time τ , respectively. These probability-generating functions satisfy the Kolmogorov backward equations (Harris, 1963) for $1 \leq i \leq k - 1$,

$$\frac{d\Phi_i}{d\tau}(y_1, y_2, \dots, y_k; \tau, t) = [\alpha_i(\tau) + \beta_i(\tau) + \mu_i(\tau)]\Phi_i(y_1, y_2, \dots, y_k; \tau, t) - \alpha_i(\tau)\Phi_i^2(y_1, y_2, \dots, y_k; \tau, t) - \mu_i(\tau)\Phi_i(y_1, y_2, \dots, y_k; \tau, t)\Phi_{i+1}(y_1, y_2, \dots, y_k; \tau, t) - \beta_i(\tau), \tag{3}$$

$$\frac{d\Psi}{d\tau}(y_1, y_2, \dots, y_k; \tau, t) = -\mu_0(\tau)N(\tau)\Psi(y_1, y_2, \dots, y_k; \tau, t)[\Phi_1(y_1, y_2, \dots, y_k; \tau, t) - 1], \tag{4}$$

where $\alpha_i(\tau)$, $\beta_i(\tau)$, and $\mu_i(\tau)$ are the growth rate, differentiation or death rate, and mutation rate of the cells in each compartment I_i at time τ , respectively, and $\mu_0(\tau)$ is the mutation rate of the normal cells at time τ .

Denote $\Psi(1, \dots, 1, 0; \tau, t)$ by $\Psi(\tau, t)$, and $\Phi_i(1, \dots, 1, 0; \tau, t)$ by $\Phi_i(\tau, t)$, then the probability, $P(t)$, of at least one malignant cell by time t , starting with only normal cells at time 0, can be written as $P(t) = 1 - \Psi(0, t)$. The associated hazard function, $h(t)$, can be written as $h(t) = \Psi'(0, t)/\Psi(0, t)$, where a ‘‘prime’’ denotes a derivative with respect to t . For Equations (3) and (4), they follow that

$$\frac{d\Psi(\tau, t)}{d\tau} = -\mu_0(\tau)N(\tau)\Psi(\tau, t)[\Phi_1(\tau, t) - 1], \tag{5}$$

$$\frac{d\Phi_i(\tau, t)}{d\tau} = [\alpha_i(\tau) + \beta_i(\tau) + \mu_i(\tau)]\Phi_i(\tau, t) - \alpha_i(\tau)\Phi_i^2(\tau, t) - \mu_i(\tau)\Phi_i(\tau, t)\Phi_{i+1}(\tau, t) - \beta_i(\tau). \tag{6}$$

By the mentioned probability-generating functions we have $\Psi(t, t) = 1$, $\Phi_i(t, t) = 1$ ($i = 1, \dots, k - 1$), and $\Phi_k(\tau, t) = 0$ for $\tau \leq t$.

By differentiating Equations (5) and (6) with respect to t we get

$$\frac{d\Psi'(\tau, t)}{d\tau} = -\mu_0(\tau)N(\tau)\Psi'(\tau, t)[\Phi_1(\tau, t) - 1] - \mu_0(\tau)N(\tau)\Psi(\tau, t)\Phi_1'(\tau, t), \tag{7}$$

$$\begin{aligned} \frac{d\Phi_i'(\tau, t)}{d\tau} &= [\alpha_i(\tau) + \beta_i(\tau) + \mu_i(\tau)]\Phi_i'(\tau, t) - 2\alpha_i(\tau)\Phi_i\Phi_i'(\tau, t) - \mu_i(\tau)\Phi_i'(\tau, t)\Phi_{i+1}(\tau, t) \\ &\quad - \mu_i(\tau)\Phi_i(\tau, t)\Phi_{i+1}'(\tau, t). \end{aligned} \tag{8}$$

Note that $\Phi_{i+1}(\tau, t)$ and $\Phi'_{i+1}(\tau, t)$ are equal to zero when $i = k - 1$. That is, the last two items of the Equation (8) are zero when $i = k - 1$.

Equations (5) and (8) constitute a system of $2k$ coupled ordinary differential equations with boundary conditions $\Phi_i(t, t) = 1$ for $1 \leq i \leq k - 1$, $\Phi'_i(t, t) = 0$ for $1 \leq i \leq k - 2$, $\Phi'_{k-1}(t, t) = -\mu_{k-1}$, $\Psi(t, t) = 1$, and $\Psi'(t, t) = 0$. To convert this boundary value system into an initial value system, we make a time transformation: $s = t - \tau$. Let $A_i(s, t) = \Phi_i(\tau, t)$, $B_i(s, t) = \Phi'_i(\tau, t)$ for $1 \leq i \leq k - 1$, $C(s, t) = \Psi(\tau, t)$, and $D(s, t) = \Psi'(\tau, t)$. When $1 \leq i \leq k - 1$, Equations (5) and (8) can be converted to the following equations:

$$\frac{dA_i(s, t)}{ds} = -[\alpha_i(t-s) + \beta_i(t-s) + \mu_i(t-s)]A_i(s, t) + \alpha_i(t-s)A_i^2(s, t) + \mu_i(t-s)A_i(s, t)A_{i+1}(s, t) + \beta_i(t-s), \tag{9}$$

$$\frac{dB_i(s, t)}{ds} = -[\alpha_i(t-s) + \beta_i(t-s) + \mu_i(t-s)]B_i(s, t) + 2\alpha_i(t-s)A_i(s, t)B_i(s, t) + \mu_i(t-s)B_i(s, t)A_{i+1}(s, t) + \mu_i(t-s)A_i(s, t)B_{i+1}(s, t), \tag{10}$$

$$\frac{dC(s, t)}{ds} = \mu_0(t-s)N(t-s)C(s, t)(A_1(s, t) - 1), \tag{11}$$

$$\frac{dD(s, t)}{ds} = \mu_0(t-s)N(t-s)D(s, t)(A_1(s, t) - 1) + \mu_0(t-s)N(t-s)C(s, t)B_1(s, t), \tag{12}$$

with initial conditions $A_i(0, t) = 1$ for $1 \leq i \leq k - 1$, $B_i(0, t) = 0$ for $1 \leq i \leq k - 2$, $B_{k-1}(0, t) = -\mu_{k-1}$, $C(0, t) = 1$, and $D(0, t) = 0$. For $i = k - 1$, $A_{i+1}(s, t) = A_k(s, t) = 0$ and $B_{i+1}(s, t) = B_k(s, t) = 0$. From the mentioned equations, we can obtain $P(t) = 1 - C(t, t)$ and $h(t) = -D(t, t)/C(t, t)$. In addition, we define $E(s, t) = 1 - \Psi(\tau, t)$ and $F(s, t) = -(\ln\Psi(\tau, t))'$. Then, Equations (11) and (12) are replaced by the equivalent equations,

$$\frac{dE(s, t)}{ds} = \mu_0(t-s)N(t-s)(1 - E(s, t))(1 - A_1(s, t)), \tag{13}$$

$$\frac{dF(s, t)}{ds} = -\mu_0(t-s)N(t-s)B_1(s, t), \tag{14}$$

with initial conditions $E(0, t) = 0$ and $F(0, t) = 0$, where $P(t) = E(t, t)$ and $h(t) = F(t, t)$. Thus, the value of $h(t)$ can be obtained by solving Equations (9) and (10) and Equations (13) and (14) directly.

2.3. Computer simulation

For the development of breast issue, it shows relatively little growth before puberty and a spurt of growth during adolescence. Thus, the growth of the breast epithelium stem cells can be presented by a logistic curve (Moolgavkar et al., 1980; Moolgavkar and Luebeck, 1990). The tissue is sensitive to changes in physiological state. For example, menopause may affect the number of premalignant cells as well as number of stem cells. Therefore, we use the value reported that the age of menopause of females is 45 years as in our previous work (Zhang and Simon, 2005; Zhang et al., 2014). The number of breast epithelium stem cells, $N(t)$, can be assumed to increase by a logistic growth curve from an initial value of 10 cells at birth to a maximum of 10^7 cells by age 20 years, whereas the number of progenitor cells decreases after age 45 years at a rate of -0.0667 per cell per year, which is 10^6 cells by age 80 years (Zhang et al., 2014). Parameters of the growth rate and differentiation (death) rate corresponding to each compartment of the intermediate cells are different values. We let α_i (β_i) and α'_i (β'_i) ($i = 1, 2, \dots, k - 1$) to denote the growth rate (differentiation or death rate) before age 45 years and after age 45 years, respectively.

We use the fourth-order Runge–Kutta method to numerically solve the mentioned differential equation system to obtain the values of the hazard function $h(t)$. The numerical optimization routine `fminsearch` in MATLAB is utilized to estimate the optimal parameters by the minimum sum of error squares, which is a very effective way for estimating the optimal parameters and is very close to the global optimum of the genetic algorithm. In this study, the error is the difference between the SEER data and the simulated values of the hazard function obtained from the models.

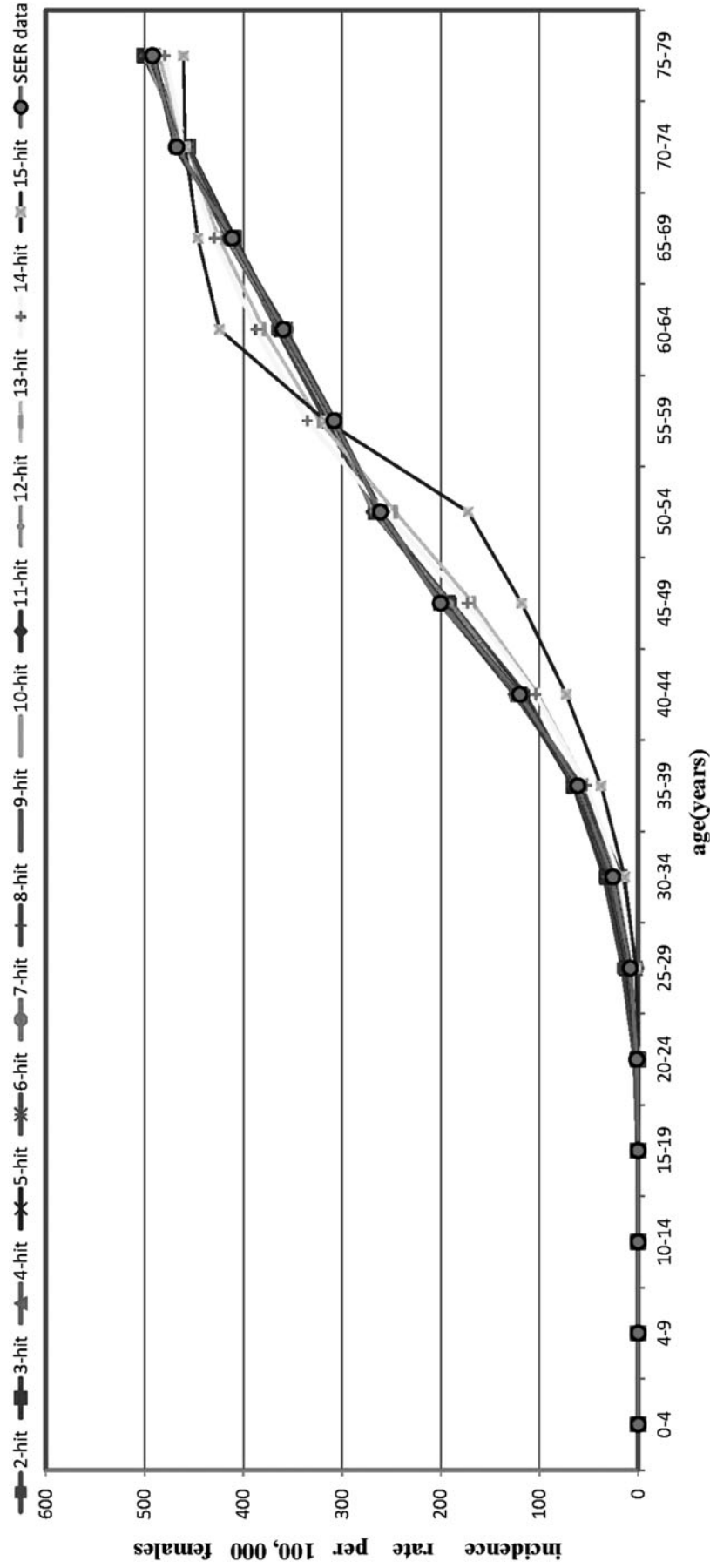


FIG. 2. The age-specific incidence rates of all races per 100,000 females for breast cancer from the SEER registry for the years 1990–1999 and rates predicted by the 2–15-stage models. SEER, surveillance, epidemiology, and end result.

3. RESULTS

3.1. Numerical results

Since there are $4(k - 1) + k = 5k - 4$ parameters in the proposed k -stage model, it contains more parameters than the information that is determined from epidemiological data. For this nonidentifiability problem, there are two popular approaches to deal with it. The first one is to set the mutation rates be equal to each other (Jeon et al., 2006; Meza et al., 2008). The other is to use a new set of parameters (Heidenreich et al., 1997; Brouwer et al., 2016) and set a fixed value for the parameter. In our models, the simulated result is insensitive to a broad range of assumed values for the growth rates, $\alpha_i, \alpha'_i (i = 1, \dots, k - 1)$. Therefore, we used the value reported by Eidemüller et al. (2015) about 12 per year.

We have only given the optimal fittings for the 2–15-stage models, which are shown in Figure 2. For the models with more than 15 stages, their fitting results are worse than the 15-stage model. From Figure 2, the 2–14-stage models fit the SEER data very well, but the 15-stage model has led to very large error.

3.2. Chi-square test of goodness of fit

We use statistical inference to test whether the optimal fittings for 2–14-stage models are available. The chi-square test is chosen to examine those optimal fittings, since the distribution observed in incidence cases is unknown. However, there are only 15 nonzero SEER data by age 79 years, which is impossible for us to examine those optimal fittings for models with more than five stages. To obtain enough data to implement the chi-square test, data set for incidence rate can be converted into data set for probability of breast cancer by the following formula:

$$P(t) = 1 - \exp\left\{-\int_0^t h(s)ds\right\}, \tag{15}$$

where $h(s)$ denotes the incidence rate of breast cancer at age s . When $P(t)$ is amplified by using a suitable multiple (i.e., 10^3), there are 49 data points that are >1 , which is adequate for performing the chi-square test. The value of the chi-square statistics is determined by

$$Chi2 = \sum_{i=31}^{79} (P_i - P_i^*)^2 * 10^3 / P_i^*, \tag{16}$$

where P_i and P_i^* are the probability of breast cancer derived from the real data and numerical simulations, respectively.

Table 1 gives the chi-square values calculated using the mentioned equation and values at the 5% significant level for the 2–15-stage models. We find that the chi-square values of the 2–14-stage models are less than the values at the 5% significant level, whereas the chi-square value of the 15-stage model is far larger than the value at the 5% significant level. These suggest that the 2–14 mutations are reasonable for normal breast stem cells to become a malignant cell. In addition, the chi-square values of the three-stage model are the smallest in all models. Thus, the optimal model is the model with three mutations for breast cancer.

TABLE 1. THE CHI-SQUARE VALUES BETWEEN THE PROBABILITY CALCULATED BY THE SURVEILLANCE, EPIDEMIOLOGY, AND END RESULT DATA AND THE ESTIMATED PROBABILITY BY THE MATHEMATICAL MODELS

<i>Chi-square test</i>	<i>2-stage</i>	<i>3-stage</i>	<i>4-stage</i>	<i>5-stage</i>	<i>6-stage</i>	<i>7-stage</i>	<i>8-stage</i>
Chi2	10.191	0.724	1.349	4.433	10.357	9.046	6.2136
df	44	41	38	35	32	29	26
S.5%	60.481	56.942	53.384	49.802	46.194	42.557	38.885
<i>Chi-square test</i>	<i>9-stage</i>	<i>10-stage</i>	<i>11-stage</i>	<i>12-stage</i>	<i>13-stage</i>	<i>14-stage</i>	<i>15-stage</i>
Chi2	18.492	18449	16.451	19.056	16.339	13.790	51.406
df	23	20	17	14	11	8	5
S.5%	35.172	31.410	27.587	23.685	19.675	15.507	11.070

Chi2 denotes the value of chi-square statistic; df (=49-1-number of parameters) denotes the degrees of freedom; S.5% denotes the value at 5% significant level.

TABLE 2. ESTIMATED OPTIMAL PARAMETER VALUES (AND 95% CONFIDENCE INTERVAL, EVALUATED USING CHI-SQUARE STATISTICS) IN THE THREE-STAGE MODEL WITH CLONAL EXPANSION FOR FEMALE

Parameters	Description	Value
β_1 [year ⁻¹]	Differentiation (death) rate in the I_1 before age 45 years	11.9840 (11.9731, 11.9956)
β_2 [year ⁻¹]	Differentiation (death) rate in the I_2 before age 45 years	11.7780 (11.7277, 11.8201)
β_1' [year ⁻¹]	Differentiation (death) rate in the I_1 after age 45 years	12.0026 (11.9750, 12.0351)
β_2' [year ⁻¹]	Differentiation (death) rate in the I_2 after age 45 years	11.8311 (11.8015, 11.8581)
μ_0 [year ⁻¹]	Mutation rate in the normal stem cells	3.19×10^{-6} (2.77×10^{-6} , 3.69×10^{-6})
μ_1 [year ⁻¹]	Mutation rate in the I_1	1.02×10^{-4} (8.84×10^{-5} , 1.18×10^{-4})
μ_2 [year ⁻¹]	Mutation rate in the I_1	1.03×10^{-4} (6.37×10^{-5} , 1.75×10^{-4})

3.3. Parameter analysis

Table 2 shows the estimated parameter values of the optimal model. The net proliferation rates of intermediate cells with one mutation approach zero. It may be explained by the fact that normal breast stem cells need two mutations in tumor suppressor genes to become the intermediate cells that expand clonally to destroy the homeostasis of cells, or the first mutation occurs in genetic instability genes. Mutations in genetic instability genes increase the rate of mutational processes such as gene amplification, chromosome rearrangements, and gain or loss of whole chromosomes (Lengauer et al., 1998). By comparing the values of β_i and β_i' ($i = 1, 2$) in Table 2, we find that the net proliferation rates of cells before age 45 years are larger than those after age 45 years. The values of parameters for the models with more than three mutations and the two-stage model have also similar results (the detailed parameter values are displayed in the Supplementary Table S1). These suggest that menopause can reduce or inhibit clonal expansion of cells.

Evidence suggests that there is a mutator phenotype that acts as a mechanism in tumor processes, such as genome instability (Loeb, 1991; Leob et al., 2003). Genome instability generates genetic diversity and can contribute to cancer development. When the stability genes are inactivated, mutations in other genes occur at a higher rate (Vogelstein and Kinzler, 2004). To test the genetic instability, we make a hypothesis that all mutation rates are equal in the models, namely $\mu_0 = \mu_1 = \dots = \mu_k$. Our results show that only the models with two and three stages can fit the SEER data, whereas the models with more than three stages cannot fit the SEER data. In Figure 3, we plot the hazards for the models with two to four stages under this hypothesis. Thus, there is a mutator phenotype in the tumorigenesis of breast cancer.

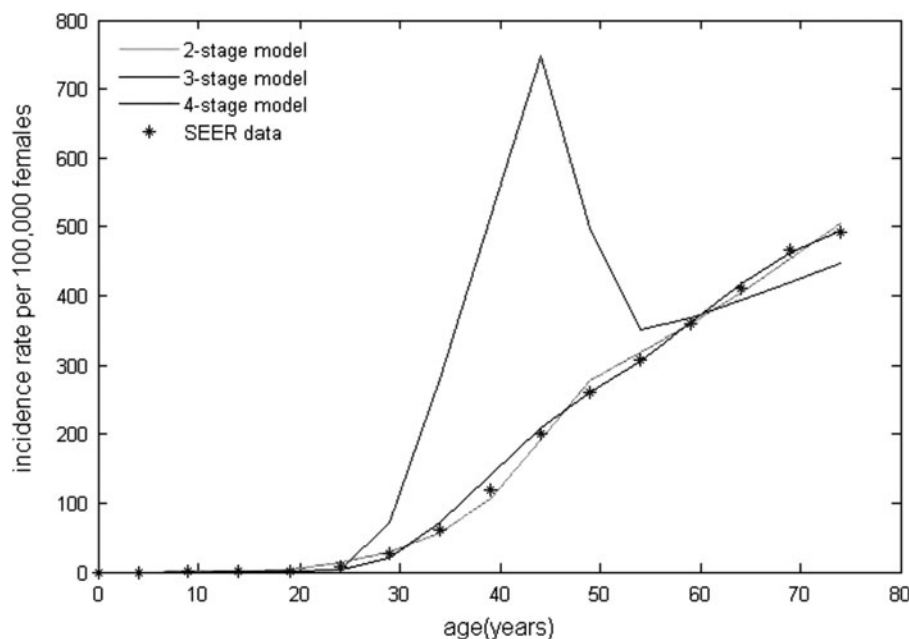


FIG. 3. The age-specific incidence rates of all races per 100,000 females for breast cancer from the SEER registry for the years 1990–1999 and rates predicted by the two to four-stage models under the hypothesis that all mutation rates are equal.

4. DISCUSSION

Our results suggest that there are 2–14 mutations in the genome of breast stem cells to become a malignant cell, which supports the results of Wood et al. (2007). Using our previous simulation for the data of breast cancer in women carrying germline BRCA1 and BRCA2 mutations, two gene mutations are required in the development of breast cancer (Simon and Zhang, 2008). The model with one mutation cannot fit the data of breast cancer in women carrying germline BRCA1 and BRCA2 mutations. Therefore, there are at least two somatic mutations such that a normal breast stem cell becomes a malignant cell in the genome of breast stem cell. Two to 14 genes are required to be involved in the genetic network for human breast cancer. These conclusions are useful for simplifying the interpretation of individual cancer genomes and the design of future breast cancer genome-sequencing efforts.

There is a mutator phenotype in the tumorigenesis of breast cancer, which can be explained by the genetic instability. The genetic instability is an early event of human cancer, which can be used as the early diagnosis of breast cancer (Lengauer et al., 1998; Zhang and Simon, 2005). Three gene mutations are more likely to occur in the progression of breast cancer. Two mutations are required to destroy the homeostasis of cells in this optimal model. Using the estimated parameter values in the models, we find that the clonal expansion of premalignant cells has an intense relationship with the hormone expression level of females. Menopause can reduce the growth of intermediate cells. Therefore, hormone therapy such as tamoxifen should be effective for breast cancer patients before their menopause.

To explain that many mutations are observed in malignant tumor cells, Loeb et al. (2003) proposed that there is a mutator phenotype that acts as a mechanism in tumor processes. However, Tomlinson and Bodmer (1999) have challenged the mutator phenotype hypothesis and claimed that selection for clonal expansion of intermediate cells is sufficient in tumor processes. This is the general debate for the relationship between mutation and selection. Our results indicate that there is a mutator phenotype in the tumorigenesis of breast cancer. Recent evidence indicates that epigenetic changes might “addict” cancer cells to altered signal transduction pathways during the early stages of tumor development (Klutstein et al., 2016; Herbst et al., 2017). However, epigenetic changes and genomic alterations cannot be easily identified. These findings can provide the guide to the analyses of genes.

ACKNOWLEDGMENTS

This research was partially supported by the National Nature Science Foundation of China (Grant Nos. 11071275 and 11371161) and by self-determined research funds of CCNU from the colleges’ basic research and operation of MOE (Grant No. CCNU16JCZX10).

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

- Armitage, P., and Doll, R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* 8, 1–12.
- Brouwer, A.F., Eisenberg, M.C., and Meza, R. 2016. Age effects and temporal trends in HPV-related and HPV-unrelated oral cancer in the United States a multistage carcinogenesis modeling analysis. *PLoS One* 11, e0151098.
- Brouwer, A.F., Meza, R., and Eisenberg, M.C. 2016. A systematic approach to determining the identifiability of multistage carcinogenesis models. *Risk Anal.* 37, 1375–1387.
- Crump, K.S., Subramaniam, R.P., and Van Landingham, C.B. 2005. A numerical solution to the nonhomogeneous two-stage MVK model of cancer. *Risk Anal.* 25, 921–926.
- Curtius, K., Hazelton, W.D., Jeon, J., et al. 2015. A multiscale model evaluates screening for neoplasia in Barrett’s Esophagus. *PLoS Comput. Biol.* 11, e1004272.
- Eidmüller, M., Holmberg, E., Jacob, P., et al. 2015. Breast cancer risk and possible mechanisms of radiation-induced genomic instability in the Swedish hemangioma cohort after reanalyzed dosimetry. *Mutat. Res.* 775, 1–9.
- Hanahan, D., and Weinberg, R.A. 2011. The hallmarks of cancer: The next generation. *Cell* 144, 646–674.

- Harris, T.E. 1963. *The Theory of Branching Processes*. Prentice-Hall Applied Mathematics Series, Englewood Cliffs, NJ.
- Hazelton, W.D., Curtius, K., Inadomi, J.M., et al. 2015. The role of gastroesophageal reflux and other factors during progression to esophageal adenocarcinoma. *Cancer Epidemiol. Biomarkers Prev.* 24, 1012–1023.
- Heidenreich, W.F., Luebeck, E.G., and Moolgavkar, S.H. 1997. Some properties of the hazard function of the two-mutation clonal expansion model. *Risk Anal.* 17, 391–399.
- Herbst, A., Vdovin, N., Gacesa, S., et al. 2017. Methylated free-circulating HPP1 DNA is an early response marker in patients with metastatic colorectal cancer. *Int. J. Cancer* 140, 2134–2144.
- Jeon, J., Luebeck, E.G., and Moolgavkar, S.H. 2006. Age effects and temporal trends in adenocarcinoma of the esophagus and gastric cardia (United States). *Cancer Causes Control* 17, 971–981.
- Klutstein, M., Nejman, D., Greenfield, R., et al. 2016. DNA methylation in cancer and aging. *Cancer Res.* 76, 3446–3450.
- Knudson, A.G. 1971. Mutation and cancer: Statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* 68, 820–823.
- Knudson, A.G. 2001. Two genetic hits (more or less) to cancer. *Nat. Rev. Cancer* 1, 157–162.
- Lengauer, C., Kinzler, K.W., and Vogelstein, B. 1998. Genetic instabilities in human cancers. *Nature* 396, 643–649.
- Loeb, L.A. 1991. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* 51, 3075–3079.
- Leob, L.A., Loeb, K.R., and Anderson, J.P. 2003. Multiple mutations and cancer. *Proc. Natl. Acad. Sci. USA* 100, 776–781.
- Luebeck, E.G., and Moolgavkar, S.H. 2002. Multistage carcinogenesis and the incidence of colorectal cancer. *Proc. Natl. Acad. Sci. USA* 99, 15095–15100.
- Meza, R., Jeon, J., Moolgavkar, S.H., et al. 2008. Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc. Natl. Acad. Sci. USA* 105, 16284–16289.
- Moolgavkar, S.H., Day, N.E., and Stevens, R.G. 1980. Two-stage model for carcinogenesis: Epidemiology of breast cancer in females. *J. Natl. Cancer Inst.* 65, 559–569.
- Moolgavkar, S.H., Holford, T.R., Levy, D.T., et al. 2012. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975–2000. *J. Natl. Cancer Inst.* 104, 541–548.
- Moolgavkar, S.H., and Luebeck, G. 1990. Two-event model for carcinogenesis: Biological, mathematical, and statistical considerations. *Risk Anal.* 10, 323–341.
- Radmacher, M.D., and Simon, R. 2000. Estimation of Tamoxifen's efficiency for preventing the formation and growth of breast tumors. *J. Natl. Cancer Inst.* 92, 48–53.
- Ries, L.A.G., Eisner, M.P., Kosary, C.L., et al. 2002. *SEER Cancer Statistics Review, 1973–1999*. National Cancer Institute, Bethesda, MD.
- Simon, R., and Zhang, X. 2008. On the dynamics of breast tumor development in women carrying germline BRCA1 and BRCA2 mutations. *Int. J. Cancer* 122, 1916–1917.
- Tomasetti, C., Marchionni, L., Nowak, M.A., et al. 2015. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc. Natl. Acad. Sci. USA* 112, 118–123.
- Tomlinson, I.P., and Bodmer, W. 1999. Selection, the mutation rate and cancer: Ensuring that the tail does not wag the dog. *Nat. Med.* 5, 11–12.
- Vogelstein, B., and Kinzler, K.W. 2004. Cancer genes and the pathways they control. *Nat. Med.* 10, 789–799.
- Wood, L.D., Parsons, D.W., Jones, S., et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science* 318, 1108–1113.
- Zhang, X., Fang, Y., Zhao, Y., et al. 2014. Mathematical modeling the pathway of human breast cancer. *Math. Biosci.* 253, 25–29.
- Zhang, X., and Simon, R. 2005. Estimating the number of rate limiting genomic changes for human breast cancer. *Breast Cancer Res. Treat.* 91, 121–124.

Address correspondence to:

Prof. Tianhai Tian
School of Mathematical Science
Monash University
Melbourne, Vic 3800
Australia

E-mail: tianhai.tian@monash.edu

Prof. Xinan Zhang
School of Mathematics and Statistics
Central China Normal University
Wuhan Luoyu Road 152th
Wuhan 430079
P.R. China

E-mail: xinanzhang@mail.cnu.edu.cn