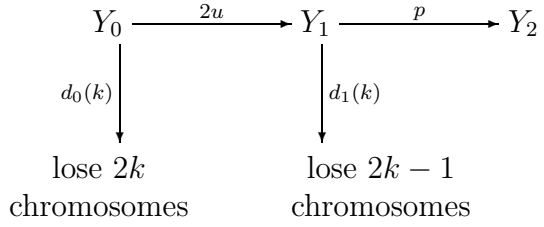
Supporting Methods

The Model Set-up. We model epithelial tissue organized into compartments. In the simples case, there is one stem cell per compartment. For example, in colon, this scenario would correspond to crypts with a stem cell situated at the base of each crypt. Stem cells divide asymmetrically, producing one (immortal) stem cell and one differentiated cell. Here, we concentrate on the dynamics of the stem cells. Each division event is equivalent to a replacement of the old stem cell with a copy of itself. Upon division of a stem cell, the immortal daughter cell might *(i)* acquire a silencing mutation in one of its alleles of the adenomatous polyposis coli (APC) gene with probability u per cell division, or *(ii)* lose one of its chromosomes, with probability p per cell division per chromosome. Once both copies of the tumor suppressor (TSP) gene have been inactivated, the cell will be able to escape homeostatic control and create a growing clone.

Optimal Rate of Loss of Heterozygozity (LOH). Suppose that a stem cell has a probability to lose a chromosome p per chromosome per cell division. First, we calculate the probability to inactivate the TSP gene by time t. The sequence of events can be expressed by the following simple diagram,



Here, Y_i is the probability for the stem cell to have i inactivated copies of the TSP gene. The first event of inactivation happens by a fine-scale genetic event (probability u times two for two alleles), and the second event is a loss of the chromosome with the remaining copy of the TSP gene (probability p). k is the total number of chromosomes ($k = 23$.)

A very important issue here is the exact extent of the damage for the cell and its reproductive potential that is induced by LOH events. In the most optimistic (for cancer) scenario *(i)*, there is no loss in fitness due to chromosome loss, i.e. $d_0(k) = 0$, except when the cell loses the mutated copy of the TSP chromosome at stage y_1, so that $d_1(k) = p$. In the most pessimistic scenario *(ii)*, a loss of any chromosome results in cell death, unless it leads to a TSP inactivation, so that $d_0(k) = 1 - (1 - p)^{2k}$ and

$d_1(k) = 1 - (1-p)^{2k-1}$. The reality is probably somewhere in between these two extremes.

We can write down the Kolmogorov forward equations for all the probabilities (skipping the argument k of d_0 and d_1),

$$\begin{aligned}
\dot{Y}_0 &= [(1-d_1)(1-2u)-1]\,Y_0, & [1]\\
\dot{Y}_1 &= (1-d_0)2uY_0 + [(1-d_1)(1-p)-1]\,Y_1, & [2]\\
\dot{Y}_2 &= (1-d_1)pY_1, & [3]
\end{aligned}$$

with the initial condition $Y_0(0) = 1$. We need to calculate the probability distribution of creating a $\mathrm{TSP}^{-/-}$ mutant as a function of time, which is given by $\dot{Y}_2$. We have,

$$\dot{Y}_2(t) = \frac{up(e^{-ut} - e^{-(p+d_1)t})}{p + d_1 - u}, \qquad [4]$$

where we assumed that $ut \ll 1$. Note that the argument given here holds without change for (constant) populations of more than one cells, as long as the number N of cells satisfies $N < 1/u$ and $N < 1/\sqrt{p}$. Otherwise, the calculations can be easily adapted to include the effect of "tunneling" (1).

Once a $\mathrm{TSP}^{-/-}$ cell has been produced, it starts dividing according to some law that is (at least, initially) close to exponential. Starting from one cell at time $t = 0$, by time t we will have $Z_y(t)$ cells, with

$$Z_y(t) = e^{a\beta[1-d_1(k)]t}. \qquad [5]$$

The parameter a is the growth rate of the initiated cells, and $0 < \beta < 1$ is the cost due to the fact that a chromosome is missing from all CIN cells because of the LOH inactivation of the TSP. The factor $[1 - d_1(k)]$ comes from the probability for a CIN cell to produce a nonviable mutant, which for scenario *(i)* happens only if only one particular chromosome is lost and, for scenario *(ii)*, if any chromosome is lost.

If we now include the mutation stage, we will need to evaluate the convolution,

$$Z_y(t) = \int_0^t \dot{y}_2(t') e^{a\beta(1-p)^{2k-1}(t-t')}\, dt'.$$

The integral yields the following laws of growth:

$$Z_y(t) = \frac{upe^{a\beta(1-d_1)t}}{a\beta(1-d_1)},$$

where we assumed that, for relevant times, $a\beta t > 1$. To find an optimal value of p that maximizes the growth, we solve $Z_y(t) = M$ for t, which gives,

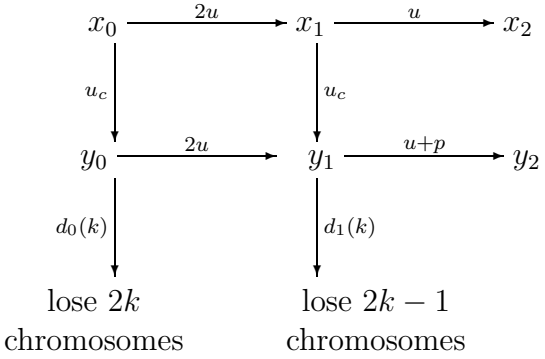
$$t(M) = \frac{1}{a\beta(1-d_1)} \log\left[\frac{a\beta M}{u}\frac{1-d_1}{p}\right],$$

and then we minimize this as a function of p. This calculation can be done easily if we assume that $p \ll 1/(2k)$ (it will turn out that the result for p_* satisfies this assumption). Expanding the expression $dt(M)/dp$ in terms of p, we obtain the equation for p,

$$\frac{1}{p} = (2k-1)\log\frac{a\beta M}{up},$$

where we formally have $k = 1$ for scenario *(i)*, and $k = 23$ for scenario *(ii)*.

Comparison of Stable and Unstable Pathways. Now, let us include the step of initiation of CIN. All pathways can be expressed by the following diagram,



Here, x_i are the probability for the stem cell to be stable and have i inactivated copies of the TSP gene, and y_i are similarly the probabilities for the cell to be CIN. u_c is the rate at which a cell may acquire CIN.

The Kolmogorov forward equations are:

$$\dot{x}_0 = [(1-2u)(1-u_c)-1]x_0, \quad [6]$$

$$\dot{x}_1 = (1-u_c)2ux_0 + [(1-u_c)(1-u)-1]x_1, \quad [7]$$

$$\dot{x}_2 = (1-u_c)ux_1, \quad [8]$$

$$\dot{y}_0 = (1-u)u_cx_0 + [(1-d_1)(1-2u)-1]\,y_0, \quad [9]$$

$$\dot{y}_1 = (1-d_0)2uy_0 + (1-u)u_cx_1 + [(1-d_1)(1-u-p)-1]\,y_1, \quad [10]$$

$$\dot{y}_2 = (1-d_1)(u+p)y_1, \quad [11]$$

with the initial condition $x_0(0) = 1$. It is easy to show that for scenario *(i)*, the two CIN pathways ($x_0 \to y_0 \to y_1 \to y_2$ and $x_0 \to x_1 \to y_1 \to y_2$) contribute equally to y_2. For scenario *(ii)*, and $p \gg u$, the second of these pathways gives a much larger contribution. The reason for this result is that losing a "wrong" chromosome will destroy the cell line in this extreme scenario. Therefore, it is much more likely to reach the state y_2 if CIN appears as late as possible. In what follows, we will ignore the first pathway entirely because it either does not contribute anything or gives a factor of 2. This omision simplifies the calculation because, now, the first step for both stable and CIN cancer is $x_0 \to x_1$, and if we only want to compare the CIN and non-CIN pathways together, this step can be ignored. This scenario is equivalent to starting from $x_1(0) = 1$ rather than $x_0(0) = 1$.

The probability distribution of creating a $\mathrm{TSP}^{-/-}$ mutant as a function of time is given by $\dot{x_2}$, for the stable pathway, and by $\dot{y_2}$, for the unstable pathway. We have,

$$\dot{x}_2(t) = ue^{-u_c t},$$

and $\dot{y_2}$ is given by Eq. 4, with u replaced by u_c.

The clonal expansion law for unstable cells is given by Eq. 5, and for stable cells we simply have $Z_x(t) = e^{at}$. Convolving the rates for the mutation and expansion stages, we arrive at the following laws of growth:

$$Z_x(t) = \frac{ue^{at}}{a}, \quad Z_y(t) = \frac{u_c(u+p)e^{a\beta(1-d_1)t}}{a\beta(1-d_1)},$$

It turns out that, unless u_c is several orders of magnitude bigger than u, Z_y grows slower than Z_x; that is, genetically activated CIN cannot be advantageous.

Large Initial Number of Cells. In the above model, the number of wild-type cells in the compartment is small ($N \ll 1/u$). In order to handle the scenario where a large population of cells is competing in a compartment, which may correspond to later stages of carcinogenesis, we numerically simulated a set of quasi-species-type equations. The estimate obtained for the optimal value of p is very similar to the ones given for the stochastic model above.

1. Komarova, N.L., Sengupta, A. & Nowak, M.A. (2003) *Jour Theor Biol* **223**(4), 433-50.