DIMACS Technical Report 2004-54
November 2004

Initiation of colorectal cancer: where do the two hits hit?

by

Natalia L. Komarova
Departments of Mathematics and Ecology & Evolution
University of California – Irvine
Irvine, CA 92697

Liming Wang[1]
Department of Mathematics
Rutgers University
Piscataway, NJ 08854

[1]supported in part by DIMACS Award Programs for Graduate Students winter 2003-2004

ABSTRACT

It is widely believed that stem cells are of special importance for colorectal cancer initiation. The earliest event being the inactivation of both alleles of the Adenomatous Polyposis Coli (APC) gene, it is thought that the stem cells are the most likely target for these two first hits. Indeed, at the first glance, short-lived differentiated cells cannot sustain a mutation long enough for the second hit to occur, because of the constant apoptosis/renewal process in epithelial tissues. Using a straightforward calculation, we show that this intuitive argument is incorrect. Our model based on the conventional view of colon crypt architecture, suggests that at least one of the two hits may occur in the migrating compartment. We suggest that a possible role of differentiating cells in cancer initiation cannot be discarded simply based on the fact that they are short–lived. More evidence is needed to understand the cellular origins of cancer and to identify whether or not a double hit in a daughter cell can be "immortalizing". In this study we discuss several scenarios and propose some experiments which can shed light on these questions.

Running title: Initiation of colorectal cancer.

Abbreviations:

"APC" stands for Adenomatous Polyposis Coli.
"SC" stands for stem cell.
"DC" stands for differentiated cell.

Key words: Stem cells, differentiated cells, APC, mutations, mathematical modeling, colon crypt.

1 Introduction

The normal functioning of colon relies on the fine-tuned balance of the epithelial cell production, differentiation and death. The regulation of the processes of cell proliferation and shedding occurs at the level of crypts - the folds of colonic epithelium which are continuously renewed by stem cell division. The appearance of dysplastic crypts in the beginning of colorectal cancers is a manifestation of the broken balance between cell division and apoptosis. At the molecular level, it has been shown that the earliest event of sporadic colorectal cancers is the inactivation of the APC gene [1], [2], or other genes involved in the Wnt pathway [3], [4]. The APC gene inhibits members of the Wnt signaling pathway, which promote the expression of β-catenin. In its turn, β-catenin acts as an enhancer of cell division [5].

It is widely believed that the relevant target cells for the first mutation are the colonic stem cells, [6], [7], [9], [10], [11], [12], [13]. The argument usually goes in the following way [14]. If the first mutation happened in a proliferative daughter cell, it would be washed away before the second hit has a chance to confer a significant phenotypic change. On the other hand, if the first mutation occurs in the "immortal" stem cell, then its mutant progeny will populate the compartment and persist for as long as it takes to accumulate further mutations which give rise to neoplasia.

The "bottom-up", the "top down" and the "migrating compartment" hypotheses. It is often assumed that the stem cells are located in a niche at the base of the crypt. The "bottom-up" model of colorectal histogenesis [15], [16], [17] states that the oncogenic mutations occur in the stem cells at the base of the crypt. A natural consequence of this model is that once such a mutation occurs, the entire crypt will be monoclonally-mutant. There has been some evidence which contradicts this view. Namely, dysplastic cells exhibiting genetic alterations in the APC gene have been found in the upper layer of crypts, whereas cells located at the bottom of the same crypt did not contain such alterations. This led to the "top-down" model [18]. The two explanations proposed were that (i) the stem cells reside near the top of the crypts, or (ii) transformed cells originate from the stem cells at the base of the crypt, then they passively move upward, after which a cycle of cell proliferation and tissue replacement starts in the top-down direction.

Another explanation has been put forward which suggests that the relevant mutations occurs in fully differentiated cells [19], or in the proliferative/migrating daughter cells [20]. This is consistent with the stem cells being located at the base of the crypt, as previously thought. At the basis of these models is the idea that a proliferating daughter cell with a silenced APC gene (or otherwise upregulated β-catenin) will acquire a stem cell phenotypic characteristic, that is, permanence in the crypt. Indeed, it has been shown that the β-catenin/T cell factor 4 complex constitutes the "master switch" that controls proliferation versus differentiation in healthy and malignant intestinal epithelial cells [21]. Inactivation of the APC gene in a migrating cell could reverse the process of differentiation and trick the cell into thinking that it is "immortal". This would lead to a continued proliferation of this cell which would avoid entering the final differentiation and programmed apoptosis stage.

Cellular origin of cancer. Even though cancer cells are often found to have many characteristics of stem cells, the precise cellular origin of cancer is still unknown for most cancers. The stem-like phenotype of cancer cells is consistent with two opposite theories: one is that the first event happens in a stem cell, and the other is that it occurs in a (partially)-differentiated cell, thus leading to its de-differentiation, or "immortalization". Experimentally this is a very difficult question, and the debate is ongoing. However, there are some examples of cancers where there is strong evidence supporting one or the other theory. One such example is leukemia [22]. Acute myeloid leukemia (AML) is assumed to reflect the transformation of a primitive stem cell compartment. On the other hand, it is thought that acute promyelocytic leukemia (APL) arises in committed myeloid progenitors. The evidence in support of the partially-differentiated-cell origin of APL has been reported in a study of clonality of leukemia [23]. Furthermore, many mouse models of APL have been designed [24], [25], [26], [27], [28], which suggest that genetic changes occurring in committed cells may lead to cancer initiation.

For some solid tumors, it has also been proposed that the first change may occur in differentiated cells; this includes skin cancers [29], and stomach cancers [30]. However, for most other cancers, including colorectal cancers, it is still unknown whether a partially differentiated cell can be a target for cancer initiation.

Despite this uncertainty, many researches feel that differentiated cells are unimportant for cancer initiation, for the following (quantitative) reason, unrelated to biological evidence. Let us concentrate on colon cancer. It is widely believed that the APC gene is a tumor suppressor gene, that is, the inactivation of both copies is required to confer phenotypic changes [1], [31] (consistent with Knudson's two-hit hypothesis [32]). Then an obvious question arises: how can the first mutation occur in the migrating compartment, without being washed away? As John Cairns writes [14], "...there are 256 exponentially multiplying cells that divide twice a day and are being replenished continually by the divisions of a single stem cell, none of these 256 cells will ever be separated from the stem cell by more than eight divisions, and the replication errors made in those eight divisions are destined, of course, to be discarded".

The point of this paper is to address exactly this issue: will migrating cell mutations be indeed discarded, or is there a chance that they will persist until the second hit comes, which immortalizes the cell and thus initiates dysplasia in the crypt? In what follows we show that the intuitive argument given above is incorrect, and that despite their finite life-span, differentiated cells can stay long enough to get the first two hits. Then, if these two hits confer an "immortalizing" phenotypic effect, this scenario can be a likely pathway of cancer initiation.

The main idea of this study. On the one hand, cells in the migrating compartment are short lived, but on the other hand, their number is very large compared to the number of stem cells. Therefore, even though most mutations occurring in the migrating compartment will be washed away, there may be enough of them to assure that some will get the second hit. Our results suggest that it is very likely that at least one of the two hits occurs in the migrating compartment. Moreover, there are circumstances where both hits are likely to happen in the migrating compartment. In particular, we show that if the rate of inactivation of the second copy of APC is elevated (e.g. due to chromosomal instability), then it is possible that both relevant hits happen in the differentiated compartment. This can have important consequences for treatment. So far many authors tended to concentrate on studying the stem cells, as the most likely targets of cancer initiation [7]. Our results suggest that differentiated cells may be as important for initial stages of colon cancer.

At this stage it is difficult to say what really happens at the very moment of cancer initiation, and which cells receive the initial hit. The research in this fascinating area is ongoing[8]. The message of this paper is that the possible role of differentiating cells in this process cannot be discarded simply based on the fact that they are short–lived. More evidence is needed to understand the cellular origins of cancer; more experiments need to be performed to identify whether or not a double hit in a daughter cell can be "immortalizing". In this study we discuss several scenarios and propose some experiments which can shed light on these questions.

This paper is organized as follows. In Section 2 we describe our mathematical model, based on a conventional view of crypt architecture. In Section 3 we identify three possible scenarios for the first two hits, depending whether they occur in stem cells or in the migrating compartment. In Section 4 we calculate their relative likelihood. In Section 5 we propose some experiments which may help to identify cellular origins of cancer.

2 The basic renewal model

We assume that the stem cells are located in a niche at the base of the crypt. They are characterized by an asymmetric division pattern resulting in one stem cell and one proliferative daughter cell. The latter cells divide and populate the migrating compartment. Cells of the migrating compartment go through a number of symmetric divisions, moving toward the crypt surface. On their way up, they go through stages of differentiation, until the fully differentiated cells stop dividing, reach the crypt surface and get shed into the lumen, to be replaced by new generations of cells coming from the bottom of the crypt. We will refer to the symmetrically dividing, migrating progeny of a stem cell (SC) as "differentiated cells" (DC), keeping in mind that the degree of differentiation increases with the number of divisions that separate the progeny from the stem cell.

Figure 1 traces the offspring of one DC created from an SC. Different levels represent consecutive moments of time (or rounds of proliferation). Another interpretation of this figure is spacial: we can think of cells of consecutive generations to be located closer and closer to the top of the crypt. The cells of "age 4" are the closest to the top, and they are shed into the lumen. This can be better seen in figure 2. The progeny of a single daughter cell is marked by the same letter. The degree of maturation/differentiation is reflected in the intensity of shading: the darker the circle is, the more mature is the cell. The apoptotic cells are presented by dashed circles. At all moments of time, the crypt contains DCs of 4 generations. In the beginning (the leftmost diagram) there are progeny of cells A, B and C, and a newly produced daughter cell, D. After some time (the middle diagram), all of the

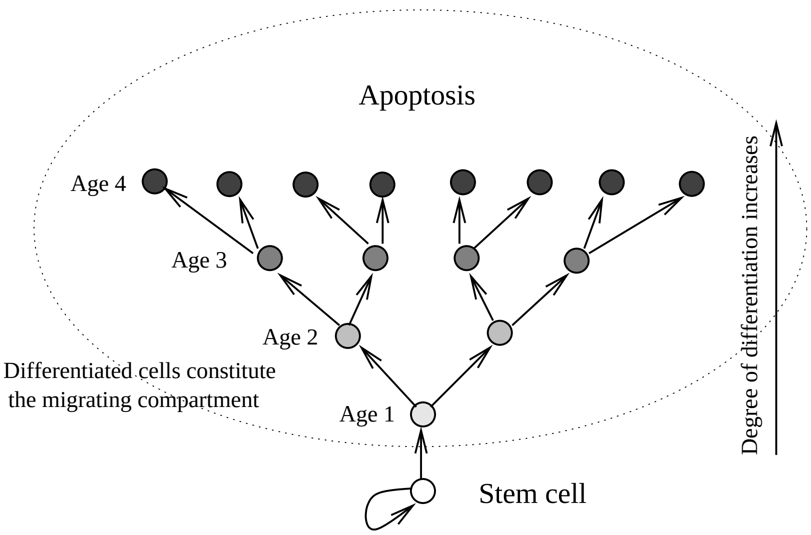


Figure 1: The history of one daughter DC. It undergoes 3 rounds of division. The number of cells in the last generation is 8.

"oldest" cells (marked with A) have been shed, the cells B and C went through one round of division, advancing their degree of maturation and moving upward, and a daughter cell, E, has been produced. The process of renewal goes on in this way, eventually replacing all DCs in the crypt.

In figure 2 only one stem cell is shown to repopulate the crypt. In reality, there are several stem cells per crypt, and thus each crypt is a composition of several clones. This can be easily included in our model: if, for example, there are four stem cells in the crypt, then this crypt can be viewed as a "superposition" of four crypts. The general rule that more mature cells are situated closer to the top of the crypt still holds, and the general dynamics of each of the clones is as in figure 2. The only difference is in the numbers that should be used in the model.

Several models of SC dynamics have been designed [33], [34]. In these papers, two main mechanisms of SC reproduction have been proposed. In the deterministic model, each SC divides asymmetrically, and the number of SC is kept constant. In more sophisticated models, each SC has a probability to produce upon division (i) two SCs, (ii) one SC and one DC, or (iii) two DCs. In this model, the number of stem cells fluctuates. The latter model seems to be more realistic. In the present paper we will use the simpler model, and note that the methods and results developed here should remain the same, with minimal changes, if the reproduction model for stem cells is refined.

3 Three scenarios

Let us first describe the process of accumulation and spread of mutations in a dynamic crypt. By "mutations" we mean any kind of genetic alteration (a point mutation, a loss

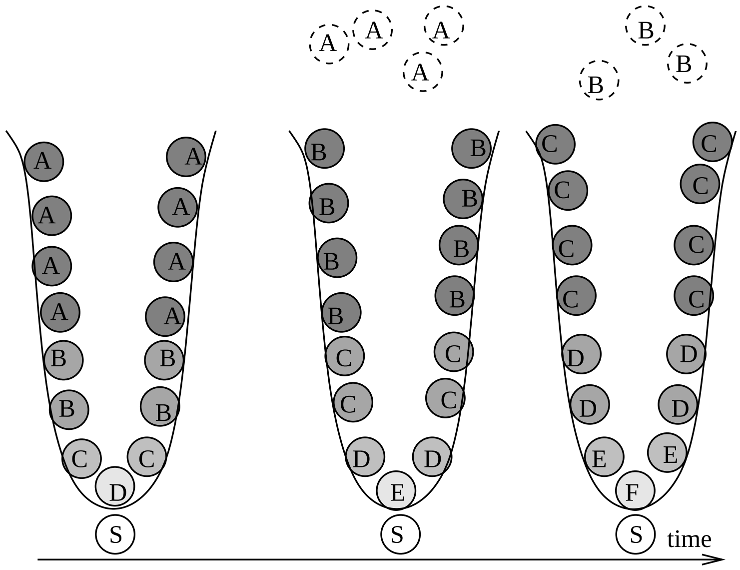


Figure 2: Schematic snapshots of one crypt at three moments of time. The stem cell is marked by “S”.

of heterozigocity (LOH) event, etc) which leads to the inactivation of an allele of the APC gene. We will assume that a cell with a single mutation has the same properties as a wild type cell, and a double-mutant has the ability to avoid apoptosis, continue divisions and thus remain and spread in the crypt.

There are three logical possibilities of how mutations can be accumulated, see figure 3.

(i) In the *ss* scenario, figure 3(a), a mutation happens in the stem cell. Then, after a few divisions, the entire crypt will consist of mutated cells. At some point, a *second* mutation occurs in the SC, shortly after which the entire crypt will consist of double mutants. This is the scenario consistent with the “bottom-up” hypothesis. It predicts that the crypt will be monoclonal with respect to double mutations.

(ii) In the *sd* scenario, figure 3(b), again a mutation occurs in the SC which then spreads throughout the crypt. However, the first double-mutant emerges in the proliferating/migrating compartment. This mutant divides and its progeny first spreads in the upward direction. At this point, the lower part of the crypt is monoclonal with respect to one mutation in the APC gene, and the upper part of the crypt is monoclonal with respect to two mutations.

(iii) In the *dd* scenario, figure 3(c), a mutation occurs in one of the migrating daughter cells. The cell divides, its progeny moves in the upward direction, but before it undergoes apoptosis, one of these cells experiences a second hit, creating a double mutant. As a result, the lower part of the crypt consists of wild-type cells, and the upper part is composed of monoclonal double-mutants.

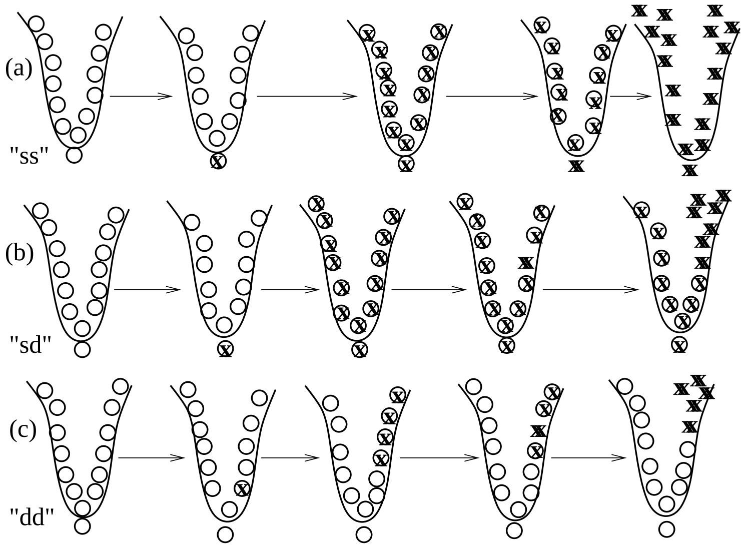


Figure 3: Three scenarios of the emergence of a double mutant. (a), the *ss* scenario. (b), the *sd* scenario. (c), the *dd* scenario. The open circles are wild-type cells, circles with an "x" contain one mutation in the tumor-suppressor gene, and double-x's are double mutants.

4 Results

It is convenient to introduce the quantity l, the total number of division rounds during the life-span of one clone, see figure 1. This number includes one asymmetric division of the stem cell and $l-1$ rounds of symmetric divisions of the DCs (for simplicity we assume that these are synchronized). The total number of progeny of an SC existing at any one time in a crypt, is $2^l - 1$. We denote $N = 2^l$. For example, in figure 1 we have $l = 4$, $N = 16$.

Since the probability of the first hit, p_1, is very small, it can be shown that most of the time the population of the crypt will be homogeneous, that is, most of the cells will either be wild-type, or will contain one mutation, see also [35]. Indeed, if a mutation arises in a DC, it gets washed out in less than l time-steps; in fact, most of the mutants are very short-lived, and only survive for one or two time-steps, because at each moment of time, the majority of the crypt consists of cells only one or two steps away from apoptosis, see figure 1. The frequency with which new mutants are created is Np_1, and only about $1/N$ of these mutations will happen in an SC. So the condition $p_1 \ll 1/N$ (the *homogeneity* condition) guarantees that the crypt contains no mutants most of the time. Unless, of course, a mutation occurs in an SC, in which case in less than l time-units, the entire crypt will consist of mutant cells. Finally, a double mutant may appear, but in this case we assume that the process is over and a dysplastic crypt has been created. Note that the homogeneity condition easily holds for the realistic values of the mutation rate[36] , $p_1 \sim 10^{-7}$, and the population size[6], [13] $N \sim 10^3$.

Let us call the probability to find the entire crypt consisting of wild-type cells, x_0. The probability that the entire crypt consists of cells with a single mutation is x_1, and the probability that the crypt is dysplastic (contains one or more double mutants) is x_2. The symbols X_0, X_1 and X_2 will be used to denote the corresponding states. Because of the homogeneity condition, we have $x_0 + x_1 + x_2 \approx 1$. A double mutant can be created via two major pathways. One pathway includes a fixation of a single mutant. First, a mutation happens in an SC (the rate is p_1), after which the entire crypt enters state X_1, and then a new mutation occurs (with rate Np_2), which brings the crypt to the state X_2 (note that this pathway combines the *ss* and *sd* scenarios). Alternatively, a second mutation can occur in a mutant clone which originates in a DC, without a prior fixation of a single mutant (the *dd* scenario). This happens with the rate R, which we calculate in Appendix A,

$$R \approx \frac{Np_1p_2(|\log p_2| - \log 2 - \gamma)}{\log 2} \tag{1}$$

(here γ is the Euler's constant, $\gamma \approx 0.577$). We have the following equations for the mutation processes:

$$\begin{aligned} \dot{x}_0 &= -p_1x_0 - Rx_0, & (2)\\ \dot{x}_1 &= p_1x_0 - Np_2x_1, & (3)\\ \dot{x}_2 &= Np_2x_1 + Rx_0. & (4) \end{aligned}$$

Solving equations (2-4), we obtain (see Appendix A) that the probability of the *dd* scenario is given by,

$$x^{dd} = \frac{R(1 - e^{-(p_1+R)t})}{p_1 + R}. \tag{5}$$

The probability of the *sd* scenario is,

$$x^{sd} \approx \frac{p_1 A}{(p_1 + R)(p_1 + R - Np_2)}, \tag{6}$$

where $A = Np_2(e^{-(p_1+R)t} - 1) - (p_1 + R)(e^{-Np_2t} - 1)$. Finally, the probability of the *ss* scenario is small compared to the *sd* scenario,

$$x^{ss} = x^{sd}/N. \tag{7}$$

We can calculate the relative contribution of the probability of two DC mutations. It turns out that if the condition

$$p_2|\log p_2| > \frac{1}{N} \tag{8}$$

is satisfied, then the probability to obtain a double mutant by two hits in the DC compartment is larger than the probability to first get a mutation in an SC. It is interesting that this condition only depends on the second mutation rate, p_2, and is independent on the first mutation rate, p_1.

To check our analytical results, we used numerical simulations to study the crypt dynamics. We trace the offspring of one stem cell; the process goes on until the first double mutant is created. We record whether each of the hits in the double mutant was the result of an SC or a DC mutation, and then we stop the simulation for the crypt. This process is repeated many times, and then the probability of different pathways is estimated. We ran the simulation with 10^3 and 10^7 realizations. The results are almost the same which suggests convergence.

Figure 4 plots the probability of having a double mutant via different pathways by time n. The curves represent the analytical prediction by formulas (5), (6) and (7). The points are results of numerical simulation and they agree with the calculations very well. Notice that for $p_1 = p_2 = 10^{-4}$ (figure 4(a)) the probability of having a double mutant initiated in an SC (pathway sd) is greater than the probability of having a double mutant initiated in a DC (pathway dd). On the other hand, when p_2 increases to $p_2 = 10^{-3}$, the situation reverses, see figure 4(b). The reversal of the two scenarios is predicted correctly by formula (8): if $N = 2^9 = 512$, this condition holds for $p_2 = 10^{-3}$ and it does not hold for $p_2 = 10^{-4}$.

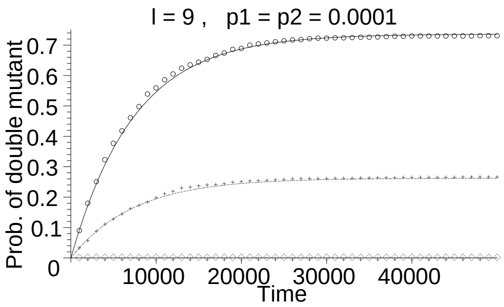
Figure 4(c) shows the situation corresponding to colon cancer initiation in the presence of chromosomal instability. The first mutation rate, $p_1 = 10^{-7}$, corresponds to the basic point mutation rate. The second mutation rate corresponds to a highly elevated rate of LOH in unstable cancers, $p_2 = 10^{-2}$. We can see that the dd scenario prevails in this case.

Simulating crypt dynamics for lower values of mutation rates, e.g., $p_1 = p_2 \approx 10^{-7}$, which corresponds to both hits occurring with the basic point-mutation rate, is very time-consuming. More sophisticated numerical methods could give a faster performance, but here we would like to emphasizes the value of analytical results. Our formulas are valid for arbitrarily small values of mutation rates, and in fact, the precision of the method grows as $p_{1,2}$ decrease!

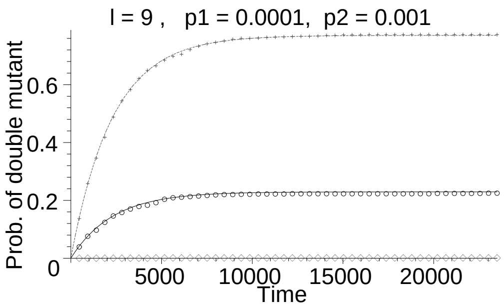
Finally, we have derived an exact analytical formula for the total probability of creating a double mutant, see Appendix B. This was used to check the simulations by comparing the numerical value of $x^{ss}+x^{sd}+x^{dd}$ with the exact formula, expression (11) in Appendix B . The formula gave a perfect agreement with simulation results (data not shown). Formula (11) works very well for large values of $p_{1,2}$. For instance, it can give the probability of producing a double mutant, where the approximation of this section breaks down, that is, in the regime where $p_1 \sim 1/N$. However, the exact calculation of Appendix B has the same disadvantage as the numerical simulations: for low values of $p_{1,2}$ it becomes difficult to implement. The approximate formulas (5), (6) and (7) can be used for all realistic values of mutation rates.

5 Conclusions

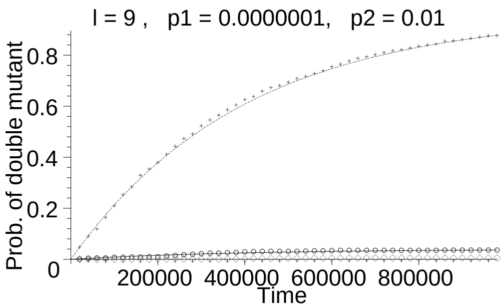
We considered the renewal dynamics of a colon crypt, repopulated by asymmetric divisions of stem cells. Dysplasia occurs as a result of inactivation of both copies of the APC gene (or other genes in the Wnt pathway). Thus, it takes two hits before phenotypic changes occur. There are three pathways for the two hits: ss (both hits occur in an SC), sd (the first hit occurs in an SC, and the next hit is acquired by one of the DC in its clone), and dd (the first hit is acquired in a DC, and the second hit occurs in one of its progeny). In this paper we used mathematical methods to calculated the probability of each of the pathways. The results can be summarized as follows:



(a)



(b)



(c)

Figure 4: The probability to acquire a double mutant as a function of time, with $l = 9$. Numerical simulations are compared with analytical results. "○" denotes the simulation for *sd* pathway, "+" denotes the simulation for *dd* pathway, and "◇" denotes the simulation for *ss* pathway. In (a), pathway *sd* is more likely. In (b), where p_2 is an order of magnitude larger, pathway *dd* prevails. In (c), *dd* pathway prevails. Pathway *ss* is always the least likely scenario.

Table 1: The most likely pathway for the two hits

	Number of stem cells per crypt					
p_2	1	2	4	8	16	32
10^{-7}	*sd*	*sd*	*sd*	*sd*	*sd*	*sd*
10^{-5}	*sd*	*sd*	*sd*	*sd*	*sd*	*sd*
10^{-4}	*dd*	*sd*	*sd*	*sd*	*sd*	*sd*
10^{-3}	*dd*	*dd*	*dd*	*dd*	*sd*	*sd*
10^{-2}	*dd*	*dd*	*dd*	*dd*	*dd*	*dd*

1. The probability of the *ss* pathway is negligible. That is, it is unlikely that both hits occur in the SC.

2. As a consequence, at least one of the hits will occur in the migrating, proliferative compartment. This is consistent with the observation that below the dysplastic layer, cells in the crypt do not exhibit the $\text{APC}^{-/-}$ phenotype. The relative importance of *sd* and *dd* pathways depends on the parameters of the system.

3. In particular, if $p_2|\log p_2| > 1/N$, the *dd* pathways becomes more important. It means that the first double mutant will appear *outside* of the SC compartment. This contradicts previous thinking that SCs are crucially important for colon cancer initiation, being the first mutational "targets".

4. If the reverse condition holds, that is, if $p_2|\log p_2| < 1/N$, then the most likely scenario is the *sd* pathway, that is, first an SC acquires a mutation, then all of its offspring contain an inactivated APC copy, and then one of these daughter cells acquires a second hit. According to the *sd* scenario, the crypt below the dysplastic cells should contain cells with one inactivated and one functional copy of the APC gene.

One remarkable property of the model is a very small number of parameters that it contains. For instance, the condition which identifies whether an SC mutation is important (that is, which of the pathways *sd* and *dd* is more likely), only depends on two parameters, p_2 and N. Table 1 identifies the most important pathway, depending on the parameter p_2 and the number of SCs per crypt. The parameter N can be found as follows. If there are n SCs per crypt, then N is the size of the crypt divided by n, that is, the "share" of each of the SCs in the crypt. We suppose that there are $2,000$ cells in a crypt, and the number of SCs per crypt[6] is varied between 1 and 32.

The parameter p_2 is the rate of inactivation of the second allele of the APC gene. In Table 1, it is varied between 10^{-7} per cell division (the basic point mutation rate [36]) and 10^{-2}. The elevated values of p_2 may be related to the phenomenon of chromosomal instability, which is commonly observed in colon cancers [37]. There is evidence that chromosomal instability arises very early in colorectal cancers [38], and the *in vitro* estimates of the corresponding elevated rate of LOH give the value as high as $p_2 = 10^{-2}$ per cell division [39]. We can see that without chromosomal instability (low values of mutation rate), the pathway *sd* is more likely. For elevated rates of APC inactivation, it is more likely that both hits will occur in the migrating compartment.

The model suggests that a possible role of the migrating compartment in cancer initiation cannot be discarded based on the argument that differentiated cells are short–lived. However, the final resolution of the question of cellular origins of cancer will require better understanding of the effects of a double hit in a differentiated cell. If both copies of the APC gene are inactivated in a DC, does it make it cancerous? This important and difficult question can be addressed in a variety of ways. Based on our "population" approach, we propose an indirect experiment which can shed light onto this issue.

It may be possible to distinguish between the three possible scenarios, *ss*, *sd* and *dd*, in the following experimental set-up. Examining cells in an initiated crypt below the dysplastic layer can yield one of the three results: (a) the cells may contain two inactivated copies of the APC gene, (b) the cells may contain one inactivated and one functional copy of the APC gene, or (c) the cells may be wild type with respect to the APC gene. In case (a) we have the *ss* scenario, that is, both hits occurred in the stem cell. Case (b) is the signature of the *sd* scenario, that is, the first mutation happened in a stem cell and the second hit was acquired in the migrating compartment. Finally, outcome (c) suggests the *dd* scenario, that is, both hits occurred in the migrating compartment. The complication of such an experiment is of course the fact that crypts may be polyclonal (that is, contain more than one SC), and therefore many cells must be checked to arrive to a decisive conclusion.

References

[1] Kinzler KW, Vogelstein, B. The Genetic Basis of Human Cancer (McGraw-Hill, Toronto 1998).

[2] The adenomatous polyposis coli (APC) tumor suppressor. Biochim Biophys Acta. 1997; 1332(3):F127-47.

[3] Katoh M. WNT2 and human gastrointestinal cancer. Int J Mol Med. 2003; 12(5):811-6.

[4] Polakis P. The oncogenic activation of beta-catenin. Curr Opin Genet Dev 1999;9(1):15-21.

[5] Behrens J, Jerchow BA, Wurtele M, Grimm J, Asbrand C, Wirtz R, Kuhl M, Wedlich D, Birchmeier W. Functional interaction of an axin homolog, conductin, with beta-catenin, APC, and GSK3beta. Science 1998;280(5363):596-9.

[6] Potten CS, Loeffler M. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties Lessons for and from the crypt. Development 1990; 110: 1001-1020.

[7] Potten CS, Booth C, Hargreaves D. The small intestine as a model for evaluating adult tissue stem cell drug targets. Cell Prolif 2003; 36: 115-129.

[8] Vogelstein B, Kinzler, K W. Cancer genes and the pathways they control. Nat Med 2004; 10: 789-799.

[9] Wong WM, Wright NA. Cell proliferation in gastrointestinal mucosa. J Clin Pathol 1999; 52: 321-333.

[10] Bach SP, Renehan AG, Potten CS. Stem cells: the intestinal stem cell as a paradigm. Carcinogenesis 2000; 21: 469-476.

[11] Fuchs E, Segre JA. Stem cells: a new lease on life. Cell 2000; 100: 143-155.

[12] Winton,JD. Stem cells in the epithelium of the small intestine and colon, in Stem Cell Biology, eds Marshak DR, Gardner RL, Gottlieb D (Spring Harbor Laboratory Press, Cold Spring Habor 2001): 515-536.

[13] Kim KM, Shibata D. Methylation reveals a niche: stem cell succession in human colon crypts. Oncogene 2002; 21:5441-5449.

[14] Cairns J Somatic stem cells and the kinetics of mutagenesis and carcinogenesis. Proc Natil Acad Sci USA 2002; 99: 10567-10570.

[15] Wright NA, Poulsom R. Top down or bottom up? Competing management structures in the morphogenesis of colorectal neoplasms. Gut 2002; 51: 306-308.

[16] Preston SL, Wong WM, Chan AO, Poulsom R, Jeffery R, Goodlad RA, Mandir N, Elia G, Novelli M, Bodmer WF, et al. Bottom-up histogenesis of colorectal adenomas: origin in the monocryptal adenoma and initial expansion by crypt fission. Cancer Res 2003; 63: 3819-3825.

[17] Wong WM, Mandir N, Goodlad RA, Wong BC, Garcia SB, Lam SK Wright NA. Histogenesis of human colorectal adenomas and hyperplastic polyps: the role of cell proliferation and crypt fission. Gut 2002; 50: 212-217.

[18] Shih IM, Wang TL, Traverso G, Romans K, Hamilton SR, Ben-Sasson S, Kinzler KW, Vogelstein B. Top-down morphogenesis of colorectal tumors. Proc Natil Acad Sci USA 2001; 98: 2640-2645.

[19] Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. Nat Rev Cancer 2001; 1: 55-67.

[20] Lamprecht SA, Lipkin M. Migrating colonic crypt epithelial cells: primary targets for transformation. Carcinogenesis 2002; 23: 1777-1780.

[21] van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, et al. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002; 111: 241-250.

[22] Grimwade D, Enver T. Acute promyelocytic leukemia: where does it stem from? Leukemia 2004; 18: 375-84.

[23] Turhan AG, Lemoine FM, Debert C, Bonnet ML, Baillon C, Picard F, Macintyre EA, Varet B. Highly purified primitive hematopoietic stem cells are PML-RARA negative and generate nonclonal progenitors in acute promyelocytic leukemia. Blood 1995; 85: 2154-2161.

[24] Brown D, Kogan S, Lagasse E, Weissman I, Alcalay M, Pelicci PG, et al. Free in PMC A PMLRARalpha transgene initiates murine acute promyelocytic leukemia. Proc Natl Acad Sci USA 1997; 94: 2551-2556.

[25] Grisolano JL, Wesselschmidt RL, Pelicci PG Ley TJ. Altered myeloid development and acute leukemia in transgenic mice expressing PML-RAR alpha under control of cathepsin G regulatory sequences. Blood 1997; 89: 376-387.

[26] He LZ, Tribioli C, Rivi R, Peruzzi D, Pelicci PG, Soares V et al. Acute leukemia with promyelocytic features in PML/RARalpha transgenic mice Proc Natl Acad Sci USA 1997; 94: 5302-5307.

[27] Westervelt P, Lane AA, Pollock JL, Oldfather K, Holt MS, Zimonjic DB et al. A high penetrance mouse model of acute promyelocytic leukemia with very low levels of PML-RAR expression. Blood 2003; 102: 1857-1865.

[28] Westervelt P, Ley TJ. Seed vs soil: The importance of the target cell for transgenic models of human leukemias. Blood 1999; 93: 2143-2148.

[29] Perez-Losada J, Balmain A. Stem-cell chierarchy in skin cancer. Nature Reviews Cancer 2003; 3: 434-443.

[30] Kirchner T, Muller S, Hattori T, Mukaisyo K, Papadopoulos T, Brabletz T Jung A. Metaplasia, intraepithelial neoplasia and early cancer of the stomach are related to dedifferentiated epithelial cells defined by cytokeratin-7 expression in gastritis. Virchows Arch 2001; 439: 512-522.

[31] Macleod K. Tumor suppressor genes. Curr Opin Genet Dev 2000; 10:81-93.

[32] Knudson AG. Hereditary cancer: two hits revisited. J Cancer Res Clin Oncol 1996; 122: 135-140.

[33] Ro S, Rannal B. Methylation patterns and mathematical models reveal dynamics of stem cell turnover in the human colon. Proc Natil Acad Sci USA 2001; 98: 10519-10521.

[34] Yatabe Y, Tavare S, Shibata D. Investigating stem cells in human colon by using methylation patterns. Proc Natil Acad Sci USA 2001; 98: 10839-10844.

[35] Komarova NL, Sengupta A, Nowak MA. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. J Theor Biol 2003; 223: 433-450.

[36] Albertini RJ, Nicklas JA, O'Neill JP, Robison SH. In vivo somatic mutations in humans: measurement and analysis. Annu Rev Genet 1990; 24: 305-326.

[37] Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998; 396: 643-649.

[38] Shih IM, Zhou W, Goodman SN, Lengauer C, Kinzler KW, Vogelstein B. Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. Cancer Res 2001; 61: 818-822.

[39] Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. Nature 1997; 386: 623-627.

[40] Moolgavkar SH, Dewanji A, Venzon DJ. A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor. Risk Anal 1998; 8: 383-392.

[41] Nowak MA, Komarova NL, Sengupta A, Jallepalli PV, Shih IM, Vogelstein B, Lengauer C. The role of chromosomal instability in tumor initiation. Proc Natil Acad Sci USA 2002; 99: 16226-16231.

[42] Iwasa Y, Michor F, Nowak MA. Stochastic tunnels in evolutionary dynamics. Genetics 2003; 166: 1571-1579.

A Calculating the probability of different scenarios

To calculate the rate R in equations (2) and (4), let us write down the probability to have at least one mutant such that both mutations happen in the DC (given that no SC mutations have happened). We assume that the mutation rate, p_1, is sufficiently small such that the clones can be treated independently (the condition is $p_1 N \ll 1$), and consider a doubly stochastic process, see also [40], [42]. We obtain

$$R = \sum_{i=1}^{l} r_i \mathcal{P}_i.$$

To see this, we write down the total rate of primary mutations, $\sum_{i=1}^{l} p_1 2^{i-1} = p_1 N$ (this is valid for $p_1 N \ll 1$). Then each term $r_i = p_1 2^{i-1}$ is a contribution corresponding to the first mutation happening in a DC of generation i. $\mathcal{P}_i$ is related to the "secondary" stochastic process which happens in the clone after the first mutation. Inside such a clone, we have $2^{l-i+1} - 2$ cell divisions, so that the total probability to get a second hit is

$$\mathcal{P}_i = 1 - (1 - p_2)^{2^{l-i+1}-2} \approx 1 - e^{-p_2(2^{l-i+1}-2)}.$$

We have

$$R = p_1 \sum_{i=1}^{l} 2^{i-1}(1 - e^{-p_2(2^{l-i+1}-2)}).$$

This expression can be calculated by replacing the summation with an integral,

$$R = \frac{N p_1}{\log 2}\left[\frac{1}{2} - \frac{1}{N} - e^{2p_2}\int_{\frac{1}{N}}^{\frac{1}{2}} e^{-p_2/x}\, dx\right].$$

In the limit of small p_2, we have expression (1).

We would like to comment on the previous results obtained for the rate at which colon cancer is initiated. [35] and [41] introduce the phenomenon of "stochastic tunneling" where a second hit occurs before the first hit has had a chance to reach fixation. There, a simpler model is considered where the population of the crypt is homogeneous and there is no distinction between stem cells and their differentiated progeny. For such models, the rate R, roughly speaking, is given by $N p_1 \sqrt{p_2}$. We can see that taking account of the structure of the colon changes these results. In particular, the new rate given by (1) is lower (because $p_2 |\log p_2| < \sqrt{p_2}$).

The probability to obtain a double mutant, x_2, is given by equation (4). It can be rewritten as a sum of three contributions,

$$x_2 = x^{dd} + x^{sd} + x^{ss}.$$

Here, the probability to obtain a double mutant by two DC mutations is given by

$$\dot{x}^{dd} = R x_0,$$

which yields equation (5). The probability to obtain a double mutant by first mutating the stem cell is

$$\dot{x}^{sd} + \dot{x}^{ss} = Np_2x_1, \tag{9}$$

where x^{ss} refers to the pathway where both mutations occur in the SC, and x^{sd} implies the second mutation in a DC. We have the following intuitive relation,

$$x^{ss}/x^{sd} \approx 1/N. \tag{10}$$

This means that the vast majority of double mutants will acquire the second mutation in the differentiated stage. We can solve equations (2), (3) and (9) and use equation (10) to eliminate x^{ss} to obtain to obtain expressions (6) and (7).

As time increases, we have

$$\lim_{t\to\infty} x^{dd} = \frac{R}{p_1 + R}.$$

This means that if the condition $R > p_1$ is satisfied, then the majority of double mutants will acquire both mutations in a DC. Using expression (1), we obtain condition (8).

A technical note: very little elaboration is required to obtain the relationship between x^{ss} and x^{sd}, equation (10). This simple relationship provides important insight that "stem cells are not the entire story, and the second hit is more likely to occur in a DC". However, we need to do more work to answer the following question: how likely is it that **both** hits fall outside the stem cell compartment? This is where the above model becomes necessary.

B Exact formula for total probability of double mutations

Here we calculate the probability of having a double mutant in a crypt, as a function of time. We focus on the progeny of one stem cell. The total number of cells in the crypt, $N(n)$, is

$$N(n) = \begin{cases} 2^n, & n < l \\ 2^l & n \geq l. \end{cases}$$

Note that $n < l$ represents the "development" stage of the crypt. After $n = l$, the size of the crypt remains constant. Each cell division may result in creating mutants. The probability of mutation is p_1, so that with probability $2p_1(1 - p_1)$ one mutant is created, and with probability p_1^2 two mutants are created. A mutated cell can acquire a second mutation with probability p_2.

Let us trace mutations in one clone. The number of mutants at time n (with one mutation) is a random variable, whose values go from 0 to 2^{l-1}. If a double mutant is created, we set this random variable to the value E (for "end"). This is an absorbing state because once a double mutant is created, it is assumed to stay in the system. We have a time-inhomogeneous Markov process for the variable j, $j \in \{0, 1, \ldots, 2^{l-1}, E\}$. The probability distribution for the variable j at time n is a row-vector $p^{(n)}$, whose evolution is given by

$$p^{(n)} = p^{(n-1)} M^{(n)}, \quad n = 1, 2, \ldots.$$

The transition matrices, $M^{(n)}$, can be easily written down. The first matrix corresponds to the asymmetric division of the SC, and is given by

$$M^{(1)}_{00} = 1 - p_1, \quad M^{(1)}_{01} = p_1, \quad M^{(1)}_{11} = 1 - p_2, \quad M^{(1)}_{1E} = p_2, \quad M^{(1)}_{EE} = 1,$$

with all the rest of the elements being zero. The matrices corresponding to symmetric divisions of DCs are given by the following:

$$M^{(n)}_{j,2j+k} = \left(\begin{array}{c} 2^{n-1} - 2j \\ k \end{array} \right) p_1^k (1-p_1)^{2^{n-1}-2j-k}(1-p_2)^{2j}, \quad 0 \le k \le 2^{n-1} - 2j, \quad 0 \le j \le 2^{n-2},$$

$$M^{(n)}_{j,E} = 1 - (1-p_2)^{2j}, \quad 0 \le j \le 2^{n-2},$$

$$M^{(n)}_{E,E} = 1,$$

with all the rest of the elements being zero. Let us suppose that the SC is not mutated at time $n = 0$. Let us consider one clone in isolation. The probability to have at least one double mutant by time n is given by

$$P_0(E;n) = \left[(1,0,\ldots) \prod_{s=1}^{n} M^{(s)} \right]_E, \quad n \le l, \quad P_0(E;n) = P_0(E;l), \quad n > l;$$

here $[\ldots]_E$ stands for the Eth entry of a row-vector. Note that after l divisions starting from SC, the cumulative probability does not change anymore because the clone is removed. Similarly, if the SC is mutated at time 0, then the corresponding probability is

$$P_*(E;n) = \left[(0,1,\ldots) \prod_{s=1}^{n} M^{(s)} \right]_E, \quad n \le l, \quad P_*(E;n) = P_*(E;l), \quad n > l.$$

This formula can be simplified to give

$$P_*(E;n) = \sum_{j=0}^{\mathcal{N}(n)-1} (1 - (1-p_2)^{2^j})(1-p_2)^{2^j - 1},$$

where the upper limit of summation is $\mathcal{N}(n) = n$ for $n \le l$, and $\mathcal{N}(n) = l$ for $n > l$.

We would like to calculate the probability to have at least one double mutant in the system by time n. We consider a sequence of clones. There are n divisions of the SC. Let us denote the event of acquiring a mutation in the SC at time k, and no further mutations in the SC before time n, by $S_{k,n}$. We have

$$Prob(S_{k,n}) = (1-p_1)^{k-1} p_1 (1-p_2)^{n-k}.$$

Given $S_{k,n}$, the probability to have no double mutant by time n is

$$Prob(\bar{E}; n | S_{k,n}) = \prod_{j=1}^{k} (1 - P_0(E; n-j+1)) \prod_{j=k+1}^{n} (1 - P_*(E; n-j+1)).$$

Each clone is originated at time j, so its "age" is $n-j+1$. Similarly, the probability to have no double mutant by time n, given that no SC mutations have occurred by time n, is

$$Prob(\bar{E};n|\text{no SC mut}) = \prod_{j=1}^{n}(1-P_0(E;n-j+1)),$$

and the probability to have no SC mutations by time n is $(1-p_1)^n$. The total probability to have at least one double mutant by time n is given by

$$Prob(E;n) = 1 - \left[\sum_{k=1}^{n} Prob(\bar{E};n|S_{k,n})Prob(S_{k,n}) + Prob(\bar{E};n|\text{no SC mut})(1-p_1)^n\right]. \quad (11)$$

This formula was compared with numerical simulations for the total probability of a double mutant, and with the approximate expression for $x^{ss}+x^{sd}+x^{dd}$ derived in Appendix A. The agreement with the numeric is perfect. The agreement with the approximation is very good, if the following consideration is taken into account. Note that the expressions for x^{ss}, x^{sd} and x^{dd} (formulas (5), (6) and (7)) were derived assuming that the size of the crypt is always N. On the other hand, in the simulations and in formula (11) we have a "development" phase where the size of the crypt is growing. The growth is exponential and thus during the first $l-2$ steps, the population size is significantly less than N, and thus mutation probabilities are negligible. This is why in order to get a perfect fit, the expressions for x^{ss}, x^{sd} and x^{dd} should be shifted by $l-2$ steps.