DISSERTATION

Computational Modeling of Cancer Initiation and Locoregional Recurrence with Spatial Structure

(空間構造を含むがん再発の数理モデル解析)

September, 2023

Abubakar, Sharafudeen Dahiru

Table of Contents

ABSTRACT

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. 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 sets 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. 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. We also estimated the mean proportion of non-malignant cells of each cancer type at the time of cancer detection and predict number of mutations required for carcinogenesis for unreported cancer types. Our study provides an insight into how to identify patients who are likely to have a shorter recurrence and where to target the therapeutic intervention.

LIST OF ABBREVIATIONS

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

SKCM - Skin Cutaneous Melanoma

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

Log MSR – Mean of Squared Logarithmic Residuals

TCGA – The Cancer Genome Atlas

COADREAD – Colorectal cancer dataset from TCGA

COAD – Colon Adenocarcinoma only (under COADREAD dataset)

READ – Colon Adenocarcinoma only

COAD_CIN – Colon adenocarcinoma with chromosomal instability

COAD_MSI – Colon adenocarcinoma with microsatellite instability

READ_CIN – Rectal adenocarcinoma with chromosomal instability

INTRODUCTION

1.1 Carcinogenesis

Cancer is the second leading cause of death (1) and a critical barrier to increasing life expectancy in every country as seen from the decline in cardiovascular related death - the leading cause of death (2). There are an estimated yearly 23.6 million new cases and 10 million deaths attributed to cancer worldwide (3). All countries will need to increase their efforts to reduce cancer burden given the ongoing delays and disruptions in cancer screenings, diagnosis and treatment worldwide due to the recent pandemic (4, 5).

Cancers are dynamic cells whose features favor cellular proliferation, differentiation and movement while restricting cell death and tissue stability (6). Several theories – the somatic mutation theory (7), the tissue organization field theory (8) and the atavistic theory of cancer (9) have contributed to our knowledge of how cancers develop (carcinogenesis). Each has its own strengths and weaknesses. Unifying these theories, however, we see that the mutations in DNA and epigenetic alterations caused by a number of factors disrupts the balance between cell proliferation, cell death and immune response (10, 11) leading to proliferative malignant tumors which can arise in a stepwise manner from any nucleated cell or tissue.

1.2 Treatment and Recurrence

Cancers, especially at their early stage can be totally cured by surgery (12), a term known as complete remission. In some cases, the cancer re-emerges at the same site - local recurrence; close to the main site - regional recurrence; or at a distant site - metastasis (13). Despite available therapeutic methods for managing cancers, loco-regional recurrence has remained a clinically

significant problem in most cancer types (13-15). Local recurrence rates could be as high as about 85% (16) for ovarian cancer, or 30% in non-small cell lung cancer (NSCLC) (17, 18) and as low as between 8% (19) to 16.5% (20) in breast cancer. Glioblastoma, one of the most aggressive cancers, have a recurrence rate of almost 100% (21). Advanced surgical techniques, chemotherapy, radiotherapy (13) as well as endocrine therapy (22) 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 (23, 24). Synthetic lethality is receiving increased attention but its effectiveness is too early to judge due to insufficient genomic data (25).

1.3 Field Cancerization

Since its introduction in 1953 (26), field cancerization has been recognized as a major cause of local recurrence (27). 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 (28, 29). 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 (30). In breast cancer, microsatellite markers, epigenetic aberrations, transcriptomic deregulations and hTERT overexpression have been detected in histologically normal mammary tissues (31, 32). In head and neck cancer, loss of heterozygosity of chromosome 9p and telomere dysregulation were commonly observed in benign squamous hyperplasia (33, 34). In colon cancer patients with Crohn's ileocolitis, the same mutations of KRAS, CDKN2A, and TP53 were observed within neoplasia and non-tumor epithelium (35, 36). In Non-small cell lung cancer, miRNA dysfunction has been shown at the level of the tumor and cancerized field (37). Findings at sites of recurrence at regions of total

excision has shown that residual tumor is not a significant cause of recurrence (38). Some researchers have proposed that adrenergic-inflammatory pathways due to choice of anesthesia could predispose to recurrence (39) but a meta-analysis have shown that anesthetic technique have no effect whatsoever (40). In addition, circulating tumor cells - CTCs (41) 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.

1.4 Theoretical Studies on Carcinogenesis

Several theoretical studies have shed light on field cancerization impacts on cancer initiation (28). Jeon et al. examined the multistage clonal expansion model by employing the Poisson process to consider the effects of premalignant cells on cancer initiation (42). The model was applied clinically to explore the efficacy and sensitivity of current biopsy-based screening and ablative methods and suggest alternative screening techniques in Barrett's esophagus (43). Foo et al. developed a spatial evolutionary framework to determine the size distribution of histologically undetectable premalignant fields during diagnosis (44). 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 (45). These findings are in agreement with bulk sequencing data that shows the accumulation of cancer-related mutations as we age (46). A 2-step tumor initiation model provides insights into the relationship between different tissue kinetic parameters and the incidence of recurrent cancers (47) by using public datasets from The Cancer Genome Atlas (TCGA), a valuable resource for genomic and clinical data analysis (48, 49) but fails to account for the varying number of mutational hits required for carcinogenesis (50-52). TCGA is a rich computational

resource for the genomic and mutational data for different cancer types (48, 49) and will be helpful in validating our understanding of field cancerization

1.5 The ‘Moran’ and ‘Branching’ Processes

In computational model of complex biological systems involving cell division and death, stochastic approaches are preferred to deterministic approaches (53). This is because they have higher sensitivity and stability than other approaches; the only disadvantage being the relatively longer run time (54). The Moran process (55) is a stochastic method that describes dynamics of a population of constant size where alleles can compete for dominance under the influence of fitness, mutation and accounts for genetic drift and natural selection (56). Using the Moran process, cell division is equal to cell death. This system outperforms other similar methods (57) like the Wright-Fisher process and is robust enough to account for heterogeneity in cell type (58). In systems where the population size is not fixed, the Branching process (59) is employed to account for growing populations where cell division outgrows cell death. Here, individual cells reproduce and mutate independently of each other, without restriction on population size. This system has been successfully applied to explain the growth dynamics of cancer cell lines (60) and phenotypic plasticity in cancer dynamics of a complex system (61). Further refinement of the branching process is applied for cancer progression models with gene mutations (62) and multi-phenotypic population of cells (63) under states of phenotypic equilibrium (64). A stochastic approach is quite useful for modeling complex biological systems involving different cell types with varying phenotypes, fitness, mutation rates and growth characteristics. On one hand, the Moran process is capable of simulating tissue dynamics where there homeostasis is maintained and cell number is kept constant despite presence of “non-lethal” mutations. On the other hand, where this

equilibrium is disrupted and cell division is uncontrolled like in cancers, the Branching process is well equipped to simulate this scenario.

1.6 Spatial Structure

Next generation sequencing has shown the wide heterogeneity of cancer cells (65). It was shown that environmental selection forces such as cell-cell, tissue architecture and cell dispersal strongly influence the molecular heterogeneity of cancer cells (66, 67). A reason for this multiplicity of clones has to do with the relative position of cells within the tissue and is termed spatial structure. A mathematical model theorized that spatial structure delays tumor emergence and could be an evolutionary mechanism to prevent cancer (68). By altering tissue architecture and cell mixing, it was disclosed that the spatial structure of the normal tissue is crucially important to the evolutionary course of cancer (69). Additional data reveal a clear difference in neutral and driver mutational evolution between a non-spatial and different types of spatial structure in a tissue (70) Here, they showed a spatial structure based on boundary growth have neutral effect on both neutral and driver mutations and could be useful in tracking mutations where tissue structure is maintained . This spatial pattern was verified to influence cell proliferation in clear cell renal carcinoma using histological and radiological data (71). Furthermore, cell cycle and mitotic activities were found to be spatially distinctive in an animal model of breast cancer (72); and a 2D lattice-based model showed how the spatial structure influences intra-tumoral competition and drug sensitivity in prostate cancer patients (73). Taking all these together, we see that cancer growth and progression are spatial processes and governs molecular heterogeneity by virtue of cellular dynamics and tissue architecture. It is proposed that spatial patterns need to be emphasized in histological grading of cancer diagnosis as well as introducing newer technologies such as spatial genomics (74), spatial transcriptomics and spatial proteomics in the study of cancer (75).

1.7 Thesis Objectives

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. 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

2.1 Computational Model

This model employs the multi-stage carcinogenesis concept. As tissues might require anywhere between 2 to 8 mutations (denoted by *S*) for malignant transformation (51), 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 (Figure 2) is employed to consider the tissue turnover dynamics, where the total number of Type 0, Type 1, Type *K* and Type *S*-1 cells is kept constant as *N* (55). 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 (Figure 2) is employed to consider the process of Type *S* proliferation (59).

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 (76). 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 $r_{S\text{-}1}$, respectively. Cell fitness, r refers to the transcriptional and metabolic potential of a cell type to "out-compete" other cell types (77). 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, $\mu_{S\text{-}1}$ (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 has 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. Instead, we can approximate intermediate mutational steps between Type 1 and Type S-1 to adjust the values of $\mu_{S\text{-}1}$ so that a low mutation rate, $\mu_{S\text{-}1}$, indicates the integration of several 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 if S is less than or equals to 3 and a mutation rate stands in for the number steps when S is greater than or equal to 4.

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 relative fitness status of dividing cell and which daughter cell are obtainable at the time of a cell division. Meanwhile, 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.

2.2 Simulation Framework

To integrate the Moran process and branching process, we adopted stochastic simulations based on Gillespie's algorithm (78) 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. The probability of each of the events happening are $\frac{1}{dN}$, $\frac{1}{r_S X_S(t)}$ and $\frac{1}{d_S X_S(t)}$ respectively. Gillepsie's Algorithm involves a random selection of an event of known

rate stated above at any given time. Therefore, the average time until any 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~~ triggered 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 $X_{S\text{-}1}$, 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 particular type (here, Type 0) of a cell among dividing cells. 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) $\frac{X_0}{N} \cdot \frac{r_{S-1}X_{S-1}(1-\mu_S)+r_1X_1\mu_{S-1}}{F}$, and (vi) $\frac{X_1}{N} \cdot \frac{r_{S-1}X_{S-1}(1-\mu_S)+r_1X_1\mu_{S-1}}{F}$. 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

$$\Pr[X_{S-1} \rightarrow X_{S-1}+1 \text{ and } X_0 \rightarrow X_0-1\,] = \frac{X_0}{N} \cdot \frac{r_{S-1}X_{S-1}(1-\mu_S)+r_1X_1\mu_{S-1}}{F}, \quad (6)$$

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

$$\Pr[X_{S-1} \rightarrow X_{S-1}+1 \text{ and } X_1 \rightarrow X_1-1\,] = \frac{X_1}{N} \cdot \frac{r_{S-1}X_{S-1}(1-\mu_S)+r_1X_1\mu_{S-1}}{F}. \quad (7)$$

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 $r_SX_S(t) \cdot \Delta T$ or $d_SX_S(t) \cdot \Delta T$, 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.

2.3 Spatial Structure

Two-dimensional lattice structure $(I \times J)$ is introduced to the 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 $< i, j >$ dies, 4 adjacent cells – $< i, j-1 >$, $< i, j+1 >$, $< i-1, j >$ and $< i+1, j >$ 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.

2.4 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 was considered based on the Branching Process. When the number of Type *S* cells exceed twice as large as the size of the normal tissue, 2*N*, growth was regarded as a deterministic process. Then, the time duration from when the number of Type *S* cells is 2*N* to 10^9, Δt_S, is given by

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

2.5 Clinical Data

The data used in our analysis were from TCGA Pan-Cancer Clinical Data Resource (48, 49) and are available in the cBio Cancer Genomics Portal (79, 80). 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 primarily due to their non-conformity to a 2-dimensional lattice structure.

2.6 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 for each patient, which corresponds to the time to recurrence in each simulation trial.

2.7 Colorectal Cancer Analysis

The TCGA colorectal cancer dataset, COADREAD available at the cBioPortal for Cancer Genomics (https://www.cbioportal.org/study/summary?id=coadread_tcga_pan_can_atlas_2018) was further categorized based on race, disease stage at diagnosis, disease subtype and tumor type. We only analyzed categories representing at least 15% of the parent classification to minimize proportional bias. For classification based on race, we analyzed Blacks and Whites (Asians accounts for only 2%); For disease stage – T3 and T4; For tumor type – colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ); For tumor subtype – colon adenocarcinoma with chromosomal instability (COAD_CIN), colon adenocarcinoma with microsatellite instability (COAD_MSI) and rectal adenocarcinoma with chromosomal instability (READ_CIN).

2.8 Statistical Analysis

The whole process of our model is conducted on C++. The survival time analysis and other statistical analysis is conducted on GraphPad 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

3.1 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 3A). 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 3B, Figure 4). 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 (Figure 4 Panel B). Higher r_1 values favored Type 1 dominance (Figure 4 Panel D and E) while equal fitness of Type 1 and Type *S*-1 cells yielded Type 1 dominance or Type 1/*S*-1 co-dominance (Figure 4 Panel C). 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 (Figure 4 Panels A - D). 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 (Figure 4, all panels)

3.2 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 (Figure 5 A, C, E, G, I, 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 (Figure 5 B, D, F, H, J, L). We found that higher fitness of Type 1 cells, r_1 increased the mean recurrence time (Figure 5 A,B), while mutation rate from Type 0 to Type 1, μ_1, had no effect on mean recurrence time (Figure 5 G,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 (Figure 5 A,B,E,F,K,L), while $r_{S\text{-}1}$ and μ_S increases in the proportion of Type *S*-1 cells (Figure 5 C,D,I,J). Mutation rate from Type 0 to Type 1, μ_1, had little effect on the proportion of Type *S*-1 cells (Figure 5 G,H).

3.3 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 6A) based on the composition of non-malignant cells population at cancer detection. Figure 6B and Figure 7 showed the distribution of the patterns in a wide parameter region. Low r_1 values showed Type 0 dominance at low $r_{S\text{-}1}$ levels with failure to detect cancer cells at very low $\mu_{S\text{-}1}$ levels (Figure 7 Panel A). With higher r_1 values, Type 1 cell types begin to dominate. When combined with high $\mu_{S\text{-}1}$, we saw Type 1/*S*-1 co-dominance (green pattern type in Figure 6B). When r_1 and $r_{S\text{-}1}$ 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 $\mu_{S\text{-}1}$ values (Figure 7 Panel C). We noticed a peculiar pattern of Type 0/*S*-1 co-dominance (black region in Figure 6B) when $r_{S\text{-}1}$, $\mu_{S\text{-}1}$, and μ_S were high with relatively lower r_1 value (Figure 7 Panel B). Type *S*-1 dominance (red pattern type in Figure 6B) 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 $r_{S\text{-}1}$, μ_1 and $\mu_{S\text{-}1}$ were high, r_1 was equal to 1.0 and μ_S was relatively small (Figure 7 Panel C). 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. The differences accounted for by integrating the spatial structure is summarized in Table 3.

3.4 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 8). Similarly, we ran 100 to 500 simulations to obtain mean recurrence time (Figure 8 A,C,E,G,I,K) and to check the effect of larger cell numbers (Figure 8 B,D,F,H,J,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 (Figure 8 C to F). However, an increase in r_1 was found to reduce the recurrence 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 (Figure 8 G - 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 8). The major difference is shown on Table 3.

3.5 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%, 8%, …, 100% of patients experienced recurrence) were employed to compare between the *in silico* and published clinical data (79, 80) of 27 cancer types. For the clinical data, the time unit was in months while our simulation did not specify any units. Fitting with clinical data would resolve this conundrum. 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 9).

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 (r_1 and r_{S-1} values in Table 1). 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 10 showed the negative correlation between mutational steps required for carcinogenesis (51) and overall mutation rates (μ_I) obtained from our studies by multiplying the mutation rates for all steps.

Finally, from the parameter sets obtained from clinical fitting (Table 1), we ran 100 simulations of each parameter set corresponding to a particular cancer type to assess the relative proportion of premalignant cells at the time of cancer detection. Figure 11 shows the mean proportion of each of the cell types when the malignant cells reach approximately 10^9 cells ($1cm^3$) in the tissue. we see a dominance of normal Type 0 cells in some cancers like breast invasive carcinoma, cervical squamous cell carcinoma, lung cancers, stomach adenocarcinoma and uterine corpus endometrial

carcinoma. Bladder urothelial cancer, colorectal adenocarcinoma and thyroid carcinoma are dominated by Type 1 lesions while co-dominance is seen in adrenocortical carcinoma, esophageal squamous cell carcinoma and ovarian serous cystadenocarcinoma.

3.6 Colorectal Cancer Analysis

We proceeded to do a deeper analysis of the colorectal cancer dataset, COADREAD to better understand cancer progression in different categories of people with the disease. We chose data points of 1% interval rather than 4% and increased the number of simulations to 100,000 to improve the sensitivity of the fitting. We compared the profiles of each category to the COAD dataset to assess how these differneces affects cancer progression and recurrence. We analyzed the distribution of parameters for each category compared to the parent dataset – COADREAD (Figure 13). From our observations, patients with a Black racial background and those with colon adenocarcinoma with microsatellite instability (COAD_MSI) had profiles which differed in every parameter. Others had variation with just one or two parameters, but the tissue turnover rate remained consistent even with categories that have different survival curves (Figure 14). The parameter combinations for the survival curves and their p values are shown in Table 4. We also analyzed proportion bias and genetic alteration frequency and realized that genetic alteration frequency accounts for differences seen in COAD_MSI but not Black patients (Figure 15).

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 6 and 7). According to the comprehensive analysis of parameter dependence, we found that field cancerization at the detection time depended on a combination of fitness of premalignant cell types and mutation rates from one cell type to the other.

We also revealed the relationship between the proportion of premalignant cells and recurrence time (Figures 5 and 8). 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 (Figure 5A, 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 (Figure 5G, H). All other fitness and mutation parameters led to shorter recurrence time as their 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 8). An exception would be the mean recurrence time with r_1 (Figure 8A, B) which was seen to shorten as r_1 increased but to get slightly extended as r_1 became much larger. This could be due to a renewed cell competition between Type 0 and Type 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, non-hematological 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 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 rS values from our clinical fitting. Interestingly, we observed high growth rates of malignant cells (r_S) in some common cancer types like lung and colorectal cancers, whereas low growth rate was estimated in breast invasive carcinoma which was also a common cancer type but was relatively asymptomatic in agreement with several studies (81, 82). 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 (51), 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 for carcinogenesis.

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 relatively "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 combined (83). This also corresponded with data that suggested that number of cell division was a significant risk factor for cancer (84).

Furthermore, we examined the proportion of premalignant cells of each cancer type at the time of cancer detection for each cancer type by using parameter sets obtained from clinical fitting. Several patterns were observed chiefly, the dominance of Type 0 cells in common cancers – breast, cervix, lung and stomach cancers. This indicated that these cancers might have a longer time to recurrence since we see a relative low number of mutated cells. Prostate cancer has Type 0 dominance but have an observable number of Type 1 cells while liver hepatocellular carcinoma and mesothelioma has Type 0 dominance with an observable Type *S*-1 cell population indicating a more likelihood of recurrence than the previously mentioned cancers. We saw a dominance of Type 1 cells for some cancer types – bladder urothelial, colorectal, adenoid cystic and thyroid cancers. These cancers are known to have common benign lesions which probably contain cancer related mutations. The major presence of Type 1 cells suggested a relatively shorter time to recurrence and should be carefully monitored. Even more alarming was kidney chromophobe and kidney renal cell carcinoma with observable Type *S*-1 cells indicating poorer prognosis. Furthermore, we saw another group of cancers with co-dominance of Type 0 and Type 1 cells – adrenocortical, esophageal, head and neck, ovarian cancer and upper tract urothelial cancers. They may pose a higher risk due to varying proportion of these cell types. Finally, all melanomas investigated had

a pattern of Type 0/Type 1 co-dominance with an observable population of Type *S*-1 cells. These could explain why there are several benign melanocytic lesions in the skin, eye and other tissues.

In the colorectal cancer dataset, we checked the cancer progression in different categories to see how certain factors contribute to cancer progression and recurrence. Our clinical fitting was partially based on Bayesian principles and an analysis of the distribution of parameter combinations disclosed interesting results (Figure 13). The values for r_S, d_S and μ_{int} showed a narrower peak than r_1 and r_{S-1} possibly indicating a higher sensitivity. As stated earlier, the parameters initially mentioned parameters could be validated by epidemiological and clinical data but the later mentioned parameters lack actionable data for validation. Of note, Blacks and COAD_MSI had a discordant profile reflecting all parameters especially d_S. Human organoid cultures estimate tissue turnover rate to be about 12 days (85) which was in line with our estimation. Even categories that show a visually different survival curve had a tissue turnover rate within these limits (Figure 14) except those 2. To further investigate, this discrepancy, we analyzed proportion bias to see if a particular category is proportionately higher in Blacks or COAD_MSI (Figure 15a) but the relative proportions were not statistically significant. We then checked whether ageing influence (86), tumor mutation burden /immune influence (87, 88) or aneuploidy (89) differed between categories but we found no statistically significant difference between groups (Figure 15b). Upon analyzing tumor alteration frequency, we found significant differences in altered genes between COAD_MSI and other categories (Figure 15c) which could explain the discordant profile. However, no difference was seen in Black patients (Figure 15d). We theorize that perhaps, some unaccounted factor(s) might play a role between cancer diagnosis and cancer recurrence in Black patients. Lifestyle changes (90) and co-morbidities (91) are potential factors that might play a role in affecting recurrence time thereby altering the profile. Another key immunological factor might

be MHC class I and II expression differences between Blacks and Whites (92) which could affect any of the parameters.

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 (93). 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 (94) and non-genetic (95) alterations. Unfortunately, studies on cell fitness with regards to known or even unknown cancer-related mutations are lacking. Also, the order of mutations in premalignant cells and a comprehensive study of cell-based or animal model mutational requirements for certain cancers are unavailable for additional validation. 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 only considered a specific order of mutation for malignant transformation in our model which gave us good fitting with clinical data; other mutational orders are disregarded. Other mutational orders could be important 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 only non-dysplastic BE and not dysplastic BE (96) thereby suggesting different mutational order (97). Reports that prior diagnosis of BE affords a better prognosis (98) with only about 5% of BE patients developing esophageal adenocarcinoma (99) further strengthens the different order of mutation concept. Above all, the source of clinical data focuses on malignant tissue and have no information on benign lesions. In this study, only 4 cell types were utilized; so, we used lower mutation rates to denote additional cell types and hereby skipped Type K cells. It is likely that this decision might underestimate the fitness of the different cell types not included even though total mutation rate should theoretically not be affected.

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 (Figure 12). 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% (100), oral squamous cell carcinoma with 90% (101), cholangiocarcinoma with 85% (102), prostate cancer with 81% (103), liver cancer with 78% (104), mesothelioma with 74% (105), head & neck squamous cell carcinoma with 69% (106), and ovarian cancer with 68% (107) 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% (108), skin cutaneous melanoma with 71% (109) and bladder urothelial cancer with 66% (110) might feel the need to complement our model with additional tools to increase its precision. Interestingly, we can gain some insight from recurrence pattern of breast cancer. In patients undergoing conservative breast surgery only, locoregional recurrence accounts for 62% of all cancer recurrence (111). 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 (112). 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 elucidates cancers which have premalignant cells with high fitness are likely to have a short recurrence time (113). The model also appears sensitive enough to distinguish categories with clear genetic differences and even recognize unaccounted factors. 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.

Table 1 – Tumor-specific carcinogenic profiles and *p* values of survival curves (**μ** values are in $\log_{10}$ while SQ are log-MSR 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

Table 2 – List of Notations used in the study and their meaning

Notation	**Meaning**
N	Total cell number in a tissue. It remains unchanged during tissue dynamics
r	Growth rate of a particular cell. It indicates cell fitness
d	Death rate of a cell at a particular time point. Death of a cell triggers cell division
μ	Mutation rate (from one cell type to another)
T	Time unit for each event
X	Number of a specific cell type at a particular time
Δ	Average of a number of events
Pr []	The probability of an event happening

Table 3 – Differences between the non-spatial structure and spatial structure model

	Non-Spatial Structure	Spatial Structure
Cancer Initiation Pattern	Only 3 possible patterns observed	7 cancer initiation patterns observed (4 additional)
Type 0 cells	Population declines no matter the parameter combination	Population of Type 0 depends on parameter combination
Effect of fitness of Type 1 cell, r_1 on cancer recurrence	Higher fitness of Type 1 cells delays cancer recurrence time	Higher fitness of Type 1 cells shortens cancer recurrence time
Effect of mutation rate from Type 0 to Type 1, μ_1 on cancer recurrence	μ_1 have no effect on cancer recurrence time	Higher μ_1 shortens cancer recurrence time
Proportion of Type S-1 cells at cancer detection	Very high particularly at low r_1 and low μ_S	Relatively very low at low r_1 and low μ_S

Table 4 – Carcinogenic Profiles of colorectal cancer categories and *p* values of survival curves (μ values are in $\log_{10}$)

		r_1	r_{S-1}	r_S	d_3	μ_1	μ_{S-1}	μ_S	μ_{int}	***p* value**
COADREAD		1.0787	1.0611	6.8331	3.5176	-4.3010	-4.3665	-3.8761	-12.5436	0.8837
Race	Black	0.9259	0.9951	2.4288	1.7592	-2.2142	-2.4909	-3.9586	-8.6635	0.9563
	White	1.0723	0.9904	6.0739	3.5306	-3.7721	-4.0969	-3.3883	-11.2573	0.8512
Stage	T3	1.0622	1.0270	7.6332	3.8752	-3.3872	-3.5072	-4.3098	-11.2041	0.9734
	T4	0.9702	0.9836	7.3659	3.9276	-3.7055	-2.5178	-3.7645	-9.9872	0.9470
Tumor	COAD	1.0644	1.0238	6.5429	3.6314	-3.3645	-4.3768	-3.5702	-11.3116	0.7144
	READ	0.9990	0.9638	4.9616	3.4045	-2.8239	-3.8761	-3.2048	-9.9066	0.9879
Subtype	COAD_CIN	1.0634	1.0536	6.7611	3.3874	-3.7122	-4.4318	-4.0862	-12.2299	0.6871
	COAD_MSI	0.9922	1.0211	2.8384	2.5118	-3.3279	-3.3830	-3.4868	-10.1979	0.7654
	READ_CIN	0.9854	0.9977	4.1588	3.0714	-3.7932	-3.3536	-3.4353	-10.5817	0.9867

FIGURES

Figure 1

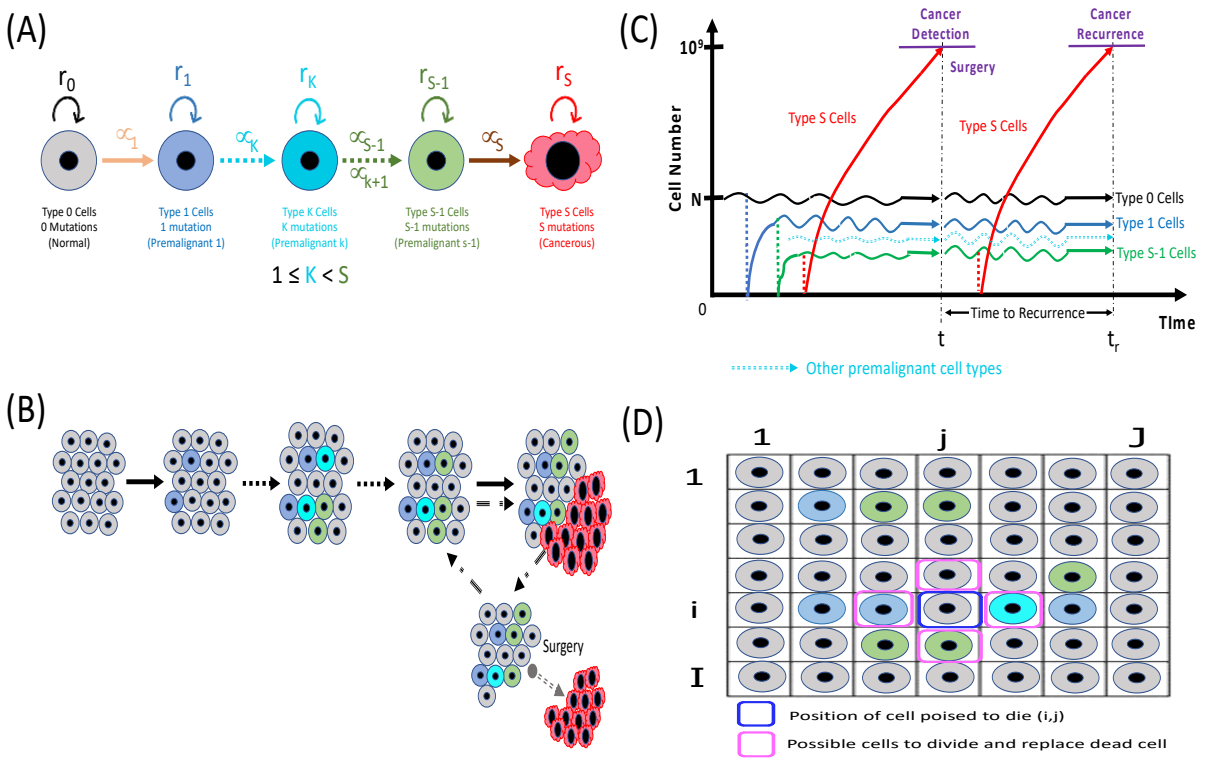


Figure 1 – The illustrative representation of our models

(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.

(113)

Figure 2

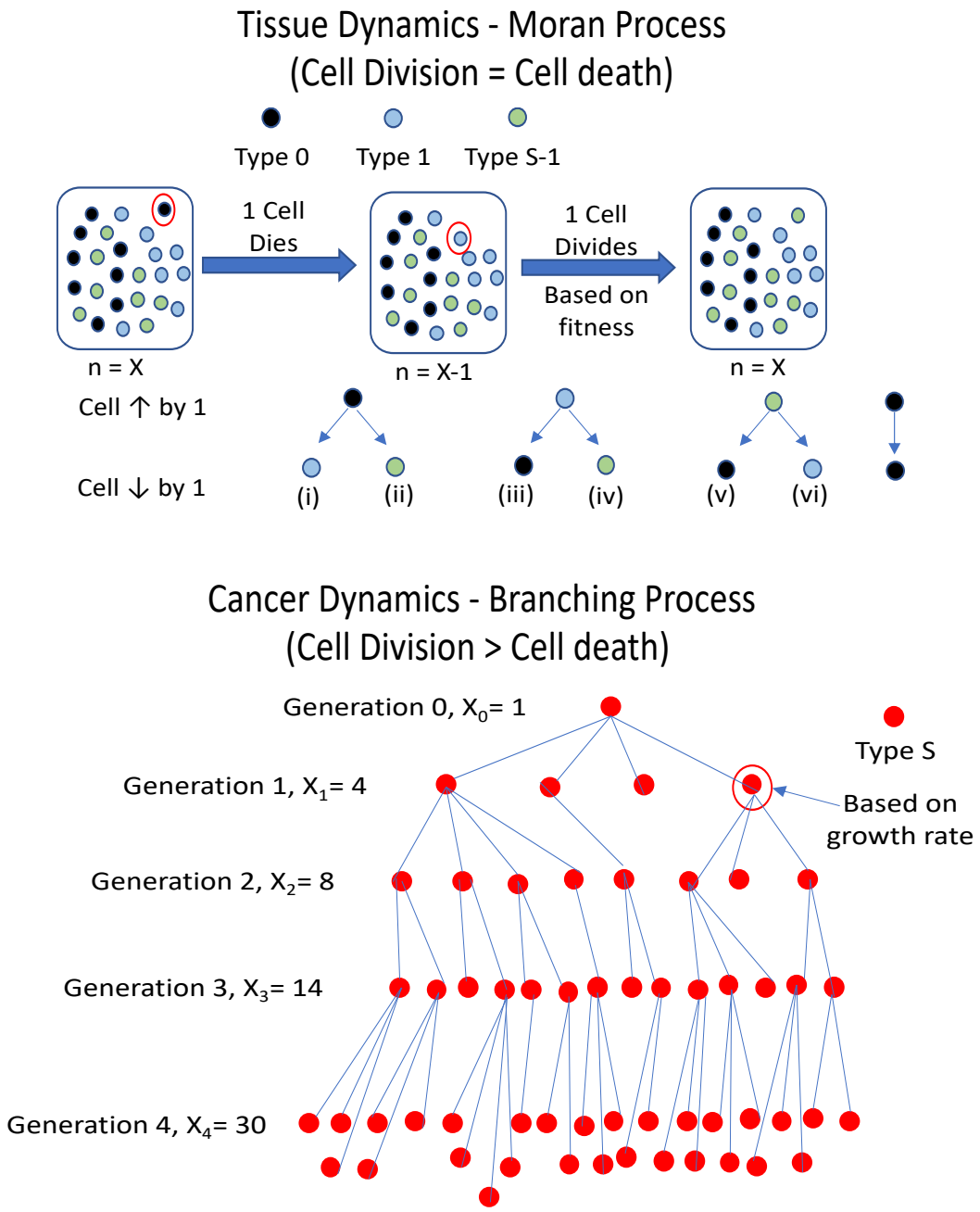


Figure 2 – Schematic description of the Moran and Branching Process.

In the Moran process, tissue turnover is in a state of equilibrium where cell division = cell death and cell number remain unchanged. In Branching process, cell division outnumbers cell death and cell number increases over time.

Figure 3

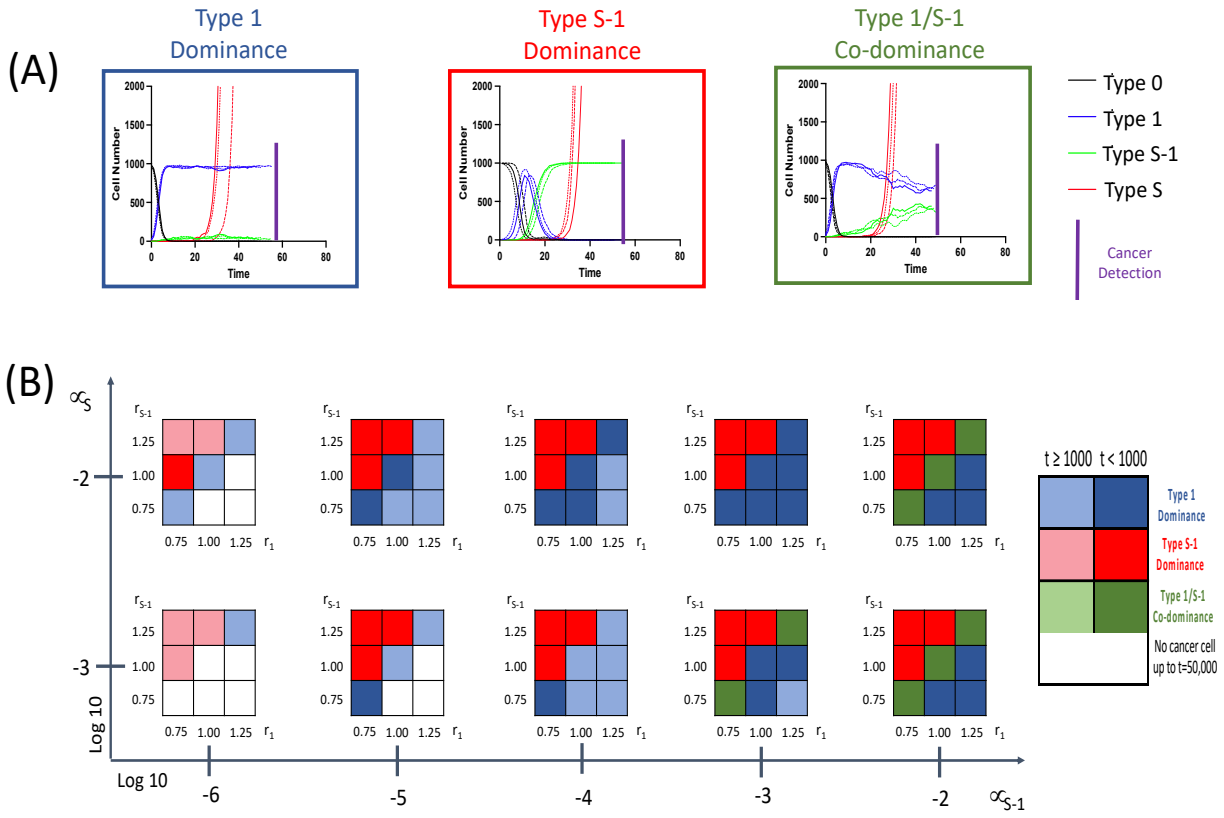


Figure 3 - Patterns of tissue composition at cancer initiation

(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. t denotes cancer detection time and are grouped based on relative length.

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; $r_{S\text{-}1}$= 0.75, 1.00 and 1.25; r_S = 1.5; μ_1 = 0.001 and 0.01; $\mu_{S\text{-}1}$ = 0.000001, 0.00001, 0.0001, 0.001 and 0.01; μ_S = 0.001 and 0.01.

Note: Mutation rate, μ_1, does not impact the results and is not shown.

(113)

Figure 4

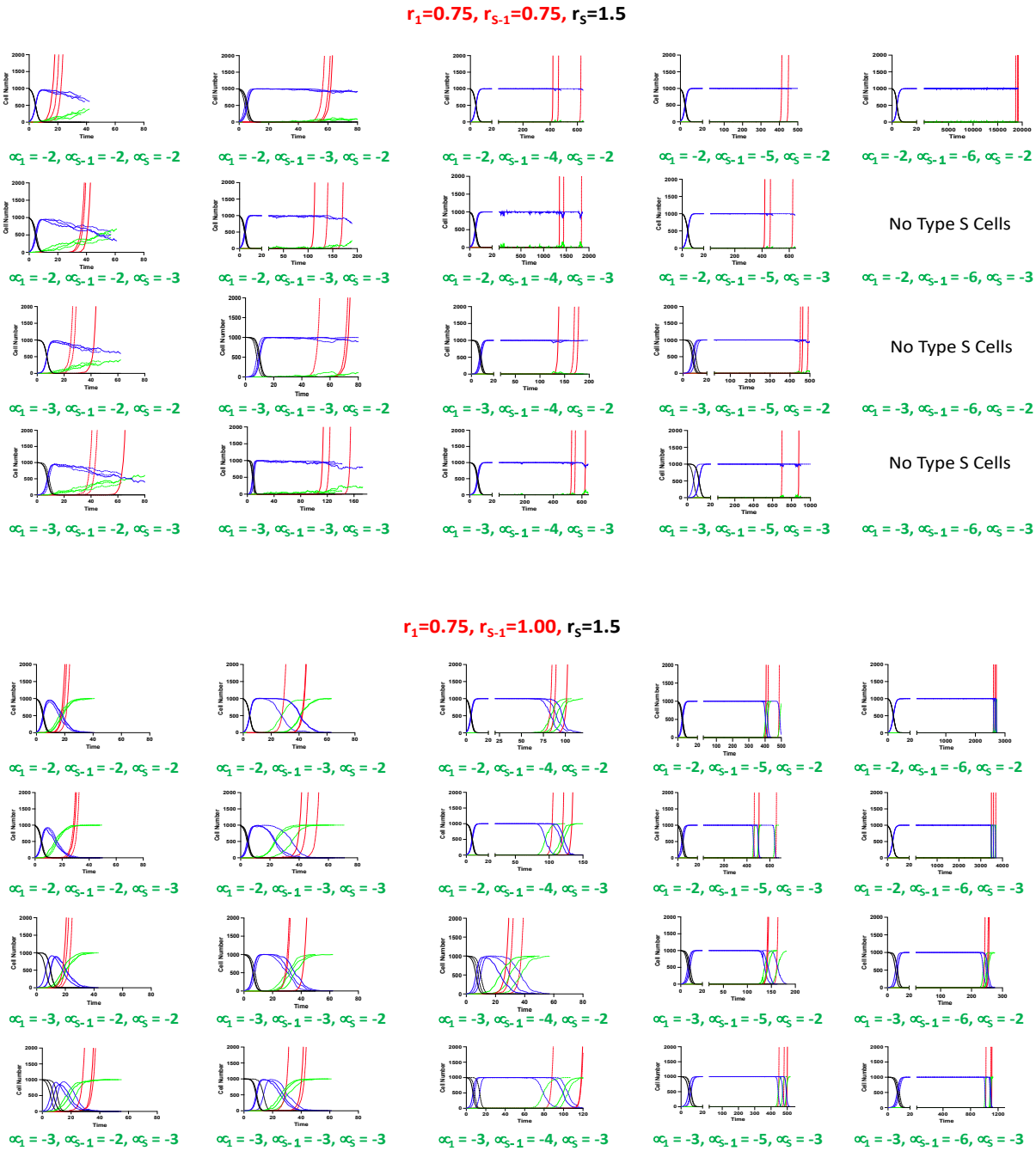


Figure 4 Panel A – Extended cancer initiation pattern for low r_1/ r_{S-1} and low r_1/normal r_{S-1} at different mutation rates

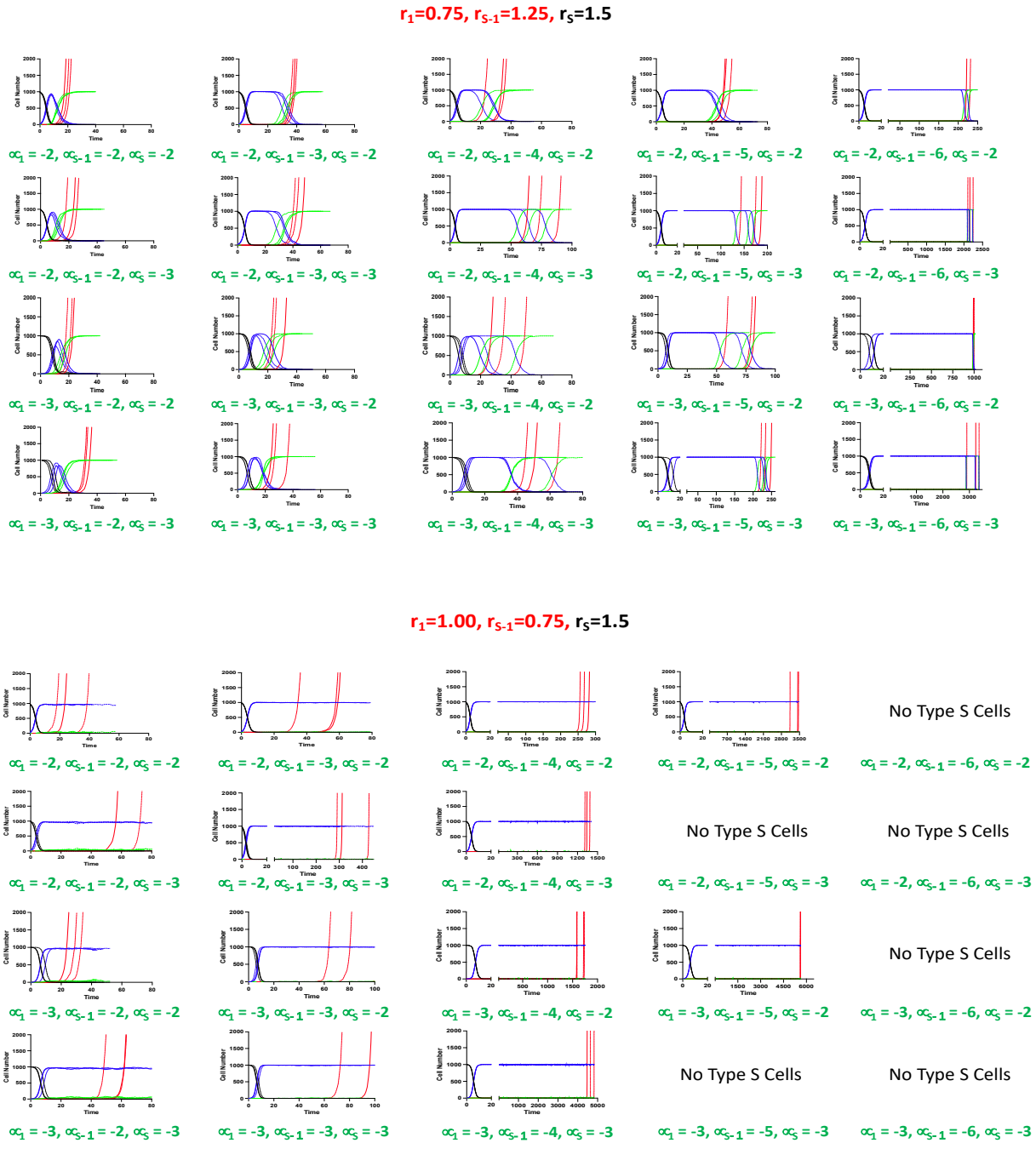


Figure 4 Panel B – Extended cancer initiation pattern for low r_1/high r_{S-1} and normal r_1/low r_{S-1} at different mutation rates

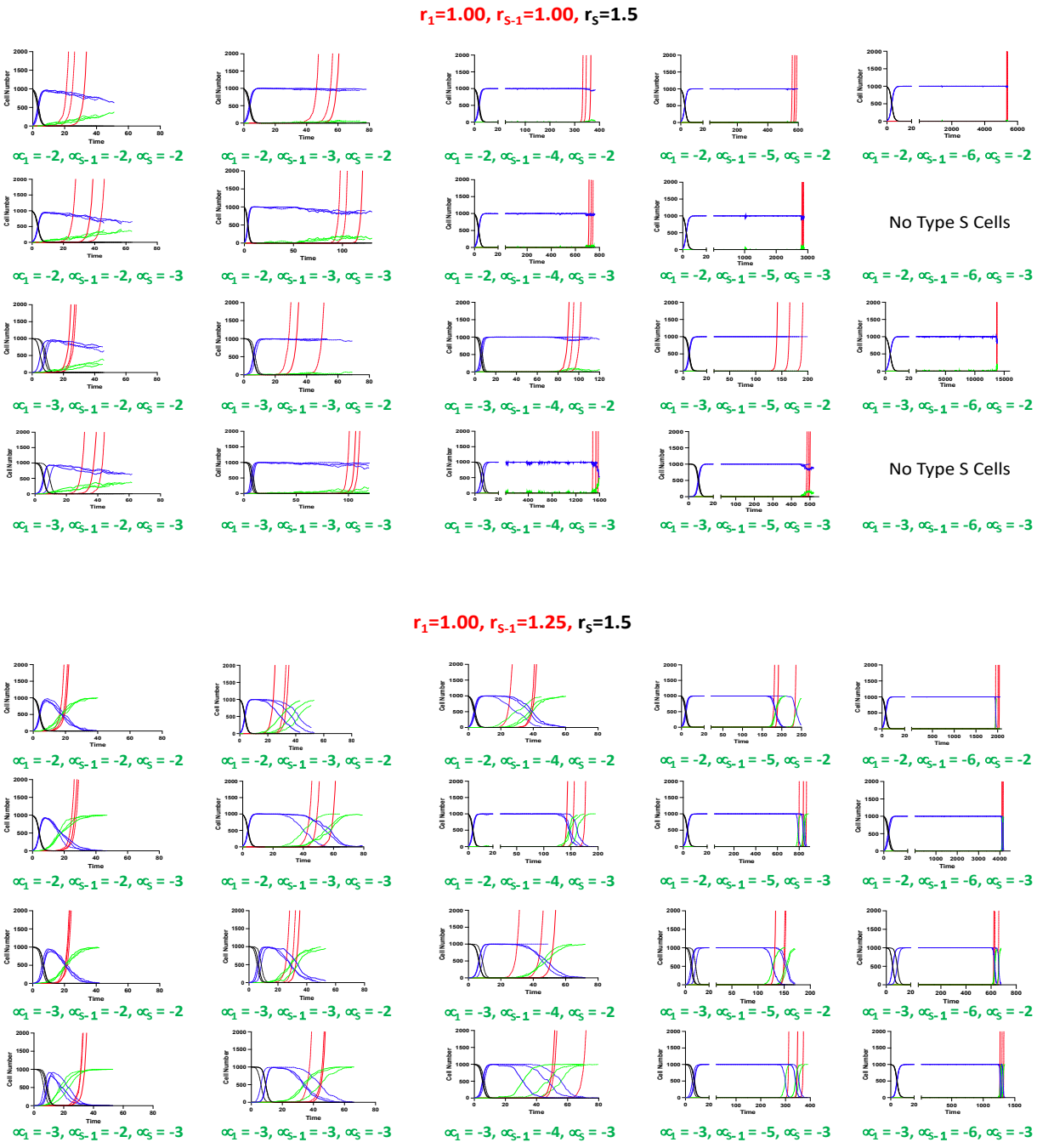


Figure 4 Panel C – Extended cancer initiation pattern for normal r_1/ r_{S-1} and normal r_1/high r_{S-1} at different mutation rates

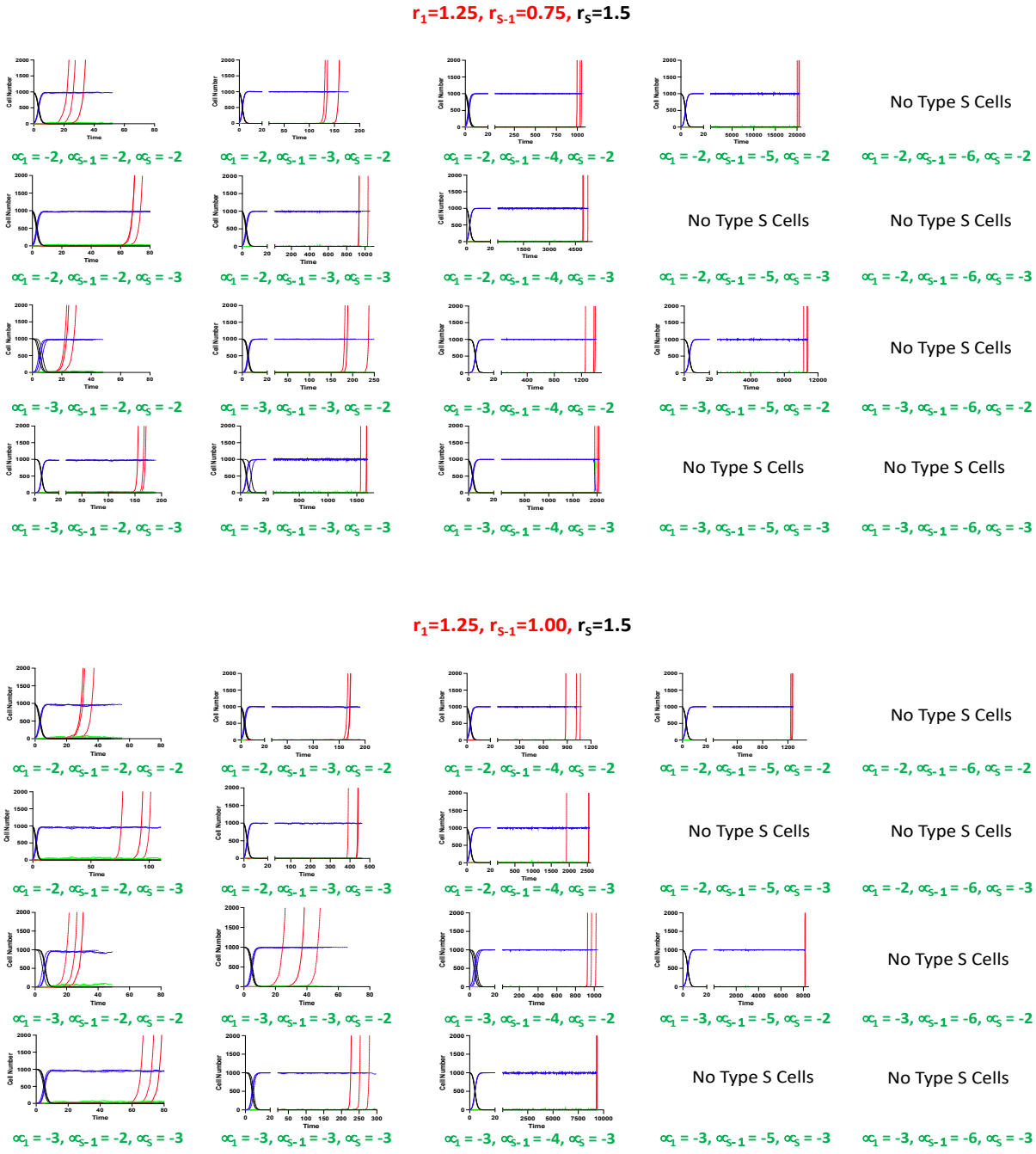


Figure 4 Panel D – Extended cancer initiation pattern for high r_1/low r_{S-1} and high r_1/normal r_{S-1} at different mutation rates

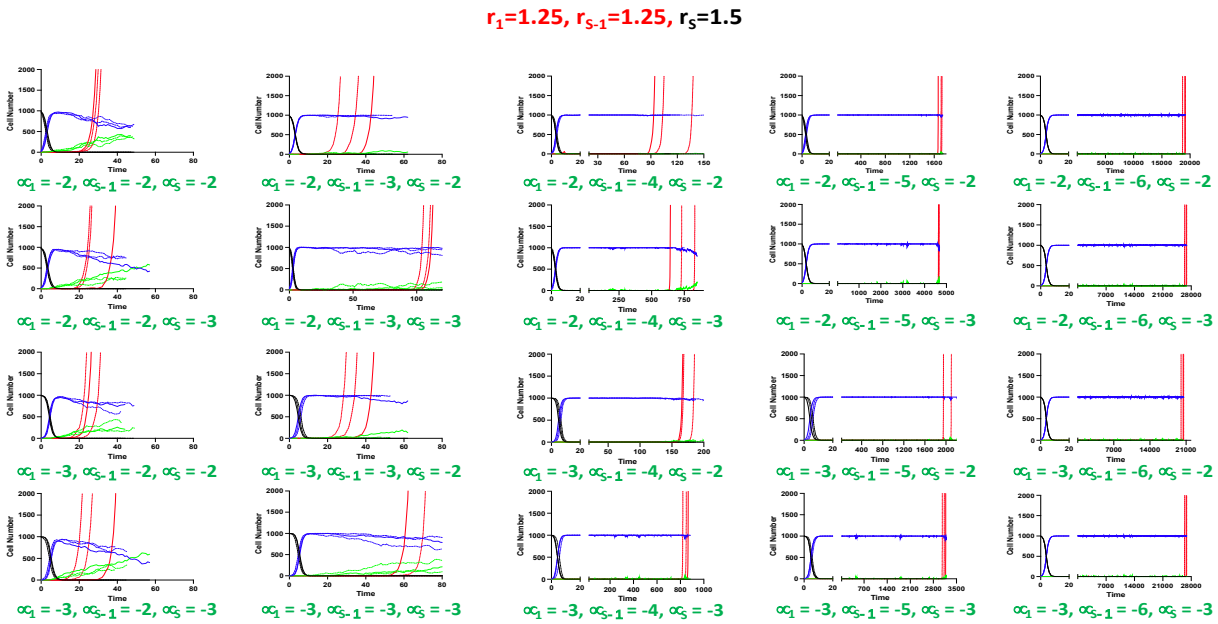


Figure 4 Panel E – Extended cancer initiation pattern for high r_1/high r_{S-1} at different mutation rates

Figure 5

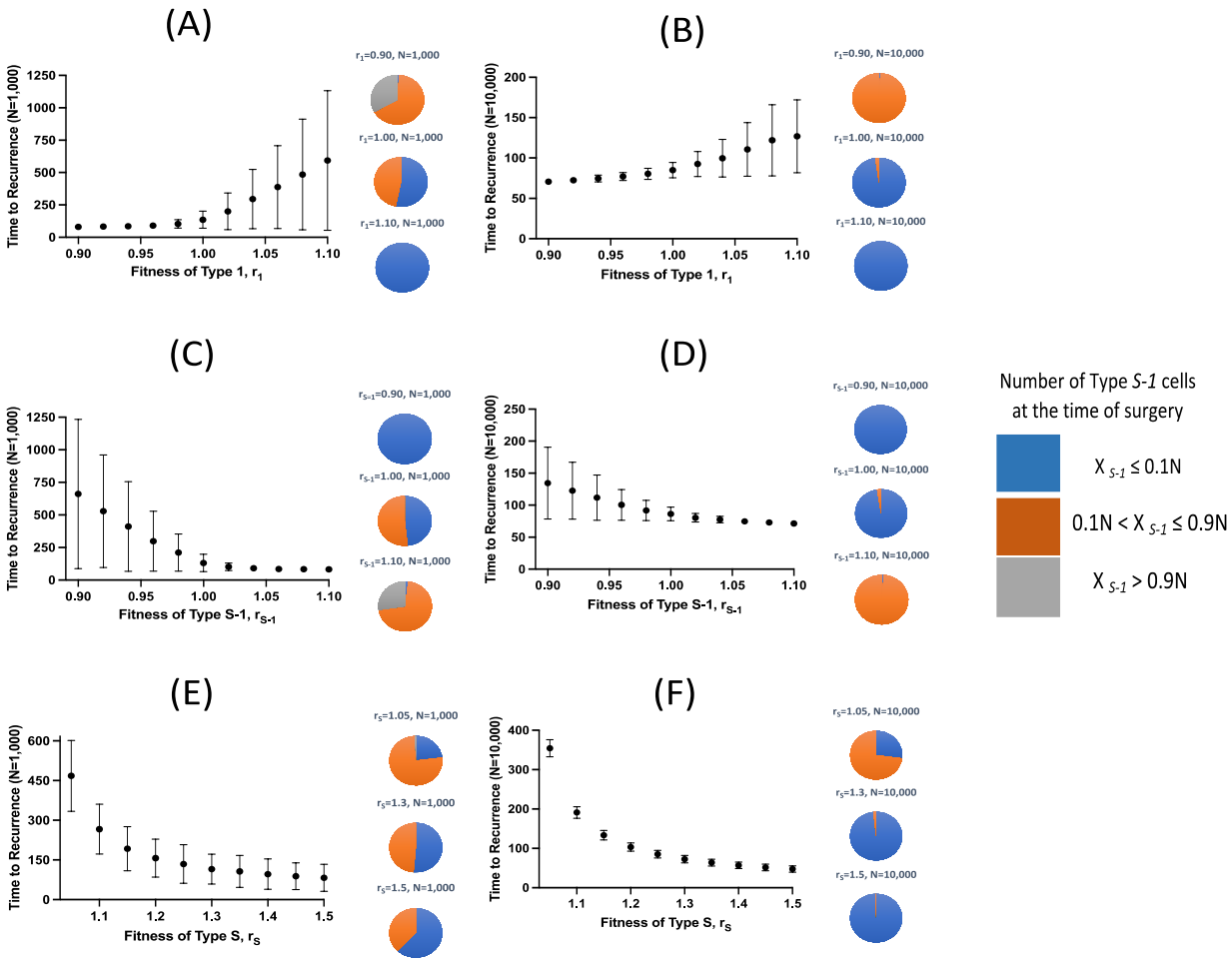


Figure 5 Panel A – Dependence of cellular fitness on recurrence time.

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 ($X_{S\text{-}1} \leq 0.1N$), intermediate ($0.1N < X_{S\text{-}1} \leq 0.9N$), and large ($X_{S\text{-}1} > 0.9N$) proportion of Type *S*-1 cells, respectively.

Standard parameter values used in (A–F) are $d = d_S = 1.0$, $r_0 = 1.0$, $r_1 = 1.0$, $r_{S\text{-}1} = 1.2$, $r_S = 1.0$, $\mu_1 = 0.001$, $\mu_{S\text{-}1} = 0.001$, $\mu_S = 0.001$; and $N = 1{,}000$ in (A, C and E); and $N = 10{,}000$ in (B, D and F).

(113)

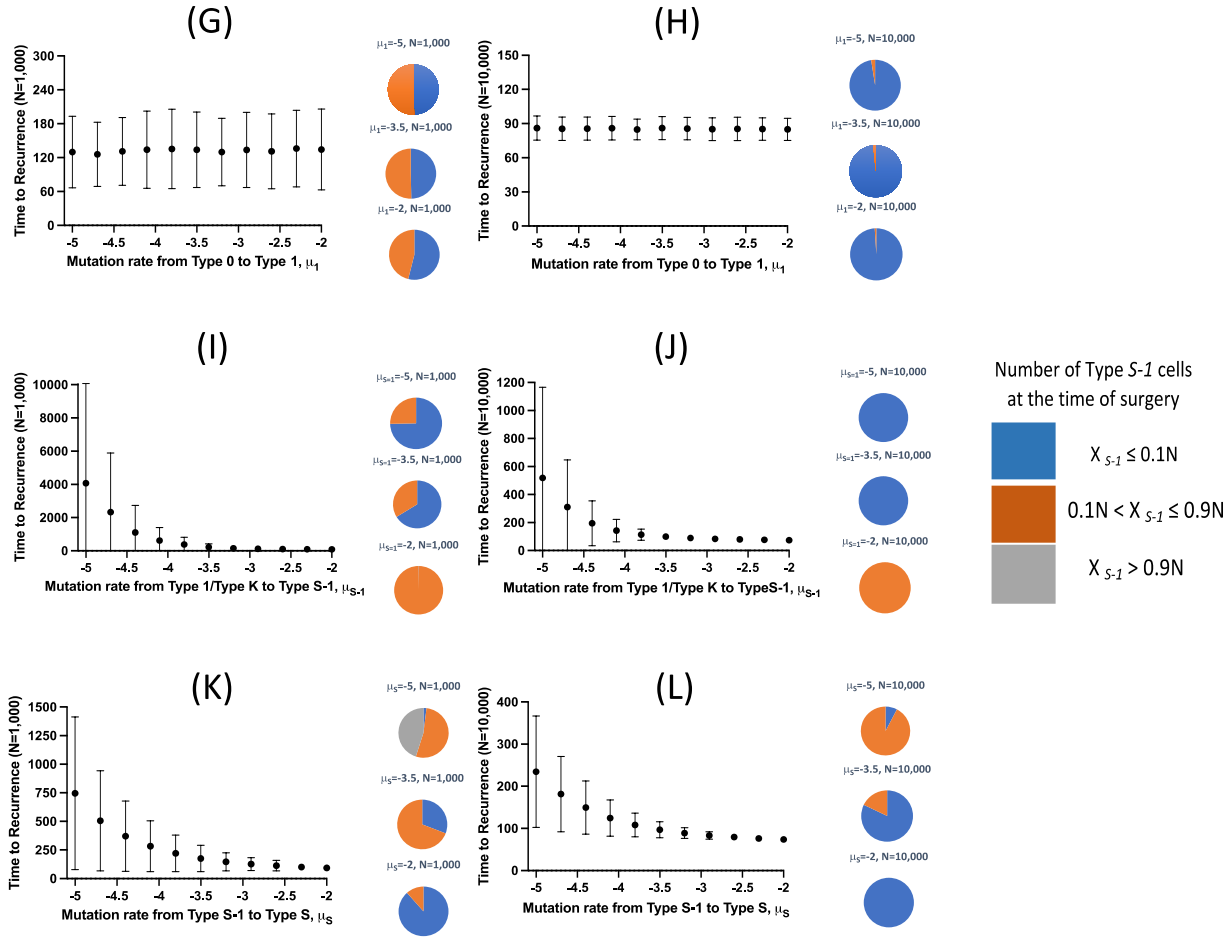


Figure 5 Panel B – Dependence of mutation rates on recurrence time.

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 ($X_{S\text{-}1} \leq 0.1N$), intermediate ($0.1N < X_{S\text{-}1} \leq 0.9N$), and large ($X_{S\text{-}1} > 0.9N$) proportion of Type *S*-1 cells, respectively.

Standard parameter values used in (G–L) are $d = d_s = 1.0$, $r_0 = 1.0$, $r_1 = 1.0$, $r_{S\text{-}1} = 1.2$, $r_S = 1.0$, $\mu_1 = 0.001$, $\mu_{S\text{-}1} = 0.001$, $\mu_S = 0.001$; and $N = 1{,}000$ in (G, I and K); and $N = 10{,}000$ in (H, J and L).

(113)

Figure 6

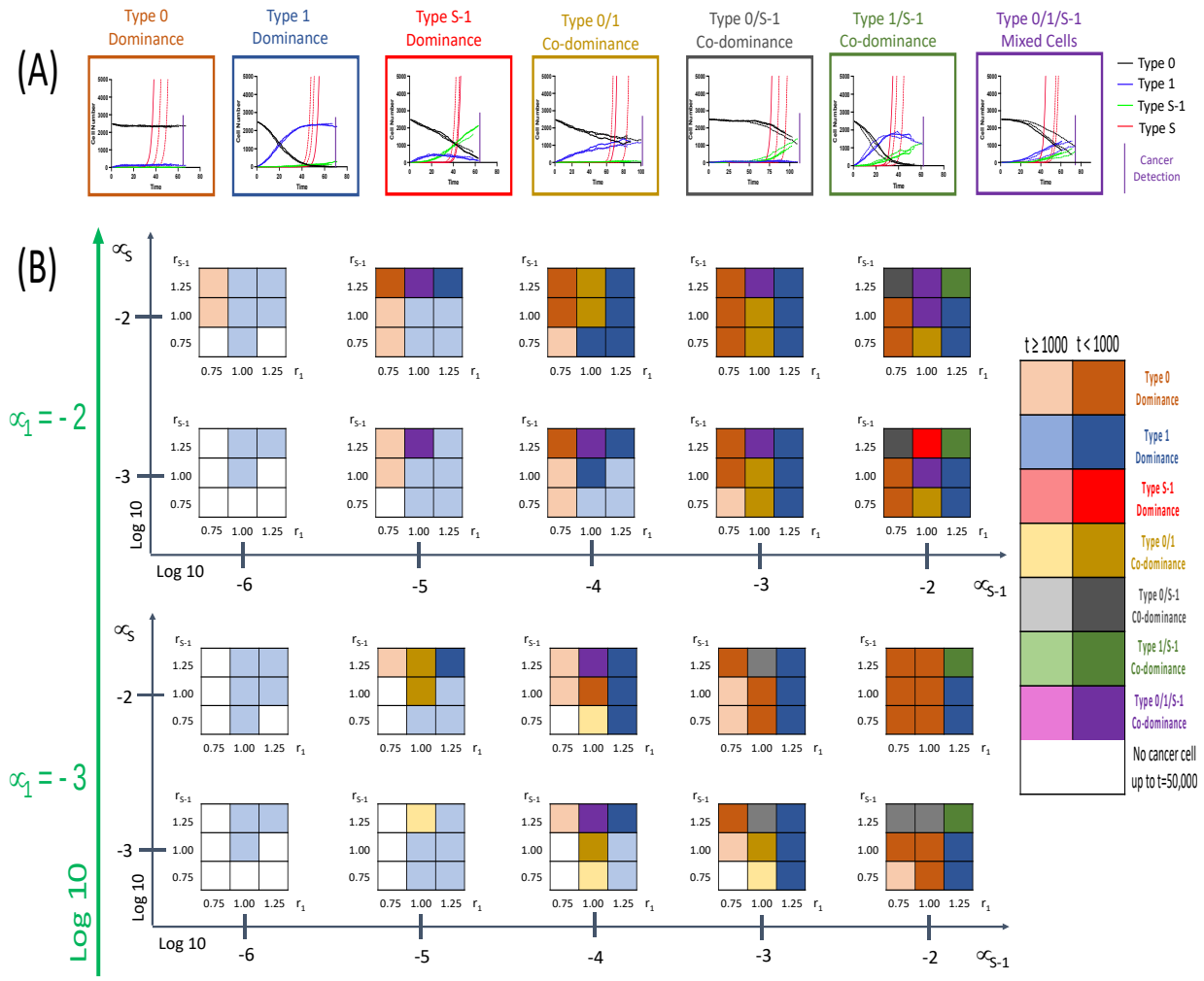


Figure 6 - Patterns of tissue composition at cancer initiation with spatial structure

(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 "t", "Dominance" and "Co-dominance" are the same as those explained in Figure 3.

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; μ_1 = 0.001 and 0.01; μ_{S-1} = 0.000001, 0.00001, 0.0001, 0.001 and 0.01; μ_S = 0.001 and 0.01. (113)

Figure 7

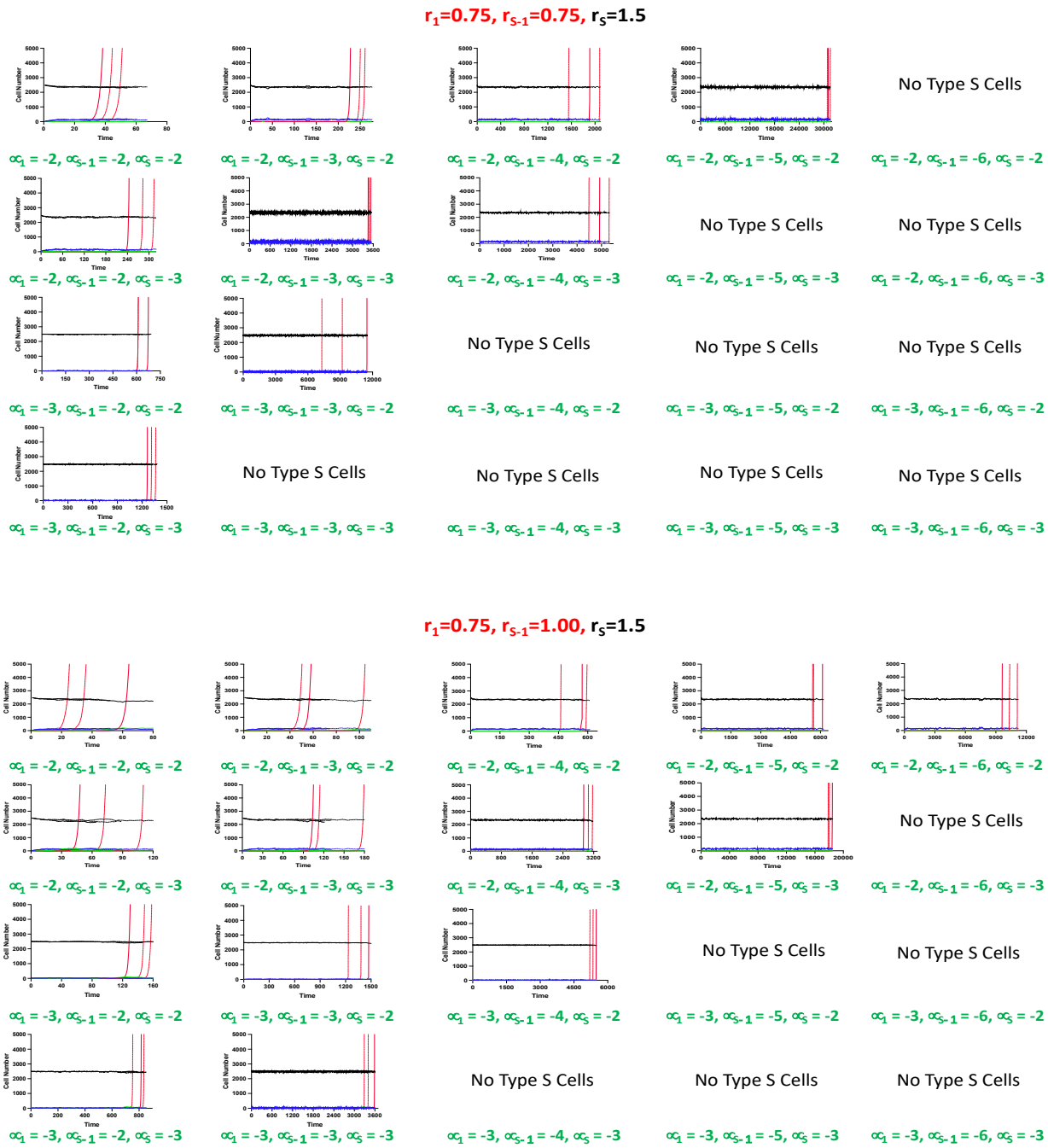


Figure 7 Panel A – Extended cancer initiation pattern with spatial structure for low r_1/ r_{S-1} and low r_1/normal r_{S-1} at different mutation rates

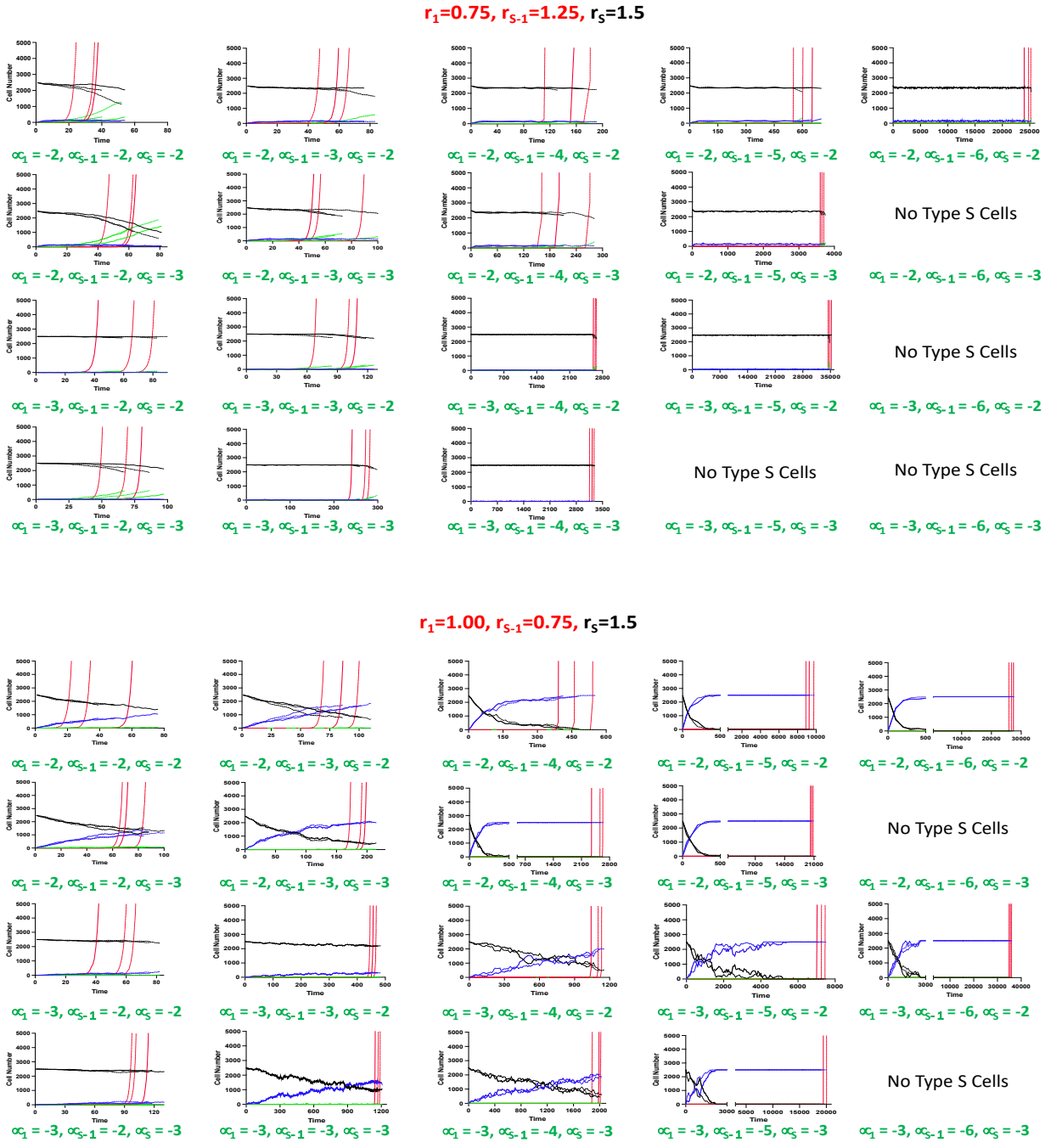


Figure 7 Panel B – Extended cancer initiation pattern with spatial structure for low r_1/high r_{S-1} and normal r_1/low r_{S-1} at different mutation rates

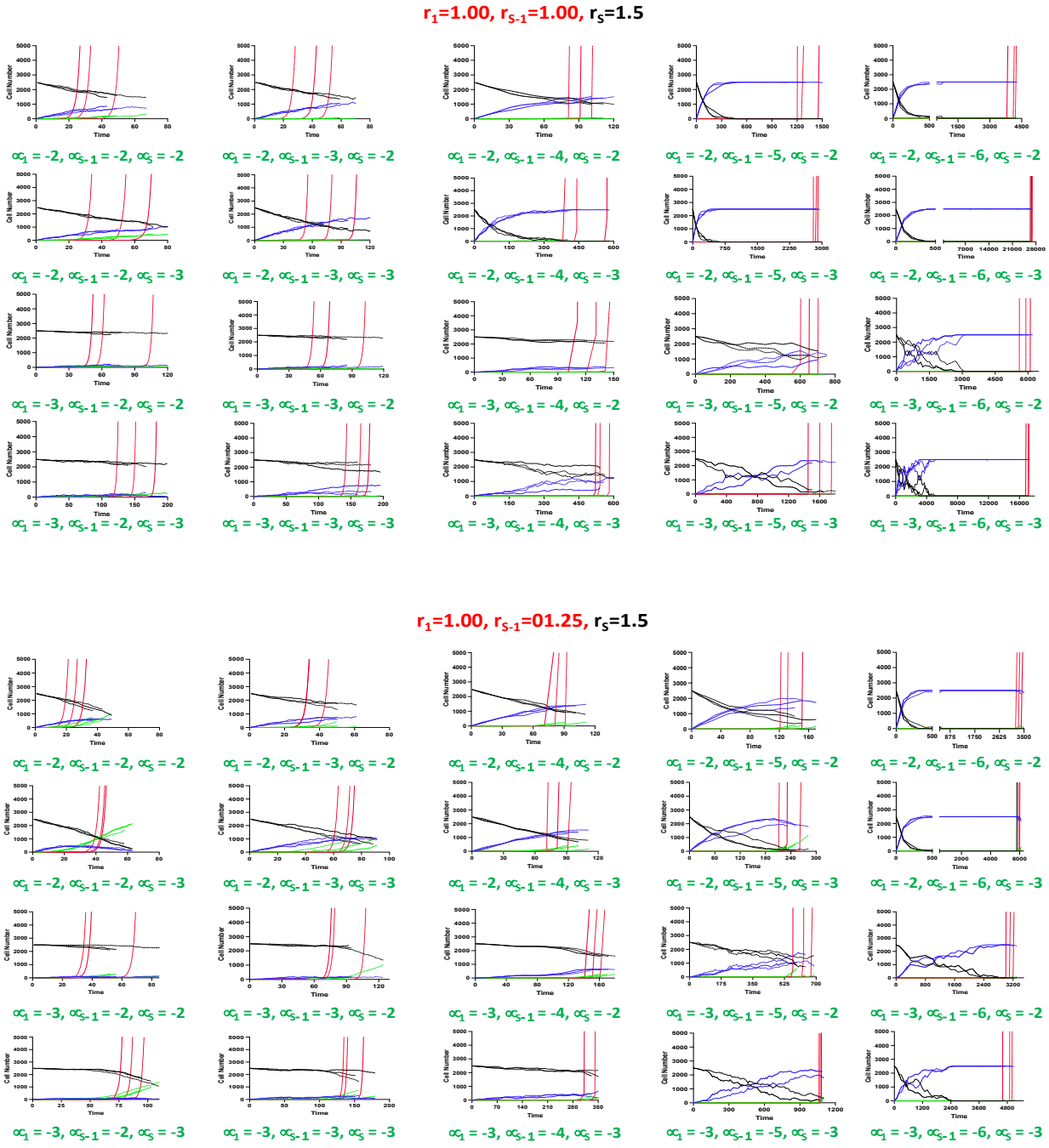


Figure 7 Panel C – Extended cancer initiation pattern with spatial structure for normal r_1/ r_{S-1} and normal r_1/high r_{S-1} at different mutation rates

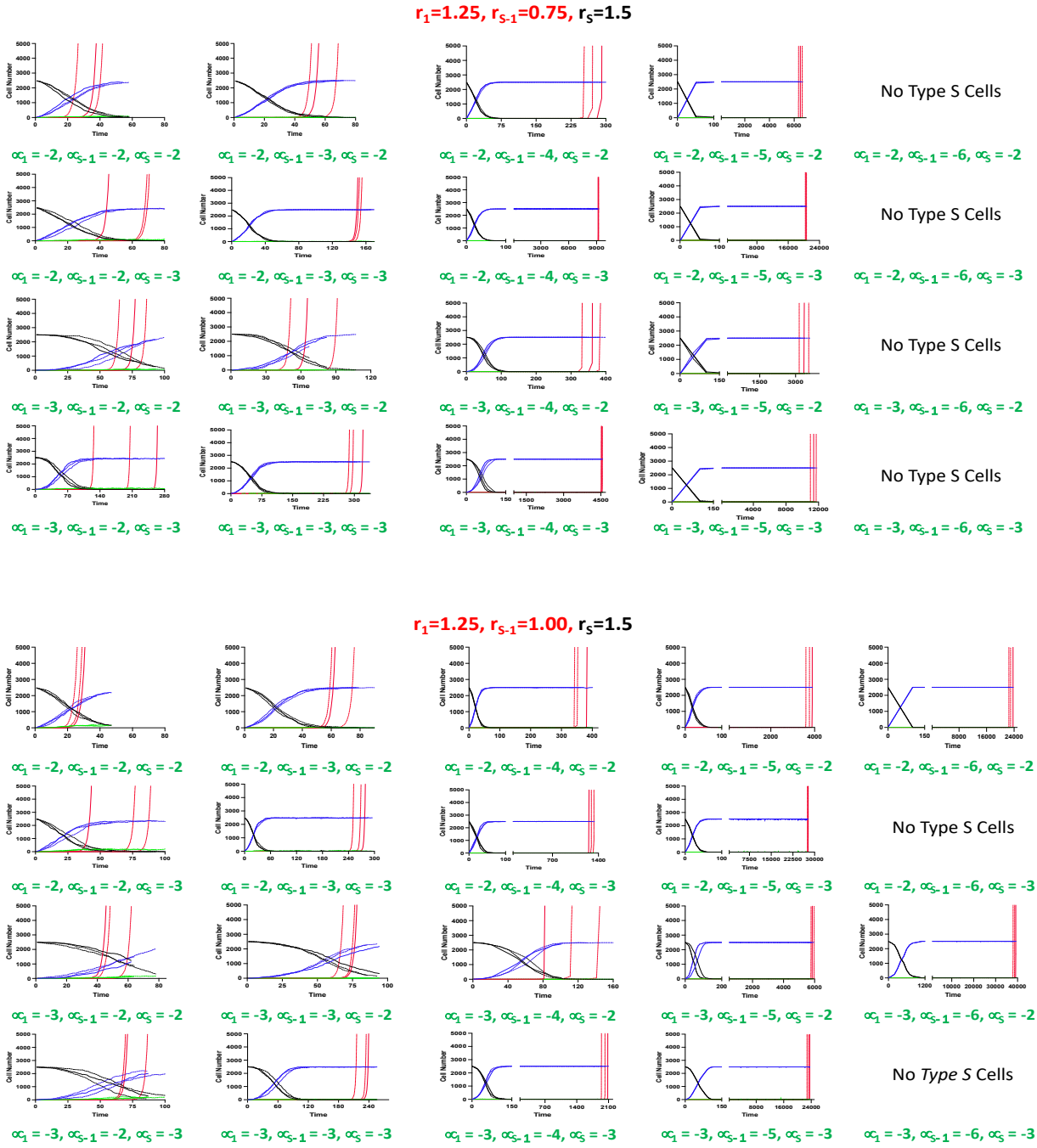


Figure 7 Panel D – Extended cancer initiation pattern with spatial structure for high r_1/low r_{S-1} and high r_1/normal r_{S-1} at different mutation rates

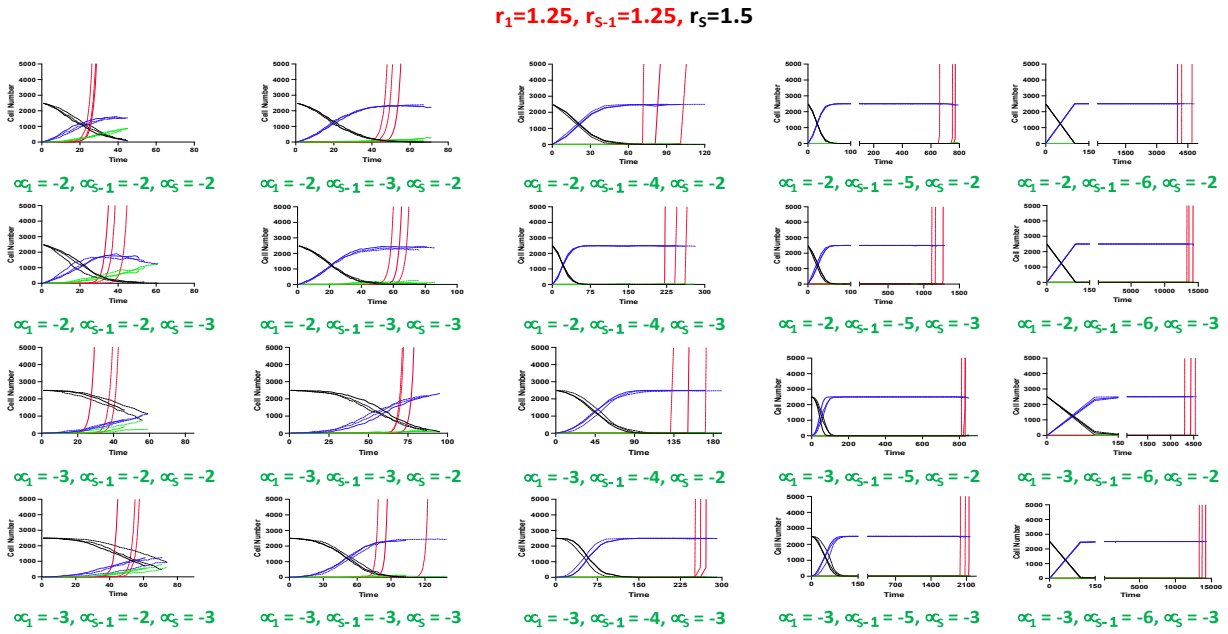


Figure 7 Panel A – Extended cancer initiation pattern with spatial structure for high r_1/ r_{S-1} at different mutation rates

Figure 8

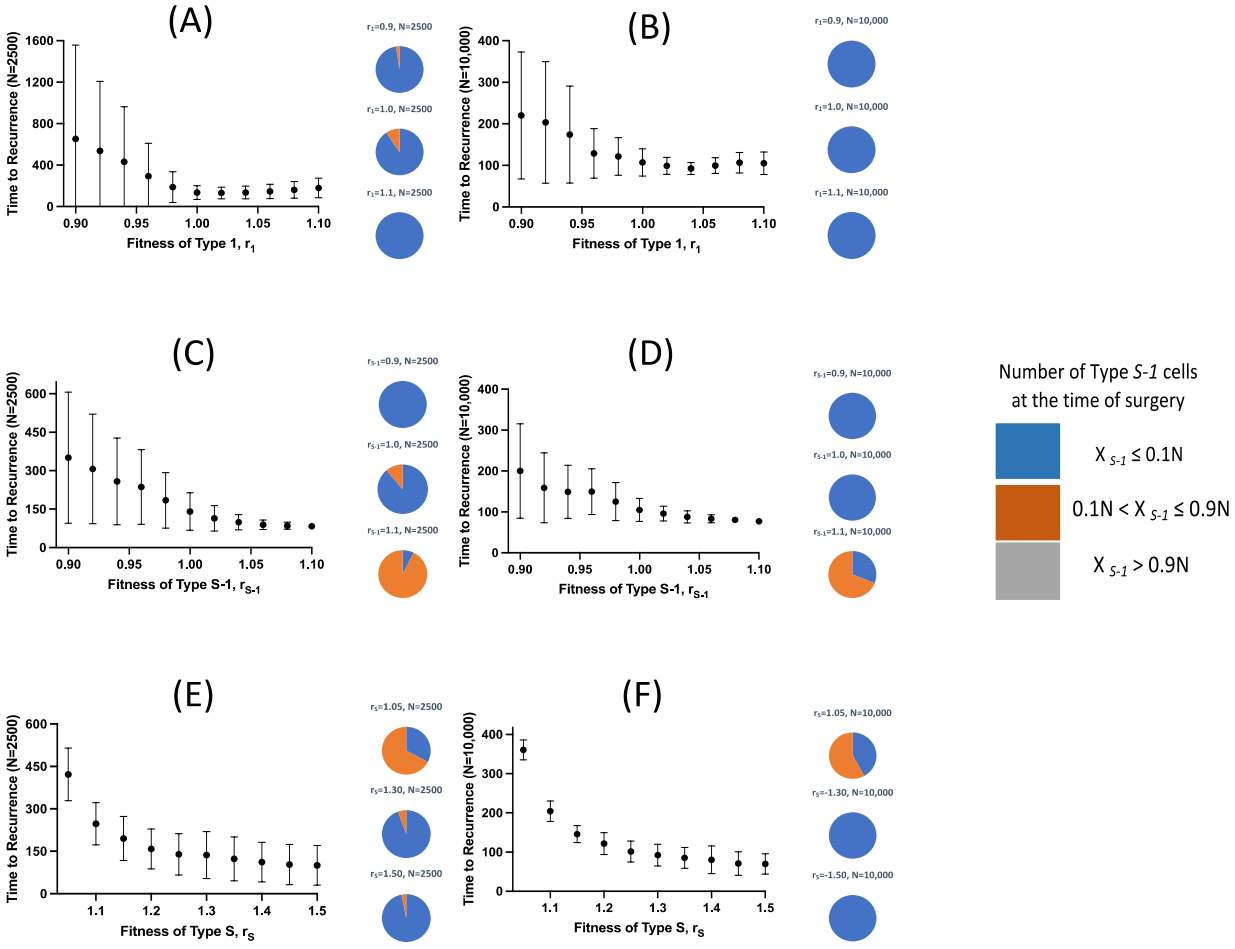


Figure 8 Panel A – Dependence of cellular fitness on recurrence time with spatial structure.

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 ($X_{S\text{-}1} \leq 0.1N$), intermediate ($0.1N < X_{S\text{-}1} \leq 0.9N$), and large ($X_{S\text{-}1} > 0.9N$) proportion of Type *S*-1 cells, respectively.

Standard parameter values used in (A–F) are $d = d_s = 1.0$, $r_0 = 1.0$, $r_1 = 1.0$, $r_{S\text{-}1} = 1.2$, $r_S = 1.0$, $\mu_1 = 0.001$, $\mu_{S\text{-}1} = 0.001$, $\mu_S = 0.001$; and $N = 2{,}500$ in (A, C and E); and $N = 10{,}000$ in (B, D and F).

(113)

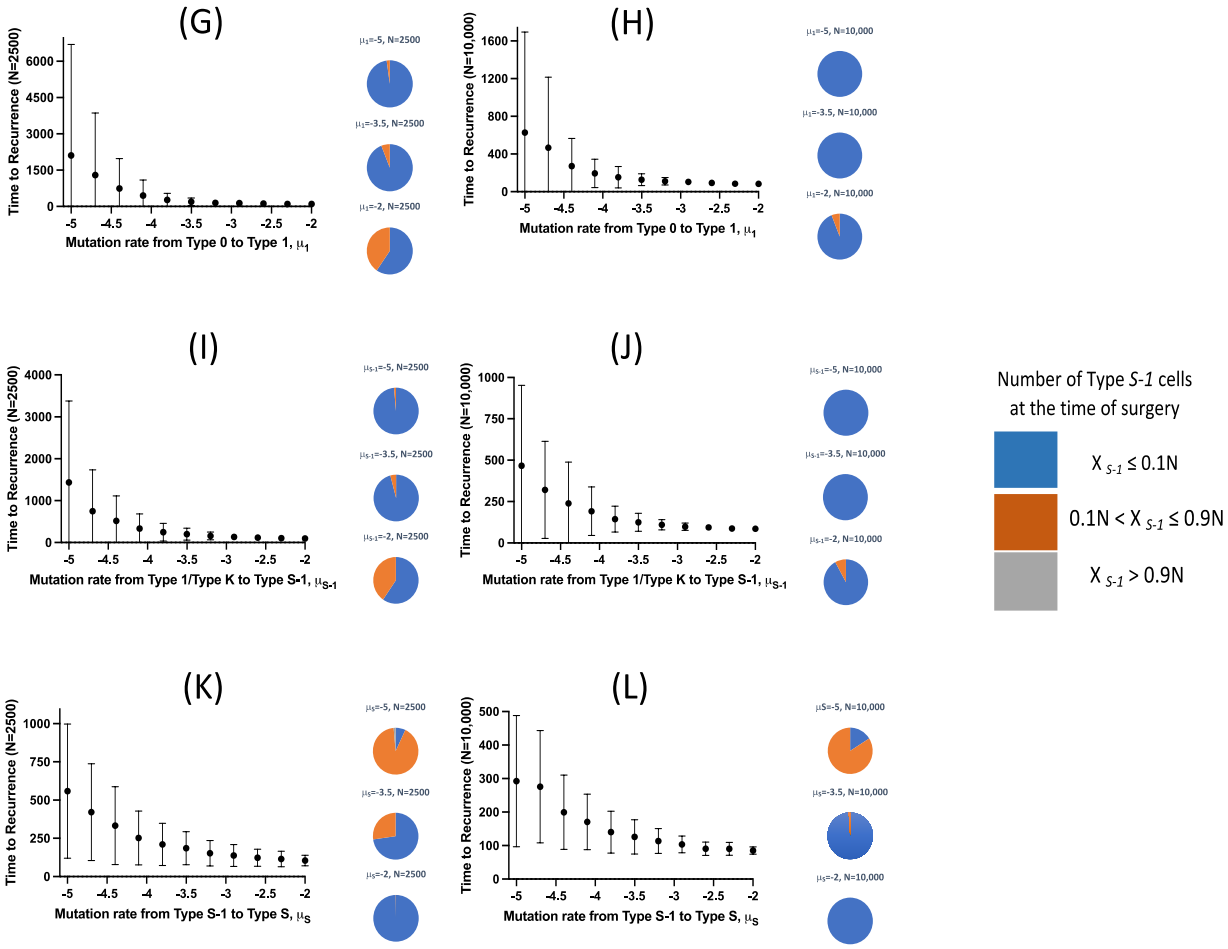


Figure 8 Panel B – Dependence of mutation rate on recurrence time with spatial structure.

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 (G–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 (G, I and K); and $N = 10{,}000$ in (H, J and L).

(113)

Figure 9

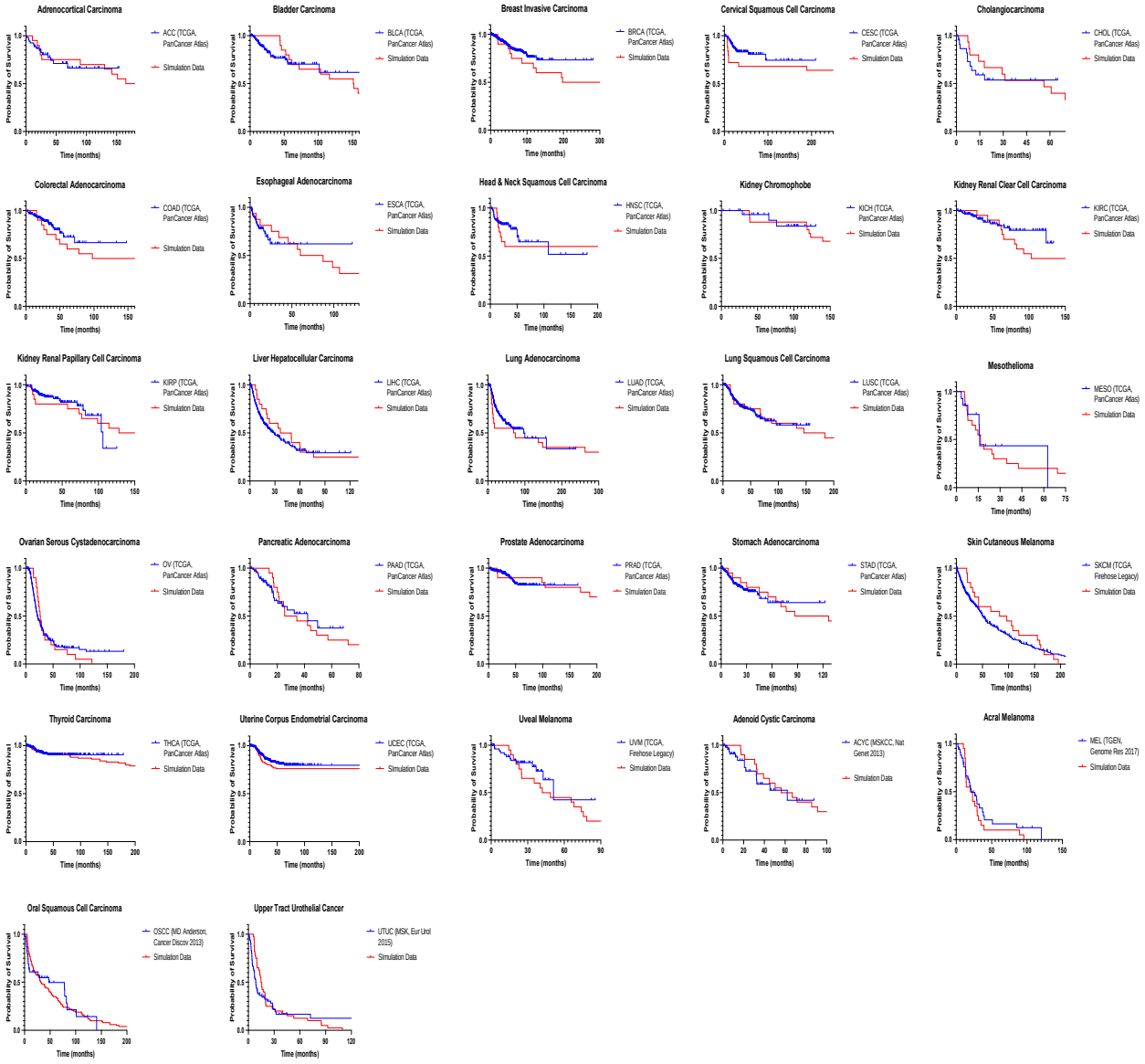


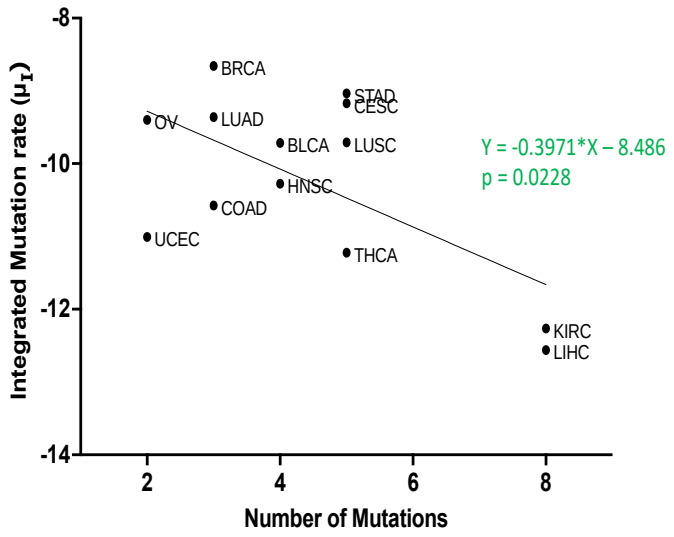
Figure 9 – Fitting of model-derived *in silico* data to published clinical data for 27 cancer types.

Thousands of stochastic runs were used to obtain parameter sets that best fit survival curves of 27 non-sarcoma cancer types. Blue curves indicate clinical data while red curves indicate simulation data survival curves. *p* values between curves are found in Table 1.

(113)

Figure 10

(A)



(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 10 – 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) Using data from (a), number of mutational hits was predicted for cancer types that were not previously reported

(113)

Figure 11

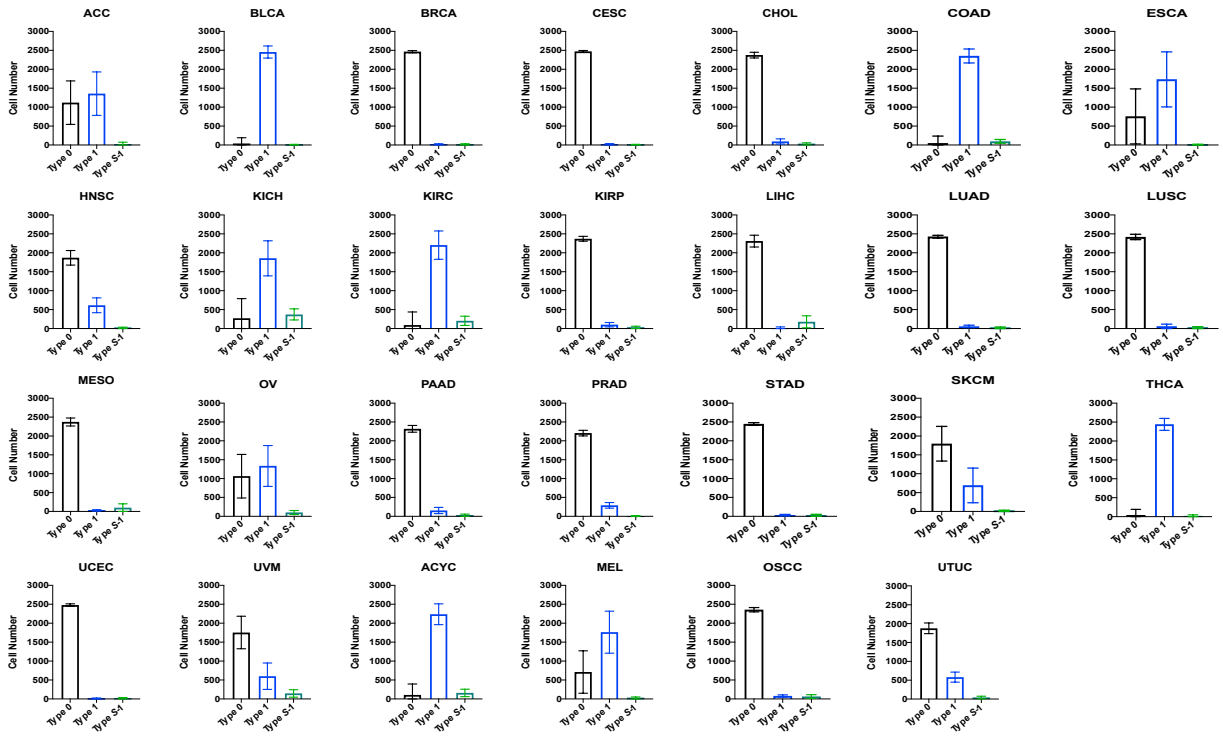
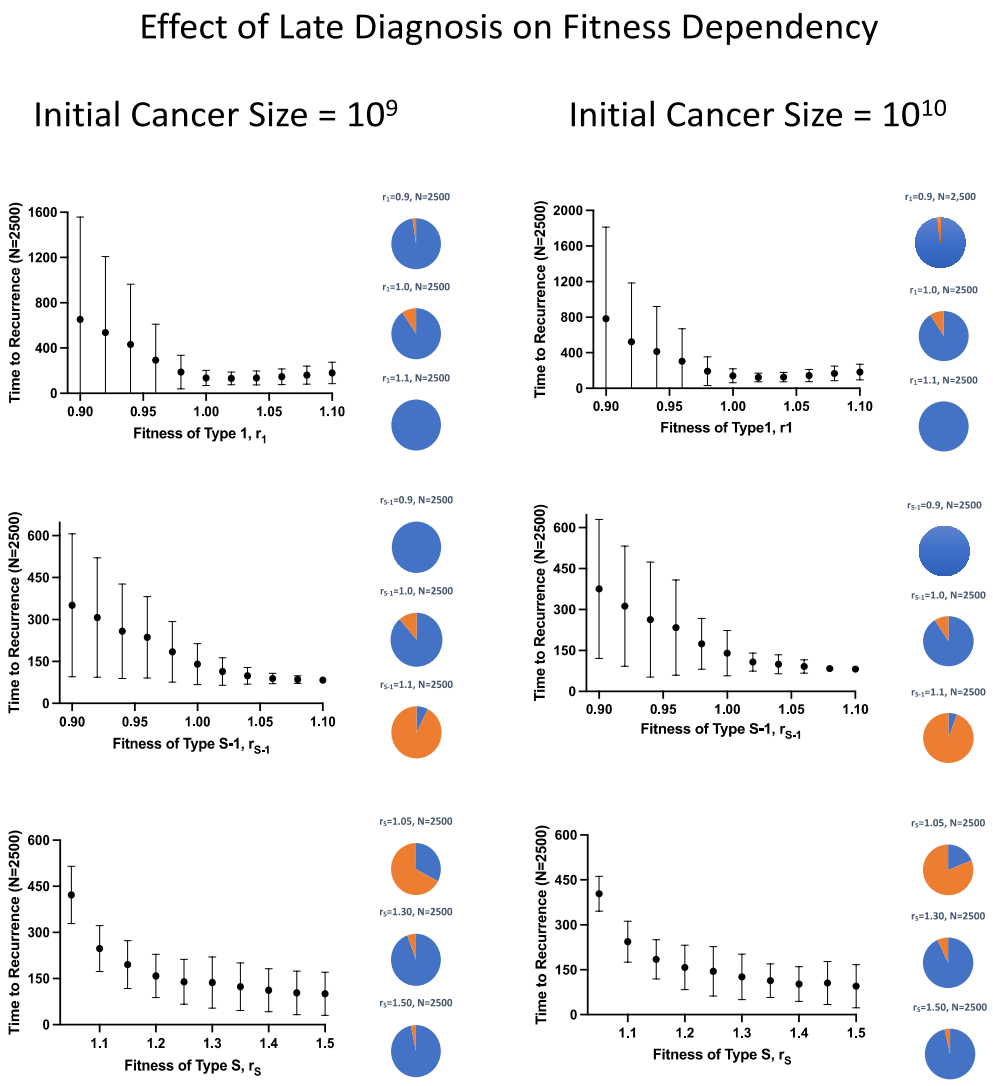


Figure 11 – Proportion of pre-malignant cells for each cancer type

60 simulations were run for each parameter set fitted from clinical data (Table 1) for each cancer type. The mean cell numbers for each cell type when malignant cells (Type *S*) reach 10^9 are shown using bars while SD are displayed as error bars. Dark-grey bars, blue bars and green bars indicate Type 0, Type 1 and Type *S*-1 cells respectively.

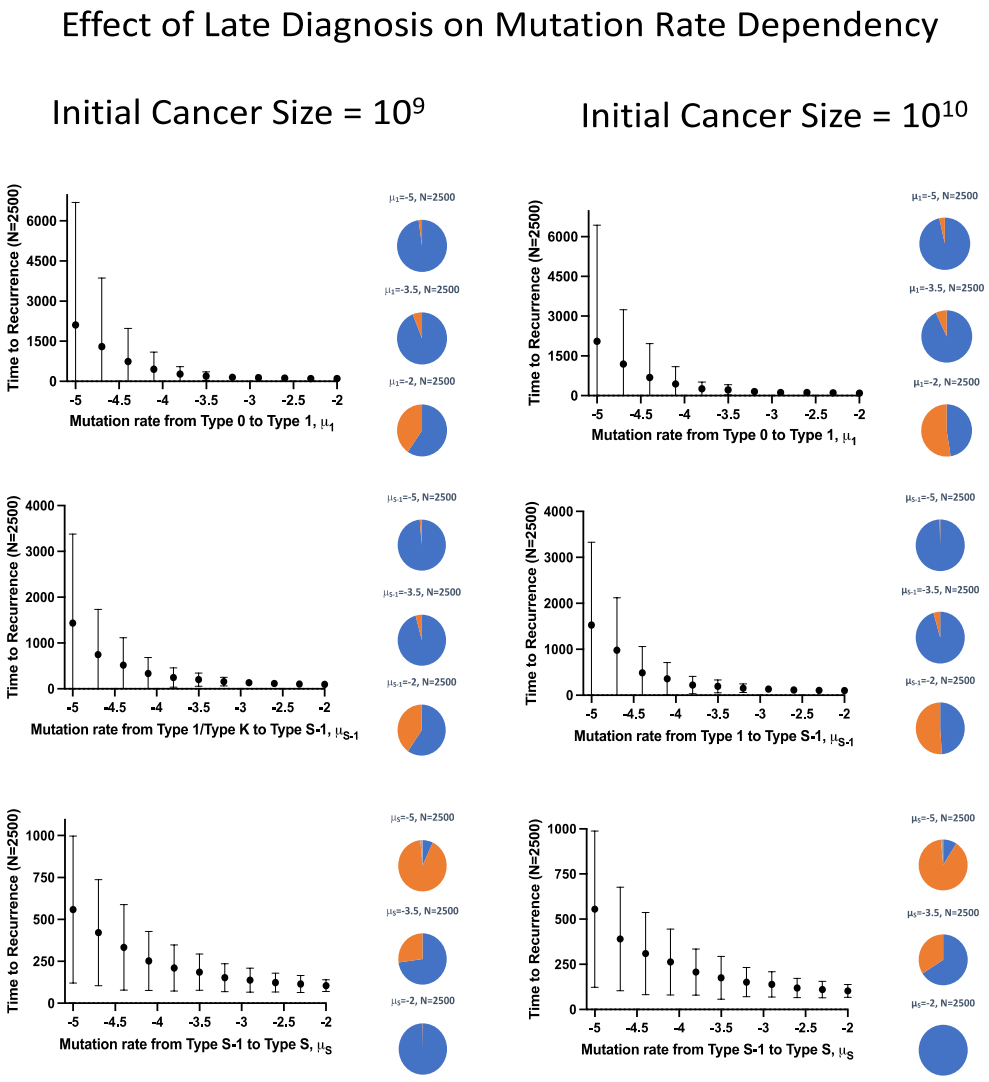
Figure 12

(a)



(a) Effect of late diagnosis on cancer recurrence time. Using the model with spatial structure, fitness parameter dependency was assessed for a scenario with late diagnosis where cancer detection was done at 10^{10} cells (right panel) rather than 10^{9} cells (left panel)

(b)



(b) Effect of late diagnosis on cancer recurrence time. Using the model with spatial structure, mutation rate parameter dependency was assessed for a scenario with late diagnosis where cancer detection was done at 10^{10} cells (right panel) rather than 10^9 cells (left panel)

Figure 13

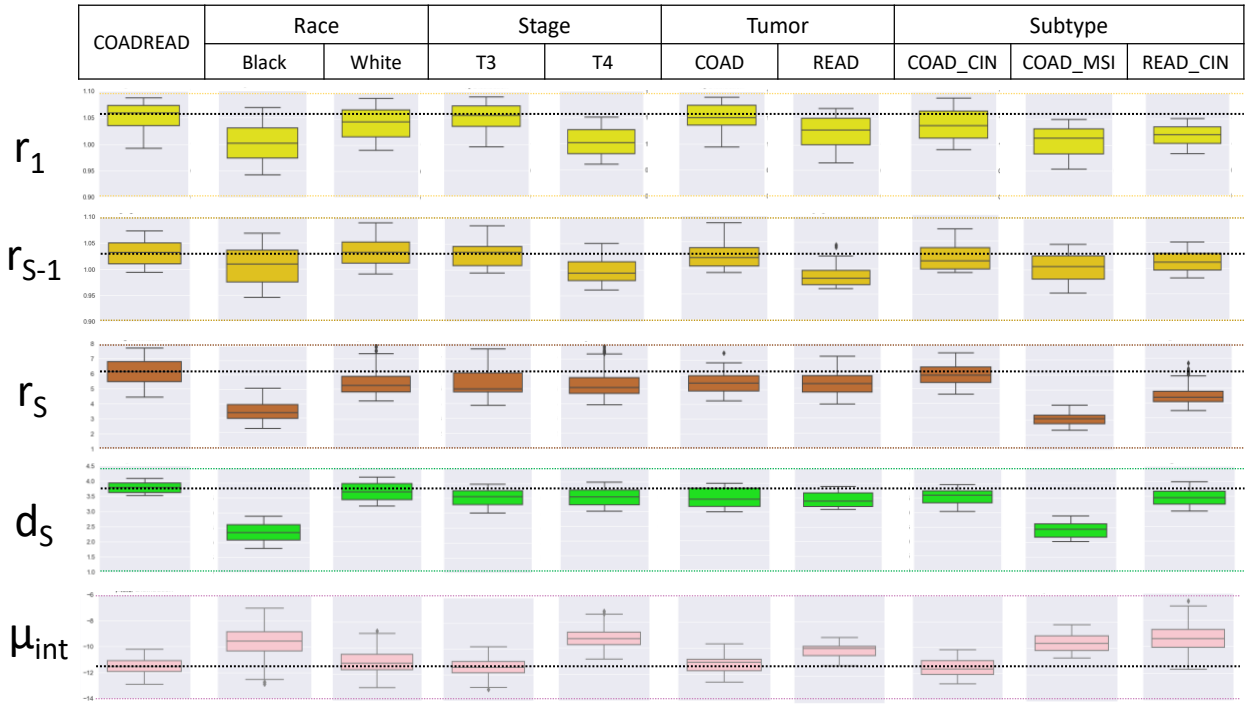


Figure 9 – Parameter distribution of colorectal cancer dataset (COADREAD) categories

Boxplots of several categories of colorectal cancer patients compared to the parent COADREAD dataset. Mutation rate, μ are in log10 values while d_3 values are in months. Dotted lines are purely for comparison / alignment purposes.

r_1 & r_{S-1} scale = 0.9 to 1.0; r_S scale = 1.0 to 8.0; d_S scale = 1.0 to 4.5; μ_{int} scale = -14 to -8

Figure 14

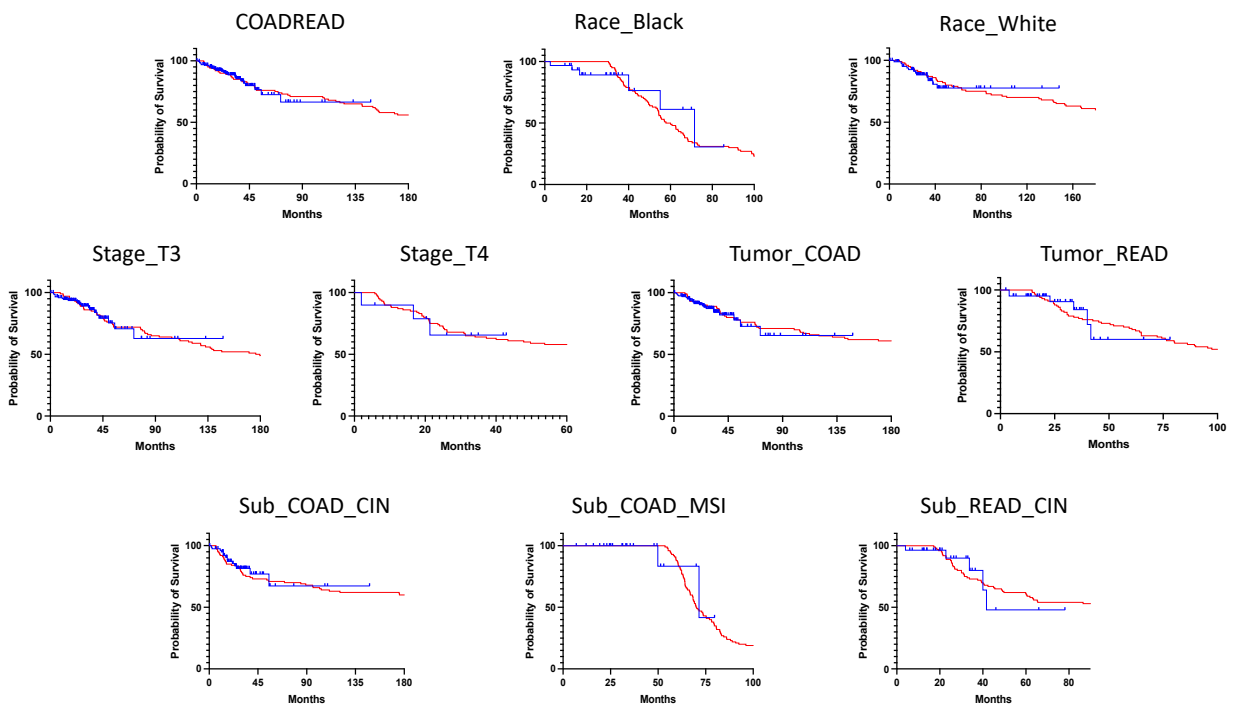


Figure 14 – Survival Curves of colorectal cancer dataset (COADREAD) categories

Survival curves (Kaplan-Meier curves) of each COADREAD category overlaid with curves obtained from computational simulations. Blue curves are for the clinical data while red curves are the simulation curves. Simulation curves are plotted based on parameter combinations in Table 4.

Figure 15

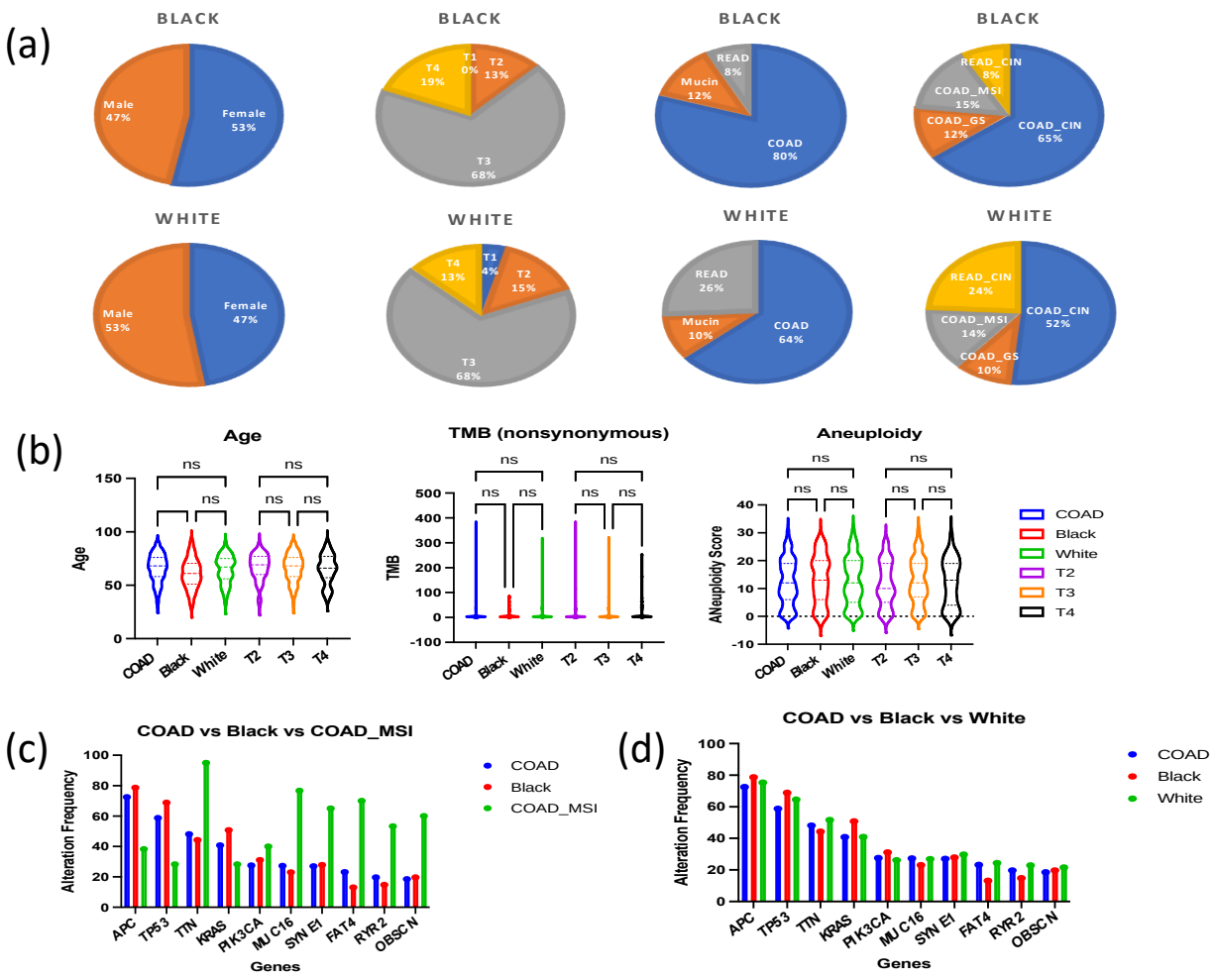


Figure 15 – Analysis of proportion bias and genetic alterations in colorectal cancer dataset categories

a. Evaluation of proportion bias in colorectal cancer datasets between blacks and whites. Blacks and Whites have vastly different profiles but have similar proportion of other categories.

b. Analysis of the differences in age, total mutation burden (TMB) and aneuploidy between different categories.

c, d. Genetic alteration frequency in top 10 mutated genes in colorectal cancer between categories with a discordant carcinogenic profile – COAD_MSI and Blacks

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: Globocan Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209-49. Epub 20210204. doi: 10.3322/caac.21660.

2. Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The Ever-Increasing Importance of Cancer as a Leading Cause of Premature Death Worldwide. *Cancer* (2021) 127(16):3029-30. Epub 20210604. doi: 10.1002/cncr.33587.

3. Global Burden of Disease Cancer C, Kocarnik JM, Compton K, Dean FE, Fu W, Gaw BL, et al. Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life Years for 29 Cancer Groups from 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *JAMA Oncol* (2022) 8(3):420-44. doi: 10.1001/jamaoncol.2021.6987.

4. Sharpless NE. Covid-19 and Cancer. *Science* (2020) 368(6497):1290. doi: 10.1126/science.abd3377.

5. Addai BW, Ngwa W. Covid-19 and Cancer in Africa. *Science* (2021) 371(6524):25-7. doi: 10.1126/science.abd1016.

6. 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.

7. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell* (2011) 144(5):646-74. doi: 10.1016/j.cell.2011.02.013.

8. Soto AM, Sonnenschein C. The Tissue Organization Field Theory of Cancer: A Testable Replacement for the Somatic Mutation Theory. *Bioessays* (2011) 33(5):332-40. doi: 10.1002/bies.201100025.

9. Davies PC, Lineweaver CH. Cancer Tumors as Metazoa 1.0: Tapping Genes of Ancient Ancestors. *Phys Biol* (2011) 8(1):015001. Epub 20110207. doi: 10.1088/1478-3975/8/1/015001.

10. Tomasetti C, Li L, Vogelstein B. Stem Cell Divisions, Somatic Mutations, Cancer Etiology, and Cancer Prevention. *Science* (2017) 355(6331):1330-4. doi: 10.1126/science.aaf9011.

11. Vyatkin AD, Otnyukov DV, Leonov SV, Belikov AV. Comprehensive Patient-Level Classification and Quantification of Driver Events in Tcga Pancanatlas Cohorts. *PLoS Genet* (2022) 18(1):e1009996. Epub 20220114. doi: 10.1371/journal.pgen.1009996.

12. Basaran D, Leitao MM, Jr. The Landmark Series: Minimally Invasive Surgery for Cervical Cancer. *Ann Surg Oncol* (2021) 28(1):204-11. Epub 20201030. doi: 10.1245/s10434-020-09265-0.

13. 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. Epub 20181017. doi: 10.3322/caac.21498.

14. 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.

15. 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. Epub 20171228. doi: 10.1038/nrclinonc.2017.194.

16. 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. Epub 20171106. doi: 10.1080/14737140.2017.1398088.

17. 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. Epub 20191220. doi: 10.1016/j.cllc.2019.12.001.

18. 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. Epub 20210308. doi: 10.1002/jso.26453.

19. Holleczek B, Stegmaier C, Radosa JC, Solomayer EF, Brenner H. Risk of Loco-Regional 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. Epub 20190530. doi: 10.1186/s12885-019-5710-5.

20. 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. Epub 20220728. doi: 10.3390/curroncol29080424.

21. Birzu C, French P, Caccese M, Cerretti G, Idbaih A, Zagonel V, et al. Recurrent Glioblastoma: From Molecular Landscape to New Treatment Perspectives. *Cancers (Basel)* (2020) 13(1). Epub 20201226. doi: 10.3390/cancers13010047.

22. 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.

23. 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. Epub 20200725. doi: 10.1016/j.asjsur.2020.07.014.

24. 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. Epub 20210210. doi: 10.1245/s10434-021-09624-5.

25. Huang A, Garraway LA, Ashworth A, Weber B. Synthetic Lethality as an Engine for Cancer Drug Target Discovery. *Nat Rev Drug Discov* (2020) 19(1):23-38. Epub 20191111. doi: 10.1038/s41573-019-0046-z.

26. 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.

27. 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. Epub 20201006. doi: 10.1158/0008-5472.CAN-20-1956.

28. Curtius K, Wright NA, Graham TA. An Evolutionary Perspective on Field Cancerization. *Nat Rev Cancer* (2018) 18(1):19-32. Epub 20171208. doi: 10.1038/nrc.2017.102.

29. 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. Epub 20200507. doi: 10.1016/j.jaad.2020.03.126.

30. Rahal Z, Sinjab A, Wistuba, II, Kadara H. Game of Clones: Battles in the Field of Carcinogenesis. *Pharmacol Ther* (2022) 237:108251. Epub 20220715. doi: 10.1016/j.pharmthera.2022.108251.

31. Heaphy CM, Griffith JK, Bisoffi M. Mammary Field Cancerization: Molecular Evidence and Clinical Importance. *Breast Cancer Res Treat* (2009) 118(2):229-39. Epub 20090815. doi: 10.1007/s10549-009-0504-0.

32. Gadaleta E, Thorn GJ, Ross-Adams H, Jones LJ, Chelala C. Field Cancerization in Breast Cancer. *J Pathol* (2022) 257(4):561-74. Epub 20220503. doi: 10.1002/path.5902.

33. 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.

34. 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.

35. 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. Epub 20111213. doi: 10.1053/j.gastro.2011.12.004.

36. 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. Epub 20190203. doi: 10.1016/j.prp.2019.02.001.

37. 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). Epub 20210113. doi: 10.3390/ijms22020746.

38. 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.

39. Dubowitz JA, Sloan EK, Riedel BJ. Implicating Anaesthesia and the Perioperative Period in Cancer Recurrence and Metastasis. *Clin Exp Metastasis* (2018) 35(4):347-58. Epub 20170911. doi: 10.1007/s10585-017-9862-x.

40. 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. Epub 20191020. doi: 10.1016/S0140-6736(19)32313-X.

41. 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. Epub 20201110. doi: 10.1016/j.ebiom.2020.103107.

42. 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. Epub 20080226. doi: 10.1016/j.mbs.2008.02.006.

43. 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. Epub 20150522. doi: 10.1371/journal.pcbi.1004272.

44. Foo J, Leder K, Ryser MD. Multifocality and Recurrence Risk: A Quantitative Model of Field Cancerization. *J Theor Biol* (2014) 355:170-84. Epub 20140413. doi: 10.1016/j.jtbi.2014.02.042.

45. 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. Epub 20161020. doi: 10.1158/0008-5472.CAN-16-1054.

46. Evans EJ, Jr., DeGregori J. Cells with Cancer-Associated Mutations Overtake Our Tissues as We Age. *Aging Cancer* (2021) 2(3):82-97. Epub 20211021. doi: 10.1002/aac2.12037.

47. Takaki M, Haeno H. Mathematical Modeling of Locoregional Recurrence Caused by Premalignant Lesions Formed before Initial Treatment. *Front Oncol* (2021) 11:743328. Epub 20211013. doi: 10.3389/fonc.2021.743328.

48. 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.

49. 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.

50. 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. Epub 20171019. doi: 10.1016/j.cell.2017.09.042.

51. 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. Epub 20190307. doi: 10.1371/journal.pcbi.1006881.

52. 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. Epub 20180612. doi: 10.1073/pnas.1803155115.

53. Wilkinson DJ. Stochastic Modelling for Quantitative Description of Heterogeneous Biological Systems. *Nat Rev Genet* (2009) 10(2):122-33. doi: 10.1038/nrg2509.

54. Simoni G, Vo HT, Priami C, Marchetti L. A Comparison of Deterministic and Stochastic Approaches for Sensitivity Analysis in Computational Systems Biology. *Brief Bioinform* (2020) 21(2):527-40. doi: 10.1093/bib/bbz014.

55. Moran P. *The Statistical Processes of Evolutionary Theory. Clarendon*. Oxford (1962).

56. Novozhilov AS, Karev GP, Koonin EV. Biological Applications of the Theory of Birth-and-Death Processes. *Brief Bioinform* (2006) 7(1):70-85. doi: 10.1093/bib/bbk006.

57. Proulx SR. The Rate of Multi-Step Evolution in Moran and Wright-Fisher Populations. *Theor Popul Biol* (2011) 80(3):197-207. Epub 20110723. doi: 10.1016/j.tpb.2011.07.003.

58. Kaveh K, McAvoy A, Nowak MA. Environmental Fitness Heterogeneity in the Moran Process. *R Soc Open Sci* (2019) 6(1):181661. Epub 20190116. doi: 10.1098/rsos.181661.

59. Athreya KB, Ney PE, Ney P. *Branching Processes*: Courier Corporation (2004).

60. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. *Cell* (2011) 146(4):633-44. doi: 10.1016/j.cell.2011.07.026.

61. Chen X, Wang Y, Feng T, Yi M, Zhang X, Zhou D. The Overshoot and Phenotypic Equilibrium in Characterizing Cancer Dynamics of Reversible Phenotypic Plasticity. *J Theor Biol* (2016) 390:40-9. Epub 20151128. doi: 10.1016/j.jtbi.2015.11.008.

62. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of Driver and Passenger Mutations During Tumor Progression. *Proc Natl Acad Sci U S A* (2010) 107(43):18545-50. Epub 20100927. doi: 10.1073/pnas.1010978107.

63. Yang G, Quan Y, Wang W, Fu Q, Wu J, Mei T, et al. Dynamic Equilibrium between Cancer Stem Cells and Non-Stem Cancer Cells in Human Sw620 and Mcf-7 Cancer Cell Populations. *Br J Cancer* (2012) 106(9):1512-9. doi: 10.1038/bjc.2012.126.

64. Jiang DQ, Wang Y, Zhou D. Phenotypic Equilibrium as Probabilistic Convergence in Multi-Phenotype Cell Population Dynamics. *PLoS One* (2017) 12(2):e0170916. Epub 20170209. doi: 10.1371/journal.pone.0170916.

65. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving Genetic Heterogeneity in Cancer. *Nat Rev Genet* (2019) 20(7):404-16. doi: 10.1038/s41576-019-0114-6.

66. Lloyd MC, Cunningham JJ, Bui MM, Gillies RJ, Brown JS, Gatenby RA. Darwinian Dynamics of Intratumoral Heterogeneity: Not Solely Random Mutations but Also Variable Environmental Selection Forces. *Cancer Res* (2016) 76(11):3136-44. Epub 20160323. doi: 10.1158/0008-5472.CAN-15-2962.

67. Pandya P, Orgaz JL, Sanz-Moreno V. Modes of Invasion During Tumour Dissemination. *Mol Oncol* (2017) 11(1):5-27. Epub 20161209. doi: 10.1002/1878-0261.12019.

68. Martens EA, Kostadinov R, Maley CC, Hallatschek O. Spatial Structure Increases the Waiting Time for Cancer. *New J Phys* (2011) 13. Epub 20111128. doi: 10.1088/1367-2630/13/11/115014.

69. West J, Schenck RO, Gatenbee C, Robertson-Tessi M, Anderson ARA. Normal Tissue Architecture Determines the Evolutionary Course of Cancer. *Nat Commun* (2021) 12(1):2060. Epub 20210406. doi: 10.1038/s41467-021-22123-1.

70. Noble R, Burri D, Le Sueur C, Lemant J, Viossat Y, Kather JN, et al. Spatial Structure Governs the Mode of Tumour Evolution. *Nat Ecol Evol* (2022) 6(2):207-17. Epub 20211223. doi: 10.1038/s41559-021-01615-9.

71. Fu X, Zhao Y, Lopez JI, Rowan A, Au L, Fendler A, et al. Spatial Patterns of Tumour Growth Impact Clonal Diversification in a Computational Model and the Tracerx Renal Study. *Nat Ecol Evol* (2022) 6(1):88-102. Epub 20211223. doi: 10.1038/s41559-021-01586-x.

72. Gaglia G, Kabraji S, Rammos D, Dai Y, Verma A, Wang S, et al. Temporal and Spatial Topography of Cell Proliferation in Cancer. *Nat Cell Biol* (2022) 24(3):316-26. Epub 20220315. doi: 10.1038/s41556-022-00860-9.

73. Strobl MAR, Gallaher J, West J, Robertson-Tessi M, Maini PK, Anderson ARA. Spatial Structure Impacts Adaptive Therapy by Shaping Intra-Tumoral Competition. *Commun Med (Lond)* (2022) 2:46. Epub 20220425. doi: 10.1038/s43856-022-00110-x.

74. Lomakin A, Svedlund J, Strell C, Gataric M, Shmatko A, Rukhovich G, et al. Spatial Genomics Maps the Structure, Nature and Evolution of Cancer Clones. *Nature* (2022) 611(7936):594-602. Epub 20221109. doi: 10.1038/s41586-022-05425-2.

75. Seferbekova Z, Lomakin A, Yates LR, Gerstung M. Spatial Biology of Cancer Evolution. *Nat Rev Genet* (2022). Epub 20221209. doi: 10.1038/s41576-022-00553-x.

76. Piraino SW, Furney SJ. Beyond the Exome: The Role of Non-Coding Somatic Mutations in Cancer. *Ann Oncol* (2016) 27(2):240-8. Epub 20151123. doi: 10.1093/annonc/mdv561.

77. 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. Epub 20190515. doi: 10.1016/j.semcancer.2019.05.010.

78. Gillespie DT. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of computational physics* (1976) 22(4):403-34.

79. 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 Discov* (2012) 2(5):401-4. doi: 10.1158/2159-8290.CD-12-0095.

80. 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. Epub 20130402. doi: 10.1126/scisignal.2004088.

81. 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. Epub 20180521. doi: 10.1007/s10147-018-1287-2.

82. 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.

83. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2022. *CA Cancer J Clin* (2022) 72(1):7-33. Epub 20220112. doi: 10.3322/caac.21708.

84. 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.

85. Zeve D, Stas E, de Sousa Casal J, Mannam P, Qi W, Yin X, et al. Robust Differentiation of Human Enteroendocrine Cells from Intestinal Stem Cells. *Nat Commun* (2022) 13(1):261. Epub 20220111. doi: 10.1038/s41467-021-27901-5.

86. Van Herck Y, Feyaerts A, Alibhai S, Papamichael D, Decoster L, Lambrechts Y, et al. Is Cancer Biology Different in Older Patients? *Lancet Healthy Longev* (2021) 2(10):e663-e77. Epub 20210929. doi: 10.1016/S2666-7568(21)00179-3.

87. Valero C, Lee M, Hoen D, Weiss K, Kelly DW, Adusumilli PS, et al. Pretreatment Neutrophil-to-Lymphocyte Ratio and Mutational Burden as Biomarkers of Tumor Response to Immune Checkpoint Inhibitors. *Nat Commun* (2021) 12(1):729. Epub 20210201. doi: 10.1038/s41467-021-20935-9.

88. Sinha N, Sinha S, Valero C, Schaffer AA, Aldape K, Litchfield K, et al. Immune Determinants of the Association between Tumor Mutational Burden and Immunotherapy Response across Cancer Types. *Cancer Res* (2022) 82(11):2076-83. doi: 10.1158/0008-5472.CAN-21-2542.

89. Ben-David U, Amon A. Context Is Everything: Aneuploidy in Cancer. *Nat Rev Genet* (2020) 21(1):44-62. Epub 20190923. doi: 10.1038/s41576-019-0171-x.

90. Choi JW, Hua TNM. Impact of Lifestyle Behaviors on Cancer Risk and Prevention. *J Lifestyle Med* (2021) 11(1):1-7. doi: 10.15280/jlm.2021.11.1.1.

91. Sarfati D, Koczwara B, Jackson C. The Impact of Comorbidity on Cancer and Its Treatment. *CA Cancer J Clin* (2016) 66(4):337-50. Epub 20160217. doi: 10.3322/caac.21342.

92. Curran T, Sun Z, Gerry B, Findlay VJ, Wallace K, Li Z, et al. Differential Immune Signatures in the Tumor Microenvironment Are Associated with Colon Cancer Racial Disparities. *Cancer Med* (2021) 10(5):1805-14. Epub 20210209. doi: 10.1002/cam4.3753.

93. 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. Epub 20220605. doi: 10.1056/NEJMoa2201445.

94. 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. Epub 20200406. doi: 10.1016/j.cell.2020.03.029.

95. 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. Epub 20211208. doi: 10.1038/s41586-021-04206-7.

96. 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. Epub 20140622. doi: 10.1038/ng.3013.

97. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal Cancer. *Nat Rev Dis Primers* (2017) 3:17048. Epub 20170727. doi: 10.1038/nrdp.2017.48.

98. Thrift AP. Global Burden and Epidemiology of Barrett Oesophagus and Oesophageal Cancer. *Nat Rev Gastroenterol Hepatol* (2021) 18(6):432-43. Epub 20210218. doi: 10.1038/s41575-021-00419-3.

99. Sharma P. Barrett Esophagus: A Review. *JAMA* (2022) 328(7):663-71. doi: 10.1001/jama.2022.13298.

100. 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. Epub 20170518. doi: 10.1089/thy.2016.0648.

101. 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. Epub 20130419. doi: 10.5732/cjc.012.10219.

102. 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. Epub 20190424. doi: 10.1245/s10434-019-07353-4.

103. 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.

104. 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. Epub 20210223. doi: 10.1007/s00464-021-08390-5.

105. 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. Epub 20151120. doi: 10.1093/ejcts/ezv398.

106. 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. Epub 20160307. doi: 10.1186/s12885-016-2229-x.

107. 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.

108. 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.

109. 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. Epub 20190307. doi: 10.1371/journal.pone.0213475.

110. 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. Epub 20140421. doi: 10.1016/j.kjms.2014.03.011.

111. 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. Epub 20160514. doi: 10.1186/s12957-016-0881-x.

112. 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. Epub 20211013. doi: 10.3389/fonc.2021.763119.

113. Abubakar SD, Takaki M, Haeno H. Computational Modeling of Locoregional Recurrence with Spatial Structure Identifies Tissue-Specific Carcinogenic Profiles. *Front Oncol* (2023) 13:1116210. Epub 20230406. doi: 10.3389/fonc.2023.1116210.

ACKNOWLEDGEMENTS

I must index my profound gratitude to my academic Supervisor, Prof Hiroshi Haeno for his guidance, patience, encouragement and microscopic scrutiny of this work. Throughout my encounter with him, he has consistently shouldered my troubles without any complaints. His dexterity in scrutinizing hundreds of lines of C++ code to correct my rookie errors are highly appreciated. The contribution of Dr Mitsuaki Takaki for the backbone of the codes is also acknowledged.

My journey in Japan would not be complete if I do not appreciate the contributions of Professor Masato Kubo. I thank him for his accommodation and guidance during the beginning of my journey here. My reflective thanks also creep up to my esteemed teachers – Professor Daisuke Kitamura, Professor Tomokatsu Ikawa, Prof Naoko Nakano and Dr Shunsuke Kon for their academic and supportive advice. I will also like to give a shout-out to Yasuyo Harada, Dr Mayumi Hirakawa, Dr Kasumi Dendo, Dr Kei Haniuda, Dr Kosuke Miyauchi, Dr Jumana Khalil and all present and former members of Kubo Lab for a good rapport and research experience.

I must also give special thanks to the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) for granting me this scholarship after several competitive rounds of selection. Indeed, the financial support in both tuition and living expenses have enabled me to smoothly undertake my studies and research. Prof Hiroshi Haeno also generously assisted after MEXT funding elapsed. I want to also acknowledge the dedication and assistance of the student support division in both Chiba University, Tokyo University of Science and the Research Institute for Biomedical Sciences for their logistic handling and administrative capabilities which simplified my stay here.

Everything could account to nothing if I don’t register my gratitude to the love, care and understanding of my family and siblings especially my loving mother – Hajiya Baraka Dahiru Abubakar, my dearest Zahraa, Ra’ees and Asma’u.

For the friends I made here in Japan who are too numerous to mention, I salute you all for your kindness, understanding and care. All the positive experience will remain close to me.

I will like to extend my sincere appreciation to the management of Ahmadu Bello University, Zaria, Nigeria for offering me a position in the College of Medical Sciences under the Department of Medical laboratory Science. The glowing recommendations from Prof Shehu Abubakar Akuyam and Dr Abdurrahman Ahmad El Fulaty were instrumental in this successful recruitment and I extend my sincere thanks to them. The kind recognition and commendations from Professor Mamoru Kato of the National Cancer Center is also sincerely appreciated. Professor Mamoru Kato and Professor Atsushi Ochiai made meaningful contributions that were instrumental to the acceptance of the research paper for publication and also to the enrichment of this dissertation.

Finally, I dedicate this work to my late father, Sheikh Dahiru Abubakar. I am what I am because of him, his support and his trust. He has lived a good life and impacted many people positively in the time he spent on this earth. May his perfect soul dwell in the highest echelons of paradise. He will be sorely missed.