TI 2020-012/II
Tinbergen Institute Discussion Paper

Preventative Cancer Treatments Through Optimizing Tissue Structure

Christopher Graser[1]

[1] University of Amsterdam, Tinbergen Institute

Discussion Paper Submission

Preventative Cancer Treatments Through Optimizing Tissue Structure

Christopher Graser

February 10, 2020

Abstract

The likelihood of cancer emergence is highly dependent on the underlying tissue structure. This article gives evolutionary explanations for why natural selection fails to select for tissue structures that would minimize the likelihood of cancer. In a second step, a mathematical framework is proposed, within which the risk of cancer emergence can be expressed and calculated dependent on a given tissue structure. This can be used to identify optimal structures and strategies for improvement. Lastly, the article explores both, ways to identify target areas for such intervention, as well as avenues towards developing treatment options.

Introduction

Models of Evolutionary Game Theory have been important in understanding the theoretical origins and dynamics of cancer emergence, and especially, in showing how the structure of different tissues affects these dynamics. This makes it possible to analyze the likelihood of cancer emergence, by expressing the probability that cancerous mutants fixate, as a function of cell reproduction dynamics in different tissues.[1]

The purpose of this paper and the framework developed in it is to propose and conceptualize an approach to preventative cancer treatment by changing tissue structure. Loosely speaking, the idea is that one would take a tissue sample of a healthy patient. This sample would be analyzed to identify the proportion of cells that carry certain mutations which could lead to cancer. This information would then be used to design an appropriate preventative treatment, which modifies the structure of the sampled tissue region and thereby decreases the likelihood of carcinogenesis within the tissue.[2]

To this aim, the framework should, firstly, formalize the evolution of tissue structures, so as to model and categorize ways in which natural selection fails to develop the tissue structures that would be optimal to prevent cancer. This will justify why there is room for improving tissue structure, further than natural selection would achieve, and it could help in identifying specific tissue regions in which treatments could be most effective. The outlined misalignments of natural selection and the minimization of cancer risk add to the equally relevant concern that natural selection might not have had sufficient time to adapt tissue structures to changing circumstances, such as the rapid increases of life expectancy, or increased exposure to radiation.

Secondly, the framework should enable one to find the optimal tissue structures for cancer prevention, so as to serve as a benchmark for treatments. Similarly, restrictions on the possible ways of tissue restructuring can be introduced to find the best achievable structure.

[1] See for example Michor et al. (2004), Hindersin et al. (2016), Altrock et al. (2015), Nowak et al. (2003) and Frank et al. (2003)

[2] Werner et al. (2016) propose a method of personalized cancer treatment, based on knowledge about the patient's *treatment trajectories*. The framework presented in this paper could be seen as the ex-ante analogue.

Obstacles to Natural Selection of Optimal Tissue Structure

There are four distinct reasons for discrepancies between tissue structures that evolve naturally and those that would minimize the risk of cancer (these will henceforth be referred to as Problems A-D). Problem (A) is *path dependence*. In many cell-types, cells need to accumulate multiple genetic mutations to become cancerous. The sequence in which these mutations are acquired need not be deterministic. Moreover, cells carrying different subsets of the required mutations can behave differently, implying that the optimal tissue structure would depend on the order in which different mutations spread through the cell population. Natural selection would favour the cell structure that minimizes cancer likelihood, given a probability distribution over different orders in which mutations spread. However, this cell structure might not be optimal, given a certain realization of some stage of the random process, i.e. once some mutation has already spread in the population.

Problem (B) is termed *age dependence*. The selection pressures to minimize cancer risk are not constant over the lifetime of organisms. Most notably, developing cancer in post-reproductive age would not be as strongly selected against as developing cancer before the end of reproductive age. Hence, if there is a trade-off between choosing a tissue structure that optimizes some other fitness enhancing trait and choosing one that would minimize cancer risk, natural selection would select some intermediate structure which is not the most cancer-resistant one. This discrepancy would suggest that there is the possibility for preventative treatments that alter this tissue structure, before these cancer-types can develop.

The third problem (C), *risk tolerance*, simply acknowledges that natural selection favours genotypes that produce the fittest phenotypes in expectation, potentially tolerating some cancer risk, if it is compensated by sufficiently large fitness increases in the cancer-free organisms. This, again, suggests that natural selection would fail to select the structure that minimizes cancer risk, which risk averse patients would prefer. The most simple example of this would be the trade off between a large body, which might confer some physical selective advantage, but also increases the amount of cells in which cancerous mutations could occur.

Problem (D), *evolvability*, refers to a path dependence in the evolutionary emergence of cell structures. Moving from a given structure to a fitness-increasing alteration of that structure could require multiple mutations, which by themselves each produce structures that are dominated by the current structure. This would imply that the tissue structure is 'stuck' in a local, but not global optimum of the fitness landscape.

Model

General Model

Consider a body modeled as a system of cells. This system of cells can be seen as a weighted digraph $G = (V, E, \partial, \Gamma)$, where V is the set of vertices (with $|V|$ being the size of the body), E is the set of edges, $\partial : E \mapsto V^2$ is a mapping from edges to ordered pairs of vertices, and $\Gamma : E \times S \mapsto (0, 1]$ assigns a weight to each

edge. An edge $e \in E$ indicates that there is a nonzero probability $\Gamma(e, s)$ that an offspring of the cell at the first vertex replaces the cell at the second vertex. Hence, a component of G is a closed system of cell reproduction, which corresponds to the notion of *compartments* in Michor et al. (2004). (See Appendix)

Each vertex $v \in V$ is inhabited by a cell of type $\theta \in \Theta$. The set of types Θ shall distinguish cells based on the cancer-relevant mutations that they carry, and the amount and respective likelihood of mutations they require to become cancerous. One can thus say, that a digraph G is in a state $s \in S_G$, where S_G contains all mappings $s : V \mapsto \Theta$. As the above notation suggests, the replacement probability $\Gamma(e, s)$ can also depend on the state s.

One can, furthermore, define mutation probabilities, which indicate the likelihood of cells transitioning between different types $\theta_1, \theta_2 \in \Theta$.

Adopting the above framework has the advantage that carcinogenesis can be modeled as a finite state Markov process. Thereby, formulating and computing fixation probabilities of cancerous cells for any pair (G, s) becomes relatively simple, which will be demonstrated in a later section.

As the focus of this framework lies on carcinogenesis and possible cancer prevention methods, it can be useful in applications to specify ∂ such that it reflects only the makeup of healthy tissue. Cancer cells have a tendency to spread much wider than their non-cancerous counterparts, implying that the number of distinct components of G could be much lower if ∂ also reflects cancerous states. Higher compartmentalization (i.e. G having more components) allows one to individually consider smaller units of the graph, which reduces the computational complexity of the analysis.[3]

Failure to evolve to optimal structure

Let $f^{nat} : \mathbf{G} \times S_G \times D \times T \mapsto \mathbb{R}_+$ be the fitness function, which assigns an expected fitness value to every quadruplet of $G \in \mathbf{G}$, $s \in S_G$, $d \in D$ and $t \in T$. $\mathbf{G}$ simply denotes the set of all weighted digraphs. In practice, one can restrict this set to reflect some conditions of biological feasibility. D shall represent the expected aggregate amount of damage caused by potentially emerging cancers across the body. T denotes time. It makes sense to regard T as the lifetime of an individual. The expected fitness value should correspond to the expected number of future offspring that are produced over some time interval by individuals who are currently specified by a pair (G, s). Note that time periods after reproductive age can, even if only slightly, also have an effect on f^{nat}, namely, if the individual can affect the reproductive success of related individuals. In the following, f^{nat} is assumed to be Riemann-integrable. Moreover, let $d(G, s, t)$ be dependent on the weighted digraph G, on the initial value s_0 of the state variable s, and on time. This captures the notion that the probabilities that cancerous cells fixate by some time t depend on the structure of reproduction within cell compartments and on the number of already mutated cells which is captured in s. Higher damage values correspond to lower fitness: $\frac{\partial f^{nat}}{\partial d} \leq 0$

[3] see Appendix for more detailed discussion of the assumption of compartmentality

Natural selection would, hence, solve the maximization problem

$$\max_{G \in \mathbf{G}} \int_T f^{nat}(G, s_0, d(G, s_0, t), t)dt \tag{1}$$

Note that the graph-dependency of d can induce a trade-off between minimizing the damage parameter and optimizing the other fitness effects of the graph structure. Problem (A) can already be formalized with this framework: Over time, the random state variable s may have diverged from its expected trajectory. Hence, Problem (A) would be a discrepancy between the graph $G^{(1)}$ that solves Equation (1) and $G^{(2)}$ that solves

$$\max_{G \in \mathbf{G}} \int_{T_A} f^{nat}(G, s_A, d(G, s_A, t), t)dt \tag{2}$$

, where $T_A \subset T$ is a subinterval of T starting at some later point in time, and s_A is the state of the cell system at this point in time.

To express Problem (B) and (C), a further function needs to be defined. $f^u : \mathbf{G} \times S_G \times D \times T \mapsto \mathbb{R}_+$ acts as the counterpart to the natural selection fitness function, in that it assigns a normative valuation to all (G, s, d, t)-quadruplets. This function can be seen as a utility function, and it will also be assumed to be Riemann-integrable.
Problems (B) and (C) would thus be discrepancies between $G^{(1)}$ and $G^{(3)}$, which solves

$$\max_{G \in \mathbf{G}} \int_T f^u(G, s_0, d(G, s_0, t), t)dt \tag{3}$$

For Problem (B), this discrepancy would arise from f^u assigning higher relative importance to later time periods than f^{nat} does. Formally, let $T_B \subset T$ be some subinterval of T, starting at a later point in time, then Problem (B) is described by the following condition:

$$\frac{\int_{T_B} f^u(G, s_0, d(G, s_0, t), t)dt}{\int_T f^u(G, s_0, d(G, s_0, t), t)dt} > \frac{\int_{T_B} f^{nat}(G, s_0, d(G, s_0, t), t)dt}{\int_T f^{nat}(G, s_0, d(G, s_0, t), t)dt} \tag{4}$$

Problem (C) is simply a violation of the assumption that f^u would be linear in the statistical expectation of f^u for different realizations of s.

$$\mathbb{E}_t[f^u(s_t)] \neq f^u(\mathbb{E}_t[s_t]) \tag{5}$$

Problem (D) can be modeled as a restriction $\hat{\mathbf{G}}$ in the set of attainable graphs $\mathbf{G}$, so that (1) would be maximized over $\hat{\mathbf{G}}$. In the example of local but not global optimality outlined in the introduction, $\hat{\mathbf{G}}$ could be of the form

$$\hat{\mathbf{G}} = \{G \in \mathbf{G} | G = \lambda^v(G^*), v \in \mathbb{N}, f^{nat}(\lambda^w(G^*)) \geq f^{nat}(\lambda^{w-1}(G^*)), 1 \leq w \leq v\}$$

where $\lambda(G^*)$ is some mutation-sequence operator on an initial graph G^*. The condition thus states that there must exist some mutation sequence from G^* to G so that all intermediate mutations are not selected against, when paired against their predecessor. Alternatively, $\mathbf{G}$ could also be a restriction on the complexity of the graph, which will be discussed in more detail for recursively coded structures.

Feasibility of Treatment and Promises of Fractal Structures

Many forms of tissue modifications might seem out of reach given today's technology, as they would require far-reaching genetic modifications to the stem cells that govern the cell division patterns in a compartment. Even if one were able to genetically engineer a stem cell which is functionally equivalent to the stem cells in the compartment, and differs only in what cell division structure it induces, transplanting this cell into a compartment would still be a challenge.

However, there is good reason to believe that inducing beneficial changes to compartment structure can be achieved by simpler means. In many models of population dynamics, fixation probabilities change dependent on the size of the population (or compartment). Hence, merging or splitting up compartments, without changing their structure, might be sufficient to reduce cancer risk in a given cluster of compartments.

The channel, through which these changes of tissue division can be achieved, could well be extracellular, i.e. by modifying the availability and spatial distribution of growth factors and other signaling circuits that affect differentiation patterns or apoptosis. For instance, compartmentalization might be naturally achieved through spatially limiting the area into which a certain growth factor is emitted. In that case, one could artificially merge compartments by supplying additional growth factors (or distributing them more evenly) to cover uncovered space, or, conversely, split up compartments artificially by extracting some of the growth factor.[4]

Similarly, if compartmentalization was achieved through physical boundaries between compartments, these could be artificially extracted or added to alter compartment sizes. Alternatively, compartmentalization could predominately be a result of cell division and differentiation patterns, for example, if there was a cluster of stem cells which each, dependent on extracellular signaling, either divide into two differentiated cells, or into one differentiated and one stem cell. Assume that the number of compartments is equal to the number of stem cells (each stem cell gives rise to a cluster(compartment) of differentiated cells). Then, reducing compartment sizes could be achieved through changing the differentiation-signal environment in a way that maintains more stem cells, which creates more compartments, and further through keeping the amount of growth factors for the cluster of compartments constant, so that each individual compartment will be smaller.[5]

Another promising route would be to make use of *fractal structures* in the body. The comparatively low amount of information stored in the genetic code, relative to the vast complexity of the resulting phenotype, indicates that there must be some usage of recursive (fractal) patterns in the genetic code. Hence, larger structures would emerge as fractal sequences of substructures that are scaled copies of one

[4] Adler et al. (2018) lay out how constant ratios of different cell types are maintained through endocytosis. It seems conceivable that a similar endocytotic processes results in a spatial containment of growth factors, which induces compartmentalization.

[5] For general notes on how the *extracellular matrix* affects carcinogenesis, see Pickup et al. (2014).

another.[6]

Applied to the framework at hand, this process could result in disjoint compartments which differ only in their size, but are identical in their cell reproduction dynamics. Stem cells in these compartments would thus be ideal candidates to identify genetic loci that determine compartment size. Analogously, analyzing the extracellular environment that gives rise to this structure would be informative about how compartment sizes can be artificially modified. The study of fractally organized structures in the body would thus be an ideal avenue to advance the understanding necessary to develop future treatment options.

Moreover, the process of cell division within the compartments of a fractal structure could itself be such that there would be an optimal compartment size. Then, by the nature of fractally repeating substructures of similar makeup, but different size, one would have tissue compartments with substantial variance around the optimal compartment size. This would fall under Problem (D), as a restriction on the complexity of G (or the required information to construct it).
Again, this suggests that modifying the size of some of the compartments could reduce fixation probabilities of possible mutations. In fractal structures it would seem more plausible that this could be achieved by stem cell transplants. The stem cells inducing different compartment sizes already exist and don't need to be artificially modified. Moreover, it seems plausible that the high genetic similarity between stem cells that are extracted and re-implanted in similar structures would increase the likelihood of successful transplantation. Similarly, on an extracellular molecular level, the concentrations that give rise to structures of different scale can simply be copied from the corresponding part of the already existing fractal structure. Indeed, this hypothesis of fractal structures being c.p. more vulnerable to carcinogenesis could be investigated empirically, by comparing the per-cell rates of cancer emergence in compartments of different size within a fractal structure.

Simple Application: Two Mutations

To illustrate that even simple tissue modifications like changes to the compartment size can drastically affect the likelihood of cancer emergence, consider the most simple model of cell reproduction: the Moran Process. In this birth-death process, in each iteration, one cell per compartment is picked to reproduce, with a probability proportional to its fitness r (which can be interpreted as its growth rate). The offspring then replaces another cell in the compartment, which is picked uniformly at random from the set of all cells in the compartment, including the parent cell.

This specification is adapted from Michor et al. (2004), who consider the case of a single mutable gene with two possible expressions. They show that the probability $P(t)$ that a compartment which is initially occupied by unmutated cells will be fully taken over by mutated cells at some time t is an increasing function of the compartment size N, if the relative fitness of the mutants is greater than that of the unmutated cells. Conversely, it is a decreasing function of N, if their relative fitness

[6] Several studies have found that genomic data can be well explained by reference to fractal paradigms, see for instance: Ghorbani et al. (2018), Moreno et al. (2011) and Petoukhov et al. (2018). For a meta-analysis finding scale-independent variability see Sapolsky and Balt (1996). For applications of fractal structure to cancer see Bizzarri et al. (2011)

is smaller.

To embed this model into the framework of the present paper, imagine that a cluster of identical compartments needs to fulfil some purpose for which it needs to consist of C cells, where $C \in \mathbb{N}$ is some constant. Assume for simplicity that cancerous cells don't metastasize or grow beyond their own compartment before they have reached fixation in their compartment. Hence, the Moran Processes in the different compartments can be viewed as independent. The task is to find the compartment size N, which maximizes $(1 - P(t))^{\frac{C}{N}}$, i.e. the probability that no cancer has emerged in any of the compartments by some time t.

The Markov process can then be set up as follows. The cells in each compartment can differ binarily on η genes, hence $\Theta = \{0,1\}^\eta$. In the Moran process, each type θ has a fitness r_θ. The fitness of the unmutated cells $r_{\{0\}^\eta} = 1$ is normalized, so that r can be interpreted as the fitness relative to unmutated cells.

A compartment $G' \subset G$ is given by the complete digraph of N nodes, and the state-space $S_{G'}$ of a compartment is characterized by all $s' : \{1, \dots, N\} \mapsto \Theta$. If reproduction occurs without mutations, the replacement probabilities $\Gamma(e, s')$ for an edge e connecting two vertices $\partial_1(e) \in V$ and $\partial_2(e) \in V$ in the compartment would be given by

$$\Gamma(e, s') = \frac{r_{s'}(\partial_1(e))}{\sum_{v \in V} r_{s'(v)}} \cdot \frac{1}{N} \tag{6}$$

The first factor denotes the fitness of the individual at the first vertex of e as a proportion of the sum of all cells' fitnesses in the compartment. The second factor reflects the uniform probability of deletion in the Moran process.

Note that with this setup, all pairs $(G', S_{G'})$, for which for all types $\theta \in \Theta$, $S_{G'}$ assigns the same number of cells to that type, are isomorphic to one another. Therefore, it is sufficient to only consider the set of isomorphism classes $S_{G'}$ as the state space. Elements of this set are simply the different ways of allocating N cells to $|\Theta|$ types. This reduces the cardinality of the considered state space drastically.

Next, to introduce mutations to the framework, let $\frac{u}{1-u}$ be the rate at which mutations occur, relative to the speed of the Moran process. Hence, in each iteration of the Markov Process, a birth-death event occurs with probability $1 - u$ and a mutation occurs with the remaining probability u. Mutation likelihoods shall be the same for all gene loci, and be small, so that $u^q \approx 0$ for $q \geq 2$, implying that cells mutate at most on one gene locus per iteration. (see Figure 1)

With this information, the Markov transition matrix can be constructed. As a small amendment to the setup of Michor et al. (2004), the state of cancer reaching fixation in the compartment (all cells are of type $\{1\}^\eta$) is assumed to be an absorbing state. This reflects the notion that cancerous cells metastasize or grow beyond their initial compartment from the time of fixation onwards. It further has the advantage that $P(t)$ shows the probability of cancer emerging at any time up to t, not only the probability of cancers arising and persisting until or throughout t. (For more notes on methodology, see Appendix.)

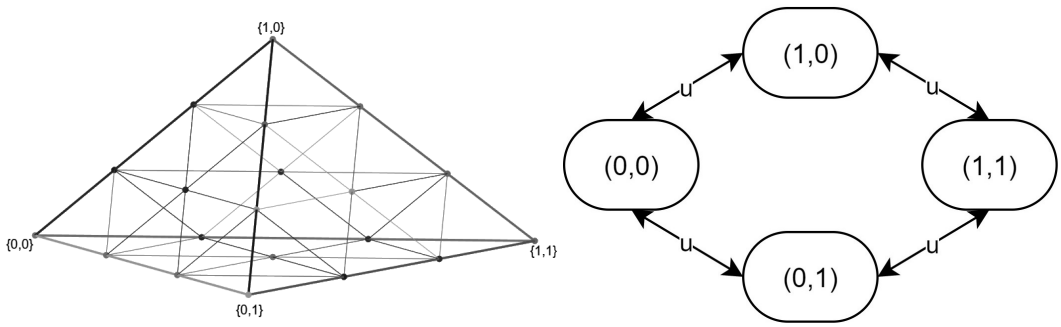

Figure 1: Illustration of the state space for a compartment of size $N = 3$ (left), and of the type space of an individual cell (right). $\{i, j\} \in S_G$ refers to a state where all cells are of type $(i, j) \in \Theta$.

The medical action-set considered is all treatments which result in some division of the N cells into compartments of equal size. Hence, one must consider all digraphs G, which are unions of identical complete components with their weighting function given by Equation (6).

Starting with the case of a cancer caused by a single mutation $\Theta = \{0, 1\}$, and an initially unmutated cluster, it turns out that $(1 - P(t))^{\frac{C}{N}}$ is increasing in N for any $t \geq 0$, independently of the relative fitnesses of the mutated and unmutated cells. This is consistent with the results of Michor et al. (2004). (Shown in the Appendix)

However, already when moving to a setting where two mutations are required for cancerous cells to emerge, the model becomes more intricate. To see this, let $r_{1,0}$, $r_{0,1}$ and $r_{1,1}$ denote the fitnesses of cells carrying the first, second or both mutations relative to unmutated cells. Consider the case of $1 < r_{1,0} = r_{0,1} < r_{1,1}$. This ordering seems quite natural, as enhanced growth capabilities are a defining feature of cancerous cells (Hanahan and Weinberg (2000)). The assumption $r_{0,1} = r_{1,0}$ is made for ease of illustration, and could be interpreted as a gene which has to mutate on two corresponding symmetric alleles so that cancer emerges.

On the one hand, the probability that any mutant cell produces a lineage which reaches fixation is a decreasing function of N. Hence, at a state in which none of the individual cells carry both mutations, it might seem that compartments should be as large as possible to decrease cancer risk. Yet, this is counterbalanced by the phenomenon that cells with higher r are favoured disproportionately by larger compartments. Hence, once cells carrying only one of the two mutations prevail in the compartment, choosing a small compartment would imply that states, in which cells carrying no mutation or carrying both mutations prevail, are reached with higher probability than if large compartments were chosen. This, of course, also includes the absorbing state, in which cancerous cells have reached fixation, and thereby affects $P(t)$. However, the relative proportions of the probability mass on states with more cancerous cells $(1, 1)$ versus on states with more cells that carry zero mutations $(0, 0)$ is more favourable, i.e. is more biased towards healthy $(0, 0)$

cells, in small compartments.[7] This induces a trade-off, which can be seen in Figure 2.

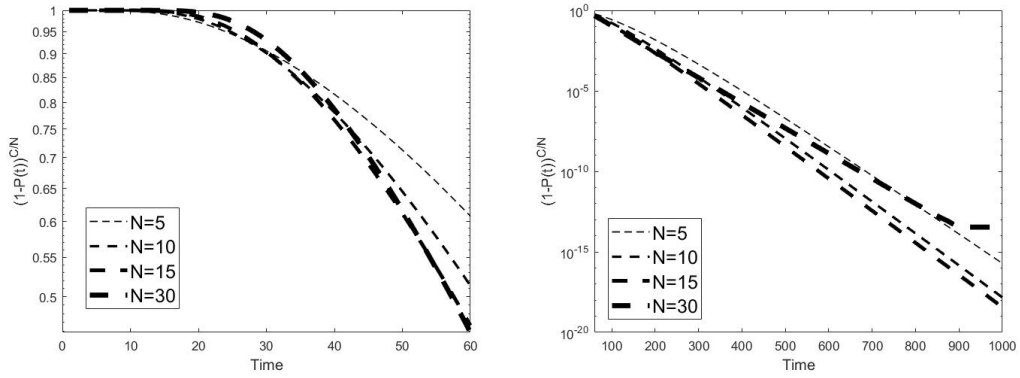

Figure 2: Plot of cancer risk over time. The highest curve corresponds to lowest cancer risk. Parameter values $C = 30$, $r_{0,0} = 1$, $r_{1,0} = 2$, $r_{0,1} = 2$, $r_{1,1} = 10$, $u = 0.01$, initial state of the system: all cells are of type $(0, 0) \in \theta$ ($\{0, 0\}$-corner of the simplex). Note the logarithmic scaling of the y-axes.

For small t, $(1 - P(t))^{\frac{C}{N}}$ is still maximized by large compartments. However, small compartments become optimal, as the probability with which the system is 'pushed back' to states with more $(0, 0)$-cells, from which the absorbing state would be 'harder' to reach, starts to matter more. This can be seen in the middle graph of Figure 2, where the line corresponding to $N = 5$ intersects with the line corresponding to $N = 30$ at $t = 35$. For very large t (≥ 775), $N = 30$ will revert to being optimal, as can be seen in the bottom right subfigure of Figure 2.

This example already highlights a variety of scenarios in which modifications of N would be beneficial. As states with large proportions of individuals that carry only one mutation are with high likelihood traversed on a path towards the fixation of a cancer cell, it follows that natural selection would take into account the 'counterbalancing'. Hence, if the lifespan of an individual would be between the equivalent of 35 and 775 time iterations, natural selection would select for some $N < C$. Let's assume the patient gets lucky, and acquires no mutations over a long time span. If the person's expected further life expectancy is less than the equivalent of 35 iterations, then the optimal preventative treatment would be to merge small compartments into larger compartments as much as possible. This is a special case of Problem (A), with $s_A = s_0$, but where the considered time horizon differs.

The obvious analogue to this for lifespans beyond 775 and patients with unmutated compartments and a remaining life expectancy of between 35 and 775 iterations would be nature selecting for large compartments, which can be improved by dividing compartments.

[7] In a similar context Michor et al. (2003b) find that the rate of cancer initiation is minimized at an intermediate compartment size N, also reflecting a trade-off between preventing the spread mutations in tumor suppressor genes (high r) versus mutations that lead to chromosomal instability (low r). However, their focus lies on explaining the ex-ante minimization of cancer risk, not on path or time dependency of the optimal N. See also Michor et al. (2003c).

However, if again the lifespan was in the interval $[35, 775]$ and the patient was diagnosed to be in a state in which mutations in sufficiently many cells have already occurred, then increasing N would be optimal. This is exemplified in Figure 3, which considers a starting state where all cells are of type $(1, 0)$. Here, at any t larger than the starting time, $(1 - P(t))^{\frac{C}{N}}$ is maximized by large compartments.

In the Appendix, it is also shown that the results presented in this section would not qualitatively change, if one were to drop the assumption that mutation rates are uniform, or the assumption that $\{1, 1\}$ is an absorbing state.

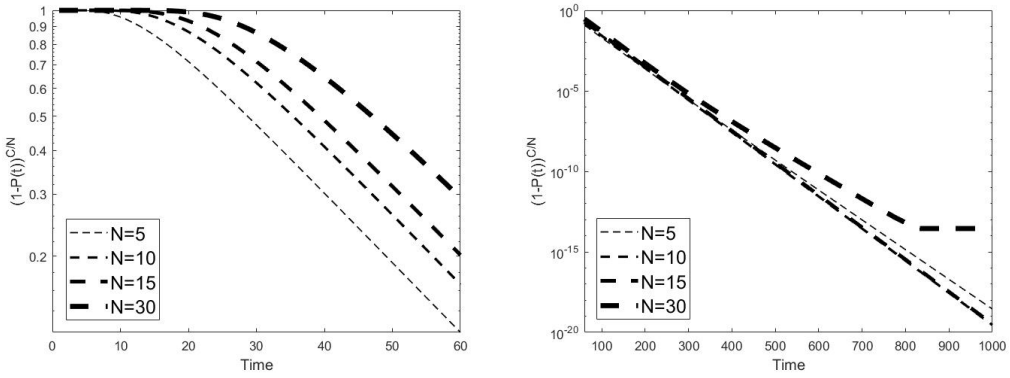

Figure 3: Plot of cancer risk over time. The highest curve corresponds to lowest cancer risk. Parameter values $C = 30$, $r_{0,0} = 1$, $r_{1,0} = 2$, $r_{0,1} = 2$, $r_{1,1} = 10$, $u = 0.01$, initial state of the system: all cells are of type $(1, 0) \in \theta$ ($\{1, 0\}$-corner of the simplex). Note the logarithmic scaling of the y-axes.

Conclusion

In this paper, a novel approach to cancer prevention, namely tissue structure modifications, has been proposed and formalized. It has been argued, that natural selection would likely fail to provide the tissue structure that minimizes the risk of cancer at any given point in time (Problem (A): *path dependence*, Problem (B): *age dependence* and Problem (D): *evolvability*) or even to start with (Problem (C): *risk toleranve*). Changes in compartment-size and the usage of fractal structures in the body have been identified as the most promising routes to feasible treatments, given current technology. Implementation may seem distant today. But the potential gains are high, as was illustrated by the simple two-hit example of a Moran process setting which has shown that optimal tissue structure to prevent carcinogenesis can be extremely path-dependent, and that restructurings can yield large decreases in cancer risk.

Appendix

Compartmentality

The term compartment is used widely and somewhat inconsistently in the literature, and emerges usually from an attempt to divide the body into clusters of cells which can be considered individually (see for example: Ledzewicz and Schättler (2002), Pérez-Caro et al. (2009), Brash et al. (2005), Zhang et al. (2014) and Grajzel et al. (2020)). Compartments are sometimes termed *modules* (as in Bellomo et al. (2008).

A precise definition, given by Teimouri et al. (2019), which also corresponds to the usage of the term in Michor et al. (2004) and other articles by the authors of that paper, is that compartments are subsets of tissues which independently achieve homeostasis.

Even if truly independent regions in the body might be rare, using this approximation has been very successful in explaining the dynamics of several different types of cancer (see: Foo et al. (2011), Michor et al. (2005a), Michor et al. (2003a) and Zhang et al. (2014)). The colonic crypts within which colorectal cancer can form are a particularly illustrative example of such compartments, which is by now exceptionally well-studied (Michor et al. (2005b) and Nowak et al. (2003)).

With regard to the framework in this paper, the assumption that G has many small disjoint components seems to be a good approximation. In applications, neglecting edges e with very small replacement probabilities $\Gamma(e, \cdot)$ might be useful, if it increases the number of components of G and thereby reduces the cardinality of the state-space that needs to be considered.

One Mutation

In Michor et al. (2004), the probability that a compartment which is initially occupied by unmutated cells, will be occupied fully by mutated cells at a time t is given by $P(t) = 1 - e^{-Nut\rho}$. Here

$$\rho = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^N}}$$

denotes the probability of a single mutant reaching fixation. Again, r denotes the relative fitness of a mutated cell. Note that $P(t)$ is increasing in N if $r > 1$ and decreasing in N if $r < 1$.

With this explicit form, $(1 - P(t))^{\frac{C}{N}}$ is simply given by

$$(1 - P(t))^{\frac{C}{N}} = (e^{-Nut\rho})^{\frac{C}{N}} = e^{-Cut\rho}.$$

This is increasing in N, which implies that having one large compartment would be optimal, independent of r. The same also holds true in the simulated Markov process with an absorbing state, as the following graphs (Figure 4) show.

Analysis of the Markov Chain

The Markov transition matrix M used in the applications is constructed by enumerating the elements of S_G and assigning transition probabilities to the appropriate

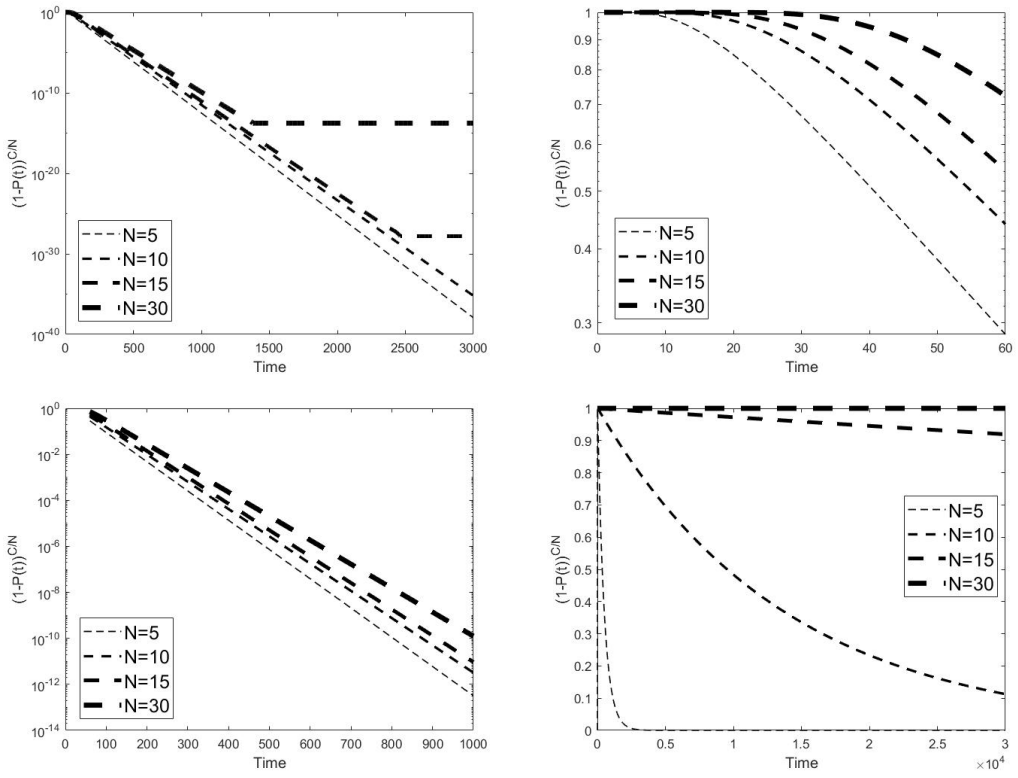

Figure 4: Cancer risk over time with one mutable gene and different values for r_1. Parameter values $C = 30$ $u = 0.01$, initial state of the system: all cells are of type $\{0\} \in \theta$. top left, top right, bottom left: $r_1 = 2$, bottom right: $r_1 = 0.5$

entries as described above. A row-stochastic specification is used. Let e_s denote a unit row vector which has entry 1 at the column corresponding to the enumeration of state s. $P(t)$ can then be computed by premultiplying e_{s_0}, with the t-th power of the Markov transition matrix, and postmultiplying the result with the transpose of $e_{s_{abs}}$, where s_0 denotes the initial state, and s_{abs} denotes the absorbing state.

$$P(t) = e_{s_0} \cdot M^t e_{s_{abs}}^T$$

To make the results of different compartment sizes in this single-compartment Markov approach comparable, one has to account for the different per-cell rates at which events (mutations or Moran-induced replacements) occur. Each iteration of the Markov process describes the occurrence of one event. Hence, in a cluster of C cells, $\frac{C}{N}$ events are described per iteration. With C being held constant, the choice of N changes the number of considered events per iteration. Hence, when comparing probabilities arising from the Markov processes of two different compartment sizes N_1 and N_2, with $N_1 = \alpha N_2$, one needs to correct for the difference in event frequency in the following way:

$$P_{N_1}(\alpha t) = P_{N_2}(t)$$

In the examples of this paper, one t (on the x-axis of the presented graphs) corresponds to the iterations of the Markov process of $N = 5$, i.e. to six events occurring.

Robustness Checks

Above it has been assumed, firstly, that mutation likelihoods u are constant across types, and secondly, that once cells of type $(1,1)$ have reached fixation, they will no longer mutate, i.e. $1,1$ was assumed to be an absorbing state. This section shows that neither of these assumptions qualitatively affect the results presented in this paper. This is to say that large compartments remain optimal at small or large time scales, and that small compartments are optimal at intermediate time scales, independent of the two assumptions.

It could be argued, that for the case of mutations in tumor suppressor genes, it would be more realistic to assume that mutations towards a harmful type are more likely than mutations from a harmful type back to a healthy type. There might simply be more ways to disable the functionality of a gene, than there are ways to repair it. This could be thought to be problematic for the following reason:

The optimality of small compartments at intermediate time scales arises from the relative success of $(0,0)$-cells versus $(1,1)$-cells in environments (states) dominated by cells of type $(0,1)$ or $(1,0)$. These $(0,0)$-cells can either arise through mutations, i.e. on a trajectory following the state $\{0,1\}$ (or $\{1,0\}$), or they can be remainders of the initial cell population, which was assumed to be at $\{0,0\}$. Hence the observed effect could be mostly caused by the performance of healthy mutants, which in reality might be very unlikely to emerge.

To see that this is not the case, consider the left side of Figure 5. The curves depicted in it correspond to a scenario in which likelihoods of 'repairing' mutations are set to zero, i.e. those from $(1,1)$ to $(1,0)$ or $(1,0)$, as well as those from $(1,0)$ or $(0,1)$ to $(0,0)$ (jointly denoted as u^-), while the remaining mutation likelihoods (u^+) are held constant. As can be seen from the figure, the patterns are similar to those in Figure 2 (with intersection points at $t = 27$ and $t = 819$), indicating that the assumption of uniform mutation rates does not affect the results qualitatively in the above sense.

Next, in order to see the effect of dropping the assumption of $\{1,1\}$ being an absorbing state, while keeping mutation likelihoods symmetric, consider the solid curves in the right subfigure of Figure 5. Different from the previous figures, the curves are more flat, as they converge to a strictly positive steady-state probability mass on the state $\{1,1\}$. However, the qualitative result of small compartments being optimal at intermediate time scales, and of large compartments being optimal at small and large time scales again remains unchanged.

Note, lastly, that in the specification presented in the left subfigure, $\{1,1\}$ naturally becomes an absorbing state, as likelihoods of mutation away from $(1,1)$ towards other types are zero, even before cells of type $(1,1)$ have reached fixation. Hence, the left subfigure already represents an interaction of both robustness checks for the extreme case of $u^- = 0$. In order to see, further, whether dropping both assumptions simultaneously would affect the results for small but positive u^-, consider the remaining lines of the right subfigure of Figure 5. Here, the intersection point of curves corresponding to large and small compartments are shifted (lower values for u^- correspond to later first intersection points and earlier second intersection

points), but again, the qualitative result remains unchanged.

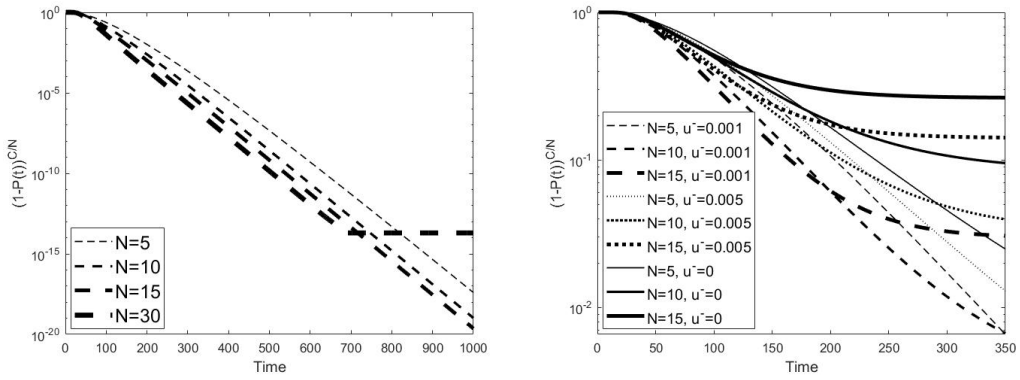

Figure 5: Parameter values $u^+ = 0.01$, $r_{0,0} = 1$, $r_{1,0} = 2$, $r_{0,1} = 2$, $r_{1,1} = 10$, initial state of the system: all cells are of type $(0,0) \in \theta$ ($\{0,0\}$-corner of the simplex). Left: $u^- = 0$, $C = 30$. Right: varying values for u^-, $C = 15$.

References

Adler, M, A Mayo, X Zhou, RA Franklin, JB Jacox, R Medzhitov, and U Alon, "Endocytosis as a stabilizing mechanism for tissue homeostasis," *Proceedings of the National Academy of Sciences*, 2018, *115* (8), E1926–E1935.

Altrock, PM, LL Liu, and F Michor, "The mathematics of cancer: integrating quantitative models," *Nature Reviews Cancer*, 2015, *15* (12), 730–745.

Bellomo, N, NK Li, and PK Maini, "On the foundations of cancer modelling: selected topics, speculations, and perspectives," *Mathematical Models and Methods in Applied Sciences*, 2008, *18* (04), 593–646.

Bizzarri, M, A Giuliani, A Cucina, F D'Anselmi, AM Soto, and C Sonnenschein, "Fractal analysis in a systems biology approach to cancer," 2011, *21* (3), 175–182.

Brash, DE, W Zhang, D Grossman, and S Takeuchi, "Colonization of adjacent stem cell compartments by mutant keratinocytes," in "Seminars in Cancer Biology," Vol. 15 Elsevier 2005, pp. 97–102.

Foo, J, K Leder, and F Michor, "Stochastic dynamics of cancer initiation," *Physical Biology*, 2011, *8* (1), 015002.

Frank, SA, Y Iwasa, and MA Nowak, "Patterns of cell division and the risk of cancer," *Genetics*, 2003, *163* (4), 1527–1532.

Ghorbani, M, EA Jonckheere, and P Bogdan, "Gene expression is not random: scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks," *Frontiers in Physiology*, 2018, *9*, 1446.

Grajzel, D, I Derényi, and GJ Szöllősi, "A compartment size-dependent selective threshold limits mutation accumulation in hierarchical tissues," *Proceedings of the National Academy of Sciences*, 2020.

Hanahan, D and RA Weinberg, "The hallmarks of cancer," *Cell*, 2000, *100* (1), 57–70.

Hindersin, L, B Werner, D Dingli, and A Traulsen, "Should tissue structure suppress or amplify selection to minimize cancer risk?," *Biology Direct*, 2016, *11* (1), 41.

Ledzewicz, U and H Schättler, "Optimal bang-bang controls for a two-compartment model in cancer chemotherapy," *Journal of Optimization Theory and Applications*, 2002, *114* (3), 609–637.

Michor, F, MA Nowak, SA Frank, and Y Iwasa, "Stochastic elimination of cancer cells," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 2003, *270* (1528), 2017–2024.

_ , SA Frank, RM May, Y Iwasa, and MA Nowak, "Somatic selection for and against cancer," *Journal of Theoretical Biology*, 2003, *225*, 377—-382.

_ , TP Hughes, Y Iwasa, S Branford, NP Shah, CL Sawyers, and MA Nowak, "Dynamics of chronic myeloid leukaemia," *Nature*, 2005, *435* (7046), 1267–1270.

_ , Y Iwasa, and MA Nowak, "Dynamics of cancer progression," *Nature Reviews Cancer*, 2004, *4* (3), 197–205.

_ , _ , C Lengauer, and MA Nowak, "Dynamics of colorectal cancer," 2005, *15* (6), 484–493.

_ , _ , NL Komarova, and MA Nowak, "Local Regulation of Homeostasis Favors Chromosomal Instability," *Current Biology*, 2003, *13*, 581–584.

Moreno, PA, PE Vélez, E Martínez, LE Garreta, N Díaz, S Amador, I Tischer, JM Gutiérrez, AK Naik, F Tobar et al., "The human genome: a multifractal analysis," *BMC Genomics*, 2011, *12* (1), 506.

Nowak, MA, F Michor, and Y Iwasa, "The linear process of somatic evolution," *Proceedings of the National Academy of Sciences*, 2003, *100* (25), 14966–14969.

Pérez-Caro, M, C Cobaleda, I González-Herrero, C Vicente-Dueñas, C Bermejo-Rodríguez, M Sánchez-Beato, A Orfao, B Pintado, T Flores, M Sánchez-Martín et al., "Cancer induction by restriction of oncogene expression to the stem cell compartment," *The EMBO journal*, 2009, *28* (1), 8–20.

Petoukhov, S, E Petukhova, and V Svirin, "New symmetries and fractal-like structures in the genetic coding system," in "International Conference on Computer Science, Engineering and Education Applications" 2018, pp. 588–600.

Pickup, MW, JK Mouw, and VM Weaver, "The extracellular matrix modulates the hallmarks of cancer," *EMBO Reports*, 2014, *15* (12), 1243–1253.

Sapolsky, R and S Balt, "Reductionism and variability in data: A meta-analysis," *Perspectives in Biology and Medicine*, 1996, *39* (2), 193–203.

Teimouri, H, MP Kochugaeva, and AB Kolomeisky, "Elucidating the correlations between cancer initiation times and lifetime cancer risks," *Scientific Reports*, 2019, *9* (1), 1–8.

Werner, B, JG Scott, A Sottoriva, A Anderson, A Traulsen, and PM Altrock, "The cancer stem cell fraction in hierarchically organized tumors can be estimated using mathematical modeling and patient-specific treatment trajectories," *Cancer Research*, 2016, *76* (7), 1705–1713.

Zhang, X, Y Fang, Y Zhao, and W Zheng, "Mathematical modeling the pathway of human breast cancer," *Mathematical Biosciences*, 2014, *253*, 25–29.