[Cancer Biology & Therapy 1:6, 685-692, November/December 2002]; ©2002 Landes Bioscience

Research Paper

Dynamics of Genetic Instability in Sporadic and Familial Colorectal Cancer

Natalia L. Komarova[1,2]
Christoph Lengauer[3,*]
Bert Vogelstein[3,4]
Martin A. Nowak[1]

[1]Institute for Advanced Study, Princeton, New Jersey USA

[2]Department of Applied Mathematics; University of Leeds; Leeds UK

[3]Sidney Kimmel Comprehensive Cancer Center; Johns Hopkins University; Baltimore, Maryland USA

[4]Howard Hughes Medical Institute; Baltimore, Maryland USA

*Correspondence to: Natalia L. Komarova; Institute for Advanced Study; New Jersey USA; Email: natalia@ias.edu

Received 08/29/02; Accepted 09/15/02

Previously published online as a CB&T "Paper in Press" at: http://landesbioscience.com/journals/cbt/

KEY WORDS

Chromosomal instability, Microsatellite instability, Mathematical modeling, Stochastic dynamics, FAP, HNPCC

ABSTRACT

Genetic instability is a defining feature of human cancer. In colorectal cancer, two specific types of genetic instabilities have been identified: microsatellite instability (MIN) leads to a 1000-fold increase in the rate of subtle DNA changes, whereas chromosomal instability (CIN) enhances the rate at which gross chromosomal changes occur during cell division. In this paper, we develop a mathematical model for the dynamics of colon cancer initiation. We outline the processes and rate constants that determine the fraction of colon cancers where MIN or CIN mutations precede the inactivation of the first tumor suppressor gene. For a wide range of parameter values, we find support for the radical hypothesis that genetic instability initiates colonic tumorigenesis. We compare sporadic and hereditary forms of colorectal cancer.

©2003 Landes Bioscience. Not for distribution.

INTRODUCTION

It is now widely accepted that alterations of at least two types of genes constitute the pathway to cancer. The first type involves genes that directly control cell birth and death processes. Such alterations may lead to a clonal expansion and dysplastic growth patterns. The second type of genetic changes affects genes that control the mutation rate of a cell. These changes result in genetic instability. The corresponding mutants may not have a direct fitness advantage compared to the wild-type, but the increased rate of genetic change enhances the production of further mutants, some of which can be advantageous and start a new wave of clonal expansion. There are at least two major types of genetic instability: microsatellite instability (MIN), and chromosomal instability (CIN).[1,2] MIN is characterized by an increased rate of point mutations and other subtle changes of the genome, and is related to defects in the mismatch repair system.[3] CIN is characterized by an increased rate of chromosomal alterations, such as loss and gain of whole chromosomes or parts of chromosomes. A consequence of CIN is an increased rate of loss of heterozygocity (LOH).

Colorectal cancer is a major cause of mortality in the Western world.[4] Approximately 5% of the population develop the disease, and about 40% of those diagnosed with it die within 5 years. Considerable progress has been made in identifying genetic events leading to colorectal cancer. Somatic inactivation of the Adenomatous Polyposis Coli (APC) gene is believed to be one of the earliest steps occurring in sporadic colorectal cancer.[5] It has been observed that the frequency of APC mutations is as high in small lesions as it is in cancers. Evidence that the APC gene plays a crucial role in colorectal cancer also comes from the study of individuals with familial adenomatous polyposis coli (FAP). FAP patients inherit a mutation in one of the copies of the APC gene; by their teens, they harbor hundreds to thousands of adenomatous polyps.

The APC gene is a tumor suppressor gene which controls cell birth and cell death processes.[6] Inactivation of only one copy of the APC gene does not seem to lead to any phenotypic changes. Inactivation of both copies of this gene appears to result in an increased cell birth to death ratio in the corresponding cell and leads to clonal expansion and the formation of a dysplastic crypt. In this paper, we define a dysplastic crypt as a crypt that consists of cells with both copies of the APC gene inactivated. Dysplastic crypts are at risk of developing further somatic mutations which will eventually lead to cancer. Dysplastic crypts are different from aberrant crypt foci, which consist of abnormal, but non-dysplastic crypts. Aberrant crypt foci often contain KRAS mutations and are not believed to progress to neoplasia. In this paper we model the inactivation of APC and therefore only consider dysplastic crypts. The typical estimate is that an average 70 year old has about 1–10 dysplastic crypts, but precise counts have never been published.

Table 1A **PARAMETERS OF THE MODEL AND THEIR POSSIBLE NUMERICAL VALUES**

Quantity	Definition	Range
M	Number of crypts in a colon	10^7
N	Effective number of cells in a crypt	1–100
τ	Effective time of cell life cycle, days	1–20
u	Probability of mutation in normal (non-MIN)cells	10-7
$\acute{u}$	Probability of mutaiton in MIN cells	10^{-4}
p_0	Rate of LOH in normal (non-CIN) cells	10^{-7}
p	Rate of LOH in CIN cells	10^{-2}
n_m	Total number of MIN genes	2–5
n_c	Total number of CIN genes	~100

Table 1B **THE THREE MAJOR CLASSES OF HOMOGENOUS STATES**

Quantity	Definition	Point mutation rate	Rate of LOH
X_O, X_1, X_2	non-CIN, non-MIN	u	p_0
Y_O, Y_1, Y_2	CIN	u	$p > p_0$
Z_O, Z_1, Z_2	MIN	$\acute{u} > u$	p_0

Paramets and notations. In Table 1A, the mutation and LOH rates are given per gene per cell division.

About 13% of all colorectal cancers have MIN and most of the rest are characterized by CIN.[1] MIN occurs in virtually all hereditary non-polyposis colorectal cancers (HNPCC), which account for about 3% of all colorectal cancers. The MIN phenotype results from defective mismatch repair. Several genes have been identified whose inactivation leads to an increased rate of subtle genetic alterations, the main ones being hMSH2 and hMLH1. Both copies of a MIN gene must be inactivated in order for any phenotypic changes to occur. HNPCC patients inherit a mutation in one of the copies of a MIN gene and often develop colorectal tumors in their forties. Unlike FAP patients, they do not have a vastly increased number of polyps, but the rate of progression from polyp to cancer is faster.

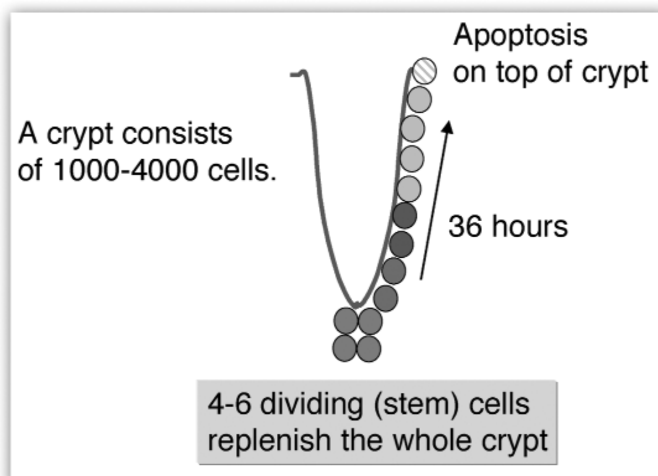

Figure 1. The epithelium of the colon is organized into crypts. Each crypt contains about 1,000–4,000 cells. A small number of stem cells, which are thought to be located at the bottom of the crypt, divide asymmetricaly to replenish the whole crypt. They give rise to differentiated cells which travel within approximately 36 hours to the top of the crypt where they undergo apoptosis. Inactivation of both copies of the APC gene is believed to prevent apoptosis. The mutated cells remain on the top of the crypt, continue to divide and ultimately take over the crypt. This process gives rise to a dysplastic crypt, which represents the first step on the way to colorectal cancer.

Molecular mechanisms leading to CIN in human cancers remain to be understood. It has been proposed that CIN may be caused by mutations in genes involved in centrosome/microtubule dynamics, or checkpoint genes that monitor the progression of the cell cycle, e.g., the spindle checkpoint or the DNA-damage checkpoint.[7] For example, heterozygous mutations in the mitotic spindle checkpoint gene hBUB1 have been detected in a small fraction of colorectal cancers with the CIN phenotype.[8-11] Also, the MAD2 gene seems to be transcriptionally repressed in various solid tumors.[12-14] Some CIN genes can act in a dominant-negative fashion: an alteration in one allele leads to CIN. There is an ongoing and heated debate about the role of genetic instability in cancer progression. In this paper we attempt to formalize the question with mathematical modeling, and formulate exact conditions under which CIN or MIN precede the inactivation of the first tumor suppressor gene. We compare sporadic colorectal cancer with FAP and HNPCC. We show which processes and rate constants are compatible with existing experimental observations and outline what needs to be measured to improve understanding

A MODEL FOR THE INITIATION OF SPORADIC COLORECTAL CANCERS

The colonic epithelium is organized in crypts covered with a self-renewing layer of cells (Fig. 1). The total number of crypts is of the order of $M=10^7$ in a human. Each crypt contains around 1000–4000 cells. A crypt is renewed by a small number of (stem) cells (perhaps 1–10).[15,16] The life cycle of stem cells is of the order of 1–20 days.[17,18] Stem cells give rise to differentiated cells which divide at a faster rate, travel to the top of the crypt where they undergo apoptosis. We start with the basic model of sporadic colorectal cancer initiation. All the relevant parameters with their respective values are summarized in Table 1(a). Let us assume that the **effective population size** of a crypt is N; this means that N cells are at risk of developing mutations which can lead to cancer. The value of N is unknown. A reasonable hypothesis is that only the stem cells are at risk of developing cancer, which gives N~1–10, and in this case the average turnover rate would be τ=1–20 days. Alternatively, we could assume that some differentiated cells are also at risk. In this case, N might be of the order of 100 and the average turnover rate could be less than 1 day. In the present paper we will concentrate on the model with N~1–10; some implications of the other model will also be discussed.

Let us denote by X_0, X_1 and X_2 the probability that the whole crypt consists of cells with 0, 1, and 2 copies of the APC gene inactivated, respectively. The simplest mutation-selection network leading from X_0 to X_1 to X_2 is shown in Figure 2. The rate of change is equal to the probability that one relevant mutation occurs times the probability that the mutant cell will take over the crypt, see Appendix for details of this approximation.

Initially, all the N cells of a crypt have two copies of the APC gene. The first copy of the APC gene can be inactivated by means of a point mutation. The probability of mutation is given by N (a mutation can occur in any of the N cells) times the mutation rate per cell division, u, times 2, because any of the two copies of the APC gene can be mutated. Because inactivation of one copy of the APC does not lead to any phenotypic changes, the rate of fixation of the

corresponding (neutral) mutant is equal to $1/N$. ("Fixation" means that the mutant cells take over the crypt.) Therefore, the rate of change from X_0 to X_1 is $2uN \times 1/N=2u$.

Once the first allele of the APC gene has been inactivated, the second allele can be inactivated either by another point mutation or by an LOH event. This process occurs with rate $N(u+p_0)$, where p_0 is the rate of LOH in normal (non-CIN) cells. We assume that mutants with both copies of the APC gene inactivated have a large selective advantage, so that once such a mutant is produced, the probability of its fixation is close to one.

There are two steps that separate the state X_0 from the state X_2. The expected number of dysplastic crypts in a person of age t is proportional to the product of the two rates and the second power of time, and is given by:

Equation 1. $$MNu(u+p_0)(t/\tau)^2$$

See the Appendix for mathematical details. Some estimates of the expected number of dysplastic crypts, based on Equation 1, are given in Table 2(a). Note that the quantity $X_2(t)$ is inversely proportional to the square of τ. Thus it is important for an organism to keep the division rate of stem cells as low as possible in order to minimize the risk of cancer. For example, if the mean number of dysplastic crypts in a 70 year old does not exceed 10-100, then τ cannot be smaller than about 10 days. This is consistent with the assumption that only the stem cells are at risk of cancer. If the effective population includes differentiated cells, then the average τ is smaller and the expected number of dysplastic crypts is too high.

Another possibility is that dysplastic crypts can be lost. Our model gives the number of dysplastic crypts that are being **produced** over time, which could be larger than the actual number of dysplastic crypts that patients have at any particular time point. Exact measurements of the incidence of dysplastic crypts will provide important information about the crucial parameters of colorectal cancer initiation.

SPORADIC COLORECTAL CANCERS, CIN AND MIN

Let us now consider the possibility of developing genetic instabilities during cancer initiation. Starting from a population of normal cells, three different events can occur: (i) inactivation of the first copy of the APC gene, (ii) inactivation of the first copy of one of n_m MIN genes, and (iii) mutation of one copy of one of n_c CIN genes.

We use the notation X_i, Y_i and Z_i, respectively, for the probability that a crypt consists of normal cells, CIN cells or MIN cells with i copies of the APC gene inactivated, see Table 2(b). Figure 3 shows the mutation-selection network of colorectal cancer initiation including CIN and MIN. All the transition rates are calculated as the relevant mutation rate times the probability that the mutant will take over the crypt.

Let us denote the rate of LOH in CIN cells as p. We assume that the crucial effect of CIN is to increase the rate of LOH,[20,21] which implies $p>p_0$. Intuitively, the advantage of CIN for the cancer cell is to accelerate the loss of the second copy of a tumor suppressor gene. Similarly, the advantage of MIN is to increase the point-mutation rate, which means that $\acute{u}>u$.

We are interested in the probability of finding the crypt in states X_2, Y_2 or Z_2 as a function of t. In other words, we want to know the probability for the dysplastic crypt to have CIN (Y_2), MIN (Z_2) or no genetic instability (X_2). The mutation-selection network of Figure 3 is more complicated than the one-dimensional network of Figure 2, but the solutions for X_2, Y_2 and Z_2 can still be written down. The key is to identify how many slow (rate-limiting) steps separate the initial state (X_0) from the state of interest. The slow

Table 2

A. **Sporadic Colorectal Cancer.** The expected number of 70 years of age, the simple model. $M = 10^7$, $N = 5$, $u = 10^{-7}$ and $t = 70$ years.

	$p_0 = 10^{-7}$	$p_0 = 10^{-6}$
$\tau = 1$	654	3,595
$\tau = 2$	73	400
$\tau = 3$	7	36
$\tau = 4$	2	9

B. **Sporadic Colorectal Cancer.** The expected number of dysplastic crypts and fractions with different instabilities, at 70 years of age, in the model with CIN and MIN. $M = 10^7$, $N = 5$, $\tau = 20$ days, $u = 10^{-7}$, $u_s = 10^{-6}$, $\acute{u} = 10^{-4}$, $p_0 = 10^{-7}$, $n_m = 3$ and $t = 70$ years

MIN gene inactivation	n_c	Total number of dysplastic drypts	Percent of CIN	Percent of MIN
mutation	1	2	28%	0.4%
mutation	10	8	80%	0.1%
mutation	100	67	98%	0.01%
silencing	1	3	21%	23%
silencing	10	9	74%	7%
silencing	100	68	97%	1%

C. **FAP.** The expected number of dysplastic crypts and the fraction CIN crypts, at16 years of age. $M = 10^7$, $N = 5$, $\tau = 20$ days, $u = 10^{-7}$, $\acute{u} = 10^{-4}$, $p_0 = 10^{-7}$, $n_m = 3$ and $t = 16$ years

n_c	Total number of dysplastic crypts	Percent of CIN	Percent of MIN
1	~3,500	17%	0%
10	~8,800	67%	0%
100	~61,300	95%	0%

D. **HNPCC.** The expected number of dysplastic crypts and the fraction MIN crypts, at40 years of age. $M = 10^7$, $N = 5$, $\tau = 20$ days, $u = 10^{-7}$, $u_s = 10^{-6}$, $\acute{u} = 10^{-4}$, $p_0 = 10^{-7}$, $n_c = 10$ and $t = 40$ years. Compared with patients with FAP and sporadic colorectal cancer

Condition	Total # of dyspl. crypts	Percent of CIN	Percent of MIN
HNPCC, mutation of MIN gene	14	15%	81%
HNPCC, silencing of MIN gene	66	3%	96%
Sporadic colorectal cancer	3	78%	3%
FAP	~22,000	67%	0%

(A) Sporadic colorectal cancer: the expected number of dysplatic crypts by the age of 70 for different values of p_0 and t (measured in days); the simple model. (B) Sporadic colorectal cancer: the expected number of dysplastic crypts and the fraction of dysplastic crypts which have CIN and MIN, for different values of n_c, and different mechanisms of MIN gene inactivation. (C) FAP: the expected number of dysplastic crypts and the fraction of CIN crypts, by the age of 16, predictions for different values of n_c. Note that the fraction of MIN crypts is zero in this case. (D) HNPCC: the exptected number of dysplastic crypts and the fraction of MIN crypts, at 40 years of age, for different mechanisms of MIN gene inactivation.

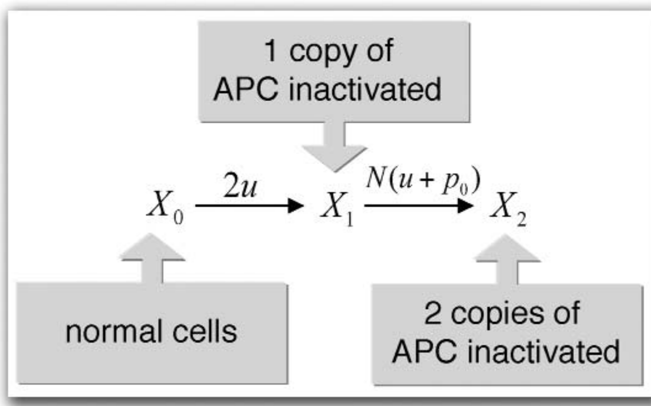

Figure 2. Mutation-selection network of sporadic colorectal cancer inititiation. Initially, the crypt is at the state X_0, i.e., all cells are wild-type. With the rate $2u$, cells with one copy of the APC gene mutated will take over the crypt (state X_1). This rate of change is calculated as N times the probablility (per cell division) to produce a mutant of X_1 ($2u$ because either of the two alleles can be mutated) times the probability of one mutant of type X_1 to get fixed ($1/N$ since there is no phenotypic change). From state X_1 the system can go to state X_2 (both copies of the APC gene inactivated) with the rate $N(u + p_0)$. This rate is calculated as N times the probability per cell division to produce a mutant of X_2 (u for an independent point mutation plus p_0 for an LOH event) times the probability of the advantageous mutant of type X_2 to take over (this is 1).

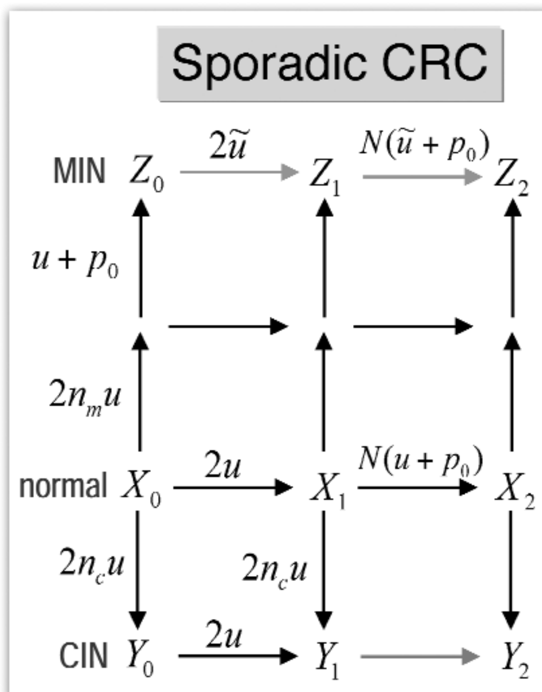

Figure 3. Mutation-selection network of sporadic colorectal cancer initiation including CIN and MIN. From the initial wild-type state, X_0, the crypt can change to state X_1 as in Figure 1, acquire a CIN mutation (the arrow down) or a MIN mutation (the arrow up). The line X_0 - X_1 - X_2 is identical to the process in Figure 1 of developing a dysplastic crypt with no genetic instabilities. The bottom row of the diagram corresponds to CIN cells acquiring the first, and then the second, mutation (loss) of the APC gene; the second copy can be lost by a point mutation or by an LOH event whose rate is much larger for CIN cells than it is for normal or MIN cells. the state Y_2 corresponds to a CIN dysplastic crypt. The top row is the development of a MIN dysplastic crypt. The MIN phenotype is characterized by an increased point mutation rate, ú. The state Z_2 is a MIN dysplastic crypt. Red arrows denote faster steps. Note that it takes only one leap (down) to go to a CIN state from a state with no genetic instability, because CIN genes are dominant-negative. It takes two steps to acquire MIN (up) becasue both copies of a MIN gene need to be inactivated before any phenotypic changes happen.

steps in our model are the ones whose rates scale with u or p_0. The step from Y_1 to Y_2 is fast, because it is proportional to the rate of LOH in CIN cells, p, which is much larger than u and p_0. The steps with the rate $ú$ are neither fast nor slow, see Appendix. For all possible pathways from the initial state to the final state of interest, we have to multiply the slow rates together times the appropriate power of τ, and divide by the factorial of the number of slow steps. Summing over all possible paths we will obtain the probability of finding the crypt in the state in question.

Applying these rules, we can see that $X_2(t)$ and $Y_2(t)$ are both quadratic in time, because it takes two rate-limiting steps to go from X_0 to X_2 and from X_0 to Y_2. The state Z_2 is separated from X_0 by two rate-limiting steps and two 'intermediate' steps (whose rate is proportional to $ú$, so the quantity $Z_2(t)$ grows as the forth power of time for $út/\tau << 1$ and as the second power of time in the opposite limit (see Appendix).

The probability that a crypt is dysplastic at time t is given by $P(t)=X_2(t)+Y_2(t)+Z_2(t)$. Therefore, the the expected number of dysplastic crypts in a person of age t is $MP(t)$. Of these dysplastic crypts, $MY_2(t)$ have CIN and $MZ_2(t)$ have MIN. This suggests that the fraction of CIN cancers is at least $Y_2(t)/P(t)$ and the fraction of MIN cancers is at least $Z_2(t)/P(t)$. The actual values may be higher because in our model, only the very first stage of cancer development is considered. At later stages of progression from a dysplastic crypt to cancer, there are more chances for cells to acquire a CIN or a MIN mutation. Some numerical examples are given in Table 2(b), where the relative fractions of dysplastic crypts with CIN, MIN and with no genetic instability are presented for different values of n_c, the number of CIN genes. Larger values of n_c lead to increased percentages of dysplastic crypts with CIN. According to observations,[13] of all sporadic colorectal cancers have MIN and 87% have CIN.[1] In terms of our model this means that we should have $Z_2(t)/P(t)<0.13$ and $Y_2/P(t)<0.87$. From Table 2(b) we can see that for values of n_c of the order of 100, the fraction of CIN crypts is higher than expected. A resolution of this can be found if we assume that CIN phenotype is disadvantageous compared to the wild-type. Indeed, genetically unstable cells can have a higher apoptosis rate because of a high frequency of mutations in essential genes. In this case, the transition rate from X_0 to Y_0 and from X_1 to Y_1 becomes lower, see Appendix for the exact expressions. For example, if the relative disadvantage of a CIN cell is 10%, then the fraction of CIN dysplastic crypts in Table 2(b) will be reduced by 20%. The fraction of MIN crypts as predicted by this model is quite low (for n_c=10 we get only 0.1% of dysplastic crypts with MIN). This could mean that MIN develops at later stages of cancer. However, there is indirect evidence that replication error phenotype precedes, and is responsible for, APC mutations in MIN cancers.[22] Our model is consistent with these data if we assume higher rates of MIN induction in a cell. This could be caused by higher mutation rates in MIN genes or the possibility of epigenetic mechanisms of gene silencing. Silencing of the hMLH1 gene is found at high frequency in sporadic MIN tumors[23-25] and is associated with methylation of the promoter region. In the diagram of Figure 3 this means that the rates from X_0 to the MIN type (vertical arrows), $2n_m u$ and $u+p_0$, should be replaced by $2n_m u_s$ and u_s+p_0, respectively, where u_s is the rate of silencing per gene per cell division. If we assume that u_s is larger than the basic mutation rate, u (say $u_s=10^{-6}$), then the expected fraction of MIN crypts predicted by our model becomes larger. Note however that at this stage there is no accurate estimation of silencing rates compared to mutation rates. Our model suggests that if epigenetic mechanisms significantly increase the APC inactivation rate, then the predicted fraction of MIN crypts is consistent with the observed frequency of MIN cancers, see Table 2B.

FAP

In FAP patients, one allele of the APC gene is inactivated in the germ line. In terms of our model, all crypts start in state X_1. The corresponding mutation-selection network is found in Figure 4A. The solutions are found in Appendix A. $X_2(t)$ and $Y_2(t)$ are linear functions of time (there is one rate-limiting step), whereas $Z_2(t)$ grows slower than the second power of time (two rate-limiting steps plus one 'intermediate' step).

Some predictions of the model are shown in Table 2(c). The expected number of dysplastic crypts and the fraction of CIN crypts are calculated for t=16 years. As the number of CIN genes, n_c, increases, we expect more dysplastic crypts, and a larger fraction of crypts with CIN. According to our model, the expected number of dysplastic crypts grows linearly with time, and by the age of 16 years is expected to be thousands to tens of thousands, see Table 2(c). This should be compared with the observation that patients with FAP have hundreds to thousands of polyps by their third decade of life.

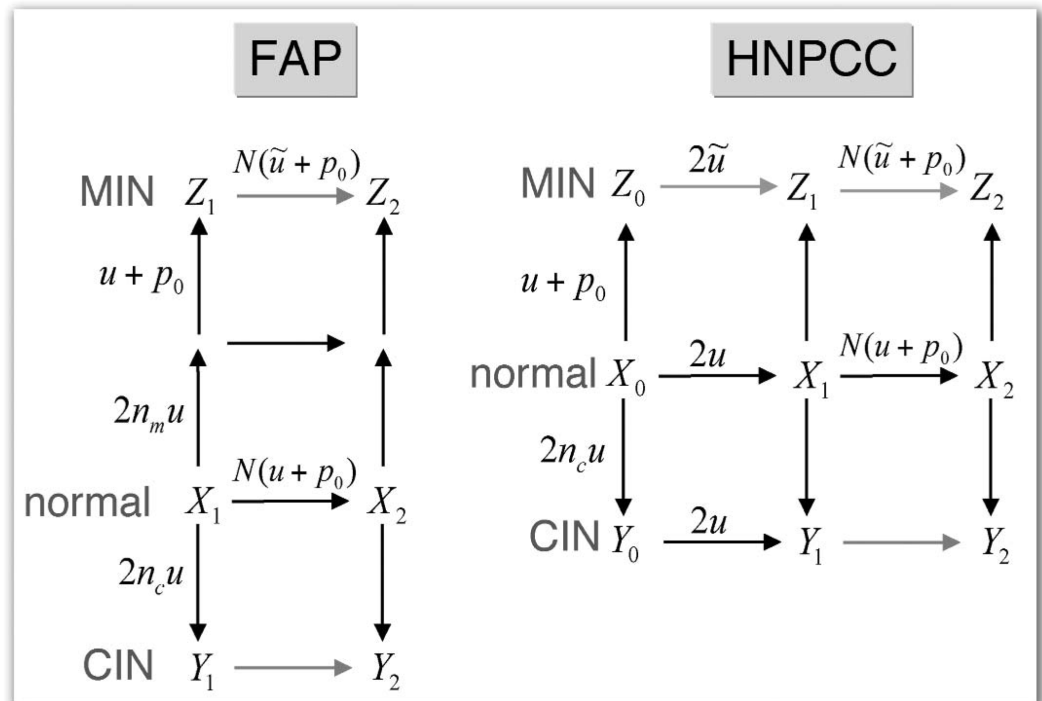

Figure 4. (A) Mutation-selection network of FAP initiation. We start with the type X_1 because the first copy of the APC gene is inactivated in the germ line. (B) Mutation-selection network of HNPCC initiation. One mutation of a MIN gene is inherited, and therefore it takes only one step (inactivation of the second copy of the MIN gene, arrows up) to develop the MIN phenotype.

The number of polyps in FAP patients does not grow linearly with time. Instead, most polyps appear 'suddenly' in the second decade of life. These observations are consistent with the predictions of our model. It is believed that polyps result from dysplastic crypts by means of further somatic mutations and clonal expansions. Therefore, the number of polyps is expected to be a higher than linear power of time, which looks like a steep increase in the number of lesions after a relatively non-eventful period. Also, the number of dysplastic crypts 10^3–10^4 in our model) is expected to be much larger than the number of polyps (10^2–10^3) consistent with the expectation that not all dysplastic crypts progress to polyps.

Another prediction of our model is that the fraction of MIN crypts in patients with FAP is negligible. This is consistent with an experimental study where MIN was found in none of the 57 adenomas from FAP patients.[26]

Finally, we note that the logical possibility exists that the second copy of the APC gene in FAP patients may be inactivated by an epigenetic event, just like the second copy of a MIN gene can be silenced. Experimental investigations[27] however suggest that this is unlikely: out of the 84 FAP tumors, only 1 exhibited hypermethylation of the APC gene.

HNPCC

Patients with HNPCC inherit one mutation in a MIN gene. The corresponding mutation-selection network is presented in Figure 4b. The solutions for X_2 and Y_2 in this case are quadratic in time (two rate-limiting steps), and the quantity $Z_2(t)$ grows slower than linear but faster than quadratic (one rate-limiting and two intermediate steps). In Table 2(d) we present the expected number of dysplastic crypts and the fraction of MIN crypts, calculated for t=40. We have explored two possibilities:

1. inactivation of the second copy of a MIN gene happens by means of a point mutation, with the rate u; and
2. inactivation of the second copy of a MIN gene happens by epigenetic silencing.

There is evidence that the second scenario is less likely in the case of HNPCC.[28] In a recent study, DNA methylation of the hMLH1 gene was found in80% of 40 sporadic MIN cancers but in 0% of 30 cancers in HNPCC patients.[29]

Our model predicts that the majority of dysplastic crypts in HNPCC patients are expected to have MIN. However, we do not find that 100% of dysplastic crypts will contain MIN. On the other hand, we know that virtually all tumors in HNPCC patients have MIN. This might suggest that selection for MIN also happens at later stages of carcinogenesis: dysplastic crypts with MIN might have a faster rate of progression to cancer than dysplastic crypts containing CIN or normal cells.

Finally we note that the total number of dysplastic crypts in HNPCC patients, as predicted by our model, is of the order 10 at age 40, which is only slightly larger than the expected number of dysplastic crypts in normal individuals and is not nearly as high as in the case of FAP 9of the order 10,000, Table 2D). This is also consistent with observations.

DISCUSSION

Presently, one of the most interesting questions in cancer research is to what extent genetic instability is an early event and hence a driving force of tumor progression. The dominant view is that neoplasia is initiated by mutations that increase the net reproductive rate of cells. These mutations occur in oncogenes and tumor suppressor genes. An alternative possibility is that neoplasia is initiated by mutations in genes that affect the mutation rate and genetic stability of a cell.

In this paper, we have provided a quantitative framework to study the dynamics of colorectal cancer initiation. We calculate the rate of dysplastic crypt formation as a consequence of inactivating both alleles of the APC tumor suppressor gene. This can either happen in normal cells or in cells that have already acquired one of the two genetic instabilities, MIN or CIN. If the rate of triggering genetic instability in a cell is high and if the cost of genetic instability is not too large, then inactivation of APC will frequently occur in cells that are genetically unstable. In this case, genetic instability is the first phenotypic modification of a cell on the way to cancer.

CIN accelerates LOH of the second allele of APC. If the number of dominant CIN genes in the human genome exceeds a critical value then CIN will precede APC inactivation. The critical number of CIN genes depends on the rate of LOH in normal and CIN cells, the number of cells in a crypt that are at risk of mutating into cancer cells and the selective disadvantage of CIN. For a wide range of parameter values, we find that as few as 10 dominant CIN genes would be enough for CIN to precede inactivation of the first tumor suppressor gene in the majority of colorectal cancers. Similarly, if inactivation of MIN genes (either by point mutation or by silencing) occurs at a sufficiently fast rate—around 10^{-6} per cell division, then MIN can precede APC inactivation in a significant number of cases. Hence, in order to decide the issue of genetic instability, the crucial question is how many dominant CIN genes can be found in the human genome, and how fast are MIN or CIN genes inactivated.

Our calculations show that important insights could be derived by carefully monitoring the incidence rate of dysplastic crypts in patients as function of age. With or without early genetic instability, the abundance of dysplastic crypts should grow approximately as a second power of time. The two rate limiting steps can either refer to two mutations of APC, or one mutation of APC and one CIN mutation. In the case of CIN, LOH of the second allele of APC is not rate limiting. Hence, two rate limiting steps for the inactivation of a tumor suppressor gene are compatible with an additional genetic instability mutation. Several further insights emerge from our analysis.

- **The Cells at Risk of Cancer.** Our calculations predict that if both stem cells and large numbers of differentiated cells in a crypt were running the risk of acquiring cancerous mutations, then the expected number of dysplastic crypts in persons of 70 years of age would be enormous. Therefore, only a small number of (stem) cells per crypt should be at risk of developing cancer.
- **Competition Among Crypts.** Another interesting possibility is that dysplastic crypts can be lost and replaced by normal crypts. In this case, many dysplastic crypts could be produced, but only a part of them is retained so that the actual number of dysplastic crypts stays low. To our knowledge, the competitive dynamics of crypts in a colon has not been investigated experimentally.
- **Fraction of Dysplastic Crypts with CIN or MIN.** About 87% of sporadic colorectal cancers have CIN while the rest have MIN. Assuming that CIN and MIN are irreversible, we conclude that the maximum fraction of dysplastic crypts with CIN should 87%, while the maximum fraction of dysplastic crypts with MIN should be 13%. This provides certain restrictions on the possible parameter values of our model (see Table 2b).
- **No MIN in FAP.** Our model predicts that the fraction of MIN dysplastic crypts in FAP patients is close to zero. A significant number of dysplastic crypts will contain CIN. This is consistent with experimental observations.
- **Epigenetic Factors.** If we assume that MIN genes in sporadic colorectal cancer are inactivated only by point mutation or LOH events, then the fraction of dysplastic crypts with MIN is very low. We get higher fractions of MIN if we assume that MIN genes can also be silenced and if such silencing of MIN genes is fast compared to point mutation or LOH. Thus epigenetic events could play a crucial role in the formation of sporadic MIN cancers.
- **Crypts with Genetic Instabilities Might Have a Faster Rate of Progression to Cancer.** There are two possibilities for how genetic instability could initiate tumor progression. (1) CIN and MIN mutations could precede APC inactivation. In this case, most dysplastic crypts would already consist of cells that have a genetic instability. (2) Even if the majority of dysplastic crypts were to consist of stable cells, it is still possible that those dysplastic crypts that have CIN or MIN have a faster rate of progression to adenoma and cancer. Indeed, all "necessary mutations" on the way from a dysplastic crypt to cancer will happen at a faster rate in genetically unstable lesions, compared to $APC^{-/-}$ lesions without CIN or MIN. In this case, most cancers would derive from cell lines where a genetic instability mutation preceded APC inactivation.
- **The Number of Dysplastic Crypts.** We calculated both the absolute numbers and relative proportions of dysplastic crypts with or without genetic instabilities. An interesting empirical project would be to measure the abundance of such dysplastic crypts as function of age. This would provide crucial information on the dynamics of colorectal cancer initiation.

METHODS

Sporadic Colorectal Cancers, the First Model. In the beginning (see Fig. 2), all cells are wild type. The fist copy of the APC gene can get inactivated by a mutation event. Because the mutation rate per gene per cell division, $u \approx 10^{-7}$, is very small and the number of cells, N, is not large, it is safe to assume that once a mutation occurs, the population typically has enough time to become homogeneous again before the next mutation occurs. The exact condition[19] is that the mutation rate, u, is much smaller than $1/N^2$. This means that most of the time, the effective population of cells in a crypt can be considered as homogeneous with respect to APC mutations. Under this assumption we have $X_0+X_1+X_2=1$.

Another assumption made in the diagram of Figure 2 is that the probability of fixation of the advantageous mutants (type X_2) is one. This is done for simplicity. More generally, the relative fitness of type X_2 is r_2, whereas the fitness of type X_0 and X_1 is 1. Then the second rate in figure 1 should be taken to be $N\rho_2(u+p_0)$, with $\rho_2=r_2^{N-1}(1-r_2)/(1-r_2^N)$. If the population size is not too large, and the relative fitness of type X_2 is much greater than 1, we have $\rho_2 \rightarrow 1$, and we obtain the expression $N(u+p_0)$. (The same argument applies to the other networks.)

The mutation-selection network of Figure 2 is equivalent to a linear system of ordinary differential equations (ODE's), where the rates by the arrows refer to the coefficients and the direction of the arrows to the sign of the terms. One (non-dimensional) time unit ($t/\tau=1$) corresponds to a generation turn-over. The calculations leading to the mutation-selection network are performed for a Moran process where the population size is kept constant by removing one cell each time a cell reproduces. Our biological time-unit corresponds to N "elementary events" of the Moran process, where an elementary event includes one birth and one death. We have:

$\dot{X}_0 = -2uX_0$,
$\dot{X}_1 = 2uX_1 - N(u + p_0)X_1$
with the constraint $X_0 + X_1 + X_2 = 1$ and the initial condition
$X_0(0) = 1,\ X_1(0) = 0$

Using $ut/\tau << 1$ and $N(p_0+u)t/\tau << 1$, we can approximate the solution for X_2 as $X_2(t)=Nu(u+p_0)(t/\tau)^2$. The quantity $X_2(t)$ stands for the probability that a crypt is dysplastic (i.e. consists of cells with both copies of the APC gene inactivated) at time t measured in days. The probability to have i dysplastic crypts by the age t is:

$$\binom{M}{i} X_2(t)^{i} (1-X_2(t))^{M-i}$$

The expected number of dysplastic crypts in a person of age t is then given by Equation 1 (p 686).

Sporadic Colorectal Cancers, CIN and MIN. Let us consider the mutation-selection network of Figure 3. Again, we will use the approximation of homogeneous states. The diagram of Figure 3 corresponds to a system of 11 linear ODE's describing the time-evolution of the probabilities to find the system in any of the 12 possible homogeneous states. An exact solution can be written down but it is a very cumbersome expression, so we will make some approximations. Let us use the fact that the quantities ut/τ and $N(u+p_0)t/\tau$ are very small compared to 1 for $t \sim 70$ years, and the quantity $Npt/\tau >> 1$. This tells us that the steps in the diagram characterized by the rates u and p_0 are slow (rate limiting) compared to the steps with the rate p. Taking the Taylor expansion of the solution in terms of ut/τ and $N(u+p_0)t/\tau$, we obtain the following result:

Equation 2. $$X_2(t) = Nu(u+p_0)(t/\tau)^2,\quad Y_2(t) = 4n_c u^2(t/\tau)^2$$

The rate $ú$ is neither fast nor slow, so the solution for Z_2 is more complicated. We have $Z_2(t)=$

$$\frac{n_m u(u+p_0)[2(b^3Ea - a^3Eb + a^3 - b^3 + ab(b^2 - a^2)\, út/\tau) + a^2b^2(a-b)ú^2)(t/\tau)^2]}{(abú)^2\,(a-b)}$$

where a = 2, b = $N(ú + p_0)/ú$ and $E_x = e^{-xút/\tau}$.

Note that the limit of this expression if the $ú$-steps are fast (i.e., if $út/\tau << 1$) is given by $Z(t)=n_m u(u+p_0)(t/\tau)^2$. In the opposite limit where $út/\tau >> 1$, we have $Z(t)=n_m Nu(u+p_0)ú(ú+p_0)(t/\tau)^4/6$. If we assume that the MIN genes are inactivated by epigenetic silencing rather than mutation events, we need to replace u by u_s in the expression for $Z_2(t)$.

In the above calculations we assumed that CIN and MIN are selectively neutral. In general, we can assume that genetic instability leads to a change of reproductive rate thus giving the mutant cells a selective disadvantage or a selective advantage. Let us suppose that the relative reproductive rate of CIN cells is r_c. Then the transition rates from X_0 to Y_0 and from X_1 to Y_1 is not $2n_c u$ but $2n_c N r_c^{N-1}(1-r_c)/(1-r_c^N)$. This quantity is larger than $2n_c u$ in the case when CIN is advantageous ($r_c > 1$) and smaller if it is disadvantageous ($r_c < 1$). An extended analysis of non-neutral CIN is performed elsewhere.[19] Note that in Figures 3, 4A and 4B, some of the rates are not marked. Including the corresponding terms only gives small corrections to the solution in the appropriate limit, so these terms can be safely ignored.

FAP. The system of equations is presented in the mutation-selection network of Figure 4A. The solution reads:

$$X_2(t) = N(u+p_0)t/\tau,\quad Y_2(t) = 2n_c ut/\tau,$$

and

$$Z_2(t) = \frac{n_m u(u+p_0)[2-2úbt/\tau + (úbt/\tau)^2 - 2E_b]}{(úb)^2}$$

In the limit where $ut/\tau \to \infty$, we have $Z_2(t)=n_m u(u+p_0)(t/\tau)^2$. If $ut/\tau << 1$, then $Z_2(t)=n_m u(u+p_0)(ú+p_0)(t/\tau)^3/3$.

HNPCC. The corresponding mutation-selection network is presented in Fig. 4b. The solutions for X_2 and Y_2 in this case are identical to those for sporadic colorectal cancers and are given by Equation 2. The solution for Z_2 is as follows:

$$Z_2(t) = \frac{(u+p_0)[a^2E_b - b^2E_a + (a-b)(úabt/\tau - a + b))]}{abú(a-b)}$$

in the limit where $ú$ is a fast rate we have $Z_2(t)=(u+p_0)t/\tau$. In the opposite limit, where $ú$ is a slow rate, $Z_2(t)=N(u+p_0)ú(ú+p_0)(t/\tau)^3/3$. If we assume that the second copy of the MIN gene is silenced by epigenetic means, we need to replace u by u_s in the expression for $Z_2(t)$.

A More Precise Description of Mutation Spectrum. The mutation spectrum of the APC gene is far from random (one reason being that the APC gene is long and multi-functional). The second APC mutation may depend on where the first APC mutation took place.[30,31] Our model is well suited to take this into account. Here is a simple way to differentiate between two kinds of point mutations. Let us assume that the total probability of a point mutation is u (as in the basic model), and there are two kinds of mutations.

1. With probability u_1, a mutation happens such that the second allele can only be inactivated by a point mutation.
2. With probability u_2, a mutation happens which can be followed by another point mutation or an LOH event. We have $u_1+u_2=u$. These two scenarios can be incorporated in our calculations adding a new level of complexity to the basic theory.

References

1. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in human cancers. Nature 1998; 396:643-9.
2. Sen S. Aneuploidy and cancer. Curr Op Oncology 2000; 12:82-8.
3. Loeb LA, Springgate CF, Battula N. Errorsi n DNA replication as a basis of malignant changes. Cancer Res 1974; 34:2311-21.
4. Parker SL, et al. Cancer Statistics, 1996. Ca Cancer J Clin 1996; 46:5-27.
5. Kinzler KW, Vogelstein B. The Genetic Basis of Human Cancer. Toronto: McGraw Hill, 1998.
6. Polakis P, the adenomatous polyposis coli (APC) tumor suppressor. Biochim Biophys Acta 1997; 1332:F127-47.
7. Kolodner RD, Putnam, CD, Myung K. Maintenance of genome stability in Saccharyomyces cerevisiae. Science 2002; 297:552-7.
8. Cahill DP, et al. Mutations of mitotic checkpoint genes in human cancer. Nature 1998; 392:300-3.
9. Imai Y, Shiratori Y, Kato N, Inoue T, Omata M. Mutational inactivation of mitotic checkpoint genes, hsMAD2 and hBUB1, is rare in sporadic digestive tract cancers. Jpn J cancer Res 1999; 90:837-40.
10. Gemma A, et al. somatic mutaiton of the hBUB1 mitotic checkpoint gene in primary lung cancer. Genes Chromosome Cancer 2000; 29:213-8.
11. Ohshima K,e t al. Mutatoin analysis of mitotic checkpoint genes (hBUB1 and hBUBR1) and microsatellite instablity in adult T-cell leukemia/lymphoma. Cancer Lett 2000; 158:141-50.
12. Li Y, Benezra R. Identification of a human mitotic checkpoint gene: hsMAD2. Science 1996; 274:246-8.
13. Wang X, et al. Correlation of defective mitotic checkpoint with aberrantly reduced expression of MAD2 protein in nasopharyngeal carcinoma cells. Carcinogenesis 2000; 21:2293-7.
14. Michel LS, et al. MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. Nature 2001; 409:355-9.
15. Ro S, Rannala B. Methylation patterns an dmathematical models reveal dynamics of stem cell turnover in the human colon. PNAS 2001; 98:10519-21.
16. Yatabe Y, Tavare S, Shibata D. Investigating stem cells in human colon by using methylation patterns. PNAS 2001; 98:10839-44.
17. Potten CS, et al. Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridiine. gut 1992; 33:71-8.\
18. Bach SP, Renehan AG, Potten CS. Stem cells: the intestinal stem cell as a paradigm. Carcinogenesis 2000; 21:469-76.
19. Komarova NL, Sengupta A, Nowak MA. Mutation selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. Submitted to J Theor Biol 2002.
20. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cances. Nature 1997; 386:623-7.
21. Bardelli A, et al. Carcinogen-specific induction of genetic instability. PNAS 2001; 98:5770-5.
22. Huang J, et al. APC mutations in colorectal tumors with mismatch repair deficiency. PNAS 1996; 93:9049-54.
23. Ahuja N, et al. Association between CpG island methylation and microsatellite instability in colorectal cancer. Cancer Res 11997; 57:3370-4.
24. Kane MF, et al. Methylation of hMLH1 promoter correlates with lack of expression of hMLH1 in a sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer res 1997; 57:808-11.
25. Cunningham JM,e ta l. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. Cancer res 1998; 58:3455-60.
26. Kelelr JJ, et al. Molecular analysis of sulindac-resistant adenomas in familial adenomatous polyposis. Clin Cancer res 2001; 7:4000-7.

27. Menigatti M, et al. Methylation pattern of different regions of the MLH1 promoter and silencing of gene expression in hereditary and sporadic colorectal cancer. Genes Chromosomes Cancer 2001; 31:357-61.
28. Yamamoto H,e t al. Differential involvement ofh te hypermethylatyor phenotype in hereditary and sporadic colorectal cancers with high-frequency microsatellite instability. Genes Chromosomes Cancer 2002; 33:322-5.
29. Esteller M, et al. DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. Human Mol Gen 2001; 10:3001-7.
30. Rowan AJ, et al. APC mutations in sporadic colorectal tumors: a mutational "hotspot" and interdependence of the "two hits". PNAS 2000; 97:3352-7.
31. Lamlum H, et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of hte germline mutation: a new facet to Knudson's 'two-hit' hypothesis. Nat Med 1999; 1071-5.