Computational modeling of locoregional recurrence with spatial structure identifies tissue-specific carcinogenic profiles

Sharafudeen Dahiru Abubakar[1], Mitsuaki Takaki[2] and Hiroshi Haeno[1]*

[1]Research Institute for Biomedical Science, Tokyo University of Science, Noda, Japan, [2]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

Introduction: Local and regional recurrence after surgical intervention is a significant problem in cancer management. The multistage theory of carcinogenesis precisely places the presence of histologically normal but mutated premalignant lesions surrounding the tumor - field cancerization, as a significant cause of cancer recurrence. The relationship between tissue dynamics, cancer initiation and cancer recurrence in multistage carcinogenesis is not well known.

Methods: This study constructs a computational model for cancer initiation and recurrence by combining the Moran and branching processes in which cells requires 3 or more mutations to become malignant. In addition, a spatial structure-setting is included in the model to account for positional relativity in cell turnover towards malignant transformation. The model consists of a population of normal cells with no mutation; several populations of premalignant cells with varying number of mutations and a population of malignant cells. The model computes a stage of cancer detection and surgery to eliminate malignant cells but spares premalignant cells and then estimates the time for malignant cells to re-emerge.

Results: We report the cellular conditions that give rise to different patterns of cancer initiation and the conditions favoring a shorter cancer recurrence by analyzing premalignant cell types at the time of surgery. In addition, the model is fitted to disease-free clinical data of 8,957 patients in 27 different cancer types; From this fitting, we estimate the turnover rate per month, relative fitness of premalignant cells, growth rate and death rate of cancer cells in each cancer type.

Discussion: Our study provides insights into how to identify patients who are likely to have a shorter recurrence and where to target the therapeutic intervention.

Introduction

Cancers are dynamic cells whose features favors cellular proliferation, differentiation and movement while restricting cell death and tissue stability (1). Surgery is a potent curative tool for managing cancers, however, local recurrence has remained a clinically significant problem in most cancer types (2–4). Local recurrence rates could be as high as about 85% (5) for ovarian cancer, or 30% in non-small cell lung cancer (NSCLC) (6, 7) and as low as between 8% (8) to 16.5% (9) in breast cancer. Advanced surgical techniques, chemotherapy, radiotherapy (4) as well as endocrine therapy (10) are being used to minimize locoregional recurrence but "minimal" improvements and treatment-related mortality has highlighted the need for a better understanding and strategy for local recurrence (11, 12).

Since its introduction in 1953 (13), field cancerization has been recognized as a major cause of local recurrence (14). Field cancerization is the presence of "histologically normal" cells surrounding cancer cells that have acquired some but not all the genetic and phenotypic traits required for malignancy in a tissue (15, 16). These cancerized cells may have a survival or growth advantage and does serve as a hotbed for recurrent tumors as only a small number of additional steps are needed for cancer initiation. Recent advances in molecular, genomic and bulk sequencing techniques have supported the role of field cancerization (17). In breast cancer, microsatellite markers, epigenetic aberrations, transcriptomic deregulations and hTERT overexpression have been detected in histologically normal mammary tissues (18, 19). In head and neck cancer, loss of heterozygosity of chromosome 9p and telomere dysregulation were commonly observed in benign squamous hyperplasia (20, 21). In colon cancer patients with Crohn's ileocolitis, the same mutations of KRAS, CDKN2A, and TP53 were observed within neoplasia and non-tumor epithelium (22, 23). In Non-small cell lung cancer, miRNA dysfunction has been shown at the level of the tumor and cancerized field (24). Residual tumor (25), anesthesia choice (26) and CTCs (27) have been shown to have minimal impact on cancer recurrence; therefore, a proper understanding of the field cancerization formation process will contribute to the estimation of the risk of locoregional recurrence and the development of optimal treatment in each tissue.

Several Theoretical studies have shed light on field cancerization impacts on cancer initiation (15). Jeon et al. examined the multistage clonal expansion model by employing the Poisson process to consider the effects of premalignant cells on cancer initiation (28). The model was applied clinically to predict the long-term impact of ablative treatments on reducing esophageal adenocarcinoma incidence in Barrett's esophagus (29). Foo et al. developed a spatial evolutionary framework to determine the size distribution of histologically undetectable premalignant fields during diagnosis (30). This model was applied to the head and neck cancer and revealed that the patient's age was a critical predictor of the size and multiplicity of precancerous lesions (31).

These findings are in agreement with bulk sequencing data that shows the accumulation of cancer-related mutations as we age (32). A 2-step tumor initiation model provides insights into the relationship between different tissue kinetic parameters and the incidence of recurrent cancers (33) by using public datasets from the cancer genome atlas (TCGA), a valuable resource for genomic and clinical data analysis (34, 35) but fails to account the varying number of mutational hits required for carcinogenesis (36–38). The cancer genome atlas (TCGA) is a rich computational resource for the genomic and mutational data for different cancer types (34, 35) and will be helpful in validating our understanding of field cancerization.

This study developed a novel computational model of multi-stage cancer initiation and recurrence with spatial structure. We employed a combined stochastic model of Moran and a branching process to represent tissue and tumor dynamics, respectively, in order to observe cancer initiation and relapse after surgical resection of the first tumor *in silico*. Particularly, we focused on the relationship between the tissue compositions at the time of surgery and the time until the emergence of recurrent tumors. This model builds upon our previous work (33) by expanding the number of mutation steps for carcinogenesis *via* adding cell types as well as incorporating the spatial structure setting. Moreover, based on the public clinical datasets for locoregional recurrence rates, we succeeded in identifying tissue-specific carcinogenic parameters for various cancer types. Our approach provided insights on how to predict the time of recurrence from the tissue dynamics at the time of surgery and how to intervene patients to prevent the recurrence.

Materials and methods

Computational model

This model employs the multi-stage carcinogenesis concept. As tissues might require anywhere between 2 to 8 driver mutations (denoted by S) for malignant transformation (37), we first identify different cell types that can lead to a malignant transformation based on number of driver mutations. Let us visualize the dynamics of 5 types of cells in a tissue (Figure 1). "Type 0", "Type 1", "Type K", "Type S-1" and "Type S" represent normal healthy cells with no mutation; premalignant cells with one cancer-related mutation; premalignant cells with K cancer-related mutation, premalignant cells with S-1 cancer-related mutations and cancer cells with S cancer-related mutations, respectively. Emergence of cancer cells must be preceded by that of premalignant cells with mutations from Type 1 cell to Type S-1 cell. Type K cell may or may not be present depending on number of mutations required for carcinogenesis. We assume that a normal healthy tissue consists of Type 0, Type 1, Type K and Type S-1 cells undergoing cellular turnover with a small probability of a mutation. Moran process is employed to consider the tissue turnover dynamics, where the total number of Type 0,
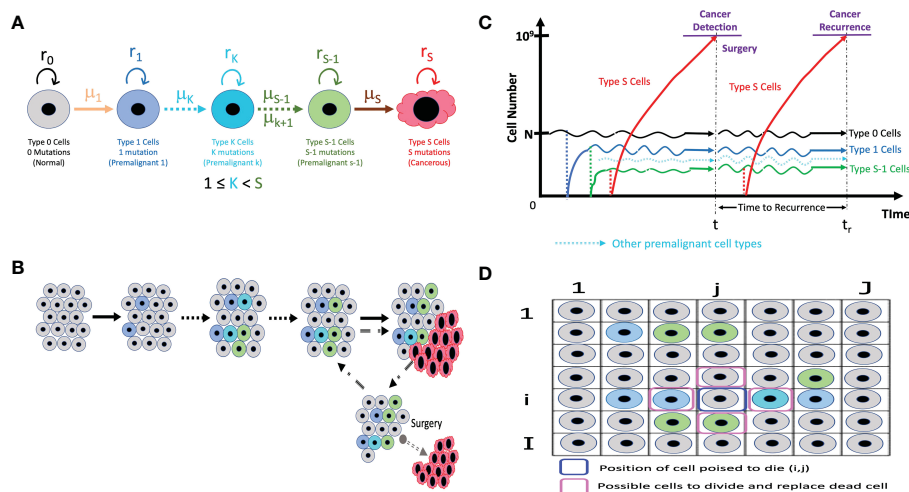
FIGURE 1

The illustrative representation of our models **(A)** The different cell types in our models with its own mutation rate (μ) and fitness (r). Type K cells may not be applicable if only 3 mutations or less are needed for carcinogenesis. **(B)** In a normal tissue composed of Type 0, Type 1, Type K, and Type S-1 cells, cell turnover is conducted according to the Moran process, and the number of cells is kept constant. If a Type S cell emerges, it proliferates without limit and can be detected and primed for surgery when the cancer cell number reaches 10^9. **(C)** At surgical intervention, all the Type S cells are resected while the number of Type 0, Type 1, Type K (if present) and Type S-1 cells remaining in a tissue are preserved. The time until the next Type S population reaches 10^9 is measured as time to recurrence. **(D)** The spatial structure integration in the model accounts for the positional relation between a cell poised to die and the possible cells that can divide to replace the dead cells.

Type 1, Type K and Type S-1 cells is kept constant as N (39). The turnover rate of a whole tissue is defined by d. Type S cells are considered as uncontrolled, highly proliferating cancer cells. The branching process is employed to consider the process of Type S proliferation (40).

Initially, N Type 0 cells occupy the tissue. There is a rare chance of a mutation every time a cell divides, and a daughter cell may change into a Type 1 cell with a mutation rate, μ_1. Mutation rate, μ, refers to the sum total of the genomic or epigenetic factors affecting change from one cell type to another (41). When a cell dies, a cell to be divided in a tissue is selected depending on the cell fitness, r. The fitness of Type 0, Type 1, Type K, and Type S-1 are denoted by r_0, r_1, r_K and r_{S-1}, respectively. Cell fitness, r refers to the transcriptional and metabolic potential of a cell type to "out-compete" other cell types (42). A cell could divide to give rise to the same cell type or mutate to another cell type. When a Type 1 cell divides with a mutation, a daughter cell may change into a Type K cell with mutation rate μ_K (if more mutations are needed) or a Type S-1 cell with a mutation rate, μ_{S-1} (if additional mutation steps are not needed). Intermediate cell type, Type K, becomes Type S-1 after sequential accumulation of mutations. Finally, a Type S-1 cell is capable of mutating to become a malignant cell – Type S cell based on the mutation rate from Type S-1 to Type S cell, μ_S. Once a Type S cell appears, the cells proliferate indefinitely based on the growth rate of Type S cells, r_S, disrupting tissue dynamics and homeostasis. Type S cancer cells are "super-competitors" with outstanding metabolic prowess and assumed to increase exponentially with a net growth rate of $r_S - d_S > 0$, where d_S is a death rate.

We propose that the most important premalignant cells are the Type 1 cell that have acquired the first driver mutation and the Type S-1 cell that needs just one more driver mutation to become a cancer cell. These cells look phenotypically normal and are not

regarded as important clinically but their genetic features are indispensable in cancer formation. As a result, the cell fitness of these 2 cell types must be taken into account in all computational analysis. For scenarios where additional mutations and more cell types are needed, we approximate intermediate mutational steps between Type 1 and Type S-1 by adjusting the values of μ_{S-1} so that a low mutation rate, μ_{S-1}, represents additional mutational steps. So, our computational analysis will be executed to account for the most important mutational events that affect cell fitness and all the mutation rates that can affect the number of steps required for carcinogenesis. In other words, we skip the state of Type K cell and a mutation rate stands in for the number steps.

The net growth of Type 0, Type 1, and Type S-1 cells is zero (equal frequency of cell division and death), while that of Type S cells is positive. Type 0, Type 1, and Type S-1 cells consist of a healthy tissue based on the Moran process, so r_0, r_1 and r_{S-1} are parameters to determine status of dividing cell and which daughter cell are obtainable at the time of a cell division. Alternatively, r_S is the growth rate, which determines the average number of increases in Type S cells during a unit time. When the number of Type S cells reaches 10^9 at the first time, all the Type S cells are discarded to represent surgical resection, whereas the number of Type 0, Type 1 and Type S-1 cells in a tissue is preserved so that the time until the emergence of the recurrent tumor is influenced by the frequency of residual Type 0, Type 1 and especially, Type S-1 cells. Since the conversion from the number of cells to the tumor volume is frequently done using the following relationship as 10^9 cells in a 1 cm^3 tumor, the time of surgery in this model is conducted when the size of the tumor becomes 1 cm^3. We describe it as "time of cancer detection". After the first treatment, the simulation continues until the next Type S cell appears from the tissue and the number reaches 10^9 again, representing the recurrence of the tumor after surgery.

Simulation framework

To integrate the Moran process and branching process, we adopted stochastic simulations based on Gillespie's algorithm (43) as follows: We firstly considered three events: (i) cell turnover in a healthy tissue as per Moran process, (ii) birth of a Type S cell as per Branching process, and (iii) death of a Type S cell. The rates of each event at time t is given by (i) dN (ii) $r_S X_S(t)$, and (iii) $d_S X_S(t)$, respectively. Here r_S, d_S, and $X_S(t)$ are a proliferation rate, a death rate, and the number of Type S cells at time t, respectively. Then an average time until one of the three events happens, ΔT, is given by

$$\Delta T = \frac{1}{dN + r_S X_S(t) + d_S X_S(t)}. \qquad (1)$$

Let us first consider the case where a cell turnover happens. The probability that a cell turnover happens in ΔT is given by $dN \cdot \Delta T$. In our model, a cell turnover in a healthy tissue is governed by a cell death. When one of N cells is randomly selected as a cell to die, and another cell is chosen to divide within the same time step to complete cell turnover. In a healthy tissue, there are three types of cells, corresponding to the number of acquired mutations, Type 0, Type 1, and Type S-1.The number of each cell type is denoted by X_0, X_1 and X_{S-1}, respectively. In brief, there are several possibilities of tissue composition transitions in the tissue dynamics and we consider the six events that affect the cell type composition of a tissue: (i) a type 0 cell increases by one while a type 1 cell decreases by one (ii) a type 0 cell increases by one while a type S-1 cell decreases by one (iii) a type 1 cell increases by one while a type 0 cell decrease by one (iv) a type 1 cell increases by one while a type S-1 cell decreases by one (v) a type S-1 cell increases by one while a type 0 cell decreases by one; or (vi) a type S-1 cell increases by one while a type 1 cell decreases by one. In such a condition, a Type 0 cell can increase by one if either a Type 1 or Type S-1 cell dies and a Type 0 cell divides without a mutation. Then the probability for these events leading to an increase in Type 0 cells are given by (i) $\frac{X_1}{N} \cdot \frac{r_0 X_0(1-\mu_1)}{F}$, and (ii) $\frac{X_{S-1}}{N} \cdot \frac{r_0 X_0(1-\mu_1)}{F}$. Here, $F = r_0 X_0 + r_1 X_1 + r_{S-1} X_{S-1}$ is a scaling factor for selecting a dividing cell. The probability of a Type 1 or Type S-1 cell death is given by $\frac{X_1}{N}$ and $\frac{X_{S-1}}{N}$, respectively. Taken together, the transition probability that the number of Type 0 cell increases by one and that of Type 1 decreases by one is given by

$$\Pr[X_0 \rightarrow X_0 + 1 \text{ and } X_1 \rightarrow X_1 - 1]$$
$$= \frac{X_1}{N} \cdot \frac{r_0 X_0(1-\mu_1)}{F}, \qquad (2)$$

and the probability that the number of Type 0 cell increases by one and that of Type S-1 decreases by one is given by

$$\Pr[X_0 \rightarrow X_0 + 1 \text{ and } X_{S-1} \rightarrow X_{S-1} - 1]$$
$$= \frac{X_{S-1}}{N} \cdot \frac{r_0 X_0(1-\mu_1)}{F}. \qquad (3)$$

A Type 1 cell can increase by one if either a Type 0 or Type S-1 cell dies, and either a Type 1 cell divides without mutation or a Type 0 cell divides with mutation to become a Type 1 cell. Then the probabilities for these events leading to an increase in Type 1 cells are given by (iii) $\frac{X_0}{N} \cdot \frac{r_1 X_1(1-\mu_{S-1})+r_0 X_0\mu_1}{F}$, and (iv) $\frac{X_{S-1}}{N} \cdot \frac{r_1 X_1(1-\mu_{S-1})+r_0 X_0\mu_1}{F}$.

Taken together, the transition probability that the number of Type 1 cell increases by one and that of Type 0 decreases by one is given by

$$\Pr[X_1 \rightarrow X_1 + 1 \text{ and } X_0 \rightarrow X_0 - 1]$$
$$= \frac{X_0}{N} \cdot \frac{r_1 X_1(1-\mu_{S-1})+r_0 X_0\mu_1}{F}, \qquad (4)$$

and the probability that the number of Type 1 cell increases by one and that of Type S-1 decreases by one is given by

$$\Pr[X_1 \rightarrow X_1 + 1 \text{ and } X_{S-1} \rightarrow X_{S-1} - 1]$$
$$= \frac{X_{S-1}}{N} \cdot \frac{r_1 X_1(1-\mu_{S-1})+r_0 X_0\mu_1}{F}. \qquad (5)$$

Similarly, a Type S-1 cell can increase by one if either a Type 0 or Type 1 cell dies, and either a Type S-1 cell divides without mutation or a Type 1 cell divides with mutation. The probabilities for the events leading to an increase in Type S-1 cells are given by (v) $\frac{X_0}{N} \cdot \frac{r_{S-1} X_{S-1}(1-\mu_S)+r_1 X_1\mu_{S-1}}{F}$, and (vi) $\frac{X_1}{N} \cdot \frac{r_{S-1} X_{S-1}(1-\mu_S)+r_1 X_1\mu_{S-1}}{F}$. Taken together, the transition probability that the number of Type S-1 cell increases by one and that of Type 0 decreases by one is given by

$$\Pr[X_{S-1} \rightarrow X_{S-1} + 1 \text{ and } X_0 \rightarrow X_0 - 1]$$
$$= \frac{X_0}{N} \cdot \frac{r_{S-1} X_{S-1}(1-\mu_S)+r_1 X_1\mu_{S-1}}{F}, \qquad (6)$$

and the probability that the number of Type S-1 cell increases by one and that of Type 1 decreases by one is given by

$$\Pr[X_{S-1} \rightarrow X_{S-1} + 1 \text{ and } X_1 \rightarrow X_1 - 1]$$
$$= \frac{X_1}{N} \cdot \frac{r_{S-1} X_{S-1}(1-\mu_S)+r_1 X_1\mu_{S-1}}{F}. \qquad (7)$$

In addition, a Type S cell can increase by one if a Type S-1 cell divides with mutation. The probability is given by $\frac{r_{S-1} X_{S-1}\mu_S}{F}$. Since a Type S cell is not a component of a tissue, once a Type S appears by mutation, another round of selection for a dividing cell is performed according to the transition probabilities described above. This is because malignant Type S cell disrupts 2D lattice structure and the Moran process is no longer applicable to it.

Next, let us consider the case where Type S cell divides or dies. The probabilities of Type S cell division or death is given by $r_S X_S(t) \cdot \Delta T$ and $d_S X_S(t) \cdot \Delta T$, respectively.

In summary, the time of one step in our simulation is calculated using Eq. (1) and in one time step, one of the following three processes occurs: (i) a cell turnover in a tissue, (ii) the birth of a Type S cell, or (iii) the death of a Type S cell. Initially, all the cells are Type 0. Once the number of Type S cells reaches 10^9, computational surgical resection sets the number of Type S cells to be 0, keeping the cell type composition in a tissue remained and computational carcinogenic process restarts again. After that, the time until the number of Type S cells reaches 10^9 is measured as recurrence time.

Spatial structure

Two-dimensional lattice structure ($I \times J$) is introduced to a tissue dynamics in our computational model. The transition probabilities

are basically the same with or without spatial structure. The difference is the choice of a dividing cell. If a cell at position $\langle i,j \rangle$ dies, 4 adjacent cells – $\langle i,j-1 \rangle$, $\langle i,j+1 \rangle$, $\langle i-1,j \rangle$ and $\langle i+1,j \rangle$ can divide to replace it. The transition probabilities are calculated according to the cell type at those positions. We assume wall boundary condition to represent an asymmetric tissue structure.

Deterministic approximation of type S cell growth

As for the calculation of the Type S growth, we assume that when the number of cells is small, the stochastic effect should be considered. When the number of Type S cells exceed twice as large as the size of the normal tissue, $2N$, growth can be regarded as a deterministic process. Then, the time duration from when the number of Type S cells is $2N$ to 10^9, Δt_s, is given by

$$\Delta t_s = \frac{1}{(r_S - d_S)}\ln\left(\frac{10^9}{2N}\right) \tag{8}$$

Clinical data

The data used in our analysis were from TCGA Pan-Cancer Clinical Data Resource (34, 35) and are available in the cBio Cancer Genomics Portal (44, 45). We adopt the clinical data of locoregional recurrence from 8,957 patients with 27 different non-sarcoma, non-hematological cancer types. From these datasets, the inclusion criterium for our study was "disease-free" survival – patients with no detectable malignant disease after surgery or total remission. We excluded data of "progression-free" survival in order to eliminate patients who survived with detectable disease possibly as a result of treatment-resistant clones; and also excluded data containing metastatic progression. We also included data from other independent publications for extra validation. Sarcomas and hematological cancers were excluded due to their non-conformity to a 2-dimensional lattice structure.

Survival time analysis

Survival time analysis of clinical data is calculated using the Kaplan–Meier method from disease-free intervals mentioned in Clinical Data section. In this study, disease-free interval is defined as the survival time without cancer recurrence of each patient, which corresponds to the time to recurrence of each simulation trial.

Statistical analysis

The whole process of our model is conducted on C++. Simulation codes have been deposited in a GitHub repository (https://github.com/sharaf501/Heano-Lab-Codes). The survival time analysis and other statistical analysis is conducted on Prism (version 9.4.1). Mantel-Cox (log-rank) test is used to compare difference between survival curves. A p value less than 0.05 is considered to be statistically significant.

Results

Cancer initiation patterns

Firstly, we conducted stochastic triplicate simulations for the cancer initiation up to the time of cancer detection. We were curious to know what effect the presence or absence of the spatial structure would have on the model. We traced the time course of 4 cell populations – Type 0, Type 1, Type S-1 and Type S cells using a combination of various parameter sets. Lower mutation rate from Type 1 to Type S-1, $\mu_{S\text{-}1}$, was additionally examined to account for additional premalignant cell types between Type 1 and Type S-1. In the model without spatial structure, we observed 3 patterns of cancer initiation based on frequency of non-malignant cell population at cancer detection (Figure 2A). Interestingly, all the patterns show a progressive decline of Type 0 cells until the entire tissue is dominated by Type 1, Type S-1 or both Type 1 and Type S-1 cells. By combining various parameter sets in our simulation, we extrapolated the varying distribution of the cancer initiation patterns (Figure 2B and Supplementary Figure 1). Lower fitness of Type 1 cells, r_1 generally favored Type S-1 cells dominance when fitness of Type S-1, $r_{S\text{-}1}$, is high. Higher r_1 values favored Type 1 dominance while equal fitness of Type 1 and Type S-1 cells yielded Type 1 dominance or Type 1/S-1 co-dominance. Mutation rates generally affected time to cancer detection and appearance of dominance. We also extended the mutation rates from Type 1 to Type S-1 cell type to denote other additional mutation steps and found a consistent increase in cancer detection times but patterns generally remained the same. In some cases where low fitness of both Type 1 and Type S-1 were coupled with lower mutation rates, Type S malignant cells failed to appear at extended times and simulations were terminated. Most curiously, mutation rate from Type 0 to Type 1, μ_1, did not affect the pattern of cancer initiation or time to cancer detection (Supplementary Figure 1).

Parameter dependence of recurrence time

Next, we examined the time to recurrence after surgical resection and the proportion of Type S-1 cells at the time of surgery in varying parameter sets. We reasoned that since Type S-1 cells needs only one more step for malignant transformation; therefore, its proportion was thought to be critical for cancer recurrence. To do this, we ran 1,000 simulations for each parameter set and calculated the mean recurrence time (Figures 3A–K). We also ran similar simulations at higher cell number in a tissue, N, between 100 to 1,000 times to assess the effect of tissue size on the parameter dependency (Figures 3B–L). We found that higher fitness of Type 1 cells, r_1 increased the mean
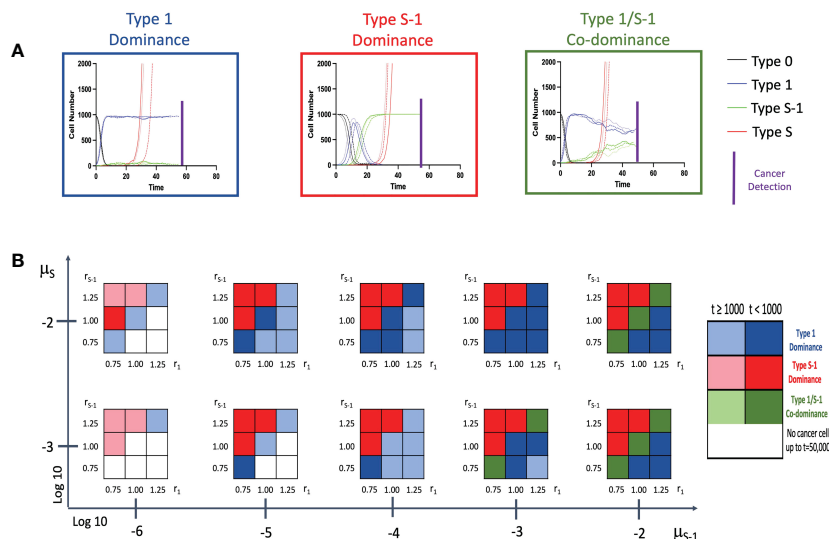
FIGURE 2
Patterns of tissue composition at cancer initiation **(A)** Simulation studies without spatial structure show three patterns of cancer initiation. For each pattern, black, blue, green and red curves indicate Type 0, Type 1, Type S-1 and Type S cells, respectively. Each parameter set was simulated in triplicate (Joined, dashed, and long-dashed lines). **(B)** Panel showing several patterns of tissue composition and time to detection using combination of various parameter sets. Cell type "Dominance" indicates >90% of a particular cell type at cancer detection. "Co-dominance" refers to 2 cell populations with >40% or 3 cell populations with >30% at cancer detection. Parameter values used are: $N = 1,000$; $d = d_s = 1.0$; $r_0 = 1.0$; $r_1 = 0.75$, 1.00 and 1.25; $r_{S-1} = 0.75$, 1.00 and 1.25; $r_S = 1.5$; $\mu_1 = 0.001$ and 0.01; $\mu_{S-1} = 0.000001$, 0.00001, 0.0001, 0.001 and 0.01; $\mu_S = 0.001$ and 0.01. Note: Mutation rate, μ_1, does not impact the results and is not shown.

recurrence time (Figures 3A, B), while mutation rate from Type 0 to Type 1, μ_1, had no effect on mean recurrence time (Figures 3G, H). Other parameters however, showed a negative correlation to the mean recurrence time – higher parameter values resulted in shorter mean recurrence time. Higher tissue cell number yielded an overall shortening of mean recurrence time but parameter dependency remained the same. We also observed a reduction in the proportion of Type S-1 cells at the time of surgery when r_1, r_S and μ_S increases (Figures 3A–L), while r_{S-1} and μ_S increases in the proportion of Type S-1 cells (Figures 3C–J). Mutation rate from Type 0 to Type 1, μ_1, had little effect on the proportion of Type S-1 cells (Figures 3G, H).

Effect of spatial structure on cancer initiation patterns

We then incorporated the spatial structure framework into the model to investigate the effect of tissue positional influence in cancer initiation patterns and time of cancer detection. After triplicate simulations using various parameters, we identified seven distinct patterns of cancer initiation (Figure 4A) based on the composition of non-malignant cell population at cancer detection. Figure 4B and Supplementary Figure 2 showed the distribution of the patterns in a wide parameter range. Low r_1 values showed Type 0 dominance at low r_{S-1} levels with failure to detect cancer cells at very low μ_{S-1} levels. With higher r_1 values, Type 1 cell types begin to dominate. When combined with high μ_{S-1}, we saw Type 1/S-1 co-dominance (green region in Figure 2B). When r_1 and r_{S-1} are equal to fitness of Type 0 (r_0), we saw Type 0/1

co-dominance or Type 0/1/S-1 co-dominance depending on μ_{S-1} values. We noticed a peculiar pattern of Type 0/S-1 co-dominance (black region in Figure 2B) when r_{S-1}, μ_{S-1}, and μ_S were high with relatively lower r_1 value. Type S-1 dominance (red region in Figure 2B) was regarded as the most undesirable scenario due to the abundance of Type S-1 cells, indicating shorter recurrence time. We saw this pattern when r_{S-1}, μ_1 and μ_{S-1} were high, r_1 was equal to 1.0 and μ_S was relatively small. Some parameter sets with low fitness failed to yield Type S cells at extended time points during the simulations. Here, the incorporation of the spatial structure to our simulation framework had remarkable alterations to the cancer initiation patterns and cancer detection time.

Effect of spatial structure on recurrence time

Subsequently, we examined the mean recurrence time after surgical resection and the proportion of Type S-1 lesions at the time of surgery in a vast parameter range with the influence of the spatial structure setting (Figure 5). Similarly, we ran 100 to 500 simulations to obtain mean recurrence time (Figures 5A–K) and to check the effect of larger cell numbers (Figures 5B–L). Generally, we saw that the integration of the spatial structure to our simulation framework had noteworthy changes to the parameter dependency to recurrence time. When the size of the normal tissue was small, the effect of fitness advantage on the proportion of Type S-1 cells in a tissue became larger. Our simulation results showed that an increase in the cell fitness shortened the mean time to recurrence (Figures 5C–F). However, an increase in r_1 was found to reduce the recurrence
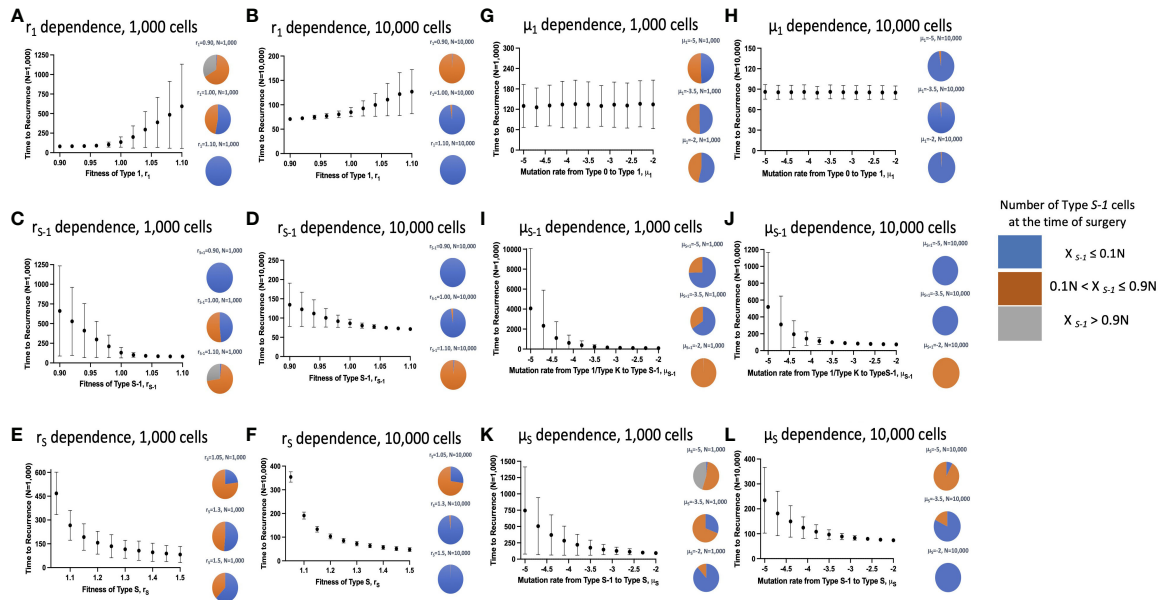
FIGURE 3

Parameter dependence on recurrence time. Simulation studies without spatial structure are shown. Mean values obtained from 100 to 1,000 simulations are shown by dots, and standard deviations are indicated by bars. Pie charts in the panels indicate the proportion of Type S-1 cells in a normal tissue at the time of first treatment. Blue, orange and grey represent small ($X_{S-1} \leq 0.1N$), intermediate ($0.1N < X_{S-1} \leq 0.9N$), and large ($X_{S-1} > 0.9N$) proportion of Type S-1 cells, respectively. Standard parameter values used in (A–L) are $d = d_s = 1.0$, $r_0 = 1.0$, $r_1 = 1.0$, $r_{S-1} = 1.2$, $r_S = 1.0$, $\mu_1 = 0.001$, $\mu_{S-1} = 0.001$, $\mu_S = 0.001$; and $N = 1,000$ in (A, C, E, G, I, K); and $N = 10,000$ in (B, D, F, H, J, L).

time but begin to increase slightly at much higher levels regardless of the tissue size. We also found a consistent reduction in the mean recurrence time as mutation rates μ_1, μ_{S-1} and μ_S increased (Figures 5G–L). We also observed a reduction in the proportion of Type S-1 cells at the time of surgery with a spatial structure. Especially, when either r_S or μ_S was small, and any of r_{S-1}, μ_1, or μ_{S-1} was large, the proportion of Type S-1 increased (Figure 5).

Fitting of recurrence time to clinical data

By using our computational model with spatial structure, multiple runs of stochastic simulations were performed with multiple parameter sets and *in silico* Kaplan–Meier curves were made. The data points about the time when 0% to 100% of patients experienced recurrence with an interval of 4% (time when 0%, 4%,
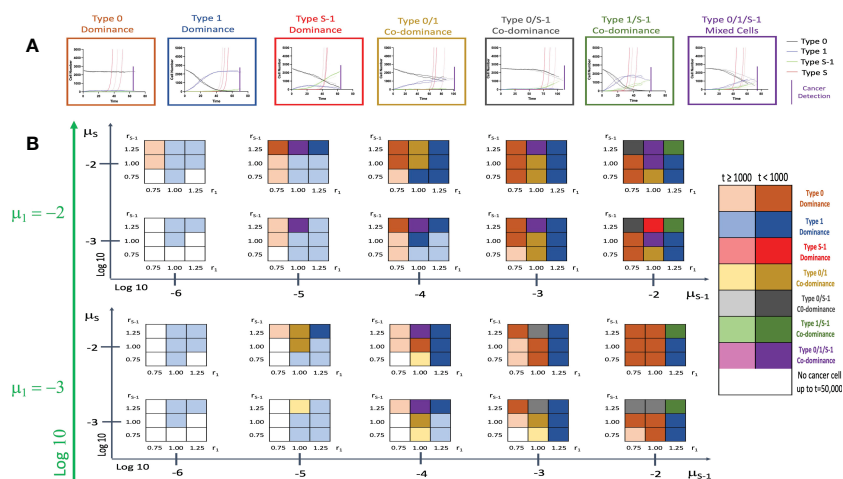
FIGURE 4

Patterns of tissue composition at cancer initiation with spatial structure. (A) Simulation studies with spatial structure show 7 patterns of cancer initiation. For each pattern, black, blue, green and red curves indicate Type 0, Type 1, Type S-1 and Type S cells, respectively. Each parameter set was simulated in triplicate (Joined, dashed, and long-dashed lines). (B) Panel showing patterns of tissue composition and time to detection using combination of various parameter sets. The definitions of "Dominance" and "Co-dominance" are the same as those explained in Figure 2. Parameter values used are: $N = 2,500$; $d = d_s = 1.0$; $r_0 = 1.0$; $r_1 = 0.75$, 1.00 and 1.25; $r_{S-1} = 0.75$, 1.00 and 1.25; $r_S = 1.5$; $\mu_1 = 0.001$ and 0.01; $\mu_{S-1} = 0.000001$, 0.00001, 0.0001, 0.001 and 0.01; $\mu_S = 0.001$ and 0.01.
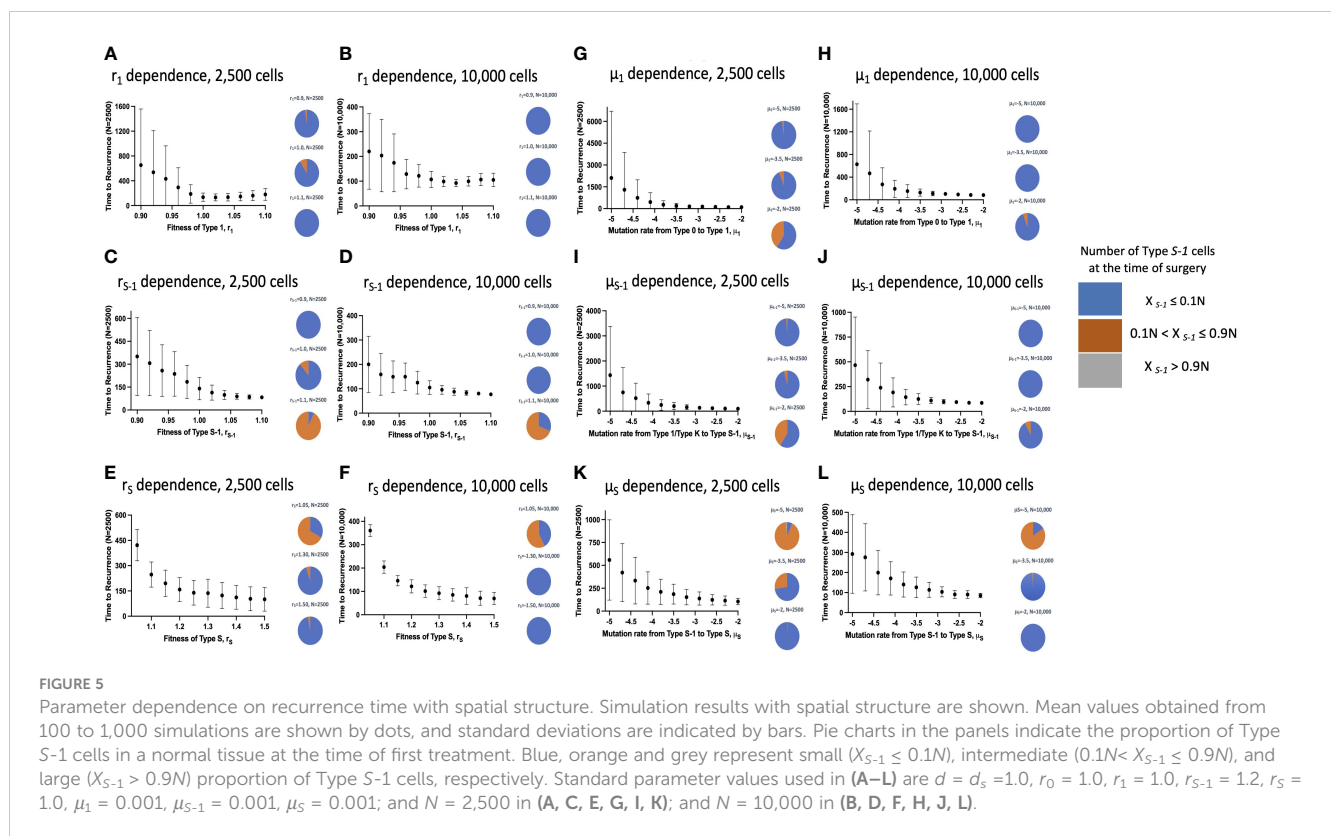
FIGURE 5
Parameter dependence on recurrence time with spatial structure. Simulation results with spatial structure are shown. Mean values obtained from 100 to 1,000 simulations are shown by dots, and standard deviations are indicated by bars. Pie charts in the panels indicate the proportion of Type S-1 cells in a normal tissue at the time of first treatment. Blue, orange and grey represent small ($X_{S-1} \leq 0.1N$), intermediate ($0.1N < X_{S-1} \leq 0.9N$), and large ($X_{S-1} > 0.9N$) proportion of Type S-1 cells, respectively. Standard parameter values used in **(A–L)** are $d = d_S = 1.0$, $r_0 = 1.0$, $r_1 = 1.0$, $r_{S-1} = 1.2$, $r_S = 1.0$, $\mu_1 = 0.001$, $\mu_{S-1} = 0.001$, $\mu_S = 0.001$; and $N = 2{,}500$ in **(A, C, E, G, I, K)**; and $N = 10{,}000$ in **(B, D, F, H, J, L)**.

8%, …, 100% of patients experienced recurrence) were employed to compare between the *in silico* and published clinical data (44, 45) of 27 cancer types. In this analysis, we adopted random sampling for parameters to obtain *in silico* recurrence data and determined the best parameter set for each cancer type that minimized the mean of squared logarithmic residuals (log-MSR) between outputs *in silico* and in public. The accepted parameter set (Table 1) was used to extrapolate recurrence time which were then fitted to clinical data and disease-free survival curves were depicted (Figure 6).

Mantel-Cox test was used to compare between the curves of simulated and clinical data revealing minimal statistical nonconformity. According to the estimated parameters (Table 1), we firstly deduced a tissue-specific turnover per month from d_S. Kidney chromophobe had the fewest cellular turnover cycles per month while bladder urothelial carcinoma and colorectal adenocarcinoma had the highest turnover cycles. Moreover, colorectal adenocarcinoma, kidney chromophobe, renal clear cell carcinoma, thyroid carcinoma, adenoid cystic carcinoma and acral melanoma showed higher fitness of all their premalignant cells than normal cells. Of note, the proliferation rate of the Type S malignant cells, r_S, was estimated to be high in cholangiocarcinoma, liver hepatocellular carcinoma, mesothelioma and upper tract urothelial cancer while being relatively low in breast invasive carcinoma, kidney chromophobe and skin cutaneous melanoma. Kidney chromophobe had the lowest mutation rate from the final premalignant cell stage to malignant cells while cervical squamous cell carcinoma and prostate adenocarcinoma had the highest mutation rate. Figure 7 showed the negative correlation between mutational steps required for carcinogenesis (37) and overall

mutation rates (μ_I) obtained from our studies by multiplying the mutation rates for all steps.

Discussion

In this study, we constructed computational models with and without spatial structure that described cell population dynamics in both normal and cancer tissues. Using our models, we clearly observed different patterns of cancer initiation and the residual premalignant cells present at the time of cancer detection or surgical intervention. Integrating the spatial structure setting to the model revealed additional patterns of cancer initiation as against just three in the model without spatial structure. Especially, the preservation of intact normal cells was observed in the model with spatial structure (Figures 2–4). According to the comprehensive analysis of parameter dependence, we found that field cancerization at the detection time depended on a combination of fitness and mutation rates.

We also revealed the relationship between the proportion of premalignant cells and recurrence time (Figures 3–5). The model without spatial structure overemphasized the power of Type 1 fitness and its ability to limit Type S-1 and Type S appearance which led to longer mean recurrence times as r_1 increases (Figures 3A, B). The same effect was seen in the mutation rate from Type 0 to Type 1 which rendered μ_1 impotent in affecting mean recurrence time (Figures 3G, H). All other fitness and mutation parameters led to shorter recurrence time as their

TABLE 1 Tumor-specific carcinogenic profiles and p values of survival curves (μ values are in log$_{10}$ while SQ are log-SSR values).

Code	Cancer Type (Data Source)	SQ	r_1	r_{S-1}	r_S	d_S	μ_1	μ_{S-1}	μ_S	μ_I	p value
ACC	Adrenocortical Carcinoma (TCGA, PanCancer Atlas)	0.616	1.003	0.958	5.578	3.436	-3.212	-3.979	-3.006	-10.197	0.9120
BLCA	Bladder Urothelial Carcinoma (TCGA, PanCancer Atlas)	0.922	1.097	0.981	5.614	3.939	-2.876	-3.616	-3.223	-9.715	0.8857
BRCA	Breast Invasive Carcinoma (TCGA, PanCancer Atlas)	0.983	0.945	0.948	2.967	1.887	-3.770	-2.228	-2.661	-8.658	0.1707
CESC	Cervical Squamous Cell Carcinoma (TCGA, PanCancer Atlas)	0.897	0.929	0.951	6.684	3.506	-3.842	-3.233	-2.098	-9.172	0.1213
CHOL	Cholangiocarcinoma (TCGA, PanCancer Atlas)	0.847	0.988	1.011	7.170	3.765	-3.504	-2.759	-2.349	-8.612	0.7716
COAD	Colorectal Adenocarcinoma (TCGA, PanCancer Atlas)	0.943	1.086	1.027	5.912	3.859	-3.176	-3.052	-4.347	-10.575	0.2783
ESCA	Esophageal Adenocarcinoma (TCGA, PanCancer Atlas)	0.936	1.048	0.919	6.767	3.655	-4.114	-3.870	-2.411	-10.394	0.8862
HNSC	Head & Neck Squamous Cell Carcinoma (TCGA, PanCancer Atlas)	0.802	0.980	0.988	5.148	3.586	-2.735	-4.886	-2.656	-10.277	0.5908
KICH	Kidney Chromophobe (TCGA, PanCancer Atlas)	0.401	1.070	1.024	2.308	1.558	-3.971	-2.542	-4.658	-11.170	0.9360
KIRC	Kidney Renal Clear Cell Carcinoma (TCGA, PanCancer Atlas)	0.786	1.081	1.064	3.655	2.713	-4.569	-3.445	-4.252	-12.265	0.1519
KIRP	Kidney Renal Papillary Cell Carcinoma (TCGA, PanCancer Atlas)	0.559	0.965	0.958	5.470	3.004	-3.182	-2.748	-3.374	-9.304	0.9880
LIHC	Liver Hepatocellular Carcinoma (TCGA, PanCancer Atlas)	0.846	0.977	1.009	7.675	3.899	-4.620	-3.690	-4.252	-12.562	0.6744
LUAD	Lung Adenocarcinoma (TCGA, PanCancer Atlas)	0.600	0.922	0.977	6.383	3.327	-3.068	-3.471	-2.821	-9.360	0.4484
LUSC	Lung Squamous Cell Carcinoma (TCGA, PanCancer Atlas)	0.524	0.997	0.952	5.220	3.568	-4.174	-2.276	-3.258	-9.708	0.9767
MESO	Mesothelioma (TCGA, PanCancer Atlas)	0.572	0.918	1.030	7.134	3.581	-3.287	-4.222	-3.561	-11.070	0.8411
OV	Ovarian Serous Cystadenocarcinoma (TCGA, PanCancer Atlas)	0.772	1.052	0.939	4.145	2.655	-3.483	-2.415	-3.502	-9.400	0.9710
PAAD	Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)	0.736	0.988	0.996	3.145	1.654	-3.218	-2.875	-2.416	-8.509	0.8931
PRAD	Prostate Adenocarcinoma (TCGA, PanCancer Atlas)	0.886	0.915	0.918	4.826	3.584	-2.286	-3.987	-2.189	-8.462	0.7336
STAD	Stomach Adenocarcinoma (TCGA, PanCancer Atlas)	0.400	0.936	0.949	5.738	3.177	-3.377	-2.268	-3.389	-9.034	0.9442
SKCM	Skin Cutaneous Melanoma (TCGA, Firehose Legacy)	0.805	1.016	0.980	2.848	1.818	-4.481	-3.607	-2.512	-10.601	0.5770
THCA	Thyroid Carcinoma (TCGA, PanCancer Atlas)	0.764	1.020	0.980	6.541	3.525	-2.824	-4.268	-4.131	-11.222	0.7340
UCEC	Uterine Corpus Endometrial Carcinoma (TCGA, PanCancer Atlas)	0.892	0.948	0.984	3.965	2.083	-4.276	-3.824	-2.908	-11.007	0.1153
UVM	Uveal Melanoma (TCGA, Firehose Legacy)	0.564	1.022	0.947	6.612	3.535	-3.397	-2.166	-4.000	-9.563	0.2663
ACYC	Adenoid Cystic Carcinoma (MSK, Nat Genet 2013)	0.521	1.082	1.065	4.523	2.492	-2.754	-3.342	-4.046	-10.142	0.6316
MEL	Acral Melanoma (TGEN, Genome Res 2017)	0.524	1.031	1.008	5.199	3.104	-2.572	-3.866	-2.487	-8.925	0.3385
OSCC	Oral Squamous Cell Carcinoma (MD Anderson, Canc. Disc 2013)	1.033	0.901	0.990	8.090	3.700	-2.721	-2.602	-3.640	-8.964	0.9997
UTUC	Upper Tract Urothelial Cancer (MSK, Eur Urol 2015)	0.628	0.938	1.009	7.470	3.800	-2.130	-3.672	-2.815	-8.617	0.5480

effects became larger. Generally, with the spatial structure setting, we found that recurrence time became shorter when mutation rates or fitness of cancer cells were large, while the time became longer when the fitness of premalignant cells or growth rate of cancer cells were low (Figure 5). An exception would be the mean recurrence time with r_1 (Figures 5A, B) which was seen to shorten as r_1 became larger but to get longer as r_1 became much larger. This could be due to a tissue competition between Type 1 and Type S-1 cells which subsequently delayed the emergence of Type S cells and hence, a more favorable recurrence time.

Moreover, we successfully estimated the characteristic parameter sets of the computational model that best reproduced the clinical data of disease-free survival in each cancer type. All the non-sarcoma cancer types were successfully fitted with no statistical deviation (Table 1). Even though some datasets like ESCA contains 2 different cancer types – esophageal adenocarcinoma and esophageal squamous cell carcinoma, we obtained p values that indicates no statistical difference. At the same time, we obtained valuable information about cellular turnover per month (d_S), relative fitness of premalignant cells (r_1, r_{S-1}), a growth rate of

FIGURE 6
Fitting of model-derived *in silico* data to published clinical data for 27 cancer types. Thousands of stochastic runs were used to obtain parameter sets that best fit survival curves of 27 non-sarcoma, non-hematologic cancer types. Blue curves indicate clinical data while red curves indicate simulation data survival curves. *p* values between curves are found in Table 1. ACC, Adrenocortical Carcinoma; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast Invasive Carcinoma; CESC - Cervical Squamous Cell Carcinoma; CHOL – Cholangiocarcinoma; COAD, Colorectal Adenocarcinoma; ESCA, Esophageal Adenocarcinoma; HNSC, Head & Neck Squamous Cell Carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney Renal Clear Cell Carcinoma; KIRP, Kidney Renal Papillary Cell Carcinoma; LIHC, Liver Hepatocellular Carcinoma; LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinoma; MESO – Mesothelioma; OV, Ovarian Serous Cystadenocarcinoma; PAAD, Pancreatic Adenocarcinoma; PRAD, Prostate Adenocarcinoma; STAD, Stomach Adenocarcinoma; THCA, Thyroid Carcinoma; UCEC, Uterine Corpus Endometrial Carcinoma; UVM, Uveal Melanoma; ACYC, Adenoid Cystic Carcinoma; MEL, Acral Melanoma; UTUC, Upper Tract Urothelial Cancer; OSCC, Oral Squamous Cell Carcinoma; SKCM, Skin Cutaneous Melanoma.

cancer cells (r_S) and mutation rates from one cell type to another (μ_1, μ_{S-1}, μ_S) for each carcinogenesis. We have specified the growth rate for each cancer using the r_S values from our clinical fitting. Interestingly, we observed relatively high growth rates of malignant cells (Type *S* cells) in some common cancer types like lung and colorectal cancers, whereas a relatively lower growth rate was estimated in breast invasive carcinoma which was also a common cancer type but was relatively asymptomatic in agreement with several studies (46, 47). From the high μ_{S-1} values, we elucidated that uveal melanoma, breast invasive carcinoma, stomach adenocarcinoma and lung squamous cell carcinoma had the shortest time to reach late premalignant cell stage from the earliest premalignant cell stage possibly indicating fewer mutational steps. On the other hand, thyroid carcinoma and head and neck carcinoma had small μ_{S-1} values, indicating the multiple steps in the carcinogenesis. Our data was in alignment with data that estimated the number of hits required for carcinogenesis (37), where liver, kidney and thyroid cancers had the lowest overall mutational rates indicating more mutational requirements while uterine, ovarian and lung cancers had higher overall mutational rates indicating fewer mutational requirements (Figure 7).

Additionally, our findings successfully revealed average cellular turnover rates per month by inferring our model with published clinical data whose measurements were in months. Kidney chromophobe and pancreatic cancer showed low turnover rates

per month (about 1.5 times), possibly indicating low incidence rate. On the other hand, bladder urothelial carcinoma, liver hepatocellular carcinoma, colorectal adenocarcinoma and upper tract urothelial carcinoma had the highest turnover rates (almost 4 times per month) which perhaps explained why they were the most common cancers in men and women (48). This also corresponded with data that suggested that number of cell division was a significant risk factor for cancer (49).

We propose from our findings that certain cell populations, specifically Type S-1 could be targeted to address the threat of locoregional recurrence. With currently available tools and advancements in personalized medicine, it is possible to prevent recurrence by targeting a particular cell type or lesion. An example in case in the outstanding success achieved using PD-1 Blockade in mismatch repair–deficient, locally advanced rectal cancer which recorded a 100% success (50). CRISPR-based mutation can also aid in cell competition studies to identify cell fitness levels among the known and unknown driver mutations to further provide actionable data for more studies.

In this study, we estimated cell fitness as a single numerical value with 1.0 indicating normal cells and other cells with ranges from normal cells. In reality, this is an "oversimplification" as cell fitness is a complex and dynamic concept which can be related to both genetic (51) and non-genetic (52) alterations. Unfortunately, studies on cell fitness with regards to known or even unknown
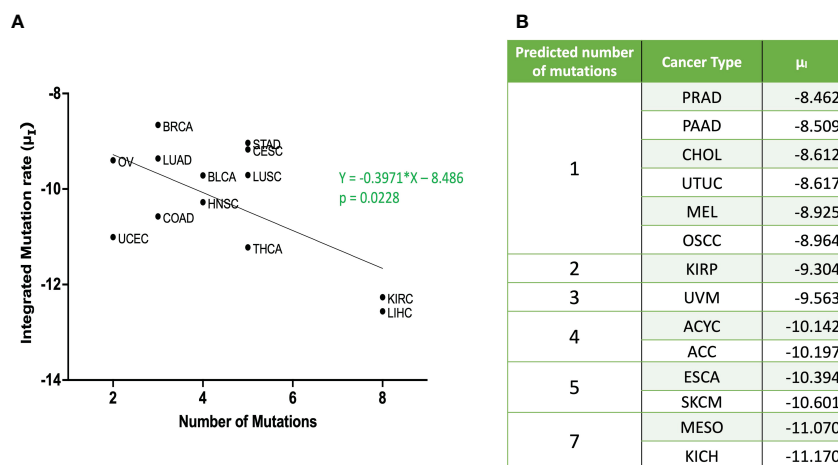
A

Y = -0.3971*X − 8.486
p = 0.0228

(Plot of Integrated Mutation rate (μ_I) vs Number of Mutations, with labeled points: BRCA, STAD, CESC, OV, LUAD, BLCA, LUSC, COAD, HNSC, UCEC, THCA, KIRC, LIHC)

B

Predicted number of mutations	Cancer Type	μ_I
1	PRAD	-8.462
	PAAD	-8.509
	CHOL	-8.612
	UTUC	-8.617
	MEL	-8.925
	OSCC	-8.964
2	KIRP	-9.304
3	UVM	-9.563
4	ACYC	-10.142
	ACC	-10.197
5	ESCA	-10.394
	SKCM	-10.601
7	MESO	-11.070
	KICH	-11.170

FIGURE 7

Relationship between integrated mutation rate and number of mutation hits required for cancer initiation **(A)** Published data for number of mutational hits required for carcinogenesis (37) in some cancer types was plotted against corresponding integrated mutation rate ($\mu_1 \cdot \mu_{S-1} \cdot \mu_S$). The linear regression was performed, and the regression line and the p value are shown. **(B)** Predicted number of mutations for unpublished cancer types as per equation from **(A)** ACC, Adrenocortical Carcinoma; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast Invasive Carcinoma; CESC, Cervical Squamous Cell Carcinoma; CHOL – Cholangiocarcinoma; COAD, Colorectal Adenocarcinoma; ESCA, Esophageal Adenocarcinoma; HNSC, Head & Neck Squamous Cell Carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney Renal Clear Cell Carcinoma; KIRP, Kidney Renal Papillary Cell Carcinoma; LIHC, Liver Hepatocellular Carcinoma; LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinoma; MESO – Mesothelioma; OV, Ovarian Serous Cystadenocarcinoma; PAAD, Pancreatic Adenocarcinoma; PRAD, Prostate Adenocarcinoma; STAD, Stomach Adenocarcinoma; THCA, Thyroid Carcinoma; UCEC, Uterine Corpus Endometrial Carcinoma; UVM, Uveal Melanoma; ACYC, Adenoid Cystic Carcinoma; MEL, Acral Melanoma; UTUC, Upper Tract Urothelial Cancer; OSCC, Oral Squamous Cell Carcinoma; SKCM, Skin Cutaneous Melanoma.

cancer-related mutations are lacking. Also, order of mutations in premalignant cells and a comprehensive study of cell-based or animal model mutational requirements for certain cancers are unavailable. These limit the tools with which we can perform additional validation of our model. Mutation rates were chosen to include processes involved with DNA repair, epigenetics, infection and role of external agents. Each of these could independently affect the model but we chose to combine them. In the current analysis, hematologic or liquid cancers were not included partly because of their dynamic nature and lack of 2D lattice arrangement but mainly the difficulty in assessing exact cell numbers. Even though certain tumor markers for certain malignancies may be used to quantify cell number, the threshold for detection and overall utility is not fully assured. The model without spatial structure might be applicable in this scenario as well as for sarcomas. Moreover, we did not stratify or independently differentiate demographic information such as age, sex or race for each cancer type. Possible extension of the analysis may be to perform age or other parameter dependent analysis. Furthermore, we did not specify the order of mutations for malignant transformation in our model, which albeit gave us a good fit with clinical data. The order of mutations is quite important as revealed from data accrued from colorectal cancer progression (53). Considering multiple mutational orders could be beneficial especially those leading to histologically 'abnormal' benign lesions. Barrett's Esophagus (BE) is a notable example where whole genome sequencing found similar mutational events between esophageal adenocarcinoma and non-dysplastic BE (54) thereby suggesting different mutational order (55). Reports that prior diagnosis of BE affords a better prognosis (56) with only about

5% of BE patients developing esophageal adenocarcinoma (57) further strengthens the different order of mutation concept.

One challenge for cancer management is late diagnosis. Our model computes a cancer detection stage of $1cm^3$ – 10^9 cells. To evaluate the effect of late diagnosis, we changed the cancer detection time to 10^{10} and assess parameter dependence on recurrence time. We observed a reduction in time to recurrence indicating that late diagnosis might contribute to shorter recurrence time (Supplementary Figure 3). Another challenge to the usage of this model is the variability of proportion of locoregional recurrence out of total recurrence rate among various cancer types. It is common knowledge that recurrence can occur at a distant area from the original issue – metastasis; our model however, does not take this into account. As a result, the utility of this model is high for certain cancer types but unfortunately, subdued for other cancer types. Consequently, malignancies where locoregional recurrence accounts for a high proportion of total recurrence such as thyroid cancer with 94% (58), oral squamous cell carcinoma with 90% (59), cholangiocarcinoma with 85% (60), prostate cancer with 81% (61), liver cancer with 78% (62), mesothelioma with 74% (63), head & neck squamous cell carcinoma with 69% (64), and ovarian cancer with 68% (65) could reap great benefit from this model. On the other hand, cancers where distant metastasis accounts for a major proportion of total recurrence such as kidney cancers with 73% (66), skin cutaneous melanoma with 71% (67) and bladder urothelial cancer with 66% (68) might feel the need to complement our model with additional tools to increase its precision. Interestingly, we can get some insight from recurrence pattern of breast cancer. In patients undergoing conservative breast

surgery only, locoregional recurrence accounts for 62% of all cancer recurrence (69). However, in a study with data for different surgical intervention types, locoregional recurrence rates were 42.9% and 19% of total recurrence in breast conservative surgery and total mastectomy respectively (70). This could perhaps be due to the elimination of the cancerized field by total mastectomy which conservative surgery is unable to achieve.

In conclusion, this model reveals parameter combinations that fit clinical data and contributes to the ever-growing knowledge about cancer initiation and recurrence. The model shows elucidate cancers which have premalignant cells with high fitness are likely to have a short recurrence time. This approach can be a valuable tool in the management of cancer especially in the field of personalized molecular medicine to target patients who are at highest risk of recurrence.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

SDA, MT, and HH conceived the idea of the study. SDA, MT, and HH developed simulation codes. SDA conducted computational simulations. SDA and HH contributed to the interpretation of the results. HH supervised the conduct of this study. All authors reviewed the manuscript draft and revised it. All authors approved the final version of the manuscript to be published.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1116210/full#supplementary-material

References

1. Tlsty TD, Coussens LM. Tumor stroma and regulation of cancer development. *Annu Rev Pathol* (2006) 1:119–50. doi: 10.1146/annurev.pathol.1.110304.100224

2. Uramoto H, Tanaka F. Recurrence after surgery in patients with nsclc. *Transl Lung Cancer Res* (2014) 3(4):242–9. doi: 10.3978/j.issn.2218-6751.2013.12.05

3. Hiller JG, Perry NJ, Poulogiannis G, Riedel B, Sloan EK. Perioperative events influence cancer recurrence risk after surgery. *Nat Rev Clin Oncol* (2018) 15(4):205–18. doi: 10.1038/nrclinonc.2017.194

4. Mahvi DA, Liu R, Grinstaff MW, Colson YL, Raut CP. Local cancer recurrence: The realities, challenges, and opportunities for new therapies. *CA Cancer J Clin* (2018) 68(6):488–505. doi: 10.3322/caac.21498

5. Corrado G, Salutari V, Palluzzi E, Distefano MG, Scambia G, Ferrandina G. Optimizing treatment in recurrent epithelial ovarian cancer. *Expert Rev Anticancer Ther* (2017) 17(12):1147–58. doi: 10.1080/14737140.2017.1398088

6. Karacz CM, Yan J, Zhu H, Gerber DE. Timing, sites, and correlates of lung cancer recurrence. *Clin Lung Cancer* (2020) 21(2):127–35.e3. doi: 10.1016/j.cllc.2019.12.001

7. Sonoda D, Matsuura Y, Kondo Y, Ichinose J, Nakao M, Ninomiya H, et al. Comparison of local therapy in patients with lung oligo-recurrence of non-Small-Cell lung cancer. *J Surg Oncol* (2021) 123(8):1828–35. doi: 10.1002/jso.26453

8. Holleczek B, Stegmaier C, Radosa JC, Solomayer EF, Brenner H. Risk of locoregional recurrence and distant metastases of patients with invasive breast cancer up to ten years after diagnosis - results from a registry-based study from Germany. *BMC Cancer* (2019) 19(1):520. doi: 10.1186/s12885-019-5710-5

9. Holloway CMB, Shabestari O, Eberg M, Forster K, Murray P, Green B, et al. Identifying breast cancer recurrence in administrative data: Algorithm development and validation. *Curr Oncol* (2022) 29(8):5338–67. doi: 10.3390/curroncol29080424

10. Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, et al. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *N Engl J Med* (2017) 377(19):1836–46. doi: 10.1056/NEJMoa1701830

11. Kim HG, Kim HS, Yang SY, Han YD, Cho MS, Hur H, et al. Early recurrence after neoadjuvant chemoradiation therapy for locally advanced rectal cancer: Characteristics and risk factors. *Asian J Surg* (2021) 44(1):298–302. doi: 10.1016/j.asjsur.2020.07.014

12. Nakauchi M, Vos E, Tang LH, Gonen M, Janjigian YY, Ku GY, et al. Outcomes of neoadjuvant chemotherapy for clinical stages 2 and 3 gastric cancer patients: Analysis of timing and site of recurrence. *Ann Surg Oncol* (2021) 28(9):4829–38. doi: 10.1245/s10434-021-09624-5

13. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* (1953) 6 (5):963–8. doi: 10.1002/1097-0142(195309)6:5<963::aid-cncr2820060515>3.0.co;2-q

14. Sinjab A, Han G, Wang L, Kadara H. Field carcinogenesis in cancer evolution: What the cell is going on? *Cancer Res* (2020) 80(22):4888–91. doi: 10.1158/0008-5472.CAN-20-1956

15. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer* (2018) 18(1):19–32. doi: 10.1038/nrc.2017.102

16. Willenbrink TJ, Ruiz ES, Cornejo CM, Schmults CD, Arron ST, Jambusaria-Pahlajani A. Field cancerization: Definition, epidemiology, risk factors, and outcomes. *J Am Acad Dermatol* (2020) 83(3):709–17. doi: 10.1016/j.jaad.2020.03.126

17. Rahal Z, Sinjab A, Wistuba II, Kadara H. Game of clones: Battles in the field of carcinogenesis. *Pharmacol Ther* (2022) 237:108251. doi: 10.1016/j.pharmthera.2022.108251

18. Heaphy CM, Griffith JK, Bisoffi M. Mammary field cancerization: Molecular evidence and clinical importance. *Breast Cancer Res Treat* (2009) 118(2):229–39. doi: 10.1007/s10549-009-0504-0

19. Gadaleta E, Thorn GJ, Ross-Adams H, Jones LJ, Chelala C. Field cancerization in breast cancer. *J Pathol* (2022) 257(4):561–74. doi: 10.1002/path.5902

20. Califano J, Ahrendt SA, Meininger G, Westra WH, Koch WM, Sidransky D. Detection of telomerase activity in oral rinses from head and neck squamous cell carcinoma patients. *Cancer Res* (1996) 56(24):5720–2.

21. Boscolo-Rizzo P, Da Mosto MC, Rampazzo E, Giunco S, Del Mistro A, Menegaldo A, et al. Telomeres and telomerase in head and neck squamous cell carcinoma: From pathogenesis to clinical implications. *Cancer Metastasis Rev* (2016) 35(3):457–74. doi: 10.1007/s10555-016-9633-1

22. Galandiuk S, Rodriguez-Justo M, Jeffery R, Nicholson AM, Cheng Y, Oukrif D, et al. Field cancerization in the intestinal epithelium of patients with crohn's ileocolitis. *Gastroenterology* (2012) 142(4):855–64.e8. doi: 10.1053/j.gastro.2011.12.004

23. Cappellesso R, Lo Mele M, Munari G, Rosa-Rizzotto E, Guido E, De Lazzari F, et al. Molecular characterization of "Sessile serrated" adenoma to carcinoma transition in six early colorectal cancers. *Pathol Res Pract* (2019) 215(5):957–62. doi: 10.1016/j.prp.2019.02.001

24. Pirlog R, Cismaru A, Nutu A, Berindan-Neagoe I. Field cancerization in nsclc: A new perspective on micrornas in macrophage polarization. *Int J Mol Sci* (2021) 22 (2):746-761. doi: 10.3390/ijms22020746

25. Syk E, Torkzad MR, Blomqvist L, Ljungqvist O, Glimelius B. Radiological findings do not support lateral residual tumour as a major cause of local recurrence of rectal cancer. *Br J Surg* (2006) 93(1):113–9. doi: 10.1002/bjs.5233

26. Sessler DI, Pei L, Huang Y, Fleischmann E, Marhofer P, Kurz A, et al. Recurrence of breast cancer after regional or general anaesthesia: A randomised controlled trial. *Lancet* (2019) 394(10211):1807–15. doi: 10.1016/S0140-6736(19)32313-X

27. Zhou KQ, Sun YF, Cheng JW, Du M, Ji Y, Wang PX, et al. Effect of surgical margin on recurrence based on preoperative circulating tumor cell status in hepatocellular carcinoma. *EBioMedicine* (2020) 62:103107. doi: 10.1016/j.ebiom.2020.103107

28. Jeon J, Meza R, Moolgavkar SH, Luebeck EG. Evaluation of screening strategies for pre-malignant lesions using a biomathematical approach. *Math Biosci* (2008) 213 (1):56–70. doi: 10.1016/j.mbs.2008.02.006

29. Curtius K, Hazelton WD, Jeon J, Luebeck EG. A multiscale model evaluates screening for neoplasia in barrett's esophagus. *PloS Comput Biol* (2015) 11(5):e1004272. doi: 10.1371/journal.pcbi.1004272

30. Foo J, Leder K, Ryser MD. Multifocality and recurrence risk: A quantitative model of field cancerization. *J Theor Biol* (2014) 355:170–84. doi: 10.1016/j.jtbi.2014.02.042

31. Ryser MD, Lee WT, Ready NE, Leder KZ, Foo J. Quantifying the dynamics of field cancerization in tobacco-related head and neck cancer: A multiscale modeling approach. *Cancer Res* (2016) 76(24):7078–88. doi: 10.1158/0008-5472.CAN-16-1054

32. Evans EJJr., DeGregori J. Cells with cancer-associated mutations overtake our tissues as we age. *Aging Cancer* (2021) 2(3):82–97. doi: 10.1002/aac2.12037

33. Takaki M, Haeno H. Mathematical modeling of locoregional recurrence caused by premalignant lesions formed before initial treatment. *Front Oncol* (2021) 11:743328. doi: 10.3389/fonc.2021.743328

34. Wang Z, Jensen MA, Zenklusen JC. A practical guide to the cancer genome atlas (Tcga). *Methods Mol Biol* (2016) 1418:111–41. doi: 10.1007/978-1-4939-3578-9_6

35. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* (2018) 173(2):400–16.e11. doi: 10.1016/j.cell.2018.02.052

36. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. *Cell* (2017) 171(5):1029–41 e21. doi: 10.1016/j.cell.2017.09.042

37. Anandakrishnan R, Varghese RT, Kinney NA, Garner HR. Estimating the number of genetic mutations (Hits) required for carcinogenesis based on the distribution of somatic mutations. *PloS Comput Biol* (2019) 15(3):e1006881. doi: 10.1371/journal.pcbi.1006881

38. Iranzo J, Martincorena I, Koonin EV. Cancer-mutation network and the number and specificity of driver mutations. *Proc Natl Acad Sci U.S.A.* (2018) 115 (26):E6010–E9. doi: 10.1073/pnas.1803155115

39. Moran P. *The statistical processes of evolutionary theory*. Oxford: Clarendon (1962).

40. Athreya KB, Ney PE, Ney P. *Branching processes*. Courier Corporation, New York: Dover Publications (2004).

41. Piraino SW, Furney SJ. Beyond the exome: The role of non-coding somatic mutations in cancer. *Ann Oncol* (2016) 27(2):240–8. doi: 10.1093/annonc/mdv561

42. Lawlor K, Perez-Montero S, Lima A, Rodriguez TA. Transcriptional versus metabolic control of cell fitness during cell competition. *Semin Cancer Biol* (2020) 63:36–43. doi: 10.1016/j.semcancer.2019.05.010

43. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* (1976) 22(4):403–34.

44. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cbio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* (2012) 2(5):401–4. doi: 10.1158/2159-8290.CD-12-0095

45. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal. *Sci Signal* (2013) 6(269):pl1. doi: 10.1126/scisignal.2004088

46. Nakabayashi N, Hirose M, Suzuki R, Suzumiya J, Igawa M. How asymptomatic are early cancer patients of five organs based on registry data in Japan. *Int J Clin Oncol* (2018) 23(5):999–1006. doi: 10.1007/s10147-018-1287-2

47. Zeng H, Ran X, An L, Zheng R, Zhang S, Ji JS, et al. Disparities in stage at diagnosis for five common cancers in China: A multicentre, hospital-based, observational study. *Lancet Public Health* (2021) 6(12):e877–e87. doi: 10.1016/S2468-2667(21)00157-2

48. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72(1):7–33. doi: 10.3322/caac.21708

49. 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. doi: 10.1126/science.1260825

50. Cercek A, Lumish M, Sinopoli J, Weiss J, Shia J, Lamendola-Essel M, et al. Pd-1 blockade in mismatch repair-deficient, locally advanced rectal cancer. *N Engl J Med* (2022) 386(25):2363–76. doi: 10.1056/NEJMoa2201445

51. Jain IH, Calvo SE, Markhard AL, Skinner OS, To TL, Ast T, et al. Genetic screen for cell fitness in high or low oxygen highlights mitochondrial and lipid metabolism. *Cell* (2020) 181(3):716–27.e11. doi: 10.1016/j.cell.2020.03.029

52. Fennell KA, Vassiliadis D, Lam EYN, Martelotto LG, Balic JJ, Hollizeck S, et al. Non-genetic determinants of malignant clonal fitness at single-cell resolution. *Nature* (2022) 601(7891):125–31. doi: 10.1038/s41586-021-04206-7

53. Yang L, Wang S, Lee JJ, Lee S, Lee E, Shinbrot E, et al. An enhanced genetic model of colorectal cancer progression history. *Genome Biol* (2019) 20(1):168. doi: 10.1186/s13059-019-1782-4

54. Weaver JMJ, Ross-Innes CS, Shannon N, Lynch AG, Forshew T, Barbera M, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet* (2014) 46(8):837–43. doi: 10.1038/ng.3013

55. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal cancer. *Nat Rev Dis Primers* (2017) 3:17048. doi: 10.1038/nrdp.2017.48

56. Thrift AP. Global burden and epidemiology of Barrett oesophagus and oesophageal cancer. *Nat Rev Gastroenterol Hepatol* (2021) 18(6):432–43. doi: 10.1038/s41575-021-00419-3

57. Sharma P. Barrett Esophagus: A review. *JAMA* (2022) 328(7):663–71. doi: 10.1001/jama.2022.13298

58. Kim H, Kim TH, Choe JH, Kim JH, Kim JS, Oh YL, et al. Patterns of initial recurrence in completely resected papillary thyroid carcinoma. *Thyroid* (2017) 27 (7):908–14. doi: 10.1089/thy.2016.0648

59. Wang B, Zhang S, Yue K, Wang XD. The recurrence and survival of oral squamous cell carcinoma: A report of 275 cases. *Chin J Cancer* (2013) 32(11):614–8. doi: 10.5732/cjc.012.10219

60. Hu LS, Zhang XF, Weiss M, Popescu I, Marques HP, Aldrighetti L, et al. Recurrence patterns and timing courses following curative-intent resection for intrahepatic cholangiocarcinoma. *Ann Surg Oncol* (2019) 26(8):2549–57. doi: 10.1245/s10434-019-07353-4

61. Wilt TJ, Jones KM, Barry MJ, Andriole GL, Culkin D, Wheeler T, et al. Follow-up of prostatectomy versus observation for early prostate cancer. *N Engl J Med* (2017) 377(2):132–42. doi: 10.1056/NEJMoa1615869

62. Zhang H, Liu F, Wen N, Li B, Wei Y. Patterns, timing, and predictors of recurrence after laparoscopic liver resection for hepatocellular carcinoma: Results from a high-volume hpb center. *Surg Endosc* (2022) 36(2):1215–23. doi: 10.1007/s00464-021-08390-5

63. Kostron A, Friess M, Crameri O, Inci I, Schneiter D, Hillinger S, et al. Relapse pattern and second-line treatment following multimodality treatment for malignant pleural mesothelioma. *Eur J Cardiothorac Surg* (2016) 49(5):1516–23. doi: 10.1093/ejcts/ezv398

64. Murakami N, Matsumoto F, Yoshimoto S, Ito Y, Mori T, Ueno T, et al. Patterns of recurrence after selective postoperative radiation therapy for patients with head and neck squamous cell carcinoma. *BMC Cancer* (2016) 16:192. doi: 10.1186/s12885-016-2229-x

65. Harter P, Sehouli J, Vergote I, Ferron G, Reuss A, Meier W, et al. Randomized trial of cytoreductive surgery for relapsed ovarian cancer. *N Engl J Med* (2021) 385 (23):2123–31. doi: 10.1056/NEJMoa2103294

66. Speed JM, Trinh QD, Choueiri TK, Sun M. Recurrence in localized renal cell carcinoma: A systematic review of contemporary data. *Curr Urol Rep* (2017) 18(2):15. doi: 10.1007/s11934-017-0661-3

67. Cho SI, Lee J, Jo G, Kim SW, Minn KW, Hong KY, et al. Local recurrence and metastasis in patients with malignant melanomas after surgery: A single-center analysis of 202 patients in south Korea. *PloS One* (2019) 14(3):e0213475. doi: 10.1371/journal.pone.0213475

68. Ozbir S, Girgin C, Kara C, Dincel C. Local and systemic recurrence patterns of urothelial cancer after radical cystectomy. *Kaohsiung J Med Sci* (2014) 30(10):504–9. doi: 10.1016/j.kjms.2014.03.011

69. Elsayed M, Alhussini M, Basha A, Awad AT. Analysis of loco-regional and distant recurrences in breast cancer after conservative surgery. *World J Surg Oncol* (2016) 14:144. doi: 10.1186/s12957-016-0881-x

70. Huang J, Tong Y, Chen X, Shen K. Prognostic factors and surgery for breast cancer patients with locoregional recurrence: An analysis of 5,202 consecutive patients. *Front Oncol* (2021) 11:763119. doi: 10.3389/fonc.2021.763119