

OPEN

Mathematical modelling the pathway of genomic instability in lung cancer

Lingling Li¹, Xinan Zhang², Tianhai Tian³ & Liuyong Pang⁴

Genomic instability plays a significant role in lung cancer. Although substantial research has been conducted using both clinical and theoretical studies, it is still a hotly debated issue to whether genomic instability is necessary or whether genomic instability precedes oncogenes activation and tumor suppressor genes inactivation for lung cancer. In response to this issue, we come up with a mathematical model incorporating effects of genomic instability to investigate the genomic instability pathway of human lung cancer. The presented model are applied to match the incidence rate data of lung cancer from the Life Span Study cohort of the atomic bomb survivors in Nagasaki and Hiroshima and the Surveillance Epidemiology and End Results registry in the United States. Model results suggest that genomic instability is necessary in the tumorigenesis of lung cancer, and genomic instability has no significant impact on the net proliferation rate of cells by statistical criteria. By comparing the results of the LSS data to those of the SEER data, we conclude that the genomic instability pathway exhibits a sensitivity to radiation exposure, more intensive in male patients.

Genomic instability (GI) is a important hallmark of almost all human cancers¹, which is a very active area of research in cancer biology. There are two main types of GI in human cancer². One is microsatellite instability (MIN) which is featured by the amplification or deletion of oligonucleotide number in microsatellite sequences^{3–5}. The other is chromosomal instability (CIN) that involves the acquisition or loss of large parts of chromosomes or entire chromosomes during the period between cell divisions. However, MIN is rarely spotted in cancers but colon cancer. CIN is widespread among solid tumors^{6,7}. As is known to all, the tumorigenesis is caused by the accumulated alterations in oncogenes genes, tumor suppressor genes and genes controlling genomic stability⁸. Alterations in tumor suppressor genes and oncogenes genes affect the clonal expansion of cells, while alterations in genes controlling genomic stability accelerate the mutational processes such as chromosome recombination, loss or gain of large parts of chromosomes or entire chromosomes, gene amplification and so on^{2,8}. Thus, GI drives the development of cancers by increasing the spontaneous mutation rate.

Biologically-based mechanistic model is a useful tool to study the initiation and progression of cancer. Nowell presented a carcinogenesis model based on clonal selection and GI in 1976⁹. Later, Tomlinson and Bodmer proposed that tumor processes is mainly result from selection of clonal expansion of mutated cells¹⁰. However, Loeb introduced the concept of mutator phenotype, which suggests that mutation in a stability gene is a necessary event in carcinogenesis^{11,12}. In addition, a large amount of experimental and theoretical studies has been conducted to analyze the mechanism of GI in the development of cancer^{13–17}. Nowak *et al.*^{13–15} presented a mathematical model with CIN to explore the dynamics of CIN in the progression of colon cancer by means of the modification to a two-stage clonal expansion model given by Moolgavkar *et al.*¹⁸. Based on this mathematical framework, the model with multiple genetic pathway was set up to study cancer^{19–22}. Zöllner *et al.*²³ applied the model incorporated GI to match the mortality data of lung cancer in the Mayak-workers cohort and study the effect of plutonium on the risk of lung cancer.

Lung cancer is the leading cause of cancer mortality all over the world^{24,25}, and the second leading cause of cancer mortality in the Life Span Study (LSS) cohort of the atomic bomb survivors in Nagasaki and Hiroshima²⁶. Cigarette smoking and radiation are two main factors to increase the risks of lung cancer^{26,27}. Evidence indicates that GI plays a curial role in the initiation and progression of lung cancer^{28–31}. Nearly all lung cancer patients

¹School of Science, Xi'an Polytechnic University, Xi'an, 710048, P.R. China. ²School of Mathematics and Statistics, Central China Normal University, Wuhan, 430079, P.R. China. ³School of Mathematical Science, Monash University, Melbourne, Vic 3800, Australia. ⁴School of Mathematics, Huanghuai University, Zhumadian, Henan, P.R. China. Correspondence and requests for materials should be addressed to L.L. (email: linglinglimath@163.com)

Age	Cases		Person years	
	Male	Female	Male	Female
30–34	1	0	686002	854695
35–39	1	7	657211	1019250
40–44	10	3	644822	1129502
45–49	13	12	560137	1028308
50–54	19	31	584624	1189838
55–59	37	33	461360	1058327
60–64	70	49	493031	947685
65–69	122	62	489571	813978
70–74	115	72	323201	614319
75–79	69	47	188111	372679

Table 1. The age-specific lung cancer incidence data of male and female from the LSS cohort of the atomic bomb survivors in Hiroshima and Nagasaki for the year 1958–1987.

exhibit GI which is resulted in an abnormal number of chromosomes known as aneuploidy^{32,33}. CIN is an important DNA alteration process for lung cancer, which is a significant poor prognostic factor^{34,35}. In addition, cancer risk may be attributed to GI, which plays a vital role in carcinogenesis and in clinical practice. Nevertheless, the mechanisms of GI in the development lung cancer are still unclear. Therefore, understanding the role of GI is very imperative to guide therapeutic interventions for lung cancer.

It is very important to determine whether GI is an early event or a late event in lung cancer development, which is a long-standing debate in cancer genetics. In this article, we describe a model with GI to analyze the pathways of GI in the initiation and progression of lung cancer. Using the age-specific incidence rate data of lung cancer from the LSS cohort in Nagasaki and Hiroshima during 1958–1987 and the Surveillance Epidemiology and End Results (SEER) registry in the United States during 1993–2012 as the study system, the model with GI is applied to match these data. We firstly discuss the impact of GI on the net reproduction rate of cells and then analyze the issues that whether GI is necessary for the process from normal stem cells into a malignant cell in the lung tissue, and whether GI precedes oncogenes activation and tumor suppressor genes inactivation for lung cancer.

Methods

The LSS data. The LSS cohort of the atomic bomb survivors in Nagasaki and Hiroshima is a primary data source for analyzing the impact of radiation on cancer risk, which involves 120,321 persons. Lung cancer occupies almost 10% of all cancers in the LSS cohort. The LSS data are stratified by region (Nagasaki and Hiroshima), sex (male and female), attained age (ages 0–85+), age at exposure (ages 0–60+), calendar time (1958–1987), colon dose (radiation dose detected in the colon), and other factors²⁶. These data can be obtained at the Radiation Effects Research Foundation (RERF) website (<http://www.rerf.jp>). The detailed description of the data can be found in the refs^{26,36}. There is no significant difference in patients between Nagasaki and Hiroshima^{36,37}. Therefore, our analyses are based on adjusted person-years and incidence cases of lung cancer grouped by sex and attained age for calendar years 1958–1987 in Nagasaki and Hiroshima (see Table 1). In our study, the lung cancer data for persons between the age of 0 and 29 years are ignored since they are equal to zero.

The SEER data. Lung cancer incidence data were obtained from the SEER registry in the United States for the years 1975–2012 (www.seer.cancer.gov)³⁸. The reported lung cancer incidence data were grouped by sex (male and female), age (0–85+), and calendar year (1975–2012) in the eighteen SEER geographic areas³⁹. The SEER population files (data from U.S. Census Bureau) provided the population bases, which was stratified by gender (male and female), calendar year (1975–2012) and age (ages 0–85+). Our analyses employed lung cancer incidence data of all races in males and females for the years 1993–2012. We considered the data for persons aged 30–79 since the lung cancer incidence rates are closed to zero for the age group 0–29. The detailed data are displayed in Table 2.

Mathematical model. CIN and the loss of heterozygosity play important roles in the progression of lung cancer⁴⁰. The predominant form of GI in lung tumors is CIN. Only a single mutational ‘hit’ is required to produce the CIN phenotype². Thus, we assume that cell mutations occur along two different pathways in lung cancer development. The model is illustrated schematically in Fig. 1. The first (upper) path does not generate GI, whereas the second (below) path is activated via transition rates v_i ($i = 0, 1$), corresponding to the alteration of genes in maintaining genomic integrity or stability. CIN accelerates the rate of inactivating tumour suppressor genes and activating oncogenes. Therefore mutation rates in the cells with GI are higher than those of the cells without GI, $\mu_{i,GI} \gg \mu_i$ ^{8,23}. When all transition rates v_i ($i = 0, 1$) equal zero, the model corresponds to the two-stage model. Our model assumes that malignant tumor occurs with probability one if a malignant cell is generated. Therefore, the mutation rates μ_1 and $\mu_{1,GI}$ are to be regarded as net or effective mutation rates, which means that various defense mechanisms are taken into account in our model such as immune surveillance. As described in the ref.²⁶, we suppose that there are N normal stem cells in the lung tissue; any normal cell mutates to premalignant cells in compartment I_1 at a rate of μ_0 ; and a type of cells with GI at a rate of v_0 . After that, any mutated cell in compartment I_1 ($I_{1,GI}$) forms two equivalent daughter cells at a rate α_1 ($\alpha_{1,GI}$); and an equivalent daughter cell and an

Age	Cases		Person years	
	Male	Female	Male	Female
30–34	517	545	45116274	43026403
35–39	1521	1649	45911554	44503264
40–44	4813	4741	46042363	45269976
45–49	11742	11041	43781154	43664229
50–54	23596	19595	38983073	39518370
55–59	37866	29581	32164300	33232561
60–64	52933	41006	25341059	26903037
65–69	66214	53061	19538332	21835643
70–74	70736	58317	15426031	18524069
75–79	65406	56679	12003384	16000285

Table 2. The age-specific lung cancer incidence data of male and female from SEER registry for the year 1993–2012.

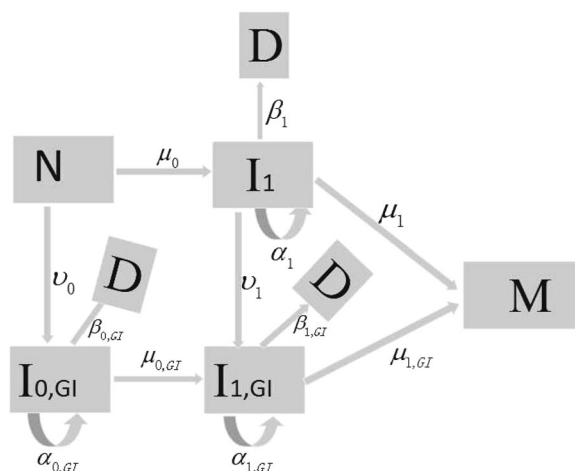


Figure 1. The schematic representation of GI model for carcinogenesis. N denotes the normal cell; I_1 the compartment of intermediate cell without GI; $I_{p,GI}$ ($i = 0, 1$) the compartment of intermediate cell with GI; D the dead or differentiated cell; M the malignant cell. $\mu_i(t)$, $\mu_{p,GI}(t)$ are the mutation rates per cell per year at time t , respectively. $\alpha_i(t)$, $\beta_i(t)$, $\alpha_{p,GI}(t)$ and $\beta_{p,GI}(t)$ are the growth rate and death rate per cell per year at time t , respectively.

mutated cell in compartment M at a rate μ_1 ($\mu_{1,GI}$); and die or differentiation at a rate β_1 ($\beta_{1,GI}$). Each mutated cell in compartment I_1 can also turn into a daughter cell and a mutated cell with GI at a rate ν_1 .

In the model, we let $X_1(t)$ to denote the number of mutated cells without GI (compartment I_1) at time t , $Y_i(t)$ ($i = 0, 1$) the number of mutated cells with GI (compartment $I_{i,GI}$) at time t , and $Z(t)$ the number of fully malignant cells (compartment M) at time t . We define the probability generating functions as follows:

$$\psi(x_1, y_0, y_1, z; t) = \sum_{i_1, j_0, j_1, k} p\{X_1(t) = i_1, Y_0(t) = j_0, Y_1(t) = j_1, Z(t) = k | X_1(0) = 0, Y_0(0) = 0, Y_1(0) = 0, Z(0) = 0\} x_1^{i_1} y_0^{j_0} y_1^{j_1} z^k, \tag{1}$$

$$\varphi(x_1, y_1, z; t) = \sum_{i_1, j_1, k} p\{X_1(t) = i_1, Y_1(t) = j_1, Z(t) = k | X_1(0) = 1, Y_1(0) = 0, Z(0) = 0\} x_1^{i_1} y_1^{j_1} z^k, \tag{2}$$

$$\phi_0(y_0, y_1, z; t) = \sum_{j_0, j_1, k} p\{Y_0(t) = j_0, Y_1(t) = j_1, Z(t) = k | Y_0(0) = 1, Y_1(0) = 0, Z(0) = 0\} y_0^{j_0} y_1^{j_1} z^k, \tag{3}$$

and

$$\phi_1(y_1, z; t) = \sum_{j_1, k} p\{Y_1(t) = j_1, Z(t) = k | Y_1(0) = 1, Z(0) = 0\} y_1^{j_1} z^k, \quad (4)$$

where x_1, y_0, y_1 and z are the arguments for these probability generating functions.

These probability generating functions satisfy the Kolmogorov backward equations as follows^{41–43}:

$$\begin{cases} \frac{d\varphi}{dt}(t) = -[\alpha_1 + \beta_1 + \mu_1 + v_1] \varphi(t) + \alpha_1 \varphi^2(t) + v_1 \varphi(t) \phi_1(t) + \mu_1 \varphi(t) z + \beta_1 \\ \frac{d\phi_0}{dt}(t) = -[\alpha_{0,GI} + \beta_{0,GI} + \mu_{0,GI}] \phi_0(t) + \alpha_{0,GI} \phi_0^2(t) + \mu_{0,GI} \phi_0(t) \phi_1(t) + \beta_{0,GI}, \\ \frac{d\phi_1}{dt}(t) = -[\alpha_{1,GI} + \beta_{1,GI} + \mu_{1,GI}] \phi_1(t) + \alpha_{1,GI} \phi_1^2(t) + \mu_{1,GI} \phi_1(t) z + \beta_{1,GI} \end{cases} \quad (5)$$

and

$$\frac{d\psi}{dt}(t) = \mu_0 N \psi(t) [\varphi(t) - 1] + v_0 N \psi(t) [\phi_0(t) - 1], \quad (6)$$

where we have dropped the arguments x_1, y_0, y_1 and z in ψ, φ, ϕ_0 and ϕ_1 for simplicity. The detailed derivation for above equations can be seen in Appendix.

Here, the hazard function, $h(t)$, signifies the cancer incidence rate at time t . It can be written as^{18,39,41}

$$h(t) = - \frac{d\psi(1, 1, 1, 0; t)}{dt} / \psi(1, 1, 1, 0; t). \quad (7)$$

It follows from Eq. (6) that

$$h(t) = -\mu_0 N [\varphi(1, 1, 0; t) - 1] - v_0 N [\phi_0(1, 1, 0; t) - 1]. \quad (8)$$

Hence, the solutions of the hazard function, $h(t)$, can be obtained by solving $\varphi(1, 1, 0; t)$ and $\phi_0(1, 1, 0; t)$. By above probability generating functions, we have $\varphi(1, 1, 0; 0) = 1$ and $\phi_i(1, 1, 0; 0) = 1$ ($i=0, 1$). Thus, the solutions of $\varphi(1, 1, 0; t)$ and $\phi_i(1, 1, 0; t)$ can be obtained by Eq. (5) with the initial conditions $\varphi(1, 1, 0; 0) = 1$ and $\phi_i(1, 1, 0; 0) = 1$.

For the two-stage model ($v_0 = v_1 = 0$), however, there is a closed-form solution of hazard function given by,

$$h(t) = \frac{\mu_0 N}{\alpha_1} \left(\frac{pq(e^{-qt} - e^{-pt})}{qe^{-pt} - pe^{qt}} \right), \quad (9)$$

where $p, q := \frac{1}{2}(-(\alpha_1 - \beta_1 - \mu_1) \mp \sqrt{(\alpha_1 - \beta_1 - \mu_1)^2 + 4\alpha_1\mu_1})$.

Parameter estimation. Given set of lung cancer cases $\{o_a\}$ with corresponding adjusted person-years $\{n_a\}$, we derive the likelihood function for the model in the following way^{39,44}. The number of lung cancer incidence cases is assumed to be a Poisson distribution with mean $\lambda_a = n_a h(a)$, where $h(a)$ is the hazard function with parameter set $\Theta = (N, \alpha_1, \beta_1, \alpha_{i,GI}, \beta_{i,GI}, \mu_{i,GI}, \mu_{i,GI}, v_i)$ ³⁹. Lung cancer cases are supposed to be independent, and then the likelihood function for entire data set of lung cancer cases $\{o_a\}$ can be written as

$$L(\Theta) = \prod_a \frac{\exp\{-\lambda_a\} \lambda_a^{o_a}}{o_a!}. \quad (10)$$

The negative log likelihood (NLL) function can be written as

$$NLL(\Theta) = -\sum_a (-\lambda_a + o_a \log \lambda_a - \log o_a!). \quad (11)$$

The NLL of the model is minimized by using the optimization routine `fminsearch` in MATLAB. The optimal parameters are obtained by minimizing the deviance, $Dev = 2NLL$. We use the AIC to measure the goodness of model fit, which can available avoid overfitting of models⁴⁵. The AIC is equal to $2(NLL + n)$ where n denotes the number of model parameters. Hence, a smaller AIC represents a better fitting result.

Analysis of genetic instability. GI has long been hypothesized to be a critical feature of cancer. However, the mechanism of GI is still not completely clear for lung cancer. In this article, we mainly pay attention to the questions that whether GI is necessary for lung cancer, and whether GI precedes oncogenes activation and tumor suppressor genes inactivation if GI occurs in the development of lung cancer. Hence, we consider the model with GI. For the model with GI, the following two cases are discussed: namely (1) GI is identified in the earliest stage of tumorigenesis ($v_0 \neq 0$); (2) GI occurs after inactivation of tumor suppressor genes or activation of oncogenes ($v_0 = 0$). The LSS data and the SEER data are applied to analyze these issues. For the LSS data, the lung cancer cases are strongly correlated to radiation exposure, which can be used to examine the effect of radiation exposure on GI by comparing with the fitting results of the SEER data.

Parameters	The LSS data		The SEER data	
	Male patients	Female patients	Male patients	Female patients
$\frac{\mu_0 N}{\alpha_1}$	0.0031	0.0015	0.0356	0.0253
p	-0.1584	-0.1226	-0.1718	-0.1610
q	0.9622×10^{-5}	5.3857×10^{-5}	0.6232×10^{-5}	1.2465×10^{-5}

Table 3. The estimated values of parameter combinations in the two-stage model by fitting the SEER data and the LSS data respectively.

Model	No. Parameter	The LSS data				The SEER data			
		Male patients		Female patients		Male patients		Female patients	
		Deviance	AIC	Deviance	AIC	Deviance	AIC	Deviance	AIC
$\gamma_1 > \gamma_{1>GI}$	5	52.684	62.684	65.677	75.677	1094.5	1104.5	1169.3	1179.3
$\gamma_1 = \gamma_{1>GI}$	4	52.615	60.615	61.272	69.272	1065.5	1073.5	1132.6	1140.6
$\gamma_1 < \gamma_{1>GI}$	5	52.399	62.399	61.409	71.409	1032.6	1042.6	1132.4	1142.4

Table 4. Comparison of the GI models with the three different hypotheses for the effect of GI on clonal expansion of cells in the development of lung cancer.

Results

Results from the two-stage model are shortly presented. The changes of clonal expansion in premalignant cells due to GI are investigated. Furthermore, the GI model with different pathway is discussed in detail.

The two-stage model. For the two-stage model ($v_0 = v_1 = 0$), we derive the closed-form solution of the hazard function (9). This model has three identifiable parameter combinations, $\frac{\mu_0 N}{\alpha_1}$, p and q . We can obtain the following formulas by the expressions of p and q ,

$$\begin{cases} p + q = -(\alpha_1 - \beta_1 - \mu_1) \\ pq = -\alpha_1 \mu_1 \end{cases}, \quad (12)$$

Therefore, the $\alpha_1 - \beta_1 - \mu_1$, $\alpha_1 \mu_1$ and $\frac{\mu_0 N}{\alpha_1}$ can be determined from the data. Here, we assume that the latent time from a malignant cell to clinical detection is 5 years^{23,26}. By fitting the LSS data and the SEER data, we obtain the values of these parameter combinations, which is displayed in Table 3. By these values of parameters in Table 3, we obtain that the net proliferation of premalignant cells in the SEER data is slightly larger than that in the LSS data, and the transformation rate from the premalignant cells to malignant cell for LSS data is larger than that for SEER data. If we fix the values of α_1 and N , the other parameter values can be determined. Similarly, we set the values of cell mutation rates to be equal, $\mu_0 = \mu_1$, and know the number of stem cells in the lung tissue, N , this identifiability problem can also be solved. The hazard function mainly rely on the net growth rate of cells, $\alpha_1 - \beta_1$, and is insensitive to a broad range of assumed value for α_1 . Thus we set the growth rate to $\alpha_1 = 12$ per year^{39,46}. If we accept the estimate of N provided by Hazelton *et al.*⁴⁷, about 10^7 , the other parameters in the two-stage model can be identified.

The model with GI. For the model with GI, there is no analytical solution and more non-identifiability parameters. Generically, the widely used approach to deal with this non-identifiability problem is to set the background mutation rates equal to each other and assume a reasonable value for some parameters or use a new set of parameters^{48,49}. Therefore, we give the following limitations:

- the mutation rates in cells are assumed as $\mu_0 = \mu_1$ and $\mu_{0>GI} = \mu_{1>GI}$, which is reasonable by fitting the data.
- the birth rate and death (differentiation) rate in the compartment $I_{0>GI}$ are the same, namely $\alpha_{0>GI} = \beta_{0>GI}$, since cells without mutation of oncogene or tumor suppressor gene have not growth advantage.
- the instability transition rates are equal, namely $v_0 = v_1$, since they belong to the inactivation of the same type genomic-integrity gene.

In the GI model, the transition rates to achieve GI mainly depend on the number of dominant CIN genes, a reasonable range of the rate is from 10^{-7} to 10^{-5} ⁵⁰. Here, we discuss the effect of GI on clonal expansion of cells. The following three scenarios are considered:

- GI decreases the rate of the cell clonal expansion, that is $\gamma_1 = \alpha_1 - \beta_1$ larger than $\gamma_{1>GI} = \alpha_{1>GI} - \beta_{1>GI}$.
- GI has no effect on clonal expansion of cells, that is $\gamma_1 = \alpha_1 - \beta_1$ equals $\gamma_{1>GI} = \alpha_{1>GI} - \beta_{1>GI}$.
- GI increases the rate of the cell clonal expansion, that is $\gamma_1 = \alpha_1 - \beta_1$ smaller than $\gamma_{1>GI} = \alpha_{1>GI} - \beta_{1>GI}$.

The comparison of the models with these three scenarios is shown in Table 4. By Table 4, we find that the values of AIC for the scenario, $\gamma_1 = \gamma_{1>GI}$ are the smallest apart from male patients in the SEER data. However, the

Model	No. Parameter	The LSS data				The SEER data			
		Male patients		Female patients		Male patients		Female patients	
		Deviance	AIC	Deviance	AIC	Deviance	AIC	Deviance	AIC
Without GI ($\vartheta_i=0$)	2	77.415	81.415	64.118	68.118	3217.5	3221.5	8648.3	8652.3
Early GI ($\vartheta_i \neq 0$)	4	52.615	60.615	61.272	69.272	1065.5	1073.5	1132.6	1140.6
Early GI ($\vartheta_1=0$)	4	52.560	60.560	65.199	73.199	1110.3	1118.3	1196.5	1204.5
Late GI ($\vartheta_0=0$)	4	55.167	63.167	57.242	65.242	691.5	699.5	600.1	608.1

Table 5. Comparison of the GI model with the different pathway. $i=0, 1$.

values of AIC for the three scenarios do not change very much for male patients in the SEER data. This suggests that GI has no significant effect on clonal expansion of cells in the lung cancer progression. Hence, we choose the GI model with $\gamma_1 = \gamma_{1,GI}$.

Model comparison. A comparison of the GI model for different pathway is presented in Table 5. The results indicate that the model with GI is a significant improvement on model accuracy compared to the model without GI for the SEER data. It has been shown that the model with GI has better goodness-of-fit than the model without GI for the SEER data. For the LSS data, the fitting result of the model with GI is better than that of the model without GI for male patients, and the optimal model is the model with $v_1=0$ for female patients. Hence, we can obtain that GI is needed in the development of lung cancer. By the values of AIC in Table 5, we find that there are significant difference in the fitting results for the SEER data and those for the LSS data. For the SEER data, the AIC value of the model with $v_1=0$ is far less than the other models for male patients and female patients. It turns out that GI occurs after oncogenes activation or tumor suppressor genes inactivation in the lung cancer development. For the LSS data, however, the model with $v_0=v_1 \neq 0$ and the one with $v_1=0$ have the smaller AIC than the one with $v_0=0$ for male patients. The preferred model is the model with $v_1=0$ for male patients. For female patients in the LSS data, however, the GI model with $v_1=0$ has the smaller AIC than the other models, and the AIC values of the model with different pathway do not get too much of a bump. This difference between the fitting results for the SEER data and those for the LSS data is due to radiation exposure, since the LSS data is strongly associated with radiation exposure. Therefore, we infer that radiation exposure can induce the mutation of genes in maintaining genomic integrity or stability for lung cancer, especially for male lung cancer patients. The fitting results of the model with different pathway are displayed in Fig. 2, which indicates that the model with GI is superior to the model without GI for the fitting of the lung cancer incidence data.

Discussion

whether GI is necessary for lung cancer or what stage of lung cancer development it arises remain hotly debated. In this paper, we have proposed a model with GI to investigate the issue. The LSS data closely related with radiation exposure⁵¹ are used to study the pathway of GI in the development of lung cancer. To eliminate the effect of radiation on GI, the SEER data are also applied to address this issue. Our results suggest that GI is highly significant for lung cancer in the SEER data. By the comparing the inference results from the model with different pathway, we obtain that GI is a late event in lung cancer development for the SEER data, while for male patients in the LSS data GI is most likely to occur before oncogenes activation and tumor suppressor genes inactivation. In addition, for female patients in the LSS data, the model with $v_1=0$ has no more significant improvement than the other models. Therefore, we conclude that the pathway of GI is sensitive to radiation in lung cancer development, and the sensitivity is more intensive for male patients.

CIN is the most common type of GI in most cancers, which leads to the high gene mutation rate by chromosome number and structure alterations over time in mutated cells with CIN compared with the cells without GI⁵. CIN is the dominating form of GI in lung tumors. Thus the mutation rate of GI in the model typically relies on the number of genes controlling chromosome stability in the genome. Our work mainly pays attention to the relationship between GI and lung cancer and the question that whether GI precedes oncogenes activation and tumor suppressor genes inactivation. Two different data cohorts are chosen to address these issues. The LSS data is a main data source used for analyzing the cancer risks from radiation exposure^{26,52,53}, and the SEER data is widely used to study the development of cancer^{26,54–56}. By the fitting of these data, we obtain that GI has no significant influence on the clonal expansion of cells, and the model with GI offers an improved description of the data. In addition, radiation exposure can affect the mechanism of GI in the progression of lung cancer. These conclusions will not only contribute to understanding the mechanisms of lung tumorigenesis, but also provide constructive suggestions for the prediction, diagnosis and treatment of lung cancer.

Although a number of works have been designed to study lung cancer, there is limited work to discuss the mechanism of GI in lung cancer by using the detailed mathematical framework. The major contribution of our work is to discuss the effect of GI on the net reproduction rate of cells and the pathway of GI in lung cancer development by the GI model. However, our proposed model can not account for the specific genes involving the GI pathway in the initiation and progression of lung cancer, and the results are obtained only by considering the incidence rate of lung cancer. More detailed biological and medical data are needed to further support our results. Besides, the GI pathway in different types of lung cancer may be different, which requires some additional information to analyze it. Recently, research demonstrated that radiation exposure is mainly affecting the pathway

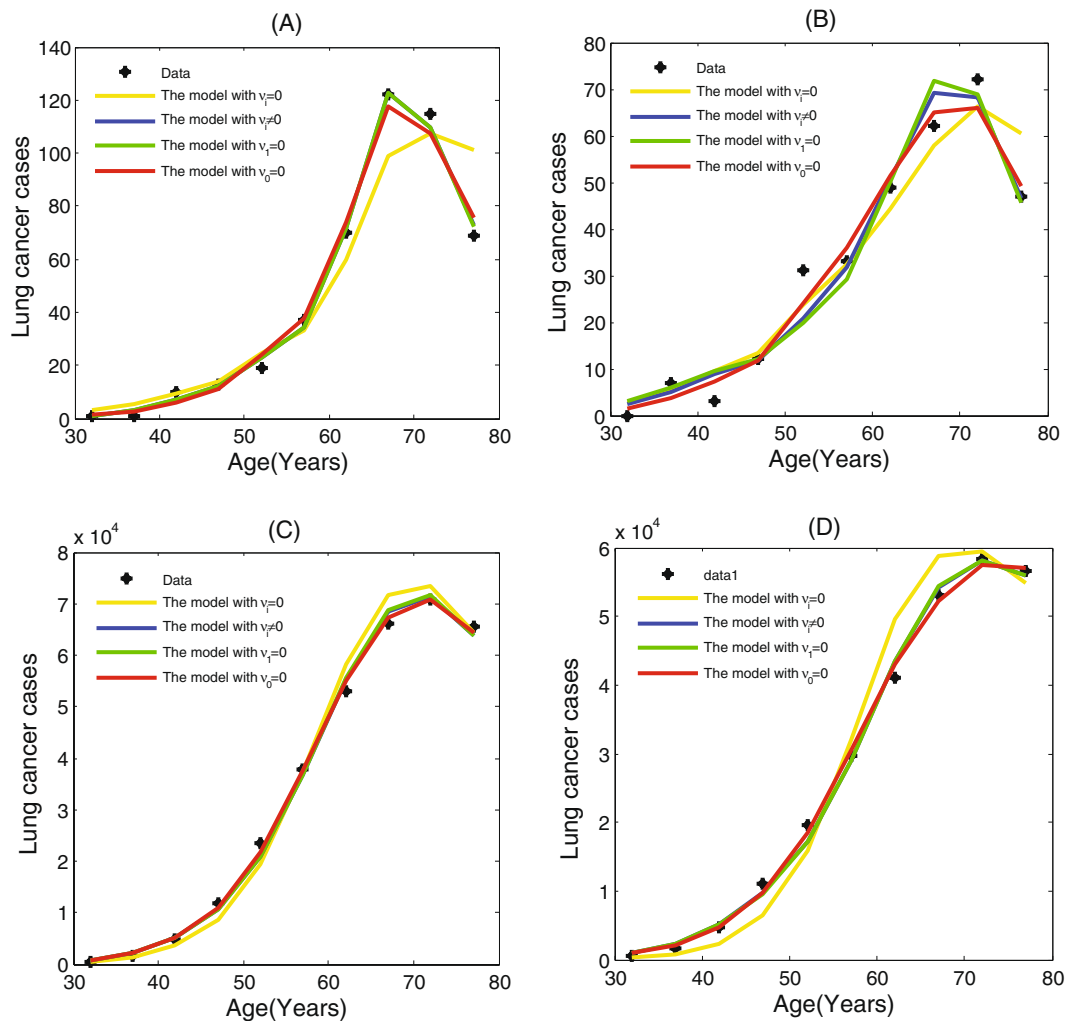


Figure 2. The age-specific lung cancer cases data for male patients and female patients from the LSS cohort for the year 1958–1987 and the SEER registry for the year 1993–2012, and cases predicted by the models. (A) Prediction for male patients in the LSS data. (B) Prediction for female patients in the LSS data. (C) Prediction for male patients in the SEER data. (D) Prediction for female patients in the SEER data.

with transmembrane receptor–mutant for lung adenocarcinoma by connecting molecular biology with epidemiology⁵³. Hence, the model incorporated growing knowledge on carcinogenesis processes is required to further study the mechanism of GI in lung cancer.

References

- Hanahan, D. & Weinberg, R. A. The hallmarks of cancer: the next generation. *Cell* **144**(5), 646–674 (2011).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**(6712), 643–649 (1998).
- Fishel, R. *et al.* The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**(5), 1027–1038 (1993).
- Leach, F. S. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**(6), 1215–1225 (1993).
- Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability—an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**(3), 220–228 (2010).
- Solomon, E., Borrow, J. & Goddard, A. D. Chromosome aberrations and cancer. *Science* **254**(5035), 1153–1160 (1991).
- Vogelstein, B. & Kinzler, K. W. *The genetic basis of human cancer* (McGraw-Hill, Toronto, Ontario, 1998).
- Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10**(8), 789–799 (2004).
- Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**(4260), 23–28 (1976).
- Tolinson, I. P. & Bodmer, W. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat. Med.* **5**(1), 11–12 (1999).
- Loeb, L. A. A mutator phenotype in cancer. *Cancer Res.* **61**(8), 3230–3239 (2001).
- Loeb, L. A. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat. Rev. Cancer* **11**(6), 450–457 (2011).
- Nowak, M. A. *et al.* The role of chromosomal instability in tumor initiation. *Proc. Natl. Acad. Sci. USA* **99**(25), 16226–16231 (2002).
- Komarova, N. L., Sengupta, A. & Nowak, M. A. Mutation–selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J. Theor. Biol.* **223**(4), 433–450 (2003).
- Nowak, M. A., Michor, F. & Iwasa, Y. Genetic instability and clonal expansion. *J. Theor. Biol.* **241**(1), 26–32 (2006).
- Abbas, I. *et al.* *In vitro* short-term exposure to air pollution pm 2.5–0.3 induced cell cycle alterations and genetic instability in a human lung cell coculture model. *Environ. Res.* **147**, 146–158 (2016).

17. Wang, G. & Vasquez, K. M. Effects of replication and transcription on DNA structure-related genetic instability. *Genes* **8**(1), 17–34 (2017).
18. Moolgavkar, S. H., Dewanji, A. & Venzon, D. J. A stochastic two-stage model for cancer risk assessment. I: The hazard function and the probability of tumor. *Risk Anal.* **8**(3), 383–392 (1988).
19. Little, M. P. & Wright, E. G. A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. *Math. Biosci.* **183**(2), 111–134 (2003).
20. Eidemüller, M., Holmberg, E., Jacob, P., Lundell, M. & Karlsson, P. Breast cancer risk and possible mechanisms of radiation-induced genomic instability in the Swedish hemangioma cohort after reanalyzed dosimetry. *Mutat. Res.* **775**, 1–9 (2015).
21. Tan, W. Y. & Xiao, W. Y. A new stochastic and state space model of human colon cancer incorporating multiple pathways. *Biology Direct* **5**(1), 1–16 (2010).
22. Asatryan, A. D. & Komarova, N. L. Evolution of genetic instability in heterogeneous tumors. *J. Theor. Biol.* **396**, 1–12 (2016).
23. Zöllner, S., Sokolnikov, M. E. & Eidemüller, M. Beyond two-stage models for lung carcinogenesis in the Mayak Workers: implications for plutonium risk. *PLoS One* **10**(5), 1–20 (2015).
24. World Health Organization, www.who.int/cancer/en/ (2013).
25. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA. Cancer J. Clin.* **63**(1), 11–30 (2013).
26. Li, L., Tian, T. & Zhang, X. The impact of radiation on the development of lung cancer. *J. Theor. Biol.* **428**, 147–152 (2017).
27. Dela Cruz, C. S., Tanoue, L. T. & Matthay, R. A. Lung cancer: epidemiology, etiology, and prevention. *Clin. Chest Med.* **32**(4), 605–644 (2011).
28. Ong, T. M. *et al.* Detection of genomic instability in lung cancer tissues by random amplified polymorphic DNA analysis. *Carcinogenesis* **19**(1), 233–235 (1998).
29. Choi, C. M. *et al.* Chromosomal instability is a risk factor for poor prognosis of adenocarcinoma of the lung: Fluorescence *in situ* hybridization analysis of paraffin-embedded tissue from Korean patients. *Lung Cancer* **64**(1), 66–70 (2009).
30. Ali, A. H. K. *et al.* Aberrant DNA methylation of some tumor suppressor genes in lung cancers from workers with chromate exposure. *Mol. Carcinog.* **50**(2), 89–99 (2011).
31. Minina, V. I. *et al.* Chromosome aberrations in peripheral blood lymphocytes of lung cancer patients exposed to radon and air pollution. *Eur. J. Cancer Prev* (2016).
32. Thompson, S. L., Bakhoun, S. F. & Compton, D. A. Mechanisms of Chromosomal Instability. *Curr. Biol.* **20**(6), R285–R295 (2010).
33. Gomes, C. J., Centuori, S. & Martinez, J. D. Abstract 3509: Overexpression of 14-3-3g contributes to chromosomal instability in human lung cancer. *Cancer Res.* **74**(19 Supplement), 3509–3509 (2014).
34. Nakamura, H. *et al.* Chromosomal instability detected by fluorescence *in situ* hybridization in surgical specimens of non-small cell lung cancer is associated with poor survival. *Lung Cancer* **9**(6), 2294–2299 (2003).
35. Paillet, E. *et al.* High level of chromosomal instability in circulating tumor cells of ROS1-rearranged non-small-cell lung cancer. *Annals of Oncology* **26**(7), 1408–1415 (2015).
36. Thompson, D. E. *et al.* Cancer incidence in atomic bomb survivors. Part II: Solid tumors. 1958–1987. *Radiat. Res.* **137**(2 Suppl), S17–S67 (1994).
37. Grant, E. J. *et al.* Solid cancer incidence among the Life Span Study of atomic bomb survivors: 1958–2009. *Radiat. Res.* **187**(5), 513–537 (2017).
38. Howlader, N. *et al.* SEER cancer statistics review, 1975–2012, national cancer institute. Bethesda, MD, based on november 2014 SEER data submission, posted to the SEER web site. *Proc. Natl. Acad. Sci. USA* (2015).
39. Li, L., Tian, T. & Zhang, X. Stochastic modelling of multistage carcinogenesis and progression of human lung cancer. *J. Theor. Biol.* **479**, 81–89 (2019).
40. Ninomiya, H. *et al.* Genetic instability in lung cancer: concurrent analysis of chromosomal, miniand microsatellite instability and loss of heterozygosity. *Br. J. Cancer* **94**(10), 1485–1491 (2006).
41. Harris, T. E. *The theory of branching processes* (Englewood Cliffs, NJ: Prentice-Hall Applied Mathematics Series, 1963).
42. Portier, C. J. & Sherman, C. K. A. Calculating tumor incidence rates in stochastic models of carcinogenesis. *Math. Biosci.* **135**(2), 129–146 (1996).
43. Kimmel, M. & Axelrod, D. E. *Branching Processes in Biology* (Springer Nature, 2015).
44. McCullagh, P. & Nelder, J. A. *Generalized linear models*, 2nd Ed., (Chapman and Hall, London, 1989).
45. Seghouane, A. K. & Amari, S. I. *The AIC criterion and symmetrizing the Kullback–Leibler divergence* (IEEE Press, 2007).
46. Fakir, H. *et al.* Stochastic population dynamic effects for lung cancer progression. *Radiat. Res.* **172**(3), 383–393 (2009).
47. Hazelton, W. D., Clements, M. S. & Moolgavkar, S. H. Multistage carcinogenesis and lung cancer mortality in three cohorts. *Cancer Epidemiol. Biomarkers Prev.* **14**(5), 1171–1181 (2005).
48. Heidenreich, W. F., Luebeck, E. G. & Moolgavkar, S. H. Some properties of the hazard function of the two-mutation clonal expansion model. *Risk Anal.* **17**(3), 391–399 (1997).
49. Brouwer, A. F., Meza, R. & Eisenberg, M. C. A systematic approach to determining the identifiability of multistage carcinogenesis models. *Risk Anal* (2016).
50. Michor, F., Iwasa, Y. & Nowak, M. A. Dynamics of cancer progression. *Nat. Rev. Cancer* **4**(3), 197–205 (2004).
51. Preston, D. L. *et al.* Solid cancer incidence in atomic bomb survivors: 1958–1998. *Radiat. Res.* **168**(1), 1–64 (2007).
52. Ozasa, K. Epidemiological research on radiation-induced cancer in atomic bomb survivors. *Radiat. Res.* **57**(Suppl 1), i112–i117 (2016).
53. Castelletti, N. *et al.* Risk of lung adenocarcinoma from smoking and radiation arises in distinct molecular pathways. *Carcinogenesis*, <https://doi.org/10.1093/carcin/bgz036> (2019).
54. Luebeck, E. G. & Curtius, K. Impact of tumor progression on cancer incidence curves. *Cancer Res.* **73**(3), 1086–1096 (2013).
55. Zhang, X. *et al.* Mathematical modeling the pathway of human breast cancer. *Math. Biosci.* **253**, 25–29 (2014).
56. Li, L., Tian, T. & Zhang, X. Mutation Mechanisms of Human Breast Cancer. *J. Comput. Biol.* **25**(4), 1–9 (2018).

Acknowledgements

This report makes use of data obtained from the Radiation Effects Research Foundation (RERF) in Hiroshima, Japan. RERF is a private foundation funded equally by the Japanese Ministry of Health and Welfare and the U.S. Department of Energy through the U.S. National Academy of Sciences. The conclusions in this report are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies. We acknowledge financial support from Natural Science Foundation of Shaanxi Provincial Department of Education (No. 19JK0359), PHD Research Start-up Fund of Xi'an Polytechnic University (No. BS201874) and science and technology planning project of Henan province (No. 182102410021). We thank Dr. Richard Simon in Biometric Research Branch of National Cancer Institute, NIH, USA for initiating the project and helpful guidance. We thank the reviewers very much for their constructive suggestions on the manuscript.

Author Contributions

L.L. processed the data and carried out computational simulations. T.T., X.Z. and L.P. helped on data collections and analyses. L.L., T.T., X.Z. and L.P. analyzed the results. L.L., T.T. and X.Z. drafted the manuscript. All authors reviewed and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-50500-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019