Context-dependent selection as the keystone in the somatic evolution of cancer

Vibishan B.[1,2] **and Milind G. Watve**[1,2,*]

[1] Department of Biology, Indian Institute of Science Education and Research (IISER), Pune

This manuscript was compiled on December 12, 2018

Please provide an abstract of no more than 250 words in a single paragraph. Abstracts should explain to the general reader the major contributions of the article. References in the abstract must be cited in full within the abstract itself and cited in the text.

Somatic evolution | Mutation accumulation | Epidemiology | Cancer etiology |

Over the past 60 years or so, ideas in the field of cancer epidemiology have evolved significantly, with the first models starting from the Armitage-Doll multi-stage models (1), which were essentially statistical fits to available data of age-specific incidence, and predicted a power law relationship of cancer risk with age. The connection between the multiple stages and sequential genetic mutation events was established definitively for retinoblastoma with the two-hit hypothesis (2). This finding has since directed a lot of attention to mutational processes and genetic instability within cells as fundamental forces in cancer, and their signatures in population-level datasets. Tomasetti et al. have made the argument for cancer risk being largely determined by random mutations (3, 4). Others have studied the impact of selectively-neutral or deleterious passenger mutations on the expansion and progression of advantageous mutant clones (5), mutation accumulation rates across tissue types (6), or dependencies between mutations (7). On the other hand, an old debate in the theory of evolution is how the simple process of random mutations and natural selection can lead to compled structures such as the eye, that need coordinated action of several genes. This is often perceived as a monkey-on-a-typewriter paradox (8)-how likely is it that a monkey sitting at a typewriter and hitting keys at random would end up typing a meaningful sentence? The problem of cancer is qualitatively similar to this, but quantitatively even more difficult, as no single mutation is known to make a cell cancerous. All cancers are necessarily a combination of different types of genomic changes including point mutations, aneuploidy, and other chromosomal aberrations. The cancer phenotype has a large number of distinguishing characeters, encapsulated by the notion of the "hallmarks" of cancer (9–11), and the wide range of characterisitics that these hallmarks include make it astonishing that so many alterations in cell properties come together in cancers purely out of chance, especially since most cancers must evolve independently in each individual organism.

Within the level of the organism, clonal expansion is another process that is implicated in carcinogenesis. Every component mutation on the way to a cancerous phenotype causes the mutant clone to expand, and as the mutant population increases, the probability of a second component mutation increases proportionately (12). Implicit in this theory is the assumption that every component mutation has a selective advantage over the normal cell. Since most changes involved in carcinogenesis relate to evading growth regulatory mechanisms, it is considered logical that any mutation that allows for such evasion will have a natural selective advantage. However, the final word is far from conclusive on the view that cells with component mutations are always at an advantage, and evidence has been accumulating over the past few years that the fitness advantage of a mutant is largely dependent on the tissue micro-environment (13, 14). Studies in mice (15) and humans (16) have demonstrated the effect of contingent factors, such as behavioural profiles and lifestyle parameters, on cancer progression. Such findings provide clear indications that the selective forces which determine mutant clone fitness can vary considerably across individual organisms, leading to *context-dependent clonal expansion* of potentially oncogenic mutants. This aspect of cancer progression, that moves beyond mutagenesis, has also seen some modeling effort in the past few years, towards attributing cancer risk to environmental factors (17), or other forms of selection (18, 19). These efforts reveal the incorporation of additional complexity to causal factors underlying cancer, but so far, what we identify here as context-dependent selection on somatic mutants has not been incorporated explicitly in models of cancer incidence.

Across biological levels, we identify at least three processes that seem to play a major role in cancer progression and etiology: (1) random mutagenesis, or the "bad luck" hypothesis, (2) expansion of mutant clones within organisms, and (3) context-dependent selection acting across organisms. Since well-curated data is available for human cancer incidence patterns, we develop models of these three processes, and compare their predictions with the epidemiological picture of cancer in the human population. We briefly summarize the essential epidemiological features that must be part of any comparative modeling framework:

1. Total and age-specific incidence: Across cancer types, total population-level incidence lies around 20 to 30%, while age-specific and cumulative incidence patterns show more

Significance Statement

Authors must submit a 120-word maximum statement about the significance of their research paper written at a level understandable to an undergraduate educated scientist outside their field of speciality. The primary goal of the Significance Statement is to explain the relevance of the work in broad context to a broad readership. The Significance Statement appears in the paper itself and is required for all research papers.

The authors declare that there are no conflicts of interest

[2] V.B. and M.G.W. contributed equally to this work.

[*] To whom correspondence should be addressed. E-mail: milind@iiserpune.ac.in

variations (20). Interestingly, recent analyses have shown for several cancer types that the age-specific incidence rates decline late in life, in contrast with general model predictions of a power law increase in incidence with age (21). The late-life decline causes the cumulative incidence to saturate with age at a small percentage of the population size, which is an important detail. No matter the lifespan, the proportion of cancer in the population can never reach 100%, which represents a finite limit that is not determined by time.

2. Incidence vs cell number: As mentioned earlier, the relationship between cancer risk and cell number (as the lifetime number of cell divisions, *lscd*) has been kept in the spotlight by recent work by Tomasetti et al. (3, 4). On the whole, the linearity of the relationship between *lscd* and cancer risk is still under debate, although an explanation of this non-linearity remains incomplete. We think that the relationship is non-linear, and possibly saturating, as is clear from an examination of the slopes of the cancer risk-*lscd* association. Linear or not, there are still real-life data for the relationship with which model predictions may be compared; as we show later, the veracity of the linearity claim is not directly relevant to the conclusions we draw from our models.

3. Incidence vs mutation rate: Empirical data on this relationship are less common, as mutation rates are difficult to measure reliably, although some efforts have been made (22). However, a general notion exists that higher mutation rates increases cancer risk, and this remains to tested rigorously, barring theoretical work dealing with the effect of mutagens on patterns in incidence data.

4. Peto's paradox, and similar observations: This relates to the incidence-cell number relationship, as cancer risk is seen not to scale with body size or cell number across species [citations], with does correlate with the latter within a species [citations]. A wide range of explanations have been offered for this observation [citations], but a comprehensive explanation has been elusive. Nevertheless, it remains an key feature of cancer etiology with regard to the modeling effort.

5. Non-mutagenic carcinogens: There are several agents, including hormones and growth factors, that increase cancer risk without affecting the basal mutation rate [citations]. The activity of these agents, and their signature in epidemiological patterns, are both important in building a complete framework of explaining cancer etiology.

The "bad luck" model

This hypothesis assumes that the required set of driver mutations accumulate in a cell by chance alone. This may happen over a period of time, or in a single large-scale event, like chromothripsis (23).

Consider an organism with a population of n stem cells, each with a mutation rate per cellular generation per genome, p. The probability that at least one cell acquires one mutation at a given point of time can be given as $1-(1-p)^n$. If k such mutations are requried for cancer onset, the probability of cancer according to the bad luck model can be given as below, based on an algebraic formulation (24):

$$p_{can} = 1-(1-p^k)^n \quad [1]$$

Given the probability of cancer per unit time, p_{can} from equation 1, the cumulative incidence of cancer for age, A, can be expressed as below:

$$p_A = 1-(1-p_{can})^A \quad [2]$$

From equation 1, it is clear that the probability of cancer has a threshold relationship with both n and p, such that incidence rises from near zero to 100% over a small range of p and n, as shown in Figure 1. Within this narrow range, total incidence as reflected by the cumulative probability, p_{can}, occasionally lies in the realistic range of about 30%, but these are the exceptions.

From equation 2, the relationship of p_{can} with age is a monotonically increasing function with a maximum at one. Figure 2 shows this relationship across the entire parameter range of n and p, for which this prediction holds. Cancer probability increases monotonically with age, and only saturates at 100% incidence, which stands in stark contrast to the observed late-life decline in age-specific rates.

For the sake of simplicity, we have ignored the cost of lethal and/or passenger mutations, and assume that the mutations occur together at any given point of time; although the model predictions change marginally, we see that this assumption does not affect the general inferences we draw. Taken together, this formulation of the "bad luck" model predicts a sharp threshold relationship of cancer probability with both n and p, and a monotonically increasing relationship with age, both of which can be falsified based on a simple comparison with published epidemiological data [citation].

While there are ways of incorporating additional complexities within the formulation presented here, we choose to explore the effects of clonal expansion and context-dependent selection through a simulation-based framework. As we show below, it allows for easier exploration of the parameters under study, and creates a logical progression in terms of biological scale; the "bad luck" model deals with cell-level properties, clonal expansion is a tissue- or organ-level phenomenon, while context-dependent selection occurs between organisms at the level of the population.

The simulation framework

We use a linear process to model the sequential accumulation of mutations in a population of stem cells. we begin by considering an organism's development starting from one stem cell, with mutation rate per cell generation per nucleotide, p, growing logistically to a carrying capacity, n, following the discrete logistic equation below:

$$m_{i,t} = m_{i,t-1} + m_{i,t-1} * g_i * \left(\frac{n - \sum_{i=0}^{k} m_{i,t-1}}{n}\right) - m_{i,t-1} * d \quad [3]$$

Here, $m_{i,t}$ is the size of the ith mutant population at time, t, with $i=0$ being the non-mutant cell population, g_i is the corresponding logistic growth rate, d is the common death rate, and k is the threshold number of oncogenic mutations required for cancer onset. As the organism develops into an adult, net growth in the stem cell compartment saturates, but reaches a dynamic equilibrium between cell death and renewal. The stem cell population can be reduced, either by death of

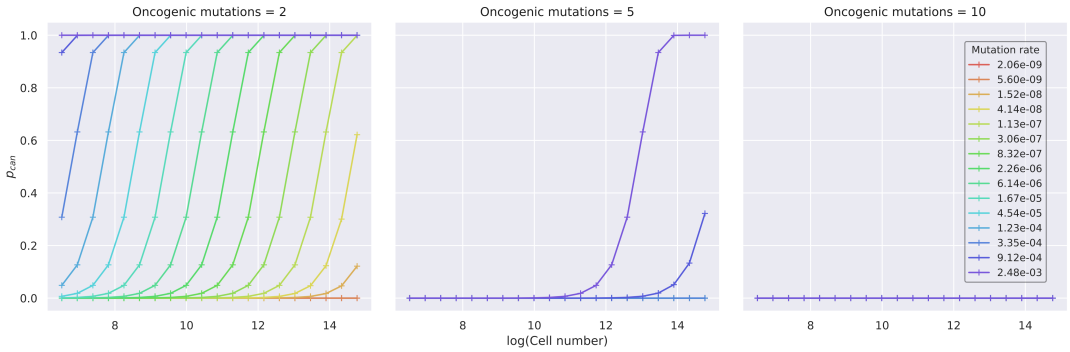

Fig. 1. Cancer probability, p_{can} vs mutation rate and cell number for the "bad luck" model; p_{can} remains near zero in part of the parameter range, and rises to one over a narrow region of the corresponding parameter. This proability is cumulative and therefore reflects total incidence in the population. For the "bad luck" model, the total incidence is rarely in the observed range of around 30%. The number of oncogenic mutations required for cancer onset does not affect the existence of a threshold with n and p, but does affect where the threshold occurs in the parameter space. The legend shows values of p for each curve.

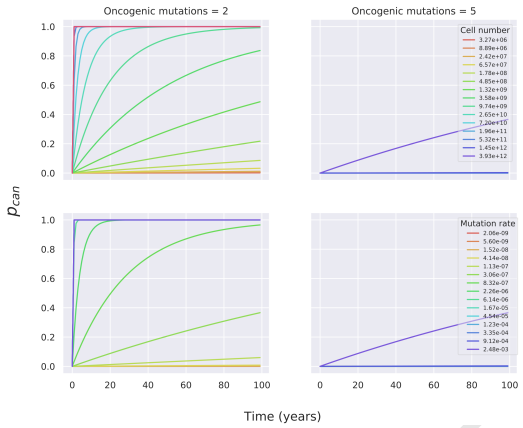

Fig. 2. Cancer probability vs age, as given by equation 2; cancer probability increases monotonically with age, saturating only at one in most cases. Where the probability does approach realistic values of total incidence, it still does not reflect the late-life decline in incidence rates observed epidemiologically. As in Figure 1, the number of oncogenic mutations required for cancer onset does not change the nature of the incidence-age relationship.

stem cells, or assymetric division to produce differentiated cells. The death rate in equation 3 is this constant rate of cell removal from the stem cell population. Assuming a common death rate for all cell populations, the replacement of the lost cells by either mutants or non-mutants is a function of their growth rates. We simulate new mutation events stochastically; the probability of at least one cell accumulating a mutation is given by $1-(1-p)^{m_{i,t}}$, and if this probability exceeds a random number between 0 and 1, a new $(i+1)$th cell population is initiated. We assume that each new oncogenic mutation offers a growth advantage over older cell populations, leading to successive cycles of clonal expansion in which the newer population gradually replaces older cells through competitive exclusion. We simulate this linear evolution process until k mutations have been accumulated, which is the assumed threshold for cancer onset. Death of the organism occurs either at cancer onset when the kth mutation occurs, or at the end of the natural lifespan of 100 years, whichever happens first. This simulation is repeated independently for 10000 organisms, and the population-level cancer incidence is recorded, along with the age of onset.

Choice of parameter range. In order to standardize the discrete logistic simulation, we assume the time unit to be one day per logistic growth step. Most human organs complete development and maturation wihtin the first 10-20 years of the lifespan [citation], and the final carrying capacity achieved is the adult stem cell number, ranging between 10^6 and 10^{11} across different tissues. Given the final population size and the time taken to reach it, a simple calculation based on the logistic equation shows the required growth rate for a non-mutant stem cell to be in the range of 0.00383-0.0131 [details in the supplement]. Starting from the non-mutant growth rate, g_0, growth rates are assumed to increase linearly for each subsequent mutant population. For all simulations, we assume $g_0 = 0.007$ to obtain cancer incidence within an organism's lifespan. Ranges of n and p are retained as in the "bad luck" model.

The context-independent selection case. While the clonal expansion theory introduced the notion of selective advantages to oncogenic mutants, it makes the implicit assumption that identical mutations have the same selective advantage in every organism in which they occur; stated otherwise, individual organisms do not differ in their propensity for mutant clonal expansion. To capture this in the context-independent selection case, we use the same linear progression of growth rates for all organisms in the simulation.

The context-dependent selection case. As argued earlier, it is becoming increasingly clear that the competitive outcomes of identical mutations can depend strongly on the micro-environmental context in which cell competition occurs. In order for selection on mutants to be context-dependent in our model, we randomize the progression of growth rates during mutation accumulation. Each organism begins with the same g_0, but the progression of growth rates is randomized across individual organisms, such that organisms with large gs would progress faster towards cancer onset, while those with small, or negative values of mutant gs would never progress to a cancerous state as the mutant gets selected against. This produces variation across organisms for cancer propensity, leading to a more realistic model.

Results

As Figure 1 shows, under the assumption of context-independent selection, the incidence of cancer shows a strong threshold relatioship with age, where cancer is unlikely or rare up to a certain age, and increases rapidly to 100% with a relatively short span of time. This is indicated by the fact that the age-specific crude incidence falls sharply to zero at some point in the lifespan, when the cumulative incidence also reaches 100%. As with the "bad luck" model, incidence has a threshold relationship wtih both n and p, in terms of the age at half-maximum incidence (Figure 3C and G) and the maximum cumulative incidence, I_{max} (Figure 3D and H). Moreover, saturation of incidence occurs only at 100%, which is identical to the "bad luck" model, despite the inherently stochastic implementation of mutation occurrence. Where the incidence of cancer is near the realistic range, for small values of p and n, the late-life decrease in incidence is still not reflected in the context-independent selection case. The prediction of 100% incidence is due to the fact that all organisms in the population share the same growth rate progression for mutants. No other parameter in the model inherently precludes the accumulation of all k oncogenic mutations in some organisms, and the dynamics of accumulation is entirely a function of p and n, along with stochastic variation. The context-independent selection model thus predicts that with increasing p and/or n, cancer incidence increases, and reaches saturation only at 100% incidence. Therefore, clonal expansion accounts for some physiologically-relevant phenomena, like competitive growth of mutants, its description of cancer incidence and the effects of model parameters are either unrealistic, or incomplete.

As opposed to the context-independent selection model, the context-dependent model produces a saturating trend in cumulative incidence that begins to saturate at a level much lower than 100%. As Figure 4 shows, the saturation limit for many values of n and p is quite close to the epidemiological estimate of cancer risk (20-30%). This is an important feature of the context-dependent model, as it allows the model to generate more realistic patterns in age-specific cancer incidence. The realistic saturation of population incidence can be explained by the fact that propensity for clonal expansion varies across organisms, such that cancer progression occurs very quickly in some organisms, and not at all in others. Of the three models analysed so far, only the context-dependent model captures a trend similar to the late-life decline observed in many cancers in humans as the crude incidence curves in Figure 4A and E show, which suggests alternative explanations for trends in cancer incidence, independent of possible evovled cancer defenses.

From Figure 3, there are some indications that n and p could have quantitatively different effects on incidence parameters. To explore this possibility, we co-randomized k with either n or p while maintaining the same growth rate progression, and looked at the association of $log(n)$ or $log(p)$ with the time taken for cancer onset. Although the data is not shown here, we find that cancer incidence is significant only up to maximum $k = 10$. Generally, increasing k shifts the observed time to cancer to later in life, as to be expected from mutation accumulation. As Figure 5 shows, for small k in this range, time to cancer onset is not strongly dependent on n, but is associated significantly with p. For moderately large k, the association of time to cancer with both n and p strengthens, and weakens again for larger k. For the context-independent selection case then, p is a more important determinant of time to cancer than n for any most of the range of k where cancer is observed. Doing the same for the context-dependent case, we randomized g as explained earlier, along with either k and n, or k and p. Remarkably, when g and n are randomized together, most of the variation in time to cancer is explained by g (Figure 6) with n having a marginal effect only (data not shown), while p retains some effect on the time to cancer (data not shown). However, the effect of g on the time to cancer is modulated by the required k as seen with n and p in Figure 5. Together, this indicates that independent of the kind of selection assumed in the model, k influences the expected effect of n, p and g. A clear difference emerging between the context-independent and -dependent selection cases lies in the parameters that explain a significant part of the variation in time to cancer. Without context-dependent selection, p, and to a smaller extent, n have a clear effect on time to cancer, but including g randomization leads to time to cancer being largely dictated by g.

Discussion

Returning to the epidemiological patterns in the introduction, we now review the predictions of the three models. The "bad luck" model predicts a strong threshold of cancer probability with n, p and with age. In all three cases, incidence increases in a roughly sigmoid fashion, and saturates only at 100% incidence. This is also broadly true of the context-independent selection model, which similarly fails to capture the saturation of incidence at around 30% in real populations, along with the late-life decline in age-specific incidence observed in several cancer types. Along the same lines, the sharp threshold relationships from both of these models of incidence with both n and p stand in clear contrast to real data. As stated earlier, the relationship between incidence and cell number is claimed to be linear by man in the field (3, 4). While it remains to be conclusively determined if this relationship indeed is linear or non-linear, there is sufficient basis to reject the threshold prediction from the "bad-luck" and context-independent selection model. Assuming that the relationship is non-linear leads to an obvious rejection of predictions from both models. On the other hand, even if we were to assume linearity over a large range of n, the first two models do not fare any better. While a similar detailed look is not yet for the relationship of incidence with p as robust data are still lacking, there is again sufficient basis to reject a sharp threshold and/or 100% incidence as predicted by the "bad luck" and context-independent models.

On all these counts, we see that the context-dependent selection offers much more realistic predictions for the relationship of incidence with n, p and age. It can produce saturation of incidence at around observed rates of 20-30%, the observed late-life decline in incidence, and more realistic relationships with n and p. It has also indicated interesting effects of the number of oncogenic mutations required for cancer. While it is expected that higher k increases the time taken for cancer onset, the value of k is also seen to modulate how strongly n, p and g affect the time to cancer onset, particularly in the moderate to higher range of k. This is an important detail that pan-cancer analyses of incidence trends must take note of. For small values of k, n does not affect incidence dynamics, but g and p do. The context-dependent model also offers

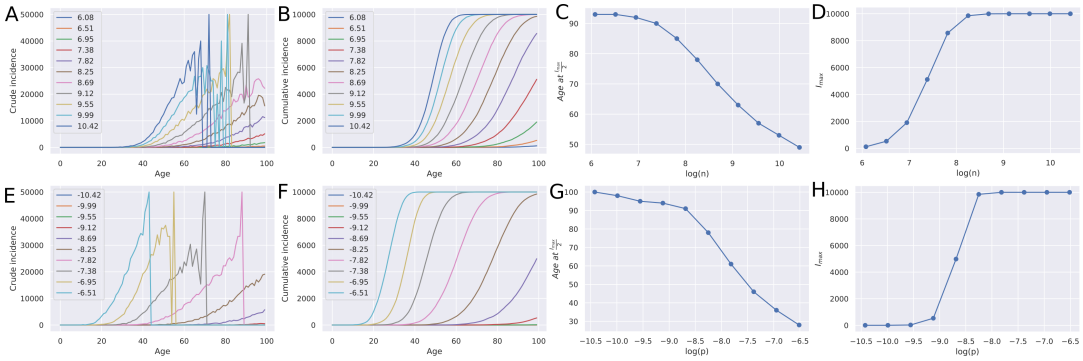

Fig. 3. Incidence patterns from the context-independent selection model over the range of (A-D) n, and (E-H) p. From left to right in each row, the plots are of (A, E) age-specific crude incidence per 100000 vs age, (B, F) cumulative incidence for the simulated population vs age, (C, G) age at which half the maximum incidence is reached vs $log(n)$ or $log(p)$, and (D, H) the maximum cumulative incidence, I_{max} vs n or p. Inset legends for the age curves are $log(n)$ and $log(p)$ in the top and bottom row respectively. For A-D, $p = 5.603 * 10^{-9}$, and for C-H, $n = 1.785 * 10^{8}$. Insets for the incidence curves show $log(n)$ and $log(p)$ in the top and bottom rows respectively. Growth rates progress linearly in the general form, $g_i = 0.007 * (i + 1)$, where $i = 0, ..., k$ and $k = 5$.

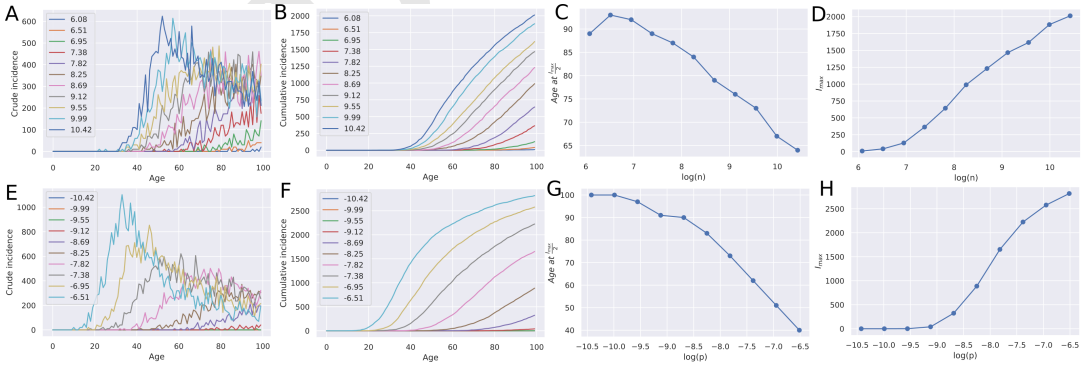

Fig. 4. Incidence patterns from the context-dependent selection model over the range of (A-D) n, and (E-H) p. From left to right in each row, the plots are of (A, E) age-specific crude incidence per 100000 vs age, (B, F) cumulative incidence for the simulated population vs age, (C, G) age at which half the maximum incidence is reached vs $log(n)$ or $log(p)$, and (D, H) the maximum cumulative incidence, I_{max} vs n or p. Inset legends for the age curves are $log(n)$ and $log(p)$ in the top and bottom row respectively. For A-D, $p = 5.603 * 10^{-9}$, and for C-H, $n = 1.785 * 10^{8}$. Insets for the incidence curves show $log(n)$ and $log(p)$ in the top and bottom rows respectively. Growth rates progress linearly from $g_0 = 0.007$ to $g_k = 0.007 * \mu$, where μ is normally-distributed random variable with $\overline{\mu} = 0$ and $\sigma = 3$, and $k = 5$.

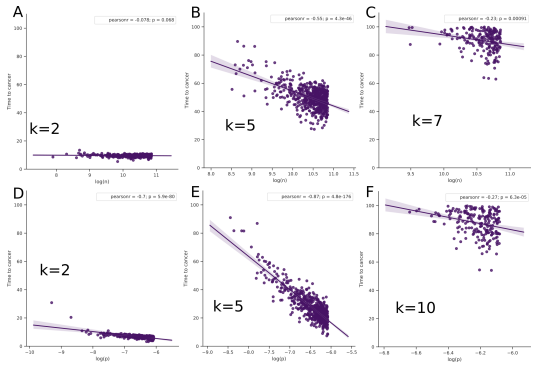

Fig. 5. Effect of k in the context-independent selection case. The plots are time to cancer onset against $log(n)$ or $log(p)$, with k randomized with (A-C) n, or (D-F) p; value of k in the inset corresponds to the number of threshold oncogenic mutations assumed for the corresponding points. From A to C, for higher threshold of oncogenic mutations, the effect of n on time to cancer gets stronger, as shown by the improvement in the association. For small k however, n does not affect the age of cancer onset. On the other hand, p has a strong effect on the time to cancer at every value of k considered. k, n and p were uniformly-distributed random variables with ranges $[0, 20]$, $[1.203 * 10^6, 2.649 * 10^{10}]$, and $[3.775 * 10^{-11}, 3.059 * 10^{-7}]$ respectively. For (A-C), $p = 5.603 * 10^{-9}$. For (D-F), $n = 1.785 * 10^8$.

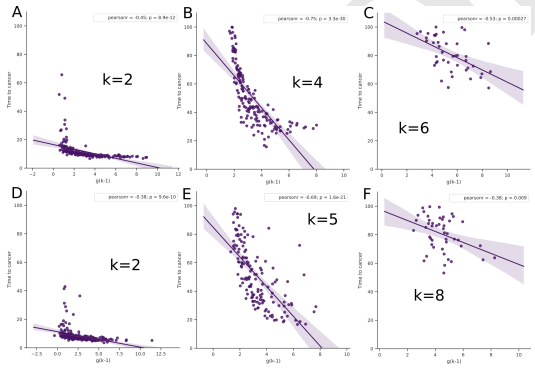

Fig. 6. Effect of k in the context-dependent selection case. The plots are time to cancer onset against $g(k-1) = 0.007 * \mu$ for the last oncogenic mutation, with k randomized with (A-C) n, or (D-F) p; value of k in the inset corresponds to the number of threshold oncogenic mutations assumed for the corresponding points. Compared to Figure 5, $g(k-1)$ explains variance in time to cancer much better than either n or p. This is true of both (A-C) when n and k are also randomized, and (D-F) when p and k are also randomized. The effect of $g(k-1)$ is nevertheless modulated by the required k. Ranges of k, n and p are the same as in Figure 5. For (A-C), $p = 5.603 * 10^{-9}$. For (D-F), $n = 1.785 * 10^8$.

potential explanations for the late-life decline in age-specific incidence. In the model, we see that cancer incidence is determined to a large extent by g, suggesting a selection-limited process. Late-life incidence in the model comes from a slow growth rate progression across mutations, and a late-life decline can therefore be described in terms of the distribution of the growth rate progression. In a population where slow growth rate progression for mutants is rare, organisms with active mutagenesis progress to cancer relatively earlier in life, and cancer is rarer later in life. It becomes important therefore, to consider the local or evolutionary factors underlying such a temporal pattern of mutation accumulation.

On the whole, the better prediction profile of the context-dependent model stems from the distribution of g in the population, and this has interesting implications for the kind of causal factors that are important in explaining cancer etiology. One of these implications concern the mutation-centric thinking that characterizes a good portion of current opinion in cancer biology. For instance, several growth factors and hormones are known to increase cancer risk, not the least of which is insulin, without increasing the basal mutation rate or the cell number. The action of such "non-mutagenic carcinogens" is not compatible with a mutation-centric approach to carcinogenesis. We propose instead that parameters like g, which reflect context-dependent selection on mutants, offer better explanative scope. This selection is imposed by the micro-environment in the tumour or pre-cancerous niche, and includes all the factors that determine the selective advantage of oncogenic or pre-cancerous mutant. As mentioned earlier, empirical evidence of such context dependence has been accumulating on multiple scales, from cell competition *in vitro* (25), to cancer progression in mouse models (15). The likelihood of cancer development is therefore a function of these context-dependent factors, whose regulation and dynamics then should become the focus of translational cancer research.

1. Armitage P, Doll R (1954) The Age Distribution of Cancer and a Multi-Stage Theory of Carcinogenesis. *British journal of cancer* 8(1):1–12.
2. Knudson AG (1971) Mutation and Cancer: Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Sciences* 68(4):820–823.
3. Tomasetti C, Vogelstein B (2015) Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347(6217):78–81.
4. Tomasetti C, Li L, Vogelstein B (2017) Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* 355(6331):1330–1334.
5. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA (2013) Impact of deleterious passenger mutations on cancer progression. *Proceedings of the National Academy of Sciences* 110(8):2910–2915.
6. Blokzijl F, et al. (2016) Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538(7624):260–264.
7. Mina M, et al. (2017) Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. *Cancer Cell* 32(2):155–168.e6.
8. Dawkins R (1996) *The blind watchmaker : why the evidence of evolution reveals a universe without design.* (W. W. Norton & Company), pp. xvii, 358.
9. Hanahan D, Weinberg R (2000) The hallmarks of cancer. *Cell* 100(1):57–70.
10. Schäfer M, Werner S (2008) Cancer as an overhealing wound: An old hypothesis revisited. *Nature Reviews Molecular Cell Biology* 9(8):628–638.
11. Hanahan D, Weinberg R (2011) Hallmarks of Cancer: The Next Generation. *Cell* 144(5):646–674.
12. Nowell P (1976) The clonal evolution of tumor cell populations. *Science* 194(4260):23–28.
13. Hanahan D, Coussens LM (2012). *Cancer Cell* 21(3):309–322.
14. Pietras K, Östman A (2010) Hallmarks of cancer: Interactions with the tumor stroma. *Experimental Cell Research* 316(8):1324–1331.
15. Cao L, et al. (2010) Environmental and Genetic Activation of a Brain-Adipocyte BDNF/Leptin Axis Causes Cancer Remission and Inhibition. *Cell* 142(1):52–64.
16. Rundqvist H, et al. (2013) Effect of Acute Exercise on Prostate Cancer Cell Growth. *PLoS ONE* 8(7):e67579.
17. Hochberg ME, Noble RJ (2017) A framework for how environment contributes to cancer risk. *Ecology Letters* 20(2):117–134.
18. Nagy JD, Victor EM, Cropper JH (2007) Why don't all whales have cancer? A novel hypothesis resolving Peto's paradox. *Integrative and Comparative Biology* 47(2):317–328.
19. Caulin AF, Maley CC (2011) Peto's Paradox: Evolution's prescription for cancer prevention.

20. American Cancer Society (2016) Cancer Facts & Figures 2016. *Cancer Facts & Figures 2016* pp. 1–9.
21. Harding C, Pompei F, Wilson R (2012) Peak and decline in cancer incidence, mortality, and prevalence at old ages. *Cancer* 118(5):1371–1386.
22. Hao D, Wang L, Di Lj (2016) Distinct mutation accumulation rates among tissues determine the variation in cancer risk. *Nature Publishing Group* (January):1–5.
23. Stephens PJ, et al. (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144(1):27–40.
24. Calabrese P, Shibata D (2010) A simple algebraic cancer equation: calculating how cancers may arise with normal mutation rates. *BMC Cancer* 10(1):3.
25. Archetti M, Ferraro DA, Christofori G (2015) Heterogeneity for IGF-II production maintained by public goods dynamics in neuroendocrine pancreatic cancer. *Proc Natl Acad Sci U S A* 112(6):1–6.