

# Sensitivity Analysis of Colorectal Cancer Initiation

Yuan Gao, Benji Brown, Dora Layanto, Julien Goldstick

December 11, 2024

## 1 Introduction

Colorectal cancer (CRC) causes the second most cancer deaths in the United States (USA). This year in the USA, it is expected that more than 150,000 people will be newly diagnosed, and more than 50,000 people will die from CRC (Siegel et al., 2024). Due to the severity of this issue to human health and society, biologists and mathematicians have turned their attention to understanding the mechanism of initiation and progression of this disease.

The biological mechanism of colorectal cancer's initiation and propagation has been well studied. Several mutations must accumulate in a colonic crypt stem cell before giving rise to cancerous tumors. These include loss of chromosome maintenance, tumor suppressor loss of function (LOF), and proto-oncogene activation. Several pathways are known to contribute to colorectal cancer, but one dominant pathway occurs in 60%-70% of sporadic (non-inherited) tumors. The pathway begins with genomic instability, followed by two LOF mutations in each of two tumor-suppressing genes, APC and  $P_{53}$  and one gain of function (GOF) mutation in both of the proto-oncogenes KRAS and PIK<sub>3</sub>CA (Nguyen et al., 2020).

As is typical of biological networks, this general pathway may involve several other events up and downstream of these mutations, and experiment may not reveal satisfying explanations for the observed trends. Here, several questions emerge for mathematical biologists such as determining which processes must be included in a model of colorectal cancer initiation to align with observed trends, in what order do these events occur, and what preventative steps may be taken to slow cancer incidence.

This system has been studied mathematically by several groups. Methods to do so may be broadly grouped into deterministic or stochastic models, and spatial or compartmental models. (Mamis et al., 2023) Deterministic models have the advantage of being exactly reproducible, and are generally more simple in implementation, whereas stochastic models trade complexity to capture more variation in biological behavior. Spatial models of CRC capture a known biological element, which is the migration of cells outwards from stem cells. However, these models require many more parameters which are harder to estimate experimentally. Contrastingly, compartmental models ignore spatial correlations in favor of

rigid transitions between states or compartments. Such models are useful in describing the dynamics of cell populations and their resistance or vulnerability to perturbation.

The paper "Aspirin's effect on kinetic parameters of cells contributes to its role in reducing incidence of advanced colorectal adenomas, shown by a multi scale computational study" by authors Wang, Boland, Goel, Wodarz, and Komarova, seeks to address open biological questions on the initiation of colorectal cancer. To do this, they build a compartmental deterministic model to investigate the dynamics of colonic crypts as they acquire mutations, compare their model predictions to epidemiological incidence, and test predictions regarding the mechanism of aspirin's effect on cancer incidence. This model consists of a system of ordinary differential equations (ODEs) to track unique crypt populations and an ODE describing the incidence probability of a fully mutated cancerous crypt. Lastly, they describe the effect of aspirin as a punishment on proliferation rate parameters (Wang et al., 2022).

To extend this work, we first built an understanding of the model's existing construction and behavior. Next, we linearized the ODE system about an equilibrium point by computing partial derivatives of each differential equation with respect to its dependent variables. Finally we conducted sensitivity analysis on the linearized system, and verified the results empirically.

## 2 Description of model equations

### 2.1 Mathematical formulation of a crypt-based model of adenoma initiation

The paper creates a mathematical model of advanced adenoma formation based on the multi-step oncogenic properties of colorectal cancer. The model is parametrized with experimentally obtained data so the epidemiological predictions can better fit the incidence data that document advanced adenoma occurrence as a function of age. This allows for accurate fine-tuning of the model predictions.

The associated selection mutation diagram is shown in Figure 1. There are six distinct crypt populations indicated as types 1 through 6. The mutation rate between types are specified as follows:

Inactivation of the first copy of the APC gene  $APC^{+/-}$ :  $u_{1 \rightarrow 2} = u_{4 \rightarrow 5} = 2u$ .

Inactivation of the remaining copy of the APC gene  $APC^{-/-}$ :  $u_{2 \rightarrow 3} = u_{5 \rightarrow 6} = u$ .

Finally, activation of the KRAS gene  $KRAS^+$ :  $u_{1 \rightarrow 4} = u_{3 \rightarrow 6} = \mu$ .

$R_{ij}$  is the conversion rate from crypt type  $i$  to type  $j$ .

In Figure 1, the populations in the top row, types 1-3, are characterized by an unmutated KRAS oncogene while the populations of the bottom row, types 4-6, have the KRAS mutation activated. Moving from left to right on this diagram, the number of inactivated copies of the APC gene increases from 0 to 2, such that populations of types 1 and 4 are  $APC^{+/+}$ , populations of types 2 and 5 are both  $APC^{+/-}$ , and populations of types 3 and 6 are  $APC^{-/-}$ .

Corresponding with the possible two mutations in APC and one mutation in KRAS there are a

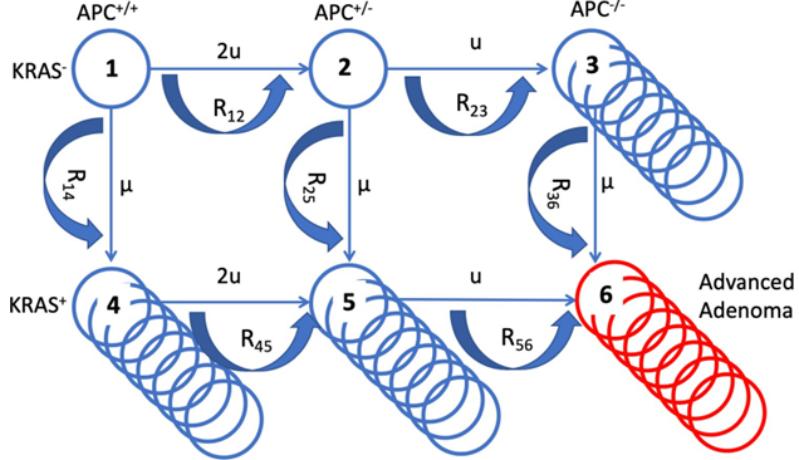


Figure 1: Schematic of the mathematical model. The six cell types are denoted by circles, the mutation rates ( $u, \mu$ ) that give rise to different types are marked by the straight arrows. Type 6 late adenoma is marked in red. Crypt conversion rates ( $R_{ij}$ ) are indicated by circular arrows. Crypt fission ( $\gamma_i$ ) are indicated by multiple circles.

total of 6 different types of crypts. The population of each crypt is denoted as  $n_i$  with  $1 \leq i \leq 6$ . The rate of population change of each crypt is denoted as  $\dot{n}_i$  with  $1 \leq i \leq 6$ .  $\delta$  is the death rate for crypts and the same for all because these are similar cells.  $\gamma_i$  is the corresponding growth rate for crypt fission and varies significantly between crypts. It is reasonable to assume that  $\gamma_2 = 0$ , because  $APC^{+/-}$  crypts do not divide, and that  $\gamma_4 = \gamma_5$  because  $APC^{+/+}$  and  $APC^{-/-}$  crypts with and without a KRAS mutation divide at the same rate.

The paper builds on previous models from (Paterson et al., 2020) by accounting for direct competition between crypts. By assuming that the crypts that undergo crypt fission, types 3, 4, and 5, are in direct competition with each other the equations are changed to models of logistic growth with carrying capacities instead of straight exponential growth. Thus the carrying capacities  $K_A$  and  $K_R$  are added to the differential equations. Where  $K_A$  is the carrying capacity for KRAS<sup>-</sup> crypts types 1, 2 and 3, and  $K_R$  is the carrying capacity for KRAS<sup>+</sup> crypts types 4, 5, and 6. This turns linear differential equations into nonlinear differential equations. In reference to (Paterson et al., 2020) where no crypt competition was included, such that  $K_R = K_A = \infty$  in that model. Due to nonlinearity an analytical solution is not available. However, a solution procedure similar for a linear system for the nonlinear equations is still possible and can be performed numerically to solve the nonlinear differential equations.

$$\begin{cases} \dot{n}_1 = -(R_{12} + R_{14})n_1, \\ \dot{n}_2 = R_{12}n_1 - (R_{23} + R_{25})n_2, \\ \dot{n}_3 = R_{23}n_2 - R_{36}n_3 + \gamma_3 n_3 \left(1 - \frac{n_3 + n_4 + n_5}{K_A}\right) - \delta n_3, \\ \dot{n}_4 = R_{14}n_1 - R_{45}n_4 + \gamma_4 n_4 \left(1 - \frac{n_3 + n_4 + n_5}{K_R}\right) - \delta n_4, \\ \dot{n}_5 = R_{25}n_2 - R_{45}n_4 - R_{56}n_5 + \gamma_5 n_5 \left(1 - \frac{n_3 + n_4 + n_5}{K_R}\right) - \delta n_5 \end{cases} \quad (1)$$

$n_1(0) = N_{crypt}, \quad n_i(0) = 0, \quad 1 \leq i \leq 5.$

The probability that by time  $t$  at least one crypt of type 6 has been created denoted as  $P(t)$ , using the mean-field approximation from (Cole et al., 2009), is given by the solution of the following equation. This equation can be used to calculate incidence which is discussed later in this paper.

$$\dot{P}(t) = (R_{56}n_5 + R_{36}n_3)(1 - P(t)), \quad P(0) = 0$$

## 2.2 Model parameters

The remaining model parameters are defined in this section. The number of stem cells per crypt is denoted as  $K$ . The division rate of wildtype stem cells per year denoted  $(r_i)$ . Corresponding to the crypt type number the relative fitness of mutated cells is given as follows. The relative fitness of APC<sup>+-</sup> cells is  $F_2 = F_{APC^{+/-}}$ , and the relative fitness of APC<sup>-/-</sup> cells is  $F_3 = F_{APC^{-/-}}$ . Similarly, the relative fitness of KRAS<sup>+</sup> cells is  $F_4 = F_{KRAS^+}$ . The division rate of APC<sup>-/-</sup> KRAS<sup>-</sup> crypts per year is  $\gamma_3$ , and the division rate of APC<sup>+/+</sup> KRAS<sup>+</sup> crypts per year is  $\gamma_4$ .

The relative fitness of the mutated cells relative to wildtype are given using the cell replacement data from Komarova et al, using  $Pr(j)$  as the probability of replacement of the wild type by type  $j$  (Komarova et al., 2002).

The death rates are assumed to be the same among crypt types, so that  $r_i = F_i r_1$ , for  $1 \leq i \leq 6$ . Wild-type crypts and those with a single copy of the APC gene do not proliferate so that:  $\gamma_1 = \gamma_2 = 0$ . Crypts with a KRAS mutation and APC<sup>+/+</sup> and APC<sup>+-</sup> phenotypes proliferate at the same rate so that:  $\gamma_4 = \gamma_5$ . The conversion rates  $R_{ij}$  are given by the equation  $R_{ij} = r_i K u_{i \rightarrow j} \rho_{ij}$ , where  $r_i$  is the division rate of type  $i$ ,  $u_{i \rightarrow j}$  is the mutation rate from type  $i$  to type  $j$ , and  $\rho_{ij}$  is the probability that one cell of type  $j$  becomes fixated in a compartment of size  $K$  with the host type.

The values of each parameter utilized are in the Appendix - Table 4

### 3 Reproduction of results in paper

We chose to reproduce figure 2c from the Wang et. al paper which highlights the behavior of the model and biological interpretation. For a description of our attempt to recreate figure 3c, see the appendix.

Figure 2c is a plot of cancerous crypt probability against time with varying parameters.

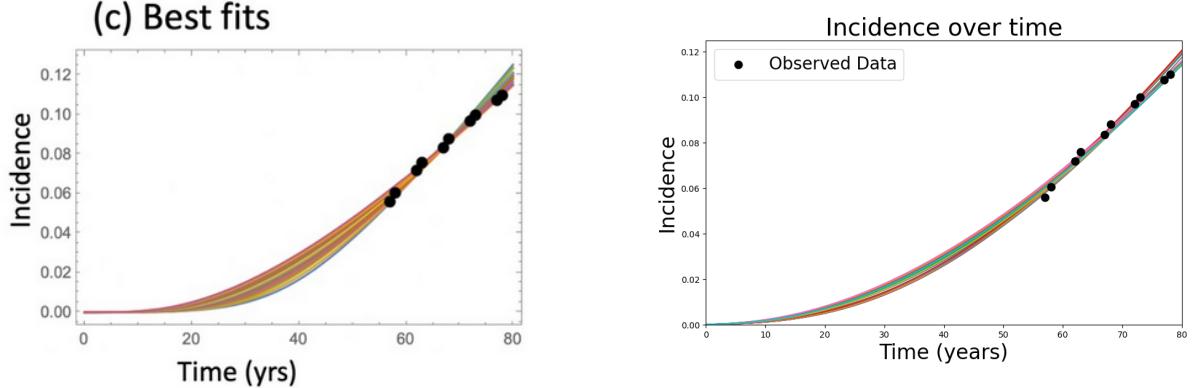


Figure 2: Left, figure 2c. from Wang et. al paper. Right, recreation of figure 2c. Plot of cancerous crypt incidence curves.

To obtain this probability, we use the solution of the last ODE in our system which gives the probability of having at least one crypt with all the necessary cancerous mutations at time t,

$$\dot{P} = (R_{56} + R_{36}n_3)(1 - P), \quad P(0) = 0 \quad (2)$$

However, small populations of type 6 crypts are not detectable. Thus epidemiological data is limited to the incidence of having a detectable number of type 6 crypts, by which time a patient is diagnosed with cancer. The paper uses a time shift to define the incidence,

$$I(t) = P(t - \Delta T) * (1 - \delta/\gamma_6) \quad (3)$$

Where  $\delta$  is the cell death rate,  $\gamma_6$  is the fission rate for type 6 crypts, and  $\Delta T$  is the estimated time to detection, 4.79 years.

From this figure, we see that the ODE model is in good agreement with observed epidemiological data as we predict very similar incidence probabilities as the sample experimental incidence. The spectrum of curves is obtained by varying stem cell division rate ( $r_1$ ),  $K_A$ , and  $K_R$  values simultaneously. Because discrete incidence probabilities are used to fit a continuous curve, our model allows predictive estimates of the number of individuals who have at least one cancerous crypt at ages outside our discrete range. For example, the graph highlights the importance of early screening for CRC as by age forty, approximately two in every one hundred individuals will have acquired at least one cancerous crypt, even

though they may not be experiencing symptoms. It is worth noting that while the figure description asserts that the figure only varies  $r_1$ , the supplied mathematica code seems to suggest division rate and carrying capacity were varied in tandem, and this procedure was necessary to produce the same diverging behavior at short and long times, and converging behavior around the incidence data.

SSE Values for Best Parameter Sets					
Set \ Params	$r$	$K_A$	$K_R$	$SSE$	
<i>Best</i>	156	562	1780	$4.63 * 10^{-5}$	
2	168	17	1330	$5.22 * 10^{-5}$	
3	168	13	1330	$5.22 * 10^{-5}$	
4	168	10	1330	$5.22 * 10^{-5}$	
5	168	23	1330	$5.22 * 10^{-5}$	

Table 1: 5 best parameter sets from varing  $r, K_A$ , and  $K_R$ .

Table 1 displays the sum of squared error (SSE) between a given curve and the observed data points for the five best parameter sets from our fitting procedure. Cell division rate was varied between 150 and 270 and K values were varied from 10 to 3000. From this table, we can see that for a crypt size of  $10^7$ , predicted stem cell division rates are fairly low, and in most cases the carrying capacity for  $K_A$  APC first cells was lower than the  $K_R$ , the carrying capacity for KRAS first cells.

These incidence curves and associated parameters verify the accuracy of our model, highlights it's predictive power, and may reveal the behavior of intracrypt dynamics.

## 4 Research Question

After exploring the model presented in the Wang et al paper, some questions emerged about its behavior. The question we chose to pursue is, which crypt transition has the greatest impact of long-term cancer incidence in the mathematical model proposed in the Wang et al paper? To address this question, we adopted a sensitivity analysis approach.

## 5 Linearization

Our system of ODEs is nonlinear due to the presence of the carrying capacity terms. This prevents us from doing the sensitivity analysis as the entries in our transition matrix are correlated with each other in a non-linear way. Hence, we need to perform a linearization of our ODE system to calculate the sensitivity matrix.

The idea of linearization is straightforward. We believe that our nonlinear system behaves linearly in a small region around its equilibrium points. We can approximate such linear behavior through Taylor expansion around its equilibrium point.

$$f(\vec{x}) \approx f(\vec{x}_{eq}) + \frac{\partial f}{\partial x} \Big|_{\vec{x}_{eq}} (\vec{x} - \vec{x}_{eq}) + O(h^2) \quad (4)$$

Notice that the first term  $f(\vec{x}_{eq})$  is the system behaviour at the equilibrium point which by definition is zero. Hence we have,

$$f(\vec{x}) \approx \frac{\partial f}{\partial x} \Big|_{\vec{x}_{eq}} (\vec{x} - \vec{x}_{eq}) = J * (\vec{x} - \vec{x}_{eq}) \quad (5)$$

Where  $J$  is the Jacobian matrix which is computed by taking the derivative with respect to all variables.

$$J = \left[ \frac{\partial f}{\partial x} \Big|_{\vec{x}_{eq}} \right] = \begin{bmatrix} \frac{\partial \dot{n}_1}{\partial n_1} & \frac{\partial \dot{n}_1}{\partial n_2} & \frac{\partial \dot{n}_1}{\partial n_3} & \frac{\partial \dot{n}_1}{\partial n_4} & \frac{\partial \dot{n}_1}{\partial n_5} \\ \frac{\partial \dot{n}_2}{\partial n_1} & \frac{\partial \dot{n}_2}{\partial n_2} & \frac{\partial \dot{n}_2}{\partial n_3} & \frac{\partial \dot{n}_2}{\partial n_4} & \frac{\partial \dot{n}_2}{\partial n_5} \\ \frac{\partial \dot{n}_3}{\partial n_1} & \frac{\partial \dot{n}_3}{\partial n_2} & \frac{\partial \dot{n}_3}{\partial n_3} & \frac{\partial \dot{n}_3}{\partial n_4} & \frac{\partial \dot{n}_3}{\partial n_5} \\ \frac{\partial \dot{n}_4}{\partial n_1} & \frac{\partial \dot{n}_4}{\partial n_2} & \frac{\partial \dot{n}_4}{\partial n_3} & \frac{\partial \dot{n}_4}{\partial n_4} & \frac{\partial \dot{n}_4}{\partial n_5} \\ \frac{\partial \dot{n}_5}{\partial n_1} & \frac{\partial \dot{n}_5}{\partial n_2} & \frac{\partial \dot{n}_5}{\partial n_3} & \frac{\partial \dot{n}_5}{\partial n_4} & \frac{\partial \dot{n}_5}{\partial n_5} \end{bmatrix} \quad (6)$$

To compute  $J$ , we first expand our system of odes as follows:

$$\begin{cases} \dot{n}_1 = -R_{12}n_1 - R_{14}n_1, \\ \dot{n}_2 = R_{12}n_1 - R_{23}n_2 - R_{25}n_2, \\ \dot{n}_3 = R_{23}n_2 - R_{36}n_3 + \gamma_3n_3 - \frac{\gamma_3n_3^2}{K_A} - \frac{\gamma_3n_3n_4}{K_A} - \frac{\gamma_3n_3n_5}{K_A} - \delta n_3, \\ \dot{n}_4 = R_{14}n_1 - R_{45}n_4 + \gamma_4n_4 - \frac{\gamma_4n_4n_3}{K_R} - \frac{\gamma_4n_4^2}{K_R} - \frac{\gamma_4n_4n_5}{K_R} - \delta n_4, \\ \dot{n}_5 = R_{25}n_2 + [R_{45}n_4] - R_{56}n_5 + \gamma_5n_5 - \frac{\gamma_5n_5n_3}{K_R} - \frac{\gamma_5n_5n_4}{K_R} - \frac{\gamma_5n_5^2}{K_R} - \delta n_5. \end{cases} \quad (7)$$

Taking derivative respect to all variables,  $n_1, n_2, n_3, n_4, n_5$ , gives us the Jacobian matrix.

$$J = \begin{bmatrix} -R_{12} - R_{14} & 0 & 0 & 0 & 0 \\ R_{12} & -R_{23} - R_{25} & 0 & 0 & 0 \\ 0 & R_{23} & A & -\frac{\gamma_3n_3}{K_A} & -\frac{\gamma_3n_3}{K_A} \\ R_{14} & 0 & -\frac{\gamma_4n_4}{K_R} & B & -\frac{\gamma_4n_4}{K_R} \\ 0 & R_{25} & -\frac{\gamma_5n_5}{K_R} & R_{45} - \frac{\gamma_5n_5}{K_R} & C \end{bmatrix} \quad (8)$$

$$\begin{aligned}
A &= -R_{36} + \gamma_3 - \frac{2\gamma_3 n_3}{K_A} - \frac{\gamma_3 n_4}{K_A} - \frac{\gamma_3 n_5}{K_A} - \delta \\
B &= -R_{45} + \gamma_4 - \frac{\gamma_4 n_3}{K_R} - \frac{2\gamma_4 n_4}{K_R} - \frac{\gamma_4 n_5}{K_R} - \delta \\
C &= -R_{56} + \gamma_5 - \frac{\gamma_5 n_3}{K_R} - \frac{\gamma_5 n_4}{K_R} - \frac{2\gamma_5 n_5}{K_R} - \delta
\end{aligned}$$

Bring in our equilibrium points which is  $[\dot{n}_1, \dot{n}_2, \dot{n}_3, \dot{n}_4, \dot{n}_5] = \vec{0}$ . We have the following result.

$$J_{eq} = \begin{bmatrix} -R_{12} - R_{14} & 0 & 0 & 0 & 0 \\ R_{12} & -R_{23} - R_{25} & 0 & 0 & 0 \\ 0 & R_{23} & -R_{36} + \gamma_3 - \delta & 0 & 0 \\ R_{14} & 0 & 0 & -R_{45} + \gamma_4 - \delta & 0 \\ 0 & R_{25} & 0 & R_{45} & -R_{56} + \gamma_5 - \delta \end{bmatrix} \quad (9)$$

Then we multiply our variables back to restore the linearized system of odes. Notice that we also added back our last equation from the original nonlinear ode system. P is the probability of having a type 6 crypt and also our variable of interest.

$$\begin{cases} \dot{n}_1 = -(R_{12} + R_{14})n_1, \\ \dot{n}_2 = R_{12}n_1 - (R_{23} + R_{25})n_2, \\ \dot{n}_3 = R_{23}n_2 - R_{36}n_3 + \gamma_3 n_3 - \delta n_3, \\ \dot{n}_4 = R_{14}n_1 - R_{45}n_4 + \gamma_4 n_4 - \delta n_4, \\ \dot{n}_5 = R_{25}n_2 + R_{45}n_4 - R_{56}n_5 + \gamma_5 n_5 - \delta n_5 \\ \dot{P} = R_{36}n_3 + R_{56}n_5(1 - P) \end{cases} \quad (10)$$

We notice that this linearized ODE system looks very much the same as our original system. The only difference is that the carrying capacity term is gone. That term is the source of nonlinearity and now this resulting ODE system is linear.

## 6 Comparison between Nonlinear and Linearized System

The Wang et. al paper suggests that the carrying capacity terms involving  $K_A$  and  $K_R$  are necessary to accurately produce incidence curves that fit epidemiological data. Given that our linearized

system was necessary for sensitivity analysis, but necessarily did not include these competition terms, we were interested in verifying if accurate behavior of the model was indeed nonrecoverable.

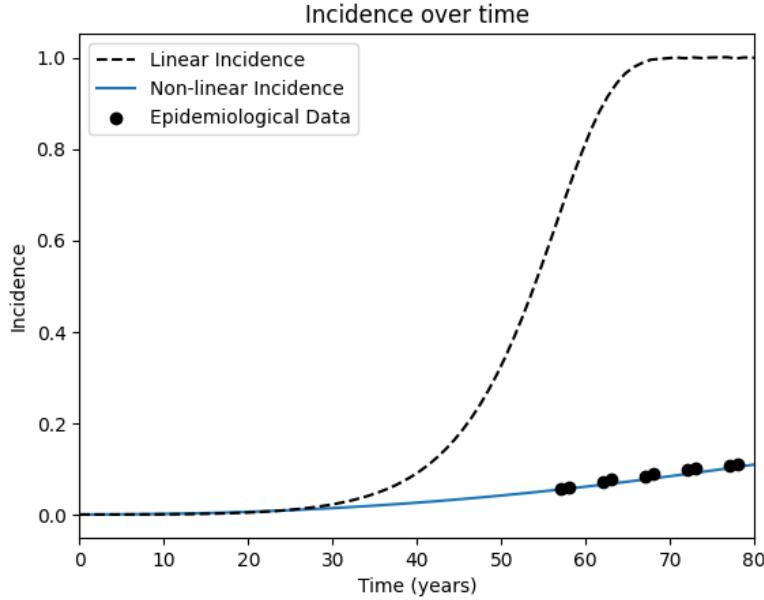


Figure 3: Comparison of incidence curves produced by the nonlinear and linearized systems.

From this figure, we see immediately that this competition term which couples mutant crypt populations is necessary to produce reliable incidence curves. Without this term, mutant crypt populations grow exponentially without limit, leading to a steep increase in the probability of having a type six crypt form. In the linear case, this probability reaches 1 within a human lifetime. This, thankfully, vastly exceeds observed incidence rates which do not climb above 0.2 within a normal human lifetime.

While we confirmed that the linearized system does not produce a relevant incidence curve, we were interested to see if the results of our sensitivity analysis could be applied to both cases.

## 7 Transition Matrix and Sensitivity Analysis

To compute the sensitivities, we first need to determine the transition matrix  $A$ . We can derive from the linearized approximation of the original ODE system, and each entry in this matrix represents the rate of crypt transition from stage  $i$  where  $i$  is the row index, to stage  $j$  where  $j$  is the column index.

$$A = \begin{bmatrix} 1 - (R_{12} + R_{14}) & 0 & 0 & 0 & 0 & 0 \\ R_{12} & 1 - (R_{23} + R_{25}) & 0 & 0 & 0 & 0 \\ 0 & R_{23} & (1 - R_{36}) + \gamma_3 - \delta & 0 & 0 & R_{35} \\ R_{14} & 0 & 0 & (1 - R_{45}) + \gamma_4 - \delta & 0 & 0 \\ 0 & R_{25} & 0 & R_{45} & (1 - R_{56}) + \gamma_5 - \delta & R_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

From that transition matrix, we computed sensitivities by first calculating the dominant left and right eigenvector of that matrix using

$$\begin{cases} A\vec{w} = \lambda\vec{w} & \text{left} \\ A^T\vec{v} = \lambda\vec{v} & \text{right} \end{cases} \quad (12)$$

Applying the sensitivity formula to find the sensitivity of each transition from stage i to j. We have

$$s_{ij} = \frac{w_i \times v_j}{\vec{w} \cdot \vec{v}} \quad (13)$$

The computed sensitivity matrix is shown below, and the max value entry is in row 3, column 3, indicating that the size of the total crypt population between stages 1 through 5 have the highest sensitivity to the stage 3 to stage 3 crypt transition. So the most relevant factor is how much the  $n_3$  crypts are persisting year-to-year in the population.

$$S = \begin{bmatrix} 0 & 0 & 2.16 * 10^{-6} & 0 & 0 & 0 \\ 0 & 0 & 1.04 * 10^{-3} & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.36 * 10^{-5} & 0 & 0 & 0 \end{bmatrix} \quad (14)$$

## 8 Empirical Verification

### 8.1 Linear model

We want to see whether actual behavior of the models matches the behavior predicted from sensitivity analysis. By manually adjusting each crypt transition rate, we can identify the transition that has the greatest influence on cancer incidence. Figure 4 displays testing in the linear model. The

red line representing  $n_3 \rightarrow n_3$  is the most diverged from the black control line. Since sensitivity analysis predicts that changes to  $n_3 \rightarrow n_3$  should cause the greatest shifts in the system, then this result is in agreement with the sensitivity analysis. This agreement is also statistically supported by the sum of squared error (SSE) analysis shown in Table 2. Biologically, it suggests that the degree to which APC-knockout crypts persist and multiply from year-to-year, is the most influential factor for cancer incidence, at least in this linear model using these specific parameter values.

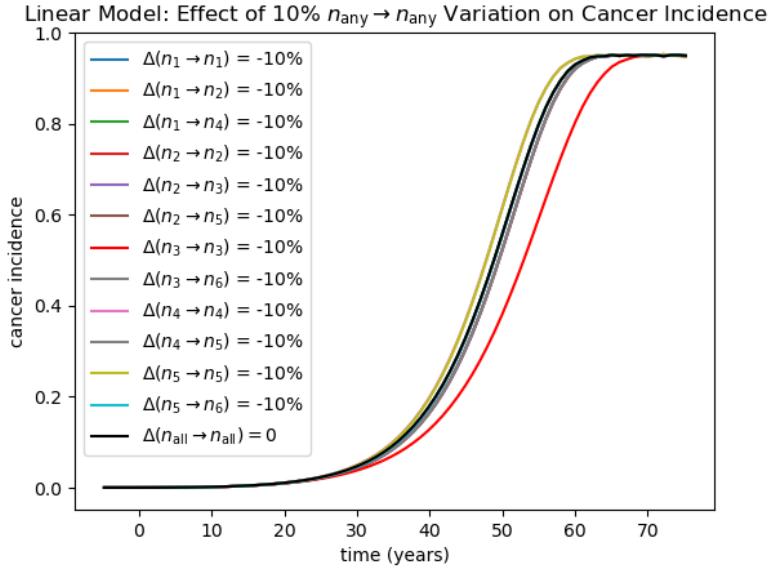


Figure 4: Describes how a 10% decrease in  $n_i \rightarrow n_j$  affects cancer incidence in the linear model. The dotted black line represents the control case where  $\Delta(n_{all} \rightarrow n_{all}) = 0$ . Each colored line represents the cancer incidence when a 10% decrease in each  $n_{any} \rightarrow n_{any}$  is applied. Increased divergence from the black line suggests greater influence on cancer incidence.

SSE Associated with Varying Crypt Transition Rates in Linear ODE						
Start \ End	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$
$n_1$	$1.88E - 2$	$1.70E - 2$		$4.48E - 6$		
$n_2$		$1.71E - 2$	$1.58E - 2$		$3.01E - 5$	
$n_3$			$5.76E - 1$			$1.58E - 2$
$n_4$				$4.29E - 2$	$4.85E - 5$	
$n_5$					$3.97E - 2$	$1.55E - 4$
$n_6$						

Table 2: SSE is a quantitative measure of difference between the black control line and each colored line in Figure 4. Larger SSE indicates a higher effect on cancer incidence. Any blank represents a transition that isn't relevant to this model. Red text marks the cell with highest SSE, indicating that this transition empirically has the greatest relative effect on cancer incidence.

## 8.2 Nonlinear model

After verifying that predicted and empirical behavior match in the linear model, we wanted to also verify this match or mismatch in the nonlinear model. Figure 5 displays testing in the nonlinear model.

Multiple colored lines that mostly correspond to transitions between stages 3, 5, and 6 appear to be tied for furthest from the black line. In contrast to the fairly clean-cut results of the sensitivity analysis, this empirical testing doesn't suggest a single particular transition responsible for most of cancer incidence. This judgment is supported by the SSE values shown on table 3. These values describe  $n_5 \rightarrow n_6$  as the most impactful crypt transition. The transition of interest  $n_3 \rightarrow n_3$  barely trails behind in fourth place. This nonlinear system might be more resistant to changes compared to the linear approximation due to the carrying capacity term. Since the carrying capacity includes the sum  $n_3 + n_4 + n_5$ , it may be "coupling" the  $n_3$ ,  $n_4$ , and  $n_5$  transitions to each other, causing these transitions to behave similarly. Biologically, this result implies that crypt transitions from the last stages of precancer, stages 3 and 5, are broadly and similarly impactful on cancer incidence rates, at least in this nonlinear model using these specific parameter values.

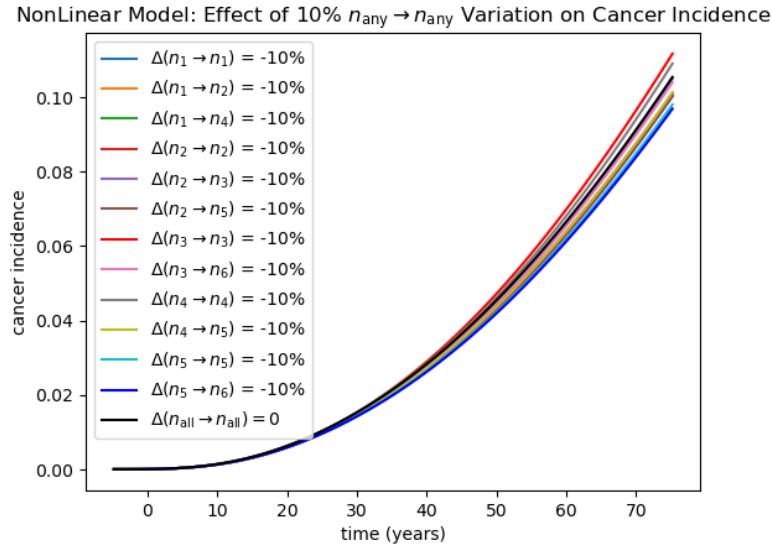


Figure 5: Describes how a 10% decrease in  $n_i \rightarrow n_j$  affects cancer incidence in the nonlinear model. The dotted black line represents the control case where  $\Delta(n_{all} \rightarrow n_{all}) = 0$ . Each colored line represents the cancer incidence when a 10% decrease in each  $n_{any} \rightarrow n_{any}$  is applied.

SSE Associated with Varying Crypt Transition Rates in NonLinear ODE						
Start \ End	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$
$n_1$	$5.22E - 4$	$4.11E - 4$		$4.48E - 6$		
$n_2$		$4.13E - 4$	$1.10E - 5$		$2.98E - 4$	
$n_3$			$4.95E - 4$			$5.22E - 5$
$n_4$				$1.35E - 4$	$2.823E - 4$	
$n_5$					$9.03E - 4$	<b><math>1.17E - 3</math></b>
$n_6$						

Table 3: SSE is a quantitative measure of difference between the black control line and each colored line in Figure 5. Larger SSE indicates a higher effect on cancer incidence. Any blank represents a transition that isn't relevant to this model. Red text marks the cell with highest SSE, indicating that this transition empirically has the greatest relative effect on cancer incidence.

## 9 Conclusion

Our goal was to identify the crypt transition with the largest relative impact on cancer incidence, given the mathematical model presented in the Wang et. al paper. We chose to approach this problem by first modeling crypt populations as a stage-structured system represented by a stage-structured matrix, and then computing the sensitivities to compare the impact of different crypt transitions on the total crypt population. There are two important caveats to this method.

First, the transition matrix requires constant values in each cell in order to compute numerical sensitivity values. To fulfill this constraint, the rate of crypt transitions must be linear. Since the model presented in the paper includes a nonlinear crypt competition parameter, we needed to find a way to eliminate or approximate this term through linearization. Biologically, this crypt competition parameter represents the competition for nutrients and space that real crypts encounter on the colonic surface, which prevents unlimited growth via fission. By calculating the Jacobian of the system around the 0 vector equilibrium point, we ended up with a linear system that was very similar to the original nonlinear system, except the crypt competition term is gone. Since we computed the sensitivity analysis on this linear approximation of the system and not the original system, another question of interest emerged from this work: can this system be faithfully approximated through the 0 equilibrium Jacobian linearization? We attempt to explore this question by plotting and comparing the cancer incidence curves generated by running each model. In this plot, we observe that the linear approximation adopts a clear sigmoidal curve within the natural human lifespan (80 years), where it predicts a 100% probability of cancer by age 60. On the other hand, the growth of the original nonlinear model is stunted by the crypt competition parameter, allowing the model to predict cancer incidence that more accurately reflects collected epidemiological data.

Second, the sensitivity analysis does not directly tell us anything about the crypt transitions' impact on cancer incidence, but rather their impact on the total crypt population. Total crypt population can be represented as  $\sum_{i=1}^6 n_i$ , while cancer incidence ( $I$ ) is represented as  $\dot{I} \propto (R_{36}*n_3 + R_{56}*n_5)*(1-P)$ , where  $P$  is the probability of cancer incidence.

Due to the fission behavior of stage 3 and stage 5 crypts that is defined by the linear model, crypt population is generally correlated with the rate of acquiring a cancerous crypt. However, to definitively compare the sensitivities of each crypt transition on cancer incidence, we needed to empirically test it in a model that directly calculates cancer incidence. We attempted to do this by plotting and comparing the cancer incidence curves described by the mathematical model, created in response to perturbing each crypt transition rate by -10%. This procedure was done in the linear model, where we found that both the empirical testing and sensitivity analysis pointed to the same conclusion: the  $n_3 \rightarrow n_3$  crypt transition is the most important player, aka how much each  $n_3$  crypt specifically is persisting from year to year in the crypt population.

To find out whether our sensitivity analysis can be applied to the original nonlinear model, we performed the same procedure in the nonlinear model. The result from this testing conflicted with the sensitivity analysis. It suggests that the general behavior of end-stage precancerous crypts,  $n_3$  and  $n_5$ , exert the most impact on overall cancer incidence, as opposed to the single  $n_3 \rightarrow n_3$  transition suggested by the sensitivity analysis. This discrepancy points to the secondary question proposed earlier in this section: can this system be faithfully approximated through the 0 equilibrium Jacobian linearization? According to this analysis, no. Since the crypt transitions in the linear approximation have demonstrably different impacts on cancer incidence compared to the original model, it suggests that the linear approximation is modelling a different system entirely that is not similar to the original model. At least in the context of plotting cancer incidence over a natural human lifetime, this linear model  $\not\approx$  the nonlinear model.

## 10 Future Directions

There are multiple gaps in the current analysis that could be explored in future work. During linearization, we chose to only calculate around a single equilibrium point of zero. There are potentially other equilibrium points that could be worth solving for linearization. One approach could be to use Python’s SymPy toolbox. With this tool and with plugging in the chosen parameter set, we have managed to find two other nonnegative equilibrium points, aside from the zero vector point. The path has been laid for anyone wanting to fully explore the linearized models offered by this Jacobian linearization method.

Another possibility is to use Lyapunov stability analysis to directly study the nonlinear ODE. Lyapunov stability is classically defined as a property of an equilibrium. When a system starts within a certain small distance from that equilibrium and it stays within that small distance forever, as opposed to converging or diverging, then that equilibrium is Lyapunov-stable (Goldhirsch et al., 1987). Instead of solving around a specific equilibrium point, we can find the Lyapunov stability around the control black line, with respect to each colored line. If the control black line has the lowest Lyapunov stability with respect to any specific colored line, then it implies that the ODE system is the most vulnerable to perturbations in this crypt transition. One caveat to this method is that there are arguments against the use of Lyapunov analysis to study perturbations of populations (Justus, 2008).

## 11 Appendix

### 11.1 Code

All of our code, including the sensitivity analysis and replication of figures from articles, are in the GitHub repository at [https://github.com/ivanygao/UW\\_AMATH422-522\\_Colorectal\\_Cancer.git](https://github.com/ivanygao/UW_AMATH422-522_Colorectal_Cancer.git).

## 11.2 Coefficients Used

Parameters used										
	$N_{crypts}$	$r$	$K_a$	$K_r$	$u$	$\mu$	$\gamma_3$	$\gamma_4$	$\gamma_5$	$\delta$
Ours	$10^7$	156	562	1780	$10^{-7} * r$	$10^{-9} * r$	0.2	0.07	0.07	0.05
Paper	$10^7$	204	1000	17	$10^{-7} * r$	$10^{-9} * r$	0.2	0.07	0.07	0.05

Table 4: Comparison of parameter values used in our model and those reported in the reference paper.

Many of our parameters match exactly, and were taken from relevant sections of the paper as well as through examining the mathematical code. Paper values for  $r$ ,  $K_A$ , and  $K_R$  can be found in appendix 1, table 3 of the Wang et. al paper. The death rate  $\delta$  is found in the figure description for figure 2. The computation of  $u$  and  $\mu$  was taken from their supplementary code.

The discrepancy in parameters results from being unable to reproduce a good fit of incidence to epidemiological data using the best parameter values reported in the paper, shown above. Thus, our own fitting procedure and parameters were used to generate all other plots.

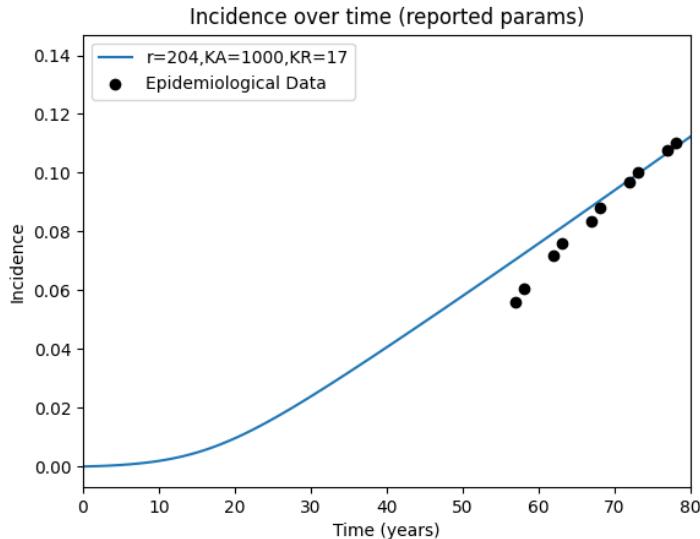


Figure 6: Poor fit results when using best fit parameters reported in Wang et. al paper.

It is important to note that their reported best parameter set contained exclusively values with  $K_A > K_R$ , while ours contained exclusively values with  $K_R > K_A$ , as such, we were unable to corroborate their figure 3c, where best fit parameters are reported to correspond with an APC first biological mechanism.

## References

- Cole, B., Logan, R., Halabi, S., Benamouzig, R., Sandler, R., Grainge, M., Stanislas, C., & Baron, J. (2009). Aspirin for the chemoprevention of colorectal adenomas: Meta-analysis of the randomized

- trials. *Journal of the National Cancer Institute*, 101, 256–66. <https://doi.org/10.1093/jnci/djn485>
- Goldhirsch, I., Sulem, P.-L., & Orszag, S. A. (1987). Stability and lyapunov stability of dynamical systems: A differential approach and a numerical method. *Physica D: Nonlinear Phenomena*, 27(3), 311–337. [https://doi.org/https://doi.org/10.1016/0167-2789\(87\)90034-0](https://doi.org/10.1016/0167-2789(87)90034-0)
- Justus, J. (2008). Ecological and lyapunov stability. *Philosophy of Science*, 75(4), 421–436. <https://doi.org/10.1086/595836>
- Komarova, N., Lengauer, C., Vogelstein, B., & Nowak, M. (2002). Dynamics of genetic instability in sporadic and familial colorectal cancer. *Cancer Biology and Therapy*, 1(6), 685–692. <https://doi.org/10.4161/cbt.321>
- Mamis, K., Zhang, R., & Bozic, I. (2023). Stochastic model for cell population dynamics quantifies homeostasis in colonic crypts and its disruption in early tumorigenesis. *Proceedings of the Royal Society B: Biological Sciences*, 290(2009), 20231020. <https://doi.org/10.1098/rspb.2023.1020>
- Nguyen, L. H., Goel, A., & Chung, D. C. (2020). Pathways of colorectal carcinogenesis. *Gastroenterology*, 158(2), 291–302. <https://doi.org/10.1053/j.gastro.2019.08.059>
- Paterson, C., Clevers, H., & Bozic, I. (2020). Mathematical model of colorectal cancer initiation. *Proceedings of the National Academy of Sciences*, 117(34), 20681–20688. <https://doi.org/10.1073/pnas.2003771117>
- Siegel, R. L., Giaquinto, A. N., & Jemal, A. (2024). Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*, 74(1), 12–49. <https://doi.org/https://doi.org/10.3322/caac.21820>
- Wang, Y., Boland, C. R., Goel, A., Wodarz, D., & Komarova, N. L. (2022). Aspirin's effect on kinetic parameters of cells contributes to its role in reducing incidence of advanced colorectal adenomas, shown by a multiscale computational study (C. S. Williams, E. Franco, E. G. Luebeck, & A. Chan, Eds.). *eLife*, 11, e71953. <https://doi.org/10.7554/eLife.71953>



# Aspirin's effect on kinetic parameters of cells contributes to its role in reducing incidence of advanced colorectal adenomas, shown by a multiscale computational study

Yifan Wang<sup>1</sup>, C Richard Boland<sup>2</sup>, Ajay Goel<sup>3</sup>, Dominik Wodarz<sup>1,4</sup>,  
Natalia L Komarova<sup>1\*</sup>

<sup>1</sup>Department of Mathematics, University of California Irvine, Irvine, United States;

<sup>2</sup>Department of Medicine, University of California San Diego School of Medicine, San Diego, United States; <sup>3</sup>Department of Molecular Diagnostics and Experimental Therapeutics, Beckman Research Institute of City of Hope Comprehensive Cancer Center, Duarte, United States; <sup>4</sup>Department of Population Health and Disease Prevention, University of California Irvine, Irvine, United States

**Abstract** Aspirin intake has been shown to lead to significant protection against colorectal cancer, for example with an up to twofold reduction in colorectal adenoma incidence rates at higher doses. The mechanisms contributing to protection are not yet fully understood. While aspirin is an anti-inflammatory drug and can thus influence the tumor microenvironment, *in vitro* and *in vivo* experiments have recently shown that aspirin can also have a direct effect on cellular kinetics and fitness. It reduces the rate of tumor cell division and increases the rate of cell death. The question arises whether such changes in cellular fitness are sufficient to significantly contribute to the epidemiologically observed protection. To investigate this, we constructed a class of mathematical models of *in vivo* evolution of advanced adenomas, parameterized it with available estimates, and calculated population level incidence. Fitting the predictions to age incidence data revealed that only a model that included colonic crypt competition can account for the observed age-incidence curve. This model was then used to predict modified incidence patterns if cellular kinetics were altered as a result of aspirin treatment. We found that changes in cellular fitness that were within the experimentally observed ranges could reduce advanced adenoma incidence by a sufficient amount to account for age incidence data in aspirin-treated patient cohorts. While the mechanisms that contribute to the protective effect of aspirin are likely complex and multi-factorial, our study demonstrates that direct aspirin-induced changes of tumor cell fitness can significantly contribute to epidemiologically observed reduced incidence patterns.

\*For correspondence:  
komarova@uci.edu

**Competing interest:** The authors declare that no competing interests exist.

**Funding:** See page 14

**Preprinted:** 12 May 2021

**Received:** 05 July 2021

**Accepted:** 15 March 2022

**Published:** 13 April 2022

**Reviewing Editor:** Christopher S Williams, Vanderbilt University, United States

© Copyright Wang et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Editor's evaluation

This work develops a multistage/component mathematical model to analyze advanced colorectal adenomas and the impact that aspirin therapy has on adenoma formation rates. This study will be interesting to the cancer evolution community and in particular those interested in colorectal cancer incidence. While the model is mainly focused on aspirin chemoprevention, the model could be adapted to test other putative preventative agents, and thus could have a broad impact.

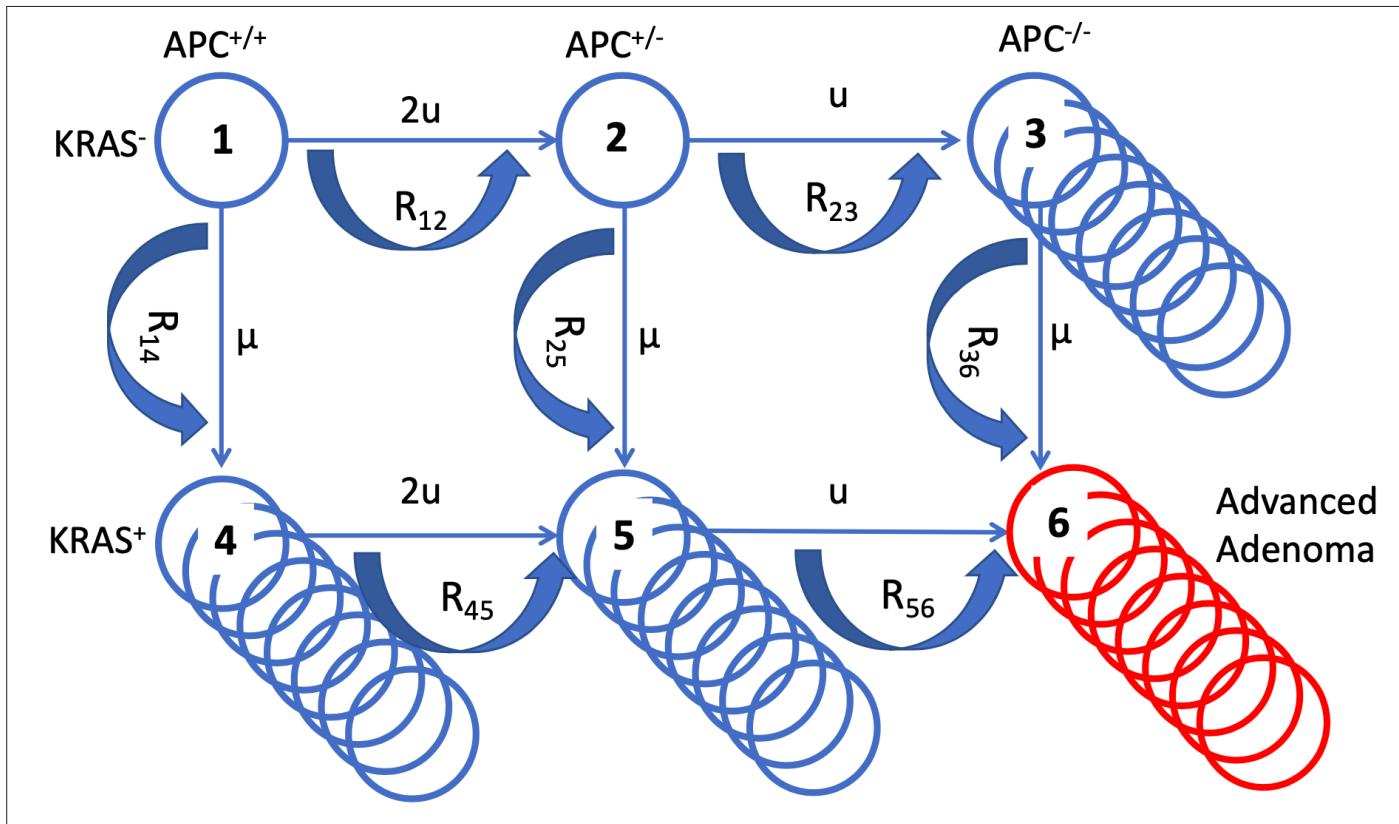
## Introduction

Colorectal cancer currently affects about 5% of the population in the USA and is a major cause of cancer-related deaths ([Siegel et al., 2020](#)). Prevention of colorectal cancer is an important goal in the quest to reduce morbidity and mortality. In this respect, long-term aspirin use has been shown to be effective ([Chan et al., 2012](#); [Thun et al., 1991](#)). Aspirin is a non-steroidal anti-inflammatory drug (NSAID) and is a cyclo-oxygenase (COX)-2 inhibitor ([Goel et al., 2003](#)). The CAPP2 trial [Burn et al., 2011](#) demonstrated that the intake of 600 mg of aspirin per day for 2 years resulted in a 63% reduction in colorectal cancer incidence in Lynch Syndrome patients. Interestingly, observation of the protective effect of aspirin required a follow-up time of more than 55 months ([Burn et al., 2011](#)). In a range of studies, aspirin has also been shown to reduce incidence of sporadic colorectal cancer ([Thun et al., 1991](#); [Tougeron et al., 2014](#); [Friis et al., 2015](#); [Lochhead and Chan, 2016](#); [Chan et al., 2007](#); [Drew et al., 2016](#); [Chan et al., 2005](#); [Rothwell, 2013](#)), and of adenomas ([Sandler et al., 2003](#); [Chan et al., 2004](#)), which are a precursor of cancer. This was evident both in observational studies and in randomized controlled trials, which are reviewed for example in [Lochhead and Chan, 2016](#); [Drew et al., 2016](#). These studies report a reduction in cancer or adenoma incidence of the order of 10–50% in aspirin-treated compared to placebo groups, depending on the exact dose and frequency of aspirin intake. While some studies failed to detect significant protective effects of aspirin, larger studies with higher aspirin doses and longer treatment times yielded significant results.

The mechanisms underlying the protective effect of aspirin are likely complex and multi-factorial. Inflammation is a possible driver of colorectal carcinogenesis ([Itzkowitz and Yio, 2004](#)), and aspirin can reduce the extent of inflammation in the cellular microenvironment, which might contribute to a reduced development of disease. Other microenvironmental effects, such as the composition of the colorectal microbiome ([Prizment et al., 2020](#); [Brennan et al., 2021](#); [Zhao et al., 2020](#)), have also been shown to determine the degree of protection provided by aspirin. Our previous in vitro and in vivo work, however, has shown that physiologically relevant aspirin concentrations can have a direct effect on tumor cells, reducing their rate of proliferation and increasing their death rate ([Shimura et al., 2020](#); [Zumwalt et al., 2017](#)). This not only results in reduced tumor growth, but can also lead to a lower probability that newly generated tumor cells successfully give rise to clonal expansion, thus increasing the likelihood that these initially transformed cells go extinct ([Wodarz et al., 2017](#)). This effect might contribute to the reduced incidence of colorectal cancer as a result of aspirin intake.

While these direct effects of aspirin on tumor cell division and death rates have been documented by us in vitro and in vivo ([Shimura et al., 2020](#); [Zumwalt et al., 2017](#)), and occurred under physiologically realistic doses, it is unclear to what extent these changes in cellular kinetics can potentially alter disease incidence. To evaluate this quantitatively, a mathematical modeling framework needs to be developed that predicts epidemiological incidence data based on cellular processes. There is a rich history of such approaches in the cancer literature in different contexts ([Fisher and Hollomon, 1951](#); [Nordling, 1953](#); [Armitage and Doll, 1954](#); [Luebeck and Moolgavkar, 2002](#); [Meza et al., 2008](#); [Moolgavkar, 1978](#); [Hornsby et al., 2007](#)), which has allowed researchers to gain fundamental insights into carcinogenic processes based on the interpretation of age-incidence data. Here, we describe a mathematical model of advanced adenoma formation and parameterize it by fitting epidemiological predictions to incidence data that document advanced adenoma occurrence as a function of age. We then use this model to test whether aspirin-mediated changes in cellular kinetics, as documented by our experiments, can result in reductions in advanced adenoma incidence that are comparable to those observed in aspirin-treated patient cohorts. We find that the magnitude of changes in the kinetics of transformed cell populations that we observed experimentally can result in a pronounced reduction of advanced adenoma incidence, and that the epidemiologically observed incidence reductions (between 10% and 50%) can be explained by our model. This indicates that the direct effects of aspirin on dividing cells can in principle explain a significant amount of the chemoprotective effect exerted by this drug. We note, however, that while this is a clear result that emerges from this mathematical modeling effort, other mechanisms of aspirin not included in this model (such as anti-inflammatory effects [[Sostres et al., 2014](#)] or modulation of the microbiome [[Prizment et al., 2020](#); [Brennan et al., 2021](#); [Zhao et al., 2020](#)]) are likely to also contribute to the observed protective effect.

We start by describing a mathematical model of advanced adenoma formation and show that when parameterized with experimentally obtained estimates, it can account for epidemiologically observed



**Figure 1.** A schematic illustrating the mathematical model. The six cell types are denoted by circles, the mutation rates that give rise to different types are marked by the straight arrows. Type 6 (late adenoma) is marked in red. Crypt conversion rates are indicated by circular arrows and crypt fission by multiple circles.

age-incidence curves, only as long as inter-crypt competition is explicitly included. We then use this model to simulate the effect of aspirin on the incidence of advanced adenomas in human populations, and compare model predictions to epidemiological data.

## Results

### Computational modeling

In order to quantify the effects of aspirin on colorectal cancer initiation and progression, we have designed a mathematical model that is rooted in the process of multistep carcinogenesis (Luebeck and Moolgavkar, 2002; Hornsby et al., 2007; Fearon, 2011; Ashley, 1969). Its assumptions are similar in principle to those in a recent study (Paterson et al., 2020), with important differences that are discussed below. There are two early molecular events that we postulate (without assuming their temporal order): (1) An inactivation of the APC gene, or a related event that affects the functioning of the beta-catenin/WNT signaling pathway, and (2) an activation of the KRAS oncogene (or another gain-of-function mutation). For simplicity, we will be referring to these mutations as APC- and KRAS-mutations, keeping in mind that the model still applies in the presence of a pair of another loss-of-function and a gain-of-function mutation (further discussed below). The inactivation of the APC tumor suppressor gene is a classic example of a loss-of-function mutation, which implies two molecular events, corresponding to the inactivation of the two copies of the gene. The associated mutation rate is therefore assumed to be  $u = 10^{-7}$  per cell division. The activation of the KRAS oncogene, on the other hand, is a gain-of-function event, whose mutation rate is about two orders of magnitude lower ( $\mu = 10^{-9}$  per cell division). The associated selection-mutation diagram is shown in **Figure 1** and contains six different cell populations, denoted as types 1 through 6. The populations occupying the top row (types 1–3) are characterized by an unmutated KRAS oncogene; the populations

of the bottom row (types 4–6) all have the KRAS mutation activated. Moving from left to right on this diagram, the number of inactivated copies of the APC gene increases from 0 to 2, such that populations of types 1 and 4 are APC<sup>+/+</sup>, populations of types 2 and 5 are both APC<sup>+/-</sup>, and populations of types 3 and 6 are APC<sup>-/-</sup>.

These different populations correspond to different stages in the pathway towards colorectal cancer. The presence of the APC<sup>-/-</sup> genotype has been related to the appearance of early adenomas (type 3) (Armaghani et al., 2012). Cells that have not inactivated APC but are characterized by KRAS activation (types 4 and 5) have been linked to aberrant crypt foci (Pretlow and Pretlow, 2005; Jass, 2006). The combination of both types of mutations (type 6) is thought to correlate with the growth of advanced adenomas (Armaghani et al., 2012). As in the previous study (Paterson et al., 2020), we assume flexibility regarding the order with which the different mutations can occur. Hence, it is assumed that the initial mutation can occur either in APC or in KRAS. It is, however, controversial whether adenoma formation can indeed be initiated by a mutation in KRAS. Some studies indicate that an initial mutation in KRAS leads to the formation of non-dysplastic polyps, which could represent an evolutionary dead end for neoplasias (Jen et al., 1994; Chan et al., 2003). On the other hand, it has been suggested that an initial KRAS mutation might be able to drive the initiation of colorectal carcinogenesis (Jass, 2006; Pretlow and Pretlow, 2005), based on mutation frequencies in aberrant crypt foci and adenomas. The assumed flexibility in the evolutionary pathway of the model accommodates these conflicting notions.

While the model assumptions about the pathways to adenoma formation are clearly defined in our model, it is important to point out that there are uncertainties in those assumptions, and that there is heterogeneity in the types of mutations that can lead to colorectal carcinogenesis. For example, it has been reported that among non-hypermutated colorectal tumors, KRAS was mutated in only about 43% of patient samples (Network, 2012), indicating the importance of a variety of evolutionary pathways. Our model, however, does not depend on the identity of particular mutations, but assumes the occurrence of mutation types; these are the inactivation of a tumor suppressor gene (which is a loss-of-function mutation, e.g. APC<sup>-/-</sup>), and a gain-of-function mutation, which can be in KRAS or an alternative gene. Our model predictions hold as long as the evolutionary pathway to advanced adenomas involves these two types of mutational events, regardless of their identity. We note that our model does not apply to potential cases of advanced adenomas that might develop via pathways characterized by a different number or different types of initiating events.

We model the population dynamics of the colon by using a colonic crypt as a basic unit, which is similar in concept to recently published work (Paterson et al., 2020). Our model is related to many previous theoretical investigations of the cell population dynamics of crypts (Komarova et al., 2002; Komarova et al., 2003; Nowak et al., 2002; Shahriyari and Komarova, 2013; Shahriyari et al., 2016), where stem cells (SC) were assumed to acquire random mutations in a constant-population turnover (birth and death) process, and selection happened at the level of individual stem cells. Once it was discovered that there were very few stem cells per crypt (Nicholson et al., 2018; Humphries and Wright, 2008), it became clear that the evolutionary dynamics can be conveniently described at the level of crypts, because crypts are likely to be homogeneous with respect to the driver mutations. The rate at which a crypt changes its mutational status from  $i$  to  $j$ , denoted by  $R_{ij}$ , depends on the population size (the number of stem cells per crypt), the mutation rate, and the relative fitness of the invading cell type compared to the resident cell type (Komarova et al., 2003; Nowak et al., 2002). The latter can be calculated from the cell displacement data reported in the literature. Cell types APC<sup>+/-</sup>, APC<sup>-/-</sup>, and KRAS + all have a selective advantage compared to the wild type, which we assume results in an increase of the SC division rate (see Section 2 of Appendix 1 for details).

Our model keeps track of crypts of different types (denoted as  $n_i$  for each type  $i$ ). Modified crypts of types APC<sup>-/-</sup> and KRAS + have been reported to undergo crypt fission; in other words, while the total population of a single crypt remains constant (even though it is populated by SCs that are fitter than the wild-type SCs), the crypt can undergo a doubling, thus increasing the total number of such modified crypts. The fission rates of different crypt types have been reported in the literature (Paterson et al., 2020; Nicholson et al., 2018; Humphries et al., 2013; Baker et al., 2014) and are denoted by  $\gamma_i$ ; we further denote by  $\delta$  the death rate of crypts, see Birtwell et al., 2020 for the role of crypt turn-over. We model these dynamics by using the following system of ordinary differential equations:

$$\begin{aligned}\dot{n}_1 &= -(R_{12} + R_{14})n_1, \\ \dot{n}_2 &= R_{23}n_1 - (R_{23} + R_{25})n_2, \\ \dot{n}_3 &= R_{23}n_2 - R_{36}n_3 + \gamma_3 n_3 \left(1 - \frac{n_3 + n_4 + n_5}{K_A}\right) - \delta n_3 \\ \dot{n}_4 &= R_{14}n_1 - R_{45}n_4 + \gamma_4 n_4 \left(1 - \frac{n_3 + n_4 + n_5}{K_R}\right) - \delta n_4 \\ \dot{n}_5 &= R_{25}n_2 - R_{45}n_4 - R_{56}n_5 + \gamma_5 n_5 \left(1 - \frac{n_3 + n_4 + n_5}{K_R}\right) - \delta n_5,\end{aligned}$$

where on the left hand side we have the rate of change for the population of crypts of each type, and the terms with the carrying capacity ( $K_A$  for KRAS<sup>-</sup> crypts and  $K_R$  for KRAS<sup>+</sup> crypts) represent competition among modified crypts that undergo crypt fission; in reference (Paterson et al., 2020) no crypt competition was included, such that  $K_R = K_A = \infty$  in that model. The initial conditions for the system above are given by  $n_i(0) = N_{\text{crypt}}$ ,  $n_i(0) = 0$ ,  $1 < i \leq 5$ , that is, initially all  $N_{\text{crypt}}$  crypts are wild type. Parameter values are presented in **Appendix 1—table 2**.

The probability to have produced a single crypt of type 6 (the APC<sup>-/-</sup>KRAS<sup>+</sup> phenotype) by time  $t$  is denoted by  $P(t)$  and is approximated by the following equation Chou and Wang, 2015,

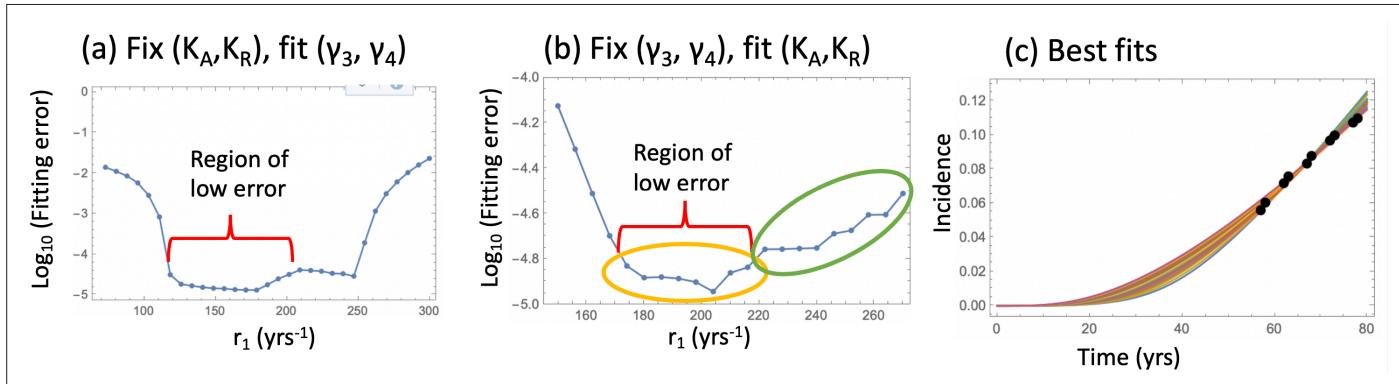
$$\dot{P} = (R_{56}n_5 + R_{36}n_3)(1 - P), \quad P(0) = 0$$

We further assume that crypts of type 6 engage in a fission-death dynamics (with the corresponding rates  $\gamma_6$  and  $\delta_6$ ). At the time of detection, an advanced adenoma is characterized by a certain size,  $N$ . If  $\Delta T$  denotes the expected time for the crypt population,  $n_6$ , to grow to the size of detection, then the value  $P(t-\Delta T)$  calculated above approximates the mathematical expression for the age-incidence curve for advanced adenoma. These approximations were checked against stochastic (Gillespie) simulations recording the incidence of size  $N$  colonies of type 6 crypts, yielding excellent agreement (see Section 5 of Appendix 1 for details).

### Fitting the adenoma incidence curve

Until recently, most of the parameters associated with cellular dynamics in colonic crypts were unknown, but presently many of the rates have been estimated with a high degree of confidence (Paterson et al., 2020), which makes it possible to parameterize the model and use it to answer questions about the process of crypt transformation and the dynamics of cancer initiation. Using the published data on the mutation rates, the total number of crypts, the number of SCs per crypt, and the relative fitness of different cell types (see **Appendix 1—table 2**), we first attempted to fit the model in the absence of crypt competition ( $K_R = K_A = \infty$ ), by varying the SC division rate within the physiological range and finding the best fitting value for crypt fission rates. The best fitting parameter combinations always corresponded to zero crypt fission rates. Non-zero crypt fission rates resulted in a much steeper rise in the adenoma incidence compared to that reported in Brenner et al., 2014. A similar result was obtained when we used different values for fitness differences (the exhaustive parameter search and a model selection procedure are described in Section 3 of Appendix 1). Finally, using the reported crypt fission rates (**Appendix 1—table 2**) we were not able to find a SC division rate within the biologically applicable range that would give the correct shape of the advanced adenoma incidence curve. The conclusion is that an unlimited exponential expansion of crypts by fission gives an unrealistically steep rise in incidence. This problem did not occur in reference (Paterson et al., 2020) because fitting of the age-specific incidence of CRC was not attempted, and instead, only the total life-time risk of CRC was compared to the model prediction.

Including crypt competition in the model has resolved this issue. We fixed the carrying capacity of type 3–5 crypts (parameters  $K_R$  and  $K_A$ ) to values much smaller than the initial number of healthy crypts,  $N_{\text{crypt}}$ , to ensure the presence of significant competition among the partially transformed crypts. Using this model, we were able to fit the data for a wide range of the SC division rates, with the non-zero best-fitting crypt fission rates that have the correct order of magnitude. Additionally, fixing the crypt fission rates to their reported values, we were able to find very well-fitting incidence curves for a wide range of SC division rates ( $r_1$ ), with the carrying capacity parameters ranging between about 100 and about 5000.



**Figure 2.** Fitting the nonlinear model to late adenoma incidence data. (a) The fitting error as a function of  $r_1$ , for the best fitting pairs ( $\gamma_3$  and  $\gamma_4$ ) with  $K_A = K_R = 1000$ ; see **Figure 2—source data 1**. (b) The fitting error as a function of  $r_1$ , for the best fitting values of ( $K_A, K_R$ ). Two groups of fits are marked with yellow and green ovals for the analysis of pathways, **Figure 3**. See **Figure 2—source data 2**. (c) The best fitting curves corresponding to increasing SC division rates,  $r_1$ , are plotted together with the epidemiological data (the values of  $r_1$  are taken from the Region of low error, panel (b)). In all panels,  $\delta = 0.05 \text{ yrs}^{-1}$ ,  $\gamma_6 = 1.01 \text{ yrs}^{-1}$ , expansion of  $n_6$  to  $N = 10^2$  crypts, and the rest of the parameters are as in **Appendix 1—table 2**. The code for panels (a–c) is provided, see **Source code 1**.

The online version of this article includes the following source data for figure 2:

**Source data 1.** Data for **Figure 2a**.

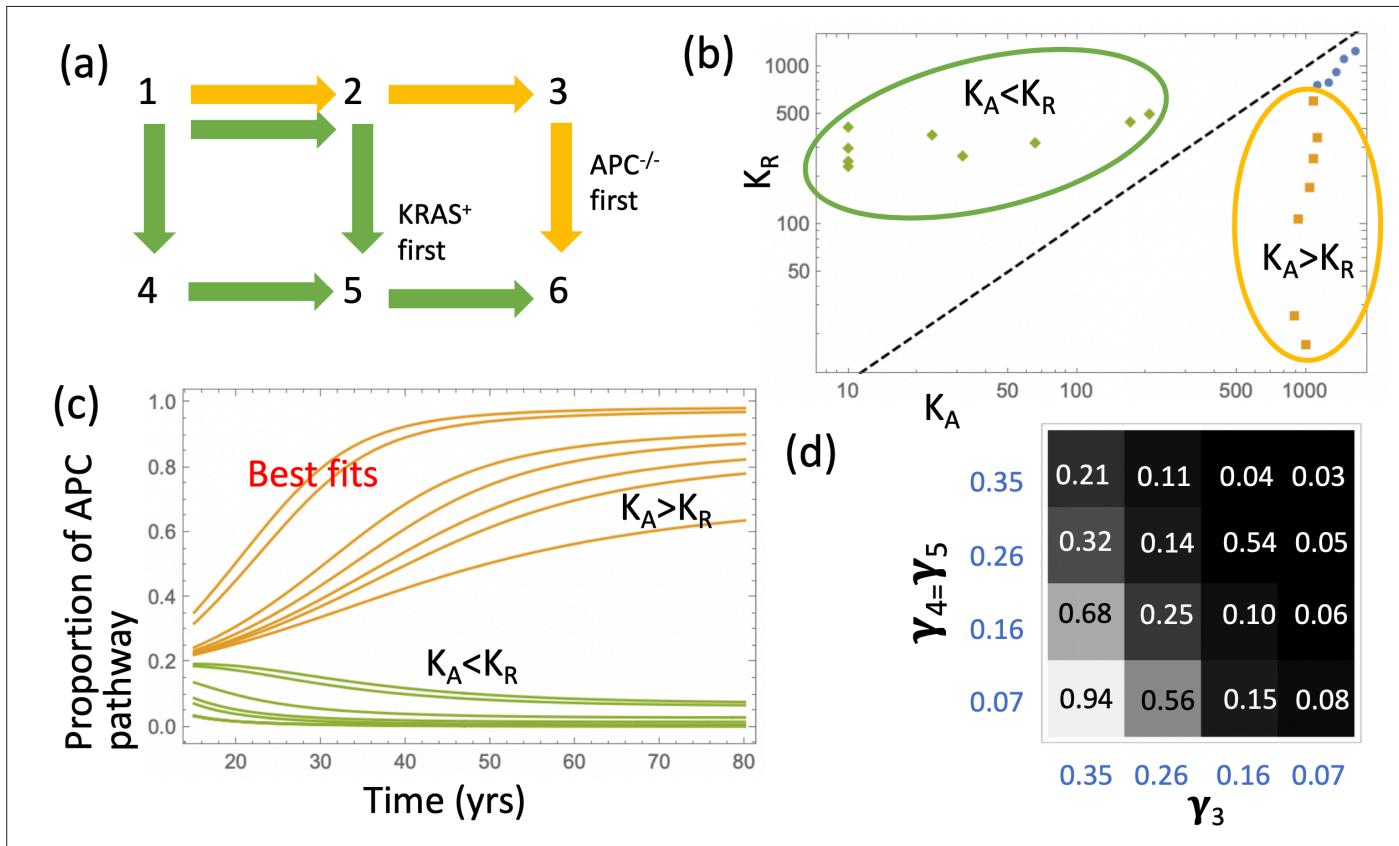
**Source data 2.** Data for **Figure 2b**.

For the model that includes crypt competition, it was possible to find nearly equally good fits for a range of biologically plausible parameter values, see **Figure 2**. The amount of data in the advanced adenoma incidence curve does not allow finding unique values for all the parameters, but instead it allows using many of the parameters fixed to their experimentally obtained values, and just fine-tuning the small number of remaining parameters whose value is unknown (such as  $K_A$  and  $K_R$ ) or only its range is known (such as the SC division rate). When using the parameterized model to study the role of aspirin, instead of selecting the best fitting parameter set, we included a number of parameter sets from the best fitting parameter ranges, to see how this variability influences the result.

### Pathways to adenoma

Next, we asked what is the most likely pathway that leads to the creation of the type 6 (advanced adenoma). It is possible that crypts of type 6 could be created by a KRAS mutation in a crypt of type 3 (we called this 'APC path'), or by an APC mutation in a crypt of type 5 ('KRAS path'), see panel (a) of **Figure 3**. We found, consistent with (Paterson et al., 2020), that the likelihood of each of these two pathways is determined by the crypt fission rates, and not by mutation rates or crypt conversion rates (**Figure 3(d)**); in addition, it is sensitive of the carrying capacity parameters,  $K_A$  and  $K_R$ . This is demonstrated in **Figure 3** by examining the best fitting parameter sets of **Figure 2(b)**. They naturally fall into two groups (circled in yellow and green): for the former group, the best fitting carrying capacity satisfy  $K_A > K_R$ , and for the latter group this inequality is reversed (see **Figure 3(b)**). We consider the former group biologically relevant, not only because it yields a smaller fitting error, but also because it corresponds to the type (APC<sup>-/-</sup>,KRAS<sup>+</sup>) crypts having a larger carrying capacity, which is consistent with this type being a more advanced stage.

The model allows for the calculation of the probabilities to develop an advanced adenoma through the APC-/ and KRAS pathways, functions  $P_{\text{APC}}(t)$  and  $P_{\text{KRAS}}(t)$ . Panel (c) of **Figure 3** plots these quantities as functions of age ( $t$ ) for the two groups of parameters. We observe that for the biologically relevant group where the carrying capacity associated with type 3 (APC<sup>-/-</sup>,KRAS<sup>+</sup>) is larger, the pathway through the inactivation of the APC gene is predominant. This is consistent with the conclusions of reference (Paterson et al., 2020). We also generated probability distributions of the numbers of type 3 and type 5 crypts at the time when the first type 6 crypt is generated (**Appendix 1—figures 12 and 13**). We observe that in the model, the number of type 3 (APC<sup>-/-</sup>,KRAS<sup>+</sup>) crypts is in the hundreds while type 5 (APC<sup>-/+</sup>,KRAS<sup>+</sup>) crypts are relatively rare. This might further argue against the importance of KRAS as an initiating event in disease evolution.



**Figure 3.** Pathways to adenoma. **(a)** A schematic representation of the two pathways. **(b)** For the two groups of fits in **Figure 2(b)**, the pairs ( $K_A, K_R$ ) are shown. The two groups are characterized by  $K_A > K_R$  and  $K_A < K_R$ , respectively. See **Figure 3—source data 1**. **(c)** The probabilities  $P_{APC}$  and  $P_{KRAS^+}$  are plotted as functions of time for these two groups of fits. **(d)** The proportion of the APC-path is shown as a function of the crypt fission rates,  $\gamma_3$  and  $\gamma_4 = \gamma_5$ . Expansion to  $10^2$  type-6 crypts is assumed (see **Appendix 1—figure 10** for expansion to  $10^5$  type-6 crypts); the rest of the parameters are as in **Appendix 1—table 2**. The code for panels **(b–c)** is provided, see **Source code 1**.

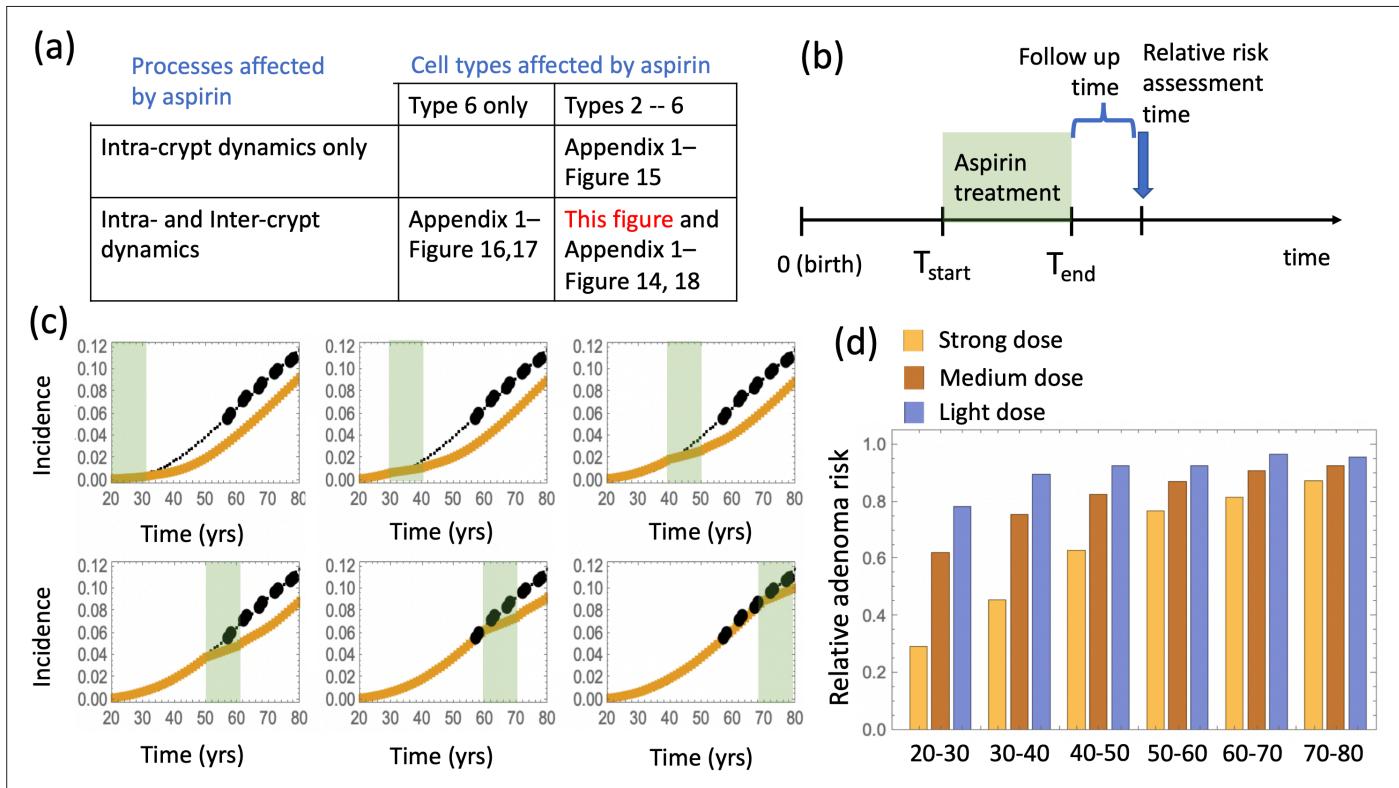
The online version of this article includes the following source data for figure 3:

**Source data 1.** Data for **Figure 3b**.

### The effect of aspirin

We asked, given that a variety of parameter values could lead to the same incidence curve, can we still say anything about the possible role of aspirin in cancer prevention/delay? To model the effect of aspirin on the relevant kinetic parameters, we used a variety of sources. One type of data was obtained by us in our earlier studies, where the effect of aspirin was quantified by measuring cells' kinetic parameters with and without aspirin treatment, *in vitro* and in xenografts (Shimura et al., 2020; Zumwalt et al., 2017). In other work, it has also been demonstrated that a related non-steroidal anti-inflammatory drug, sulindac, inhibited the fission of APC-deficient crypts and thus reduced adenoma numbers in mice.

It is, however, unclear which exact cell populations aspirin might affect *in vivo*. Therefore, we implemented the effect of aspirin in the epidemiological model by testing different sets of assumptions: (a) aspirin affects the fitness of cells within crypts (*intra-crypt dynamics*), and it may or may not affect crypt turnover dynamics through crypt fission and death rates (*inter-crypt dynamics*); (b) the fitness of type 6 cells is reduced, and the fitness of type 2–5 cells may or may not be reduced as well. In other words, only the most transformed cell type (that is, the most modified cell type that combines both the APC<sup>-/-</sup> mutation and the KRAS<sup>+</sup> mutation) is affected by aspirin, or all mutated cells, that is types 2–6, are affected. Different combinations of these assumptions have been explored, as summarized in **Figure 4(a)**.



**Figure 4.** The effect of aspirin on the advanced adenoma incidence curve. **(a)** Modeling assumptions and references to figures that present the results. **(b)** A schematic showing the timing of the model with aspirin treatment. **(c)** Predicted advanced adenoma incidence in the absence of aspirin treatment (thin black lines are the fitted curves and black dots are incidence data); and under aspirin treatment where the drug affects both inter- and intra-crypt dynamics of types 2 to -6 (yellow line). Each panel corresponds to aspirin treatment administered during one decade (20–30 years, 30–40 years, etc). The treatment period is shaded light green. See **Figure 4—source data 1**. **(d)** Relative incidence of advanced adenoma, where strong (yellow, the same as in panel **(c)**), medium (brown) and light (blue) aspirin treatment dose was used (**Table 1**). Treatment is applied for different decades (as marked under the bars), and the relative risk is evaluated at the end of the treatment decade (zero follow-up time). See **Figure 4—source data 2**. Parameter set #2 (**Appendix 1—tables 1–3**) was used.  $2.5 \times 10^5$  and  $5 \times 10^5$  independent simulations were used for each condition in panels **(c)** and **(d)** respectively. The code for panels **(c–d)** is provided, see **Source code 2** (Mathematica). See also **Source code 3** (Fortran), a stochastic simulation that produces simulated late adenoma incidence.

The online version of this article includes the following source data for figure 4:

**Source data 1.** Data for **Figure 4c**.

**Source data 2.** Data for **Figure 4d**.

The effect of aspirin on cellular kinetics is modeled by using results of our xenograft experiments (**Shimura et al., 2020**), where we documented a dose-dependent reduction in the cell division rate (fold difference  $F_r < 1$ ) and an increase in the cell death rate (fold difference  $F_d > 1$ ), see **Table 1**. The relationship between the experimentally used aspirin doses in mice (**Shimura et al., 2020**) and the number of aspirin pills per weeks in humans (**Nair and Jacob, 2016**) is also given in **Table 1**. The strongest dose we used in study (**Shimura et al., 2020**) roughly translates to 1–2 standard aspirin pills

**Table 1.** Aspirin doses in xenograft experiments, the equivalent human dose and the resulting changes in kinetic rates.

Dose in xeno- grafts (mg/kg)	Equivalent dose in humans (325 cm pills per week)	Fold difference in division rate, $F_r$	Fold difference Un death rate, $F_d$	Fitness factor
15	1.8	0.9	1.5	0.86
50	6.1	0.75	1.75	0.70
100	12.2	0.5	2.0	0.45

a day for humans, which is the second strongest dose considered in **Chan et al., 2004** (6–14 pills per week), see Section 6 of Appendix 1. When aspirin is applied in our model, we assumed the following:

### Effect on the intra-crypt dynamics

For the purposes of our model, it is the combined effect of aspirin on cell division and death rates that changes the cells' relative fitness and decreases the probability of crypt conversion. To translate this information into the fold decrease in SC fitness, we note that, while the fold-reduction in division rate could be directly implemented, an increase in death rate is less straightforward. This is because in contrast to cell lines, with SCs, cell removal can occur through a combination of apoptosis and loss through differentiation, which might be the dominant component in the colorectal tissue. Therefore, if the rate of SC apoptosis is increased, say, two-fold in the presence of aspirin, this does not translate to a two-fold reduction in SC fitness. In the extreme scenario of zero SC death in the absence of aspirin, a two-fold increase in this parameter will not lead to a change in SC fitness. To calculate the fitness factor, we assumed that the removal rate of SCs,  $d$ , is comprised of 90% differentiation and 10% apoptosis, and that it is the latter that is affected by aspirin. If in the absence of aspirin, cellular fitness is given by the ratio  $r/d$ , then in the presence of aspirin this changes to  $r/d \times F_r/(0.9 + 0.1F_d)$ , which gives the fitness factor in **Table 1**. This factor enters into the crypt conversion rate, see Section 2 of Appendix 1. In particular, if only type 6 is affected, then rates  $R_{36}$  and  $R_{56}$  will experience a reduction. If types 2 to –6 are affected, then all conversion rates will be reduced.

### Effect of inter-crypt dynamics<sup>sf</sup>

In addition to affecting cellular fitness within the crypts, it is also logical to assume that aspirin reduces crypt fission rates and increases crypt death rates (the bottom row of the table in **Figure 4(a)**). This is supported by data (**Fischer et al., 2014**), and the rationale behind this assumption is that crypt fission is ultimately connected with divisions of individual cells, and crypt death is associated with cell death. Therefore, we assume that under aspirin treatment,  $\gamma_i \rightarrow F_r \rightarrow \gamma_i$  and  $\delta \rightarrow F_d\delta$  (that is, the fold-differences apply to the crypt fission and death rates). Again, this could affect the most modified crypts only (type 6), thus reducing the rate  $\gamma_6$  and increasing the death rate  $\delta_6$ ; alternatively, this could affect to all type 2 to –6 crypts, thus reducing all the crypt fission rates and increasing all the crypt death rates.

The delaying effect of aspirin was studied by using stochastic (Gillespie) simulations, where the models were run according to the schematic in **Figure 4(b)**. In particular, model parameters were switched from their (best fitted, aspirin-free) values to their modified values for the duration of treatment from  $T_{start}$  to  $T_{end}$ . Simulations were stopped when the colony of type-6 crypts grew to its detection size (or when  $T = 80$  was reached). Many simulation runs were aggregated to derive the age-incidence curve for advanced adenoma. **Figure 4(c)** presents typical simulation results for the models where aspirin affects both intra-crypt (conversion) and inter-crypt (crypt fission and death) dynamics for all the mutated cell types. The thin black lines represent the incidence curve in the absence of treatment, as was obtained by fitting the advanced adenoma data from **Brenner et al., 2014** (black dots). The yellow curves represent the predicted age-incidence for individuals who were undergoing aspirin treatment during the time-window (of 10 years) in different decades of their lives (see the green shading in each panel representing treatment). Panel (d) of **Figure 4** shows the relative risk of advanced adenoma for patients that received aspirin treatment during different decades of their lives. This is calculated at the time-point that is referred as 'Relative risk assessment time' in panel (b), and corresponds to a zero follow up time in this case. The yellow bars correspond to the prediction for individuals treated with a relatively strong dose of aspirin (roughly 6–14 pills a week) for 10 years. We have further performed simulations to obtain model predictions pertaining to lower aspirin doses (**Table 1**), which we referred to as medium (about 6 pills a week, brown bars in **Figure 4(d)**), and light (less than two pills a week, blue bars in **Figure 4(d)**).

We can compare the predicted relative adenoma risk with the data reported in the literature, see e.g. (**Chan et al., 2004; Cole et al., 2009; Cao et al., 2016; Drew et al., 2016**). In particular, the dose-dependence of colorectal adenoma was studied (**Chan et al., 2004**), and it was shown that the relative risk for adenoma was 0.80 for women who used 0.5–1.5 standard (325 mg) tablets per week, 0.74 for those who used 2–5 tablets per week, 0.72 for those who used 6–14 tablets per week, and 0.49 for those who used more than 14 tablets per week. Comparing this with the relative advanced

adenoma risk plot in **Figure 4(d)**, we can see that the model predictions are consistent with the observed bounds: for the aspirin dose that corresponds to 6–14 tablets per week (referred to as high dose, **Figure 4(d)**), the predicted relative risk ranges between about 0.35 and 0.80, depending on the age when aspirin was administered. Consistently with (**Chan et al., 2004**), we found that the effect of aspirin decreases with smaller doses, resulting in the relative risk for intermediate doses varying roughly in the 0.6–0.9 range (brown bars in **Figure 4(d)**), while for the light dose it varied in the 0.8–0.9 range (blue bars in **Figure 4(d)**).

Note that in our simulations for different scenarios, we observed remarkable quantitative consistency of results over a very wide range of parameters where uncertainties exist. For example, in **Appendix 1—figure 14**, we compared two markedly different assumptions on the numbers of type-6 crypts that constitute advanced adenoma at detection (**Lang et al., 2020; Sun et al., 2014; Kim et al., 2007; Tsai and Strum, 2011; Jones et al., 2008; Kang and Shibata, 2013; Dewanjji et al., 2011**), with the results remaining very similar.

While in **Figure 4(c and d)** we assume that aspirin treatment lasts 10 years, in **Appendix 1—figure 18** we investigated the effect of a shorter duration of aspirin treatment for the high dose regime and found the relative risk closer to 0.80. One conclusion that follows from this and other simulations (see Appendix 1, section 6) is that the extent of the aspirin-induced reduction in the adenoma risk, resulting from the mechanisms studied here, is consistent with the reported risk reduction at least for a subset of the parameter combinations. Therefore, this mechanism cannot be rejected based on the predicted advanced adenoma incidence reduction.

Next, we examined how changing model assumptions about the effect of aspirin alter these results. While the simulations of **Figure 4** assume that both inter- and intra-crypt dynamics are affected by aspirin, **Appendix 1—figure 15** only includes aspirin's effect on intra-crypt dynamics. Under this assumption, even the strongest aspirin dose did not result in the magnitude of the effect reported in **Chan et al., 2004**: the reduction in advanced adenoma risk was within a few percent only. This suggests that including aspirin's effect on inter-crypt (fission/death) dynamics is essential to explain the data, which makes intuitive sense because crypt dynamics are thought to be drivers of disease development.

We also investigated the consequence of the assumption that aspirin only affects the most transformed (type 6) cells (**Appendix 1—figure 16**), and found that while the effect is reduced compared to the full model of **Figure 4**, one still observes a significant decrease in advanced adenoma risk. Interestingly, if assessment time follows treatment immediately (zero follow-up time) then there is almost no difference between the prediction of the model where only type 6 is affected compared to that where crypts 2–6 are affected (**Appendix 1—figure 17**). Increasing the follow-up time, however, reveals an increase of the difference between the two model predictions. For example, with a 15 years follow-up time, the model where all types 2–6 are affected shows the relative risk of about 0.6, while the model with only type 6 affected yields a relative risk of about 0.8. The reason for this is the lag-phase that exists between the generation of the first type-6 crypt and its growth to detection (which in our simulations takes between about 5 and 11 years).

There are several further patterns that emerge. We observe that risk reduction clearly depends on the age of the patients when aspirin was administered. As we see in **Figure 4(d)**, the relative risk can be as low as about 0.4 for patients that started treatment at age 20 and assessed at age 30, compared to a more modest reduction to relative risk of 0.8 for patients receiving treatment later in life (see additional discussion below). Further, we will mention that aspirin-induced risk reduction, as predicted by this model, does not disappear even decades after aspirin treatment stopped (**Figure 4(b)**).

Finally, we comment on another aspect of our model that is different from several other models used in the field (including **Luebeck and Moolgavkar, 2002; Meza et al., 2008; Paterson et al., 2020**). When predicting the age-incidence curve that results from the microscopic dynamics of selection and mutations, we explicitly included the growth of the most modified crypt type to its detectable size. While stopping the simulations once the first type-6 crypt is produced leads to qualitatively similar results, the inclusion of a relatively slow growth of the adenoma significantly changed numerical values of the fitted parameters. It appears that including this stage in the simulations helps improve the quantitative contribution (rather than a proof-of-principle) of this style of mathematical modeling.

## Discussion

We used mathematical modeling approaches to test the hypothesis that the changes in tumor cell kinetics observed during aspirin treatment *in vitro* and *in vivo* can translate into a protective effect on a population level that is consistent with epidemiological observations for late adenoma. This was done by first constructing a mathematical model of *in vivo* carcinogenesis describing evolutionary events leading up to the late adenoma stage. This model was then used to calculate expected population incidence as a function of age. Many of the model parameters have recently been estimated experimentally, which provides a solid basis for this modeling effort. Remaining parameters were estimated by fitting the incidence prediction to epidemiological data on late adenoma detection. A linear model that did not include inter-crypt competition was rejected because its best fits corresponded to zero crypt fission rates, and the more (statistically) powerful model was adopted instead, where individual mutated crypts experienced both fission and nonlinear competition dynamics. This parameterized model was used as a basis to explore how changes in the kinetics / fitness of cells, brought about by aspirin, can modify the predicted incidence of late adenomas.

Our previous *in vitro* and *in vivo* work (*Shimura et al., 2020; Zumwalt et al., 2017*) indicated that aspirin reduces the rate of colorectal tumor cell division and increases the rate of tumor cell death in a dose-dependent way, by up to twofold. In the current modeling study, three different experimental aspirin doses (converted to human aspirin intake) were explored, for which we previously measured their effect on the kinetic parameters. The mathematical analysis demonstrated that parameter changes of a magnitude that lies within our experimentally observed range can lead to significant reductions in predicted late adenoma incidence, which are consistent with the epidemiologically observed reductions (ranging between 10–50%, *Chan et al., 2004*). The model identified dose, treatment duration, and the age at which treatment was started as important determinants of protection in this context. We can conclude that the aspirin-induced changes in cellular fitness that we observed experimentally can in principle explain a significant portion of the protective effect observed on the population level.

This does of course not preclude alternative mechanisms that can further contribute to the protective effect. It is very likely that a reduction in the level of inflammation within the microenvironment of the cells can reduce the incidence of colorectal cancer, because inflammation has been identified as a driver of this disease (*Itzkowitz and Yio, 2004*). Moreover, other microenvironmental factors, such as the composition of the colorectal microbiome, have been shown to influence the ability of aspirin to reduce tumor growth (*Prizment et al., 2020; Brennan et al., 2021; Zhao et al., 2020*). This is therefore also likely to play a role in explaining the epidemiological data. Quantification of these further complexities in future work will allow us to introduce these additional aspects into the modeling framework, which would result in a refinement of predictions.

As with most mathematical modeling studies, there are uncertainties in assumptions that need to be kept in mind. Our experiments (*Shimura et al., 2020; Zumwalt et al., 2017*) were performed with tumor cell lines, both *in vitro* and in mouse xenografts. While the xenografts capture a higher degree of biological complexity than *in vitro* experiments, cellular processes in the human colon are even more complex. Colorectal tissues and tumors are characterized by stronger cell hierarchies than our experimental system, including stem, transit amplifying, and terminally differentiated cells. Our analysis was presented under the assumption that colorectal stem cells initiate and maintain tumor growth. While our experimental system did not specifically focus on stem cells, other studies indicate that the effect of aspirin on the kinetics of stem cells in particular is similar (*Chen et al., 2018*), thus justifying model assumptions. Another uncertainty concerns the cell type in which the tumor originates, and the exact identity of the cell compartment that maintains tumor growth. While we concentrated our model description around stem cells as the cell of origin that drives disease, the model defines this population as having the ability to self-renew thus maintaining the expansion of the tumor. Hence, this cell population in the model could also correspond to compartments downstream in the differentiation pathway, such as transit amplifying cells, given the marked plasticity within the intestinal epithelium. The model is thus in principle consistent with hypotheses that colorectal cancer might have a different cell of origin (*Huels and Sansom, 2015*). Interestingly, it has been shown that aspirin had a negative impact on colon organoids derived from non-neoplastic issues, and that aspirin particularly reduced the rapidly cycling transit amplifying cell population (*Devall et al., 2021*).

Another point of uncertainty concerns the identity of the cell populations that are affected by aspirin. To address this, we made several assumptions, and results remained robust. We first assumed that aspirin influences all mutated cell populations (type 2–6). Results, however, remained fairly similar in an alternative model, where only type 6 cells (characterized by APC-/ and KRAS + mutations) were affected by aspirin, although in this case the effect of aspirin is weaker, which is not surprising given that a smaller cell population loses fitness. More crucial was the assumption that aspirin influences not only the cell dynamics themselves, but also the crypt fission dynamics, reducing the rate at which crypts divide and/or increasing crypt death rate. While this is supported by data (**Fischer et al., 2014**), further experimental investigation into the exact mechanism by which aspirin affects inter-crypt dynamics is needed to back up our modeling assumptions.

An important component of all of this work is the underlying mathematical model of in vivo adenoma formation. The assumptions about the genetic events that occur during adenoma formation are consistent with our current understanding of adenoma evolution (**Fearon, 2011**), and a similar model that also includes evolutionary events beyond adenomas has recently been published (**Paterson et al., 2020**). While the model description in our and the previously published study (**Paterson et al., 2020**) was focused on APC-/ and KRAS + mutations as initial events, the same evolutionary dynamics would occur if the identity of mutations were different, as long as the evolutionary pathway involves the inactivation of a tumor suppressor gene and the activation of an oncogene, as specified in the model description.

An important difference between our and the previous model concerns assumptions about crypt fission dynamics. The previous study (**Paterson et al., 2020**) assumed that crypt fission can occur without density-dependent effects. Using experimentally available parameter estimates, this model could account for the life-time risk of colorectal cancer. When applying a similar model to late adenoma age incidence data, however, we could not obtain a good fit for the age-incidence curve, and the best fit was in fact obtained in the absence of any crypt fission. In the absence of inter-crypt competition (that is, with unlimited crypt fission), the predicted adenoma incidence rose too sharply with age compared to epidemiological data. When introducing density-dependence into the crypt fission process, however, late adenoma age incidence data could be readily fit, and so we used this model assumption to go forward. Indeed, it is likely that density-dependent effects play a role in crypt fission, because this process is probably influenced by signaling factors that become limiting as the number of crypts increases. It would be important to verify this assumption experimentally in future work.

Finally, it is interesting to discuss the results of the ASPREE trial (**McNeil et al., 2018; McNeil et al., 2021**) in the context of the work presented here. This trial investigated the effect of aspirin treatment in a cohort of older individuals, 70 years or older without cardiovascular disease, dementia, or disability. It was found that cancer incidence was not significantly changed by aspirin, but that the aspirin-treated group experienced a higher rate of cancer-induced mortality. The absence of a significant effect of aspirin on cancer incidence in this study is consistent with our model predictions. Our mathematical analysis demonstrated that the effect of aspirin treatment on cancer incidence diminished when treatment was initiated in older ages. Our modeling approach, however, cannot make predictions about cancer-induced mortality, because it describes the evolutionary process up to the stage of advanced adenoma only. Our previous work (**Wodarz et al., 2017**), however, offers an interpretation of these data. Because of their advanced age, it is likely that a certain fraction of the ASPREE participants already harbored tumors that had not been detected yet due to the absence of overt clinical symptoms. In fact, a previous history of cancer was not an exclusion criterion in the trial. As the established tumors continue to grow during aspirin treatment, they likely do so with altered kinetics (reduced division rates and increased death rates, leading to a higher turnover). This means that by the time the tumor has reached a given size (e.g. at which it becomes clinically detectable), it will have undergone more cell divisions under aspirin treatment compared to the placebo group. Hence, the tumor will on average have accumulated more mutations once this detectable tumor size is reached. This in turn means that the aspirin-treated tumor might be more virulent and less responsive to therapies, resulting in more deaths. The theoretically derived notion that upon detection, an aspirin-treated tumor is more evolved than a tumor that grows without aspirin (**Wodarz et al., 2017**) is supported by the ASPREE analysis, which found that aspirin-treated patients were more likely to have metastasized cancers and stage 4 cancers compared to the placebo group (**McNeil et al., 2018; McNeil et al., 2021**).

In conclusion, this modeling analysis suggests that a direct impact of aspirin on the kinetics and fitness of mutated cells can significantly reduce the incidence of colorectal adenomas, with a magnitude that is consistent with epidemiological data. This highlights the importance of investigating this effect of aspirin experimentally in more detail, especially under experimental conditions that approximate cell dynamics in the human colorectal tissue with greater accuracy.

## Materials and methods

### The adenoma incidence data

In order to study the incidence of advanced adenoma, we used the data reported in *Brenner et al., 2014* for the age-ranges 55–59, 60–64, 65–69, 70–74, and 75–79. While this study provides incidence data for nonadvanced adenoma, advanced adenoma, and colorectal cancer (CRC), we focused only on the combined incidence of advanced adenoma and cancer. This assumes that individuals that have developed CRC have most likely already developed an advanced adenoma by the age of testing, and further that nonadvanced adenoma likely refers to fewer mutational steps compared to our type 6, where both the APC gene is fully inactivated and the KRAS gene is mutated. The paper reports data separately for males and females; for our purposes we combined the two values to study the average, since the model is not sufficiently detailed to distinguish between the sexes.

### Mathematical modeling

The mathematical model describes stochastic dynamics of colonic crypts. There are six types of crypts that are included in the model, which differ by their mutational status. The number of crypts of each type is denoted by  $n_i$ , where  $i = 1$  corresponds to the wild-type crypts,  $i = 2$  to type  $\text{APC}^{+/-}$ ,  $i = 3$  to  $\text{APC}^{-/-}$ ,  $i = 4$  to  $\text{KRAS}^+$ ,  $i = 5$  to  $\text{KRAS}^+\text{APC}^{+/-}$ , and  $i = 6$  to  $\text{KRAS}^+\text{APC}^{-/-}$  (the most modified type associated with advanced adenoma). The model contains the processes of crypt conversion (whereby a mutation in a stem cell can fixate in a given crypt thus changing its mutational status), as well as crypt fission/death processes. Inter-crypt competition is included by way of nonlinear (logistic) terms. Given the initial condition (wild-type crypts only) and model parameters, the model outputs the probability to observe, by time  $t$ , a specified population of type-6 crypts ( $n_6 = N$ ), which is assumed to be associated with advanced adenoma detection. This represents a numerically generated age-incidence curve. The expected behavior was described by a system of ordinary differential equations (ODEs), and the prediction was fitted to the advanced adenoma incidence reported in *Brenner et al., 2014*. While the model was parameterized by using the rates found in literature and describing the kinetics in humans, a subset of parameters are unknown (or only their ranges are known); these parameters were estimated by the fitting procedure.

Using the parameterized model that is consistent with the advanced adenoma incidence, we incorporated the effect of aspirin by adjusting the kinetic parameters of cells (division and death rates of cells, which describes the effect of aspirin on intra-crypt dynamics), as well as kinetic rates of crypts (crypt fission and death rates, which describes the effect of aspirin on inter-crypt dynamics). This was done by using experimentally measured factors. Fully stochastic (Gillespie) simulations were used to quantify the predicted advanced adenoma incidence curves for patients that used different doses and durations of aspirin treatment. For further details of the modeling, see Appendix 1.

### Acknowledgements

We thank the reviewers of this paper for very valuable comments that have influenced to the current form of the manuscript.

Support of the following grants is gratefully acknowledged: NIH 1 U01 CA187956-01 (AG, RB, NK, DW); NSF-Simons Center for Multiscale Cell Fate Research (NK, YW); NIH/NCI U54-CA217378 (NK, DW, YW).

## Additional information

### Funding

Funder	Grant reference number	Author
National Cancer Institute	NIH 1 U01 CA187956-01	C Richard Boland Ajay Goel Dominik Wodarz Natalia L Komarova
National Science Foundation	NSF-Simons Center for Multiscale Cell Fate Research	Yifan Wang Natalia L Komarova
MIDAS	AWD00000238	Natalia L Komarova Yifan Wang

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Yifan Wang, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing; C Richard Boland, Ajay Goel, Conceptualization, Funding acquisition, Writing – review and editing; Dominik Wodarz, Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review and editing; Natalia L Komarova, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing

### Author ORCIDs

Yifan Wang  <http://orcid.org/0000-0003-3292-4972>  
 C Richard Boland  <http://orcid.org/0000-0002-8120-5088>  
 Ajay Goel  <http://orcid.org/0000-0003-1396-6341>  
 Dominik Wodarz  <http://orcid.org/0000-0002-8017-3707>  
 Natalia L Komarova  <http://orcid.org/0000-0003-4876-0343>

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.71953.sa1>  
 Author response <https://doi.org/10.7554/eLife.71953.sa2>

## Additional files

### Supplementary files

- Transparent reporting form
- Source code 1. Mathematica code for **Figures 2 and 3**.
- Source code 2. Mathematica code for **Figure 4**.
- Source code 3. A Fortran code for stochastic simulations (e.g. **Figure 4**).

### Data availability

Data and some relevant code are available on Dryad under DOI 10.7280/D1M11M. Code for Figures 2, 3 and 4 is uploaded in Source Code files 1-3.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Komarova NL	2022	Data files for simulated advanced adenoma age incidence, under aspirin treatment	<a href="https://doi.org/10.7280/D1M11M">https://doi.org/10.7280/D1M11M</a>	Dryad Digital Repository, 10.7280/D1M11M

## References

- Armaghani T, Wilson JD, Chu Q, Mills G. 2012. Genetic alterations in colorectal cancer. *Gastrointestinal Cancer Research* 5:19–27. DOI: [22574233](https://doi.org/10.22574/22574233), PMID: 22574233.
- Armitage P, Doll R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* 8:1–12. DOI: <https://doi.org/10.1038/bjc.1954.1>, PMID: 13172380.
- Ashley DJ. 1969. The two "hit" and multiple "hit" theories of carcinogenesis. *British Journal of Cancer* 23:313–328. DOI: <https://doi.org/10.1038/bjc.1969.41>, PMID: 5788039.
- Baker A-M, Cereser B, Melton S, Fletcher AG, Rodriguez-Justo M, Tadrous PJ, Humphries A, Elia G, McDonald SAC, Wright NA, Simons BD, Jansen M, Graham TA. 2014. Quantification of crypt and stem cell evolution in the normal and neoplastic human colon. *Cell Reports* 8:940–947. DOI: <https://doi.org/10.1016/j.celrep.2014.07.019>, PMID: 25127143.
- Birtwell D, Luebeck G, Maley CC. 2020. The evolution of metapopulation dynamics and the number of stem cells in intestinal crypts and other tissue structures in multicellular bodies. *Evolutionary Applications* 13:1771–1783. DOI: <https://doi.org/10.1111/eva.13069>, PMID: 32821281.
- Brennan CA, Nakatsu G, Gallini Comeau CA, Drew DA, Glickman JN, Schoen RE, Chan AT, Garrett WS. 2021. Aspirin modulation of the colorectal cancer-associated microbe *fusobacterium nucleatum*. *MBio* 12:e00547-21. DOI: <https://doi.org/10.1128/mBio.00547-21>, PMID: 33824205.
- Brenner H, Altenhofen L, Stock C, Hoffmeister M. 2014. Incidence of colorectal adenomas: birth cohort analysis among 4.3 million participants of screening colonoscopy. *Cancer Epidemiology, Biomarkers & Prevention* 23:1920–1927. DOI: <https://doi.org/10.1158/1055-9965.EPI-14-0367>, PMID: 25012996.
- Burn J, Gerdes A-M, Macrae F, Mecklin J-P, Moeslein G, Olschwang S, Eccles D, Evans DG, Maher ER, Bertario L, Bisgaard M-L, Dunlop MG, Ho JWC, Hodgson SV, Lindblom A, Lubinski J, Morrison PJ, Murday V, Ramesar R, Side L, et al. 2011. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. *Lancet (London, England)* 378:2081–2087. DOI: [https://doi.org/10.1016/S0140-6736\(11\)61049-0](https://doi.org/10.1016/S0140-6736(11)61049-0), PMID: 22036019.
- Cao Y, Nishihara R, Wu K, Wang M, Ogino S, Willett WC, Spiegelman D, Fuchs CS, Giovannucci EL, Chan AT. 2016. Population-wide impact of long-term use of aspirin and the risk for cancer. *JAMA Oncology* 2:762–769. DOI: <https://doi.org/10.1001/jamaoncol.2015.6396>, PMID: 26940135.
- Chan TL, Zhao W, Leung SY, Yuen ST, Cancer Genome Project. 2003. BRAF and KRAS mutations in colorectal hyperplastic polyps and serrated adenomas. *Cancer Research* 63:4878–4881. DOI: <https://doi.org/10.1158/0731-6558.CAN-02-3811>, PMID: 12941809.
- Chan AT, Giovannucci EL, Schernhammer ES, Colditz GA, Hunter DJ, Willett WC, Fuchs CS. 2004. A prospective study of aspirin use and the risk for colorectal adenoma. *Annals of Internal Medicine* 140:157–166. DOI: <https://doi.org/10.7326/0003-4819-140-3-200402030-00006>, PMID: 14757613.
- Chan AT, Giovannucci EL, Meyerhardt JA, Schernhammer ES, Curhan GC, Fuchs CS. 2005. Long-term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer. *JAMA* 294:914–923. DOI: <https://doi.org/10.1001/jama.294.8.914>, PMID: 16118381.
- Chan AT, Ogino S, Fuchs CS. 2007. Aspirin and the risk of colorectal cancer in relation to the expression of COX-2. *The New England Journal of Medicine* 356:2131–2142. DOI: <https://doi.org/10.1056/NEJMoa067208>, PMID: 17522398.
- Chan AT, Arber N, Burn J, Chia WK, Elwood P, Hull MA, Logan RF, Rothwell PM, Schröder K, Baron JA. 2012. Aspirin in the chemoprevention of colorectal neoplasia: an overview. *Cancer Prevention Research (Philadelphia, Pa.)* 5:164–178. DOI: <https://doi.org/10.1158/1940-6207.CAPR-11-0391>, PMID: 22084361.
- Chen Z, Li W, Qiu F, Huang Q, Jiang Z, Ye J, Cheng P, Low C, Guo Y, Yi X, Chen W, Yu Y, Han Y, Wu J, Jin S, Kong D, Huang J. 2018. Aspirin cooperates with p300 to activate the acetylation of H3K9 and promote FasL-mediated apoptosis of cancer stem-like cells in colorectal cancer. *Theranostics* 8:4447–4461. DOI: <https://doi.org/10.7150/thno.24284>, PMID: 30214631.
- Chou T, Wang Y. 2015. Fixation times in differentiation and evolution in the presence of bottlenecks, deserts, and oases. *Journal of Theoretical Biology* 372:65–73. DOI: <https://doi.org/10.1016/j.jtbi.2015.02.021>, PMID: 25744205.
- Cole BF, Logan RF, Halabi S, Benamouzig R, Sandler RS, Grainge MJ, Chaussade S, Baron JA. 2009. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. *Journal of the National Cancer Institute* 101:256–266. DOI: <https://doi.org/10.1093/jnci/djn485>, PMID: 19211452.
- Devall MAM, Drew DA, Dampier CH, Plummer SJ, Eaton S, Bryant J, Díez-Obrero V, Mo J, Kedrin D, Zerjav DC, Takacs-Nagy O, Jennelle LT, Ali MW, Yilmaz ÖH, Moreno V, Powell SM, Chan AT, Peters U, Casey G. 2021. Transcriptome-wide in vitro effects of aspirin on patient-derived normal colon organoids. *Cancer Prevention Research (Philadelphia, Pa.)* 14:1089–1100. DOI: <https://doi.org/10.1158/1940-6207.CAPR-21-0041>, PMID: 34389629.
- Dewanji A, Jeon J, Meza R, Luebeck EG. 2011. Number and size distribution of colorectal adenomas under the multistage clonal expansion model of cancer. *PLOS Computational Biology* 7:e1002213. DOI: <https://doi.org/10.1371/journal.pcbi.1002213>, PMID: 22022253.
- Drew DA, Cao Y, Chan AT. 2016. Aspirin and colorectal cancer: the promise of precision chemoprevention. *Nature Reviews. Cancer* 16:173–186. DOI: <https://doi.org/10.1038/nrc.2016.4>, PMID: 26868177.
- Fearon ER. 2011. Molecular genetics of colorectal cancer. *Annual Review of Pathology* 6:479–507. DOI: <https://doi.org/10.1146/annurev-pathol-011110-130235>, PMID: 21090969.

- Fischer JM**, Schepers AG, Clevers H, Shibata D, Liskay RM. 2014. Occult progression by Apc-deficient intestinal crypts as a target for chemoprevention. *Carcinogenesis* **35**:237–246. DOI: <https://doi.org/10.1093/carcin/bgt296>, PMID: 23996931
- Fisher JC**, Hollomon JH. 1951. A hypothesis for the origin of cancer foci. *Cancer* **4**:916–918. DOI: [https://doi.org/10.1002/1097-0142\(195109\)4:5<916::aid-cncr2820040504>3.0.co;2-7](https://doi.org/10.1002/1097-0142(195109)4:5<916::aid-cncr2820040504>3.0.co;2-7), PMID: 14879355
- Friis S**, Riis AH, Erichsen R, Baron JA, Sørensen HT. 2015. Low-Dose Aspirin or Nonsteroidal Anti-inflammatory Drug Use and Colorectal Cancer Risk: A Population-Based, Case-Control Study. *Annals of Internal Medicine* **163**:347–355. DOI: <https://doi.org/10.7326/M15-0039>, PMID: 26302241
- Goel A**, Chang DK, Ricciardiello L, Gasche C, Boland CR. 2003. A novel mechanism for aspirin-mediated growth inhibition of human colon cancer cells. *Clinical Cancer Research* **9**:383–390 PMID: 12538492.,  
**Hornsby C**, Page KM, Tomlinson IPM. 2007. What can we learn from the population incidence of cancer? Armitage and Doll revisited. *The Lancet. Oncology* **8**:1030–1038. DOI: [https://doi.org/10.1016/S1470-2045\(07\)70343-1](https://doi.org/10.1016/S1470-2045(07)70343-1), PMID: 17976613
- Huels DJ**, Sansom OJ. 2015. Stem vs non-stem cell origin of colorectal cancer. *British Journal of Cancer* **113**:1–5. DOI: <https://doi.org/10.1038/bjc.2015.214>, PMID: 26110974
- Humphries A**, Wright NA. 2008. Colonic crypt organization and tumorigenesis. *Nature Reviews. Cancer* **8**:415–424. DOI: <https://doi.org/10.1038/nrc2392>, PMID: 18480839
- Humphries A**, Cereser B, Gay LJ, Miller DSJ, Das B, Gutteridge A, Elia G, Nye E, Jeffery R, Poulson R, Novelli MR, Rodriguez-Justo M, McDonald SAC, Wright NA, Graham TA. 2013. Lineage tracing reveals multipotent stem cells maintain human adenomas and the pattern of clonal expansion in tumor evolution. *PNAS* **110**:E2490–E2499. DOI: <https://doi.org/10.1073/pnas.1220353110>, PMID: 23766371
- Itzkowitz SH**, Yio X. 2004. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *American Journal of Physiology. Gastrointestinal and Liver Physiology* **287**:G7–G17. DOI: <https://doi.org/10.1152/ajpgi.00079.2004>, PMID: 15194558
- Jass JR**. 2006. Colorectal cancer: a multipathway disease. *Critical Reviews in Oncogenesis* **12**:273–287. DOI: [https://doi.org/10.1615/critrevoncog.v12.i3-4\\_50](https://doi.org/10.1615/critrevoncog.v12.i3-4_50), PMID: 17425506
- Jen J**, Powell SM, Papadopoulos N, Smith KJ, Hamilton SR, Vogelstein B, Kinzler KW. 1994. Molecular determinants of dysplasia in colorectal lesions. *Cancer Research* **54**:5523–5526 PMID: 7923189.,  
**Jones S**, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, Traulsen A, Nowak MA, Siegel C, Velculescu VE, Kinzler KW, Vogelstein B, Willis J, Markowitz SD. 2008. Comparative lesion sequencing provides insights into tumor evolution. *PNAS* **105**:4283–4288. DOI: <https://doi.org/10.1073/pnas.0712345105>, PMID: 18337506
- Kang H**, Shibata D. 2013. Direct measurements of human colon crypt stem cell niche genetic fidelity: the role of chance in non-Darwinian mutation selection. *Frontiers in Oncology* **3**:264. DOI: <https://doi.org/10.3389/fonc.2013.00264>, PMID: 24133655
- Kim DH**, Pickhardt PJ, Taylor AJ. 2007. Characteristics of advanced adenomas detected at CT colonographic screening: implications for appropriate polyp size thresholds for polypectomy versus surveillance. *AJR. American Journal of Roentgenology* **188**:940–944. DOI: <https://doi.org/10.2214/AJR.06.0764>, PMID: 17377027
- Komarova NL**, Lengauer C, Vogelstein B, Nowak MA. 2002. Dynamics of genetic instability in sporadic and familial colorectal cancer. *Cancer Biology & Therapy* **1**:685–692. DOI: <https://doi.org/10.4161/cbt.321>, PMID: 12642695
- Komarova NL**, Sengupta A, Nowak MA. 2003. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Journal of Theoretical Biology* **223**:433–450. DOI: [https://doi.org/10.1016/s0022-5193\(03\)00120-6](https://doi.org/10.1016/s0022-5193(03)00120-6), PMID: 12875822
- Lang BM**, Kuipers J, Misselwitz B, Beerenwinkel N. 2020. Predicting colorectal cancer risk from adenoma detection via a two-type branching process model. *PLOS Computational Biology* **16**:e1007552. DOI: <https://doi.org/10.1371/journal.pcbi.1007552>, PMID: 32023238
- Lochhead PJ**, Chan AT. 2016. Aspirin and the Prevention of Colorectal Cancer. *Best Practice & Research. Clinical Gastroenterology* **25**:461–472.
- Luebeck EG**, Moolgavkar SH. 2002. Multistage carcinogenesis and the incidence of colorectal cancer. *PNAS* **99**:15095–15100. DOI: <https://doi.org/10.1073/pnas.222118199>, PMID: 12415112
- McNeil JJ**, Nelson MR, Woods RL, Lockery JE, Wolfe R, Reid CM, Kirpach B, Shah RC, Ives DG, Storey E, Ryan J, Tonkin AM, Newman AB, Williamson JD, Margolis KL, Ernst ME, Abhayaratna WP, Stocks N, Fitzgerald SM, Orchard SG, et al. 2018. Effect of aspirin on all-cause mortality in the healthy elderly. *The New England Journal of Medicine* **379**:1519–1528. DOI: <https://doi.org/10.1056/NEJMoa1803955>, PMID: 30221595
- McNeil JJ**, Gibbs P, Orchard SG, Lockery JE, Bernstein WB, Cao Y, Ford L, Haydon A, Kirpach B, Macrae F, McLean C, Millar J, Murray AM, Nelson MR, Polekhina G, Reid CM, Richmond E, Rodríguez LM, Shah RC, Tie J, et al. 2021. Effect of aspirin on cancer incidence and mortality in older adults. *Journal of the National Cancer Institute* **113**:258–265. DOI: <https://doi.org/10.1093/jnci/djaa114>, PMID: 32778876
- Meza R**, Jeon J, Moolgavkar SH, Luebeck EG. 2008. Age-specific incidence of cancer: Phases, transitions, and biological implications. *PNAS* **105**:16284–16289. DOI: <https://doi.org/10.1073/pnas.0801151105>, PMID: 18936480
- Moolgavkar SH**. 1978. The multistage theory of carcinogenesis and the age distribution of cancer in man. *Journal of the National Cancer Institute* **61**:49–52. DOI: <https://doi.org/10.1093/jnci/61.1.49>, PMID: 276637
- Nair AB**, Jacob S. 2016. A simple practice guide for dose conversion between animals and human. *Journal of Basic and Clinical Pharmacy* **7**:27–31. DOI: <https://doi.org/10.4103/0976-0105.177703>, PMID: 27057123

- Network CGA.** 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**:330–337. DOI: <https://doi.org/10.1038/nature11252>, PMID: 22810696
- Nicholson AM**, Olpe C, Hoyle A, Thorsen A-S, Rus T, Colombé M, Brunton-Sim R, Kemp R, Marks K, Quirke P, Malhotra S, Ten Hoopen R, Ibrahim A, Lindskog C, Myers MB, Parsons B, Tavaré S, Wilkinson M, Morrissey E, Winton DJ. 2018. Fixation and spread of somatic mutations in adult human colonic epithelium. *Cell Stem Cell* **22**:909–918.. DOI: <https://doi.org/10.1016/j.stem.2018.04.020>, PMID: 29779891
- Nordling CO.** 1953. A new theory on the cancer-inducing mechanism. *British Journal of Cancer* **7**:68–72. DOI: <https://doi.org/10.1038/bjc.1953.8>, PMID: 13051507
- Nowak MA**, Komarova NL, Sengupta A, Jallepalli PV, Shih IM, Vogelstein B, Lengauer C. 2002. The role of chromosomal instability in tumor initiation. *PNAS* **99**:16226–16231. DOI: <https://doi.org/10.1073/pnas.202617399>, PMID: 12446840
- Paterson C**, Clevers H, Bozic I. 2020. Mathematical model of colorectal cancer initiation. *PNAS* **117**:20681–20688. DOI: <https://doi.org/10.1073/pnas.2003771117>, PMID: 32788368
- Pretlow TP**, Pretlow TG. 2005. Mutant KRAS in aberrant crypt foci (ACF): initiation of colorectal cancer? *Biochimica et Biophysica Acta* **1756**:83–96. DOI: <https://doi.org/10.1016/j.bbcan.2005.06.002>, PMID: 16219426
- Prizment AE**, Staley C, Onyeagha GC, Vivek S, Thyagarajan B, Straka RJ, Demmer RT, Knights D, Meyer KA, Shaukat A, Sadowsky MJ, Church TR. 2020. Randomised clinical study: oral aspirin 325 mg daily vs placebo alters gut microbial composition and bacterial taxa associated with colorectal cancer risk. *Alimentary Pharmacology & Therapeutics* **52**:976–987. DOI: <https://doi.org/10.1111/apt.16013>, PMID: 32770859
- Rothwell PM.** 2013. Aspirin in prevention of sporadic colorectal cancer: current clinical evidence and overall balance of risks and benefits. *Recent Results in Cancer Research. Fortschritte Der Krebsforschung. Progres Dans Les Recherches Sur Le Cancer* **191**:121–142. DOI: [https://doi.org/10.1007/978-3-642-30331-9\\_7](https://doi.org/10.1007/978-3-642-30331-9_7), PMID: 22893203
- Sandler RS**, Halabi S, Baron JA, Budinger S, Paskett E, Keresztes R, Petrelli N, Pipas JM, Karp DD, Loprinzi CL, Steinbach G, Schilsky R. 2003. A randomized trial of aspirin to prevent colorectal adenomas in patients with previous colorectal cancer. *The New England Journal of Medicine* **348**:883–890. DOI: <https://doi.org/10.1056/NEJMoa021633>, PMID: 12621132
- Shahriyari L**, Komarova NL. 2013. Symmetric vs. asymmetric stem cell divisions: an adaptation against cancer. *PLOS ONE* **8**:e76195. DOI: <https://doi.org/10.1371/journal.pone.0076195>, PMID: 24204602
- Shahriyari L**, Komarova NL, Jilkine A. 2016. The role of cell location and spatial gradients in the evolutionary dynamics of colon and intestinal crypts. *Biology Direct* **11**:1–17. DOI: <https://doi.org/10.1186/s13062-016-0141-6>, PMID: 27549762
- Shimura T**, Toden S, Komarova NL, Boland C, Wodarz D, Goel A. 2020. A comprehensive in vivo and mathematical modeling-based kinetic characterization for aspirin-induced chemoprevention in colorectal cancer. *Carcinogenesis* **41**:751–760. DOI: <https://doi.org/10.1093/carcin/bgz195>, PMID: 31904094
- Siegel RL**, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, Cersek A, Smith RA, Jemal A. 2020. Colorectal cancer statistics. *CA* **70**:145–164. DOI: <https://doi.org/10.3322/caac.21601>, PMID: 32133645
- Sostres C**, Gargallo CJ, Lanas A. 2014. Aspirin, cyclooxygenase inhibition and colorectal cancer. *World Journal of Gastrointestinal Pharmacology and Therapeutics* **5**:40–49. DOI: <https://doi.org/10.4292/wjgpt.v5.i1.40>, PMID: 24605250
- Sun S**, Klebaner F, Tian T. 2014. A new model of time scheme for progression of colorectal cancer. *BMC Systems Biology* **8 Suppl** 3:1–8. DOI: <https://doi.org/10.1186/1752-0509-8-S3-S2>, PMID: 25350788
- Thun MJ**, Namboodiri MM, Heath CW Jr. 1991. Aspirin use and reduced risk of fatal colon cancer. *The New England Journal of Medicine* **325**:1593–1596. DOI: <https://doi.org/10.1056/NEJM199112053252301>, PMID: 1669840
- Tougeron D**, Sha D, Manthravadi S, Sinicrope FA. 2014. Aspirin and colorectal cancer: back to the future. *Clinical Cancer Research* **20**:1087–1094. DOI: <https://doi.org/10.1158/1078-0432.CCR-13-2563>, PMID: 24327271
- Tsai FC**, Strum WB. 2011. Prevalence of advanced adenomas in small and diminutive colon polyps using direct measurement of size. *Digestive Diseases and Sciences* **56**:2384–2388. DOI: <https://doi.org/10.1007/s10620-011-1598-x>, PMID: 21318587
- Wodarz D**, Goel A, Boland CR, Komarova NL. 2017. Effect of aspirin on tumour cell colony formation and evolution. *Journal of the Royal Society, Interface* **14**:134. DOI: <https://doi.org/10.1098/rsif.2017.0374>, PMID: 28878032
- Zhao R**, Coker OO, Wu J, Zhou Y, Zhao L, Nakatsu G, Bian X, Wei H, Chan AWH, Sung JJY, Chan FKL, El-Omar E, Yu J. 2020. Aspirin reduces colorectal tumor development in mice and gut microbes reduce its bioavailability and chemopreventive effects. *Gastroenterology* **159**:969–983.. DOI: <https://doi.org/10.1053/j.gastro.2020.05.004>, PMID: 32387495
- Zumwalt TJ**, Wodarz D, Komarova NL, Toden S, Turner J, Cardenas J, Burn J, Chan AT, Boland CR, Goel A. 2017. Aspirin-Induced Chemoprevention and Response Kinetics Are Enhanced by PIK3CA Mutations in Colorectal Cancer Cells. *Cancer Prevention Research (Philadelphia, Pa.)* **10**:208–218. DOI: <https://doi.org/10.1158/1940-6207.CAPR-16-0175>, PMID: 28154202

## Appendix 1

### Mathematical formulation of a crypt-based model of adenoma initiation

Let us enumerate the types in the way presented in Table **Appendix 1—table 1**.

**Appendix 1—table 1.** Enumeration of the different genotypes.

Mutations in APC    Mutations in KRAS    Type number

Mutations in APC	Mutations in KRAS	Type number
0	0	1
1	0	2
2	0	3
0	1	4
1	1	5
2	1	6

Then we can denote by  $n_i$  with  $1 \leq i \leq 6$  the number of crypts of type . Suppose  $R_{ij}$  is the conversion rate from crypt type to type  $j$ , and  $\gamma_i$  the growth rate (by crypt fission) of crypts of type . We have the following equations:

$$\dot{n}_1 = -(R_{12} + R_{14})n_1, \quad (1)$$

$$\dot{n}_2 = R_{12}n_1 - (R_{23} + R_{25})n_2 + \gamma_2 n_2, \quad (2)$$

$$\dot{n}_3 = R_{23}n_2 - R_{36}n_3 + \gamma_3 n_3, \quad (3)$$

$$\dot{n}_4 = R_{14}n_1 - R_{45}n_4 + \gamma_4 n_4, \quad (4)$$

$$\dot{n}_5 = R_{25}n_2 + R_{45}n_4 - R_{56}n_5 + \gamma_5 n_5, \quad (5)$$

with the initial conditions

$$n_1(0) = N_{crypt}, \quad n_i(0) = 0, \quad 1 \leq i \leq 5. \quad (6)$$

In the ODEs above, we have ignored the process of stochastic tunneling such that the crypts can only convert one step at a time. It is further possible to ignore the negative (outgoing) rates, which simplifies this linear system to the following:

$$\dot{n}_1 = 0, \quad (7)$$

$$\dot{n}_2 = R_{12}n_1 + \gamma_2 n_2, \quad (8)$$

$$\dot{n}_3 = R_{23}n_2 + \gamma_3 n_3, \quad (9)$$

$$\dot{n}_4 = R_{14}n_1 + \gamma_4 n_4, \quad (10)$$

$$\dot{n}_5 = R_{25}n_2 + R_{45}n_4 + \gamma_5 n_5. \quad (11)$$

The probability  $P(t)$  that by time  $t$  at least one crypt of type 6 has been created (using the mean-field approximation from **Cole et al., 2009**) is given by the solution of the equation

$$\dot{P} = (R_{56}n_5 + R_{36}n_3)(1 - P), \quad P(0) = 0, \quad (12)$$

see also section 5.1. The solution can be obtained exactly and is given by

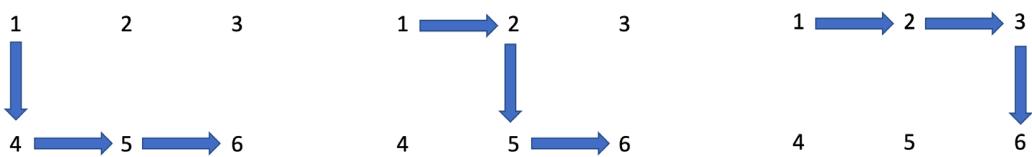
$$P = 1 - \exp \left\{ -N_{\text{crypt}} (S_{45}R_{1 \rightarrow 4 \rightarrow 5 \rightarrow 6} + S_{25}R_{1 \rightarrow 2 \rightarrow 5 \rightarrow 6} + S_{23}R_{1 \rightarrow 2 \rightarrow 3 \rightarrow 6}) \right\}, \quad (13)$$

where the quantities  $S_{ij}$  correspond to the paths  $1 \rightarrow i \rightarrow j \rightarrow 6$ ,

$$R_{1 \rightarrow i \rightarrow j \rightarrow 6} = R_{1i}R_{ij}R_{j6}, \quad (14)$$

see **Appendix 1—figure 1**. They can be written down by using the following function:

$$g(x) = \frac{e^{xt} - 1 - xt}{x^2} > 0. \quad (15)$$



**Appendix 1—figure 1.** Three pathways to adenoma.

We have

$$g(0) = \frac{t^2}{2}, \quad g'(x) \equiv \frac{\partial g}{\partial x} = \frac{2+tx+e^{xt}(tx-2)}{x^3} > 0,$$

that is, this function increases monotonically in  $x$ . We have

$$S_{ij} = \frac{g(\gamma_j) - g(\gamma_i)}{\gamma_i - \gamma_j}.$$

Note that expressions  $S_{ij}$  have a singularity if any of the quantities  $\gamma_i$ ,  $\gamma_j$  is zero and/or if  $\gamma_i = \gamma_j$ . For example, if both growth rates are zero ( $\gamma_i = \gamma_j = 0$ ), we have

$$S_{ij} = \frac{t^3}{6}.$$

In our context it is reasonable to assume that  $\gamma_2 = 0$ , that is, APC+/- crypts do not divide, and that  $\gamma_4 = \gamma_5$ , that is, APC+/+ and APC+/- crypts with an additional KRAS mutation divide at the same rate. In the  $\gamma_i = \gamma_j$  case, taking the limit as  $\gamma_j \rightarrow \gamma_i$ , we obtain

$$S_{45} = g'(\gamma_4), \quad S_{25} = \frac{g(\gamma_4) - t^2/2}{\gamma_4}, \quad S_{23} = \frac{g(\gamma_3) - t^2/2}{\gamma_3}. \quad (16)$$

It is convenient to view  $S_{25}$  as a function of  $\gamma_4$ ,  $S_{25} = F(\gamma_4)$ , where

$$F(x) = \frac{1}{x^3} \left( e^{xt} - \left( 1 + xt + \frac{(xt)^2}{2} \right) \right) = \frac{1}{x^3} \sum_{i=3}^{\infty} \frac{(xt)^i}{i!},$$

then  $S_{23} = F(\gamma_3)$ , and  $F(x)$  is an increasing function of  $x$ . That is, if  $\gamma_4 > \gamma_3$  then  $S_{25} > S_{23}$ , and if  $\gamma_4 < \gamma_3$  then  $S_{25} < S_{23}$ . We further note that  $S_{45} = \tilde{F}(\gamma_4)$  with

$$\tilde{F}(x) = \frac{1}{x^3} (e^{xt}(tx - 2) + tx + 2) = \frac{1}{x^3} \sum_{i=3}^{\infty} \frac{(xt)^i(i-2)}{i!} > F(x).$$

To summarize, we can see that  $S_{45} > S_{25}$ , and  $S_{23}$  is greater (smaller) than  $S_{25}$  if  $\gamma_3$  is greater (smaller) than  $\gamma_4$ .

## Model parameters

Here we define parameters that appear in this model. We will make the following assumptions (see also Table **Appendix 1—table 2**):

- The mutation rate between types is given as follows:  $u_{1 \rightarrow 2} = u_{4 \rightarrow 5} = 2u$  (inactivation of the first copy of the APC gene, that is, any of the two copies);  $u_{2 \rightarrow 3} = u_{5 \rightarrow 6} = u$  (inactivation of the remaining copy of the APC gene);  $u_{1 \rightarrow 4} = u_{2 \rightarrow 5} = u_{3 \rightarrow 6} = \mu$  (activation of the KRAS gene). The fitness of cells with APC+/-, APC-/-, and KRAS+ phenotypes relative to the wild type cells was determined using the cell replacement data from **Komarova et al., 2002**. In particular, we assumed that for any phenotype,

$$F_j \equiv \frac{r_i d_j}{d_j r_i} = \frac{Pr(j)}{1 - Pr(j)}, \quad j \in 2, 3, 4$$

where  $Pr(j)$  is the probability of replacement of the wild type by type  $j$  found in **Komarova et al., 2002**. Using this formula, we obtain the values given in Table **Appendix 1—table 2**. Then the fitness of other types is multiplicative:  $F_5 = F_4 F_2$ ,  $F_6 = F_3 F_4$ .

- The death rates are assumed the same among the types, such that the division rates of cells satisfy  $r_i = F_i r_1$ ,  $1 < i \leq 6$ .
- Wild-type crypts and those with only a single copy of APC gene mutated do not proliferate ( $\gamma_1 = \gamma_2 = 0$ ).
- Crypts with a KRAS mutation and APC+/+ and APC+/- phenotypes proliferate at the same rate,  $\gamma_4 = \gamma_5$ .
- All the parameters are specified in Table **Appendix 1—table 2**.

The conversion rates are given by

$$R_{ij} = r_i K u_{i \rightarrow j} \rho_{ij}, \quad (17)$$

where  $r_i$  is the division of type  $i$ ,  $u_{i \rightarrow j}$  is the mutation rate from type  $i$  to type  $j$ , and  $\rho_{ij}$  is the probability that one cell of type  $j$  becomes fixated in a compartment of size  $K$  with the host type  $i$ . To calculate this probability, let us denote by  $d_i$  the death rate of type  $i$ . Then we have

$$\rho_{ij} = \frac{1 - \frac{r_i d_j}{r_j d_i}}{1 - \left(\frac{r_i d_j}{r_j d_i}\right)^K}, \quad (18)$$

where  $\frac{r_i d_j}{r_j d_i}$  is the inverse of the relative fitness of type  $j$  with respect to type  $i$ .

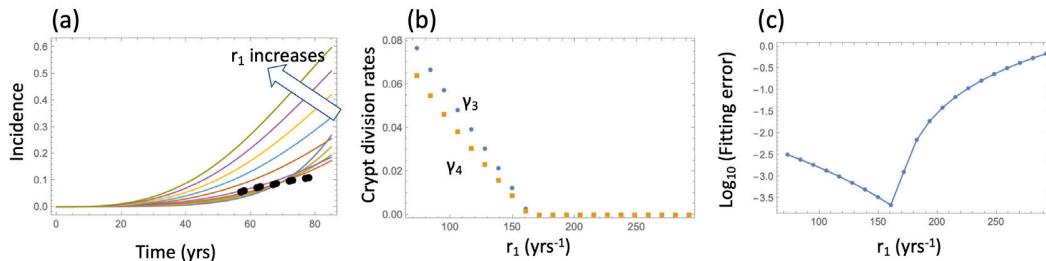
**Appendix 1—table 2.** Parameters, notations, and their values.

Parameter	Notation	Value/ Range
Number of crypts	$N_{crypt}$	$10^7$
Number of SCs per crypt	$K$	7
Rate of inactivation of APC (per cell division)	$u$	$10^{-7}$
Rate of inactivation of APC (per cell division)	$\mu$	$10^{-9}$
Division rate of WT SCs (per year)	$r_1$	(18, 365)
Relative fitness of APC +/- cells	$F_2 = F_{APC^{+/-}}$	1.6
Relative fitness of APC-/- cells	$F_3 = F_{APC^{-/-}}$	3.76
Relative fitness of KRAS+ cells	$F_4 = F_{KRAS}$	3.54
Division rate of APC-/- crypts (per year)	$\gamma_3$	0.2
Division rate of KRAS+ crypts (per year)	$\gamma_4$	0.07

### Fitting the linear model to advanced adenoma incidence data

The linear model of the adenoma incidence is given by equations (13, 14, 15, 16, 17, 18).

In order to fit this model to the data, we fix parameters  $N_{crypt}$ ,  $K$ ,  $u$ ,  $\mu$ , and vary the remaining parameters. This is done in stages. We first fix the fitness parameters  $F_2$ ,  $F_3$ ,  $F_4$  to their values in table (Appendix 1—table 2) and vary the remaining parameters  $r_1$ ,  $\gamma_3$ ,  $\gamma_4$  to find the global minimum of the error between the model and the data, see Appendix 1—figure 2a-c. Then we take other select values for the relative fitness parameters to show that the results remain qualitatively similar (not shown). Next, we describe the results of fitting and the patterns that were observed.



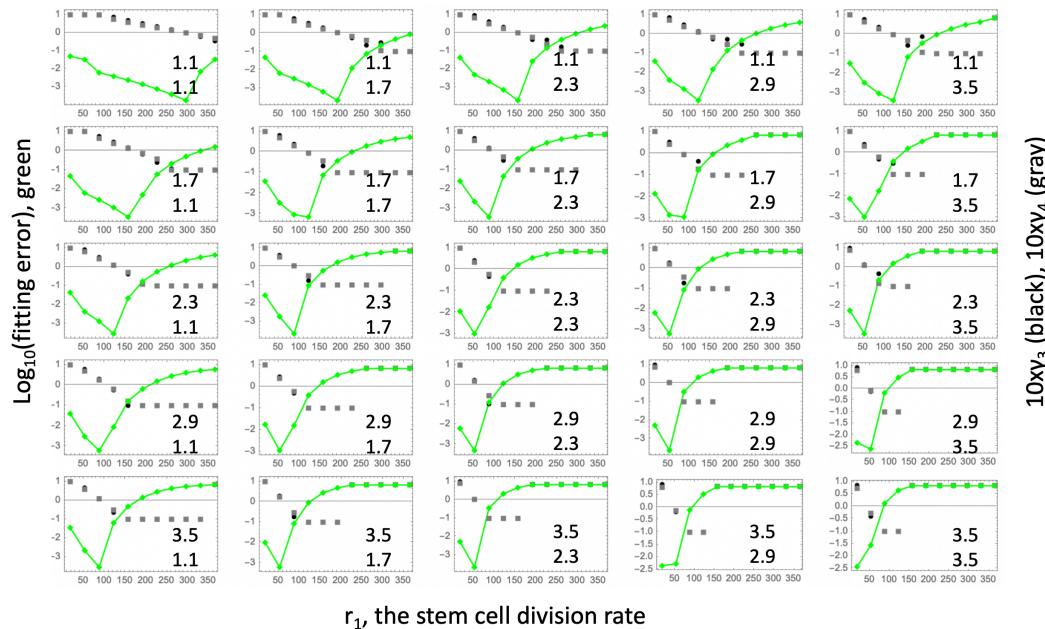
**Appendix 1—figure 2.** Model fitting to the incidence data. (a) The best fitting curves corresponding to increasing SC division rates,  $r_1$ , are plotted together with the epidemiological data. (b) The best fitting parameters  $\gamma_3$  and  $\gamma_4$  are shown for each value of  $r_1$ . (c) The fitting error as a function of  $r_1$ . The relative fitness values are fixed to  $F_{APC^{+/-}} = 1.6$ ,  $F_{APC^{-/-}} = 3.76$ ,  $F_{KRAS} = 3.54$ . The rest of the parameters are as in Table Appendix 1—table 2.

To find the global minimum of the error, we varied parameter  $r_1$  (the division rate of the wild type SCs) between once a day and once every 20 days (which corresponds to the division rate of 365 yrs $^{-1}$  and 18 yrs $^{-1}$ ). For each value of  $r_1$ , the error was minimized in the 2-dimensional parameter space ( $\gamma_3$ ,  $\gamma_4$ ), and a unique minimum was always found. The best fits corresponding to a subset of these division rates are plotted in Appendix 1—figure 2(a), with the best fitting values of  $\gamma_3$  and  $\gamma_4$  shown in panel (b) for each  $r_1$ . We can see that as  $r_1$  increases, that is, SCs are assumed to divide faster, the best fitting crypt division rates decrease (that is, crypt fission proceeds at a slower rate). For  $r_1$  greater than about 250 yrs $^{-1}$  (that is, divisions once every day and a half or faster), the best fitting crypt fission rates are zero. This parameter combination (division rate of about every 1.5 days and zero crypt fission) corresponds to the minimal error of fitting (panel (c)). This is evident from the shape of the best fitting incidence curves,  $P(t)$ , corresponding to different  $r_1$  values (panel (a)). For low rates of SC division, the fission rates are relatively high, resulting in a curve  $P(t)$  that has a steep rise, yielding a large fitting error and a qualitatively unrealistic shape of the incidence curve if compared with the data. This is the consequence of a pronounced exponential increase in the number of crypts, which sharply accelerates the generation of type 6 crypts. As  $r_1$  increases and crypt fission rates decrease, the incidence curves become less steep, until the best fitting crypt fission rate reaches zero. At this stage, the best fitting incidence curve is achieved, because further increase in  $r_1$  results in an increase in the incidence that happens too early.

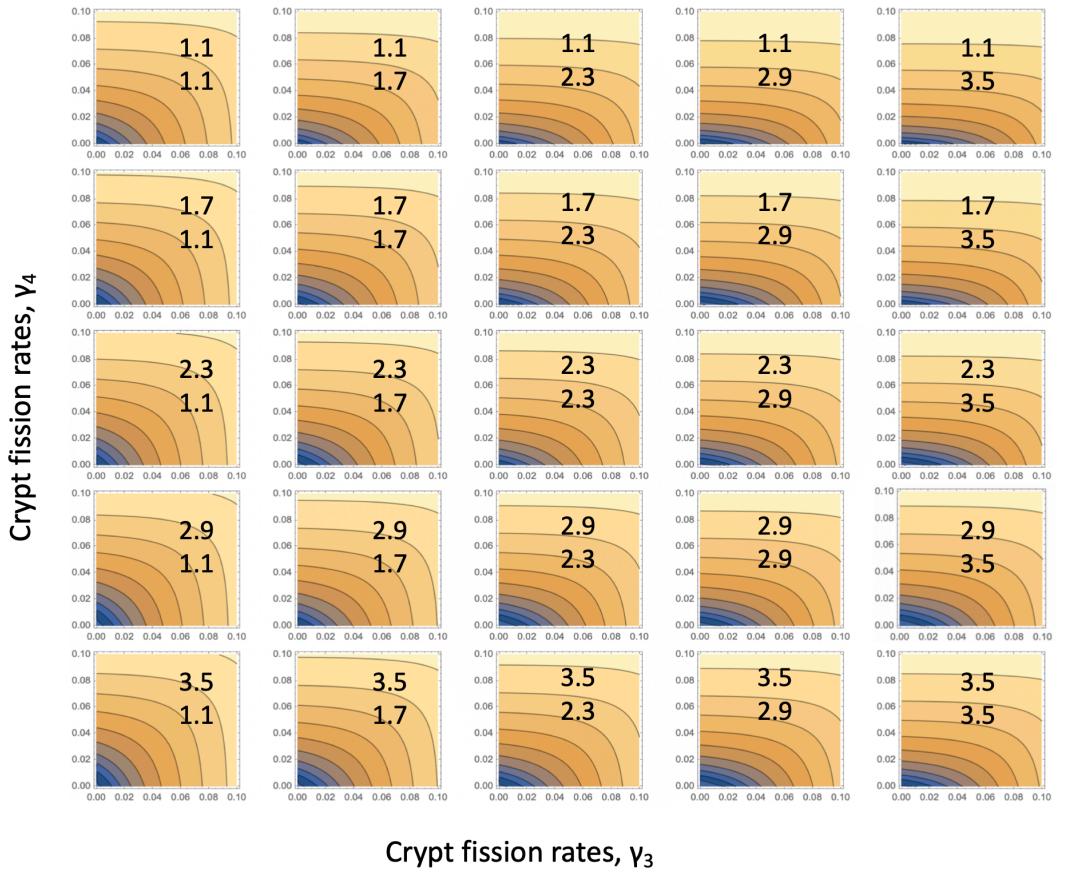
So far, in order to fit the model to the adenoma data, we fixed several of the parameters to their measured values and varied the three remaining parameters ( $r_1$ ,  $\gamma_3$ ,  $\gamma_4$ ) within realistic ranges to investigate the error landscape and find global minima. To obtain a more comprehensive picture of the model behavior, we have implemented a procedure where five parameters were varied: the division rate of stem cells,  $r_1$ , two cellular fitness parameters,  $R_{APC^{+/-}}$  and  $R_{KRAS}$  (with  $F_{APC^{++}} = 2F_{APC^{+/-}}$ ); and two crypt fission parameters,  $\gamma_3$  and  $\gamma_4$ . The rest of the parameters were set to their values in Table Appendix 1—table 2.

The fitness parameters were varied between 1.1 and 3.5 to match the measured range. For each pair ( $R_{APC^{+/-}}$ ,  $R_{KRAS}$ ), a fitting procedure identical to that of Appendix 1—figure 2 was performed, see Appendix 1—figure 3. Each graph corresponds to a unique pair ( $R_{APC^{+/-}}$ ,  $R_{KRAS}$ ), and the values of these coefficients are indicated on each panel. The horizontal axes of each panel is the stem cell division rate,  $r_1$ . The green curves show the  $\log_{10}$ (fitting error), and the black (gray) dots show the best fitting values of  $\gamma_3$  ( $\gamma_4$ ), multiplied by 10 for convenience of presentation. If the best fitting crypt fission parameter was negative, then the fitting error shows corresponded to  $\gamma_3 = \gamma_4 = 0$ . We observe that in all the cases, the best fitting parameter combination is reached when the crypt fission rates become zero, reproducing the result of Appendix 1—figure 2, but for a wide parameter range.

To observe this more clearly, we also presented the error landscape for the best fitting parameter  $r_1$ , as a function of  $\gamma_3$  and  $\gamma_4$  (the horizontal and vertical axes in each panel, respectively). The quantity  $\log_{10}(\text{fitting error})$  is shown as a heat map, with darker colors corresponding to lower error values. We can see that the lowest error corresponds to the corner  $\gamma_3 = \gamma_4 = 0$ .



**Appendix 1—figure 3.** Fitting the linear model. Each panel shows fitting results for a particular parameter combination,  $(R_{APC^{+/-}}, R_{KRAS})$ ; the values of these two fitness parameters are indicated, and  $F_{APC^{+/+}} = 2F_{APC^{+/-}}$ . The green lines show the  $\log_{10}(\text{fitting error})$  as a function of  $r_1$ . The fitted parameters  $\gamma_3$  and  $\gamma_4$  (multiplied by 10) are shown as black (gray) lines. The rest of the parameters are as in Table **Appendix 1—table 2**.



**Appendix 1—figure 4.** Fitting the linear system: for each parameter combination,  $(R_{APC^{+/-}}, R_{KRAS})$ , of **Appendix 1—figure 3**, the error landscape is shown that corresponds to the best fitting  $r_1$  (see the minima of the green lines in **Appendix 1—figure 3**). The contour plot represents  $\log_{10}(\text{fitting error})$  as a function of  $\gamma_3$  and  $\gamma_4$  (darker colors correspond to lower values). The rest of the parameters are as in Table **Appendix 1—table 2**.

### Nonlinear model: competition among crypts

Let us assume that the types of crypts that undergo crypt fission (types 3, 4, and 5) are in direct competition with each other, modeled as logistic, as opposed to straight exponential, growth. Let us denote the carrying capacity associated with the growth of the KRAS- crypt (type 3) as  $K_A$ , and that for the KRAS+ crypts (types 4 and 5) as  $K_R$ . Then the nonlinear system with crypt competition becomes

$$\dot{n}_1 = -(R_{12} + R_{14})n_1, \quad (19)$$

$$\dot{n}_2 = R_{12}n_1 - (R_{23} + R_{25})n_2, \quad (20)$$

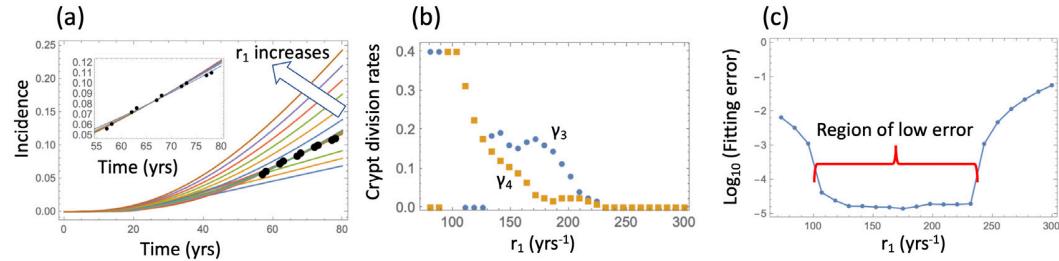
$$\dot{n}_3 = R_{23}n_2 - R_{36}n_3 + \gamma_3 n_3 \left(1 - \frac{n_3 + n_4 + n_5}{K_A}\right) - \delta n_3, \quad (21)$$

$$\dot{n}_4 = R_{14}n_1 - R_{45}n_4 + \gamma_4 n_4 \left(1 - \frac{n_3 + n_4 + n_5}{K_R}\right) - \delta n_4, \quad (22)$$

$$\dot{n}_5 = R_{25}n_2 + R_{45}n_4 - R_{56}n_5 + \gamma_5 n_5 \left(1 - \frac{n_3 + n_4 + n_5}{K_R}\right) - \delta n_5, \quad (23)$$

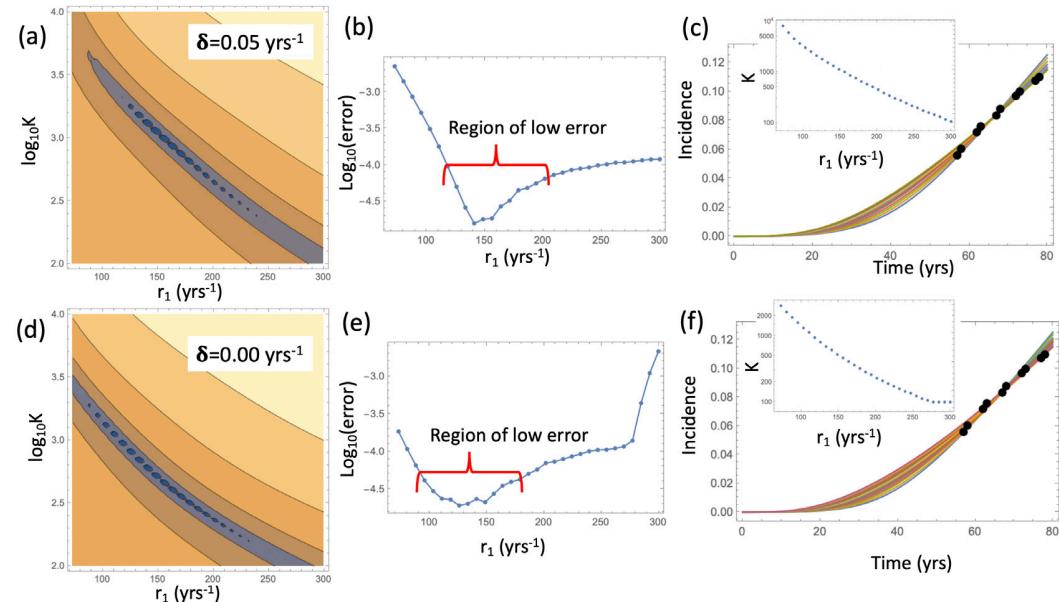
with initial conditions (6) and the probability to create a crypt of type 6 given by system (12). While an analytical solution is no longer available, a procedure similar to that performed for the linear system can be performed numerically. Results can be seen in figure **Appendix 1—figure 5**. As before, all the parameters were fixed to their values in Table **Appendix 1—table 2**, except the

cell division rate,  $r_1$ , and the crypt fission rates,  $\gamma_3$  and  $\gamma_4$ . Additionally, we assumed a crypt carrying capacity  $K_A = K_R = K = 10^3$  and crypt death rate  $\delta = 0.05 \text{ yrs}^{-1}$ . The best fitting values of  $\gamma_3$  and  $\gamma_4$  were found for each value of  $r_1$ , which was varied in the realistic range. This procedure yielded a range of low-error fits (panel (c)) that correspond to nonzero values of crypt fission rates in the realistic range (panel (b)), with the corresponding incidence curves given in the inset of panel (a).



**Appendix 1—figure 5.** Nonlinear model fitting to the incidence data (fixing the crypt carrying capacity, varying crypt fission rates). (a) The best fitting curves corresponding to increasing SC division rates,  $r_1$ , are plotted together with the epidemiological data; inset: the best fitting curves corresponding to the values of  $r_1$  from the Region of low error, see panel (c). (b) The best fitting parameters  $\gamma_3$  and  $\gamma_4$  are shown for each value of  $r_1$ . (c) The fitting error as a function of  $r_1$ . The relative fitness values are fixed to  $F_{APC^{+/-}} = 1.6$ ,  $F_{APC^{-/-}} = 3.76$ ,  $F_{KRAS} = 3.54$ ,  $K_A = K_R = K = 1000$ . The rest of the parameters are as in Table Appendix 1—table 2.

Since this procedure yielded a wide range of similarly good fits under a fixed value for the crypt carrying capacity, we next performed a fitting procedure where the crypt fission rates  $\gamma_3$  and  $\gamma_4$  were fixed to those in Table Appendix 1—table 2, and the best carrying capacity,  $K$ , was determined for each division rate,  $r_1$ , by fitting to the data. Results (under the assumption that  $K_A = K_R = K$ ) are presented in Appendix 1—figure 6a-c. We can see that for a range of values of the cell division rates,  $r_1$ , a low-error fit was found, with the carrying capacity values ranging between about  $5 \times 10^2$  and  $5 \times 10^3$ . The best fitting values are  $K = 1318$ ,  $r_1 = 141.1$  (divisions approximately every 2.5 days). Changing the crypt death rate to  $\delta = 0$  yields very similar results, see Appendix 1—figure 6d-f.



**Appendix 1—figure 6.** Nonlinear model fitting to the incidence data (fixing crypt fission rates, varying the crypt carrying capacity). Compared are the cases in the presence of a non-zero crypt death rate,  $\delta = 0.05 \text{ yrs}^{-1}$  (a–c), and in the absence of crypt death rate,  $\delta = 0.0 \text{ yrs}^{-1}$  (d–f). (a,d) The heatmap of the fitting error, where all parameters are fixed except  $K_A = K_R = K$  and  $r_1$ . Dark colors correspond to lower values of the error. (b,e) The

Appendix 1—figure 6 continued on next page

## Appendix 1—figure 6 continued

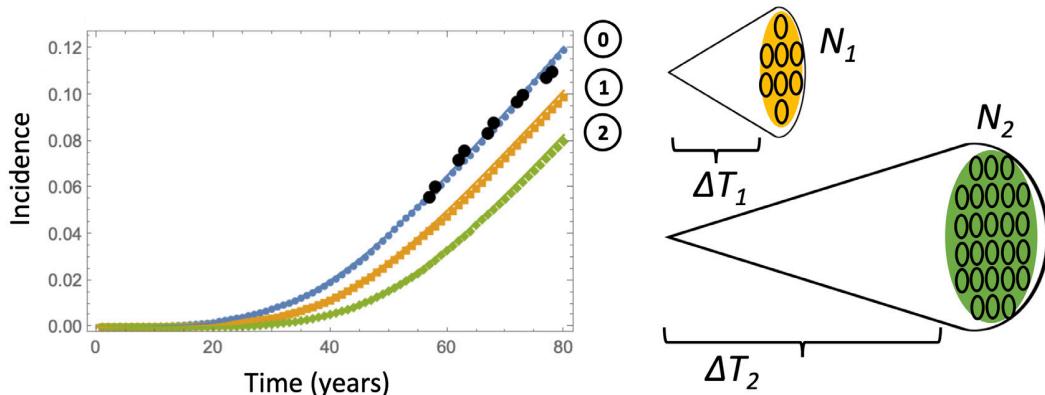
fitting error as a function of  $r_1$ . (c,f) The best fitting curves corresponding to increasing SC division rates,  $r_1$ , are plotted together with the epidemiological data the value of  $r_1$  are taken from the Region of low error. Inset: The best fitting carrying capacity  $K$  is shown for each value of  $r_1$ . The rest of the parameter values are fixed to,  $F_{APC^{+/-}} = 1.6$ ,  $F_{APC^{-/-}} = 3.76$ ,  $F_{KRAS} = 3.54$ ,  $\gamma_3 = 0.2$ ,  $\gamma_4 = 0.07$  and the rest as in Table Appendix 1—table 2.

These results suggest that biologically, the nonlinear model is a more appropriate choice because it produces the best fit for values of crypt fission rates that are within the experimentally observed range, while the linear model requires zero crypt fission rates. From the statistical point of view, the nonlinear model is a significantly more powerful model e.g. by applying the Akaike Information Criterion (AIC).

## Model predictions and advanced adenoma generation dynamics

Gillespie simulations and the mean-field approximation

**Equation 12** approximates the probability  $P(t)$  that by time  $t$ , a single type 6 crypt has been generated. In order to test its validity we have run stochastic Gillespie simulations based on system (LABEL:n1non-23), which were stopped when the first type 6 crypt was created. The resulting incidence was compared to that obtained from the deterministic system (LABEL:n1non-23, 12), see **Appendix 1—figure 7**, line (0). There, the blue symbols and the blue solid line correspond to the stochastic simulations and the deterministic approximation respectively. The deterministic approximation showed excellent agreement with Gillespie simulations.



**Appendix 1—figure 7.** Comparing deterministic solutions (system (LABEL:n1non-23, 12), solid lines) with the stochastic Gillespie simulations (symbols of the same color). The three lines correspond to the different stopping conditions: (0) 1 crypt of type 6; (1)  $N_1 = 10^2$  crypts of type 6, (2)  $N_2 = 10^5$  crypts of type 6. The rest of the parameters are as in **Appendix 1—figure 6a-c**. 100,000 independent simulations were performed for each line, and the standard deviation bars are too small to see.

Gillespie simulations described here can be used to simulate stochastic effects associated with crypt dynamics studied here, see e.g. results presented in Section 5.3. These simulations were also used to study the effect of aspirin, Section 6. Finally, this methodology can be extended to study the clonality of abnormal crypts. The ODE model does not keep track of clonality. For example, if type 5 crypt is created multiple times (by conversion) in the system, the variable  $n_5$  simply gives the total number of type-5 crypts. In a stochastic Gillespie model, however, it is possible to keep track of different clones by designating each newly generated crypt as a different “sub-type”, which can then clonally expand through crypt fission. This however goes beyond the scope of the current study.

### The growth phase of advanced adenoma

So far we have stopped the simulations as soon as the first copy of a type 6 crypt was created. Alternatively, one could explicitly account for the growth phase of the APC-/ KRAS+ crypts. Estimates for the doubling time of advanced adenomas vary in the literature. For example, (**Jones et al., 2008**) suggests that the doubling time of advanced adenoma is 250 days (that is, the net rate of expansion  $1.01 \text{ yr}^{-1}$ ), with a similar doubling time of about 1 year quoted in **Kang and Shibata**,

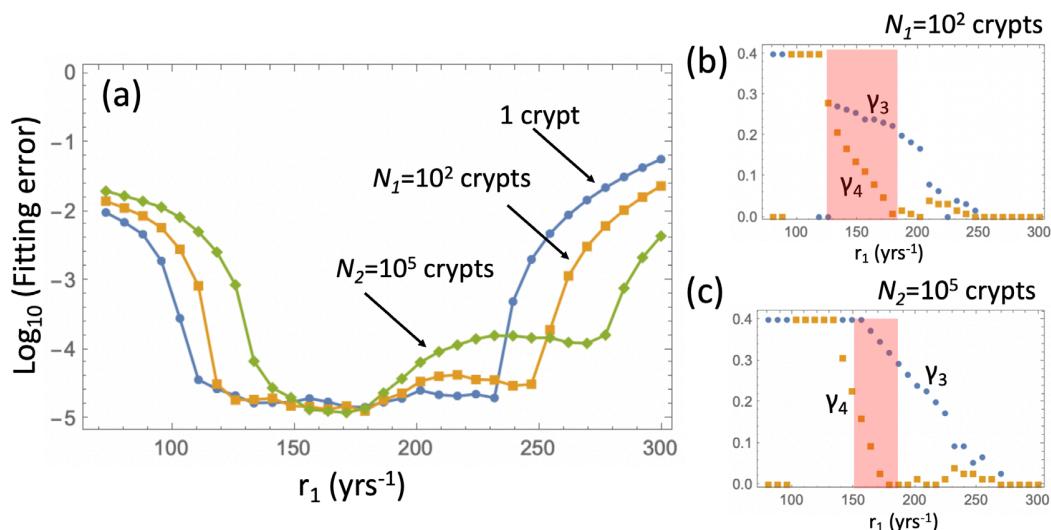
2013 (the net rate of expansion  $0.7 \text{ yr}^{-1}$ ). The number of crypts in an advanced adenoma can be estimated as follows. Given that the minimum advanced adenoma size is 10 mm (see [Dewanji et al., 2011](#) and also [Chen et al., 2018](#)), and assuming  $10^9 \text{ cells per cm}^3$  ([Huels and Sansom, 2015](#)), we obtain that each advanced adenoma contains about  $10^6 \text{ cells}$ . This comprises about  $10^2$  crypts, if we assume that type 6 crypts are  $10^4 \text{ cells each}$ , which is somewhat larger than normal colonic crypts that measure about  $2 \times 10^3 \text{ cells}$  ([Devall et al., 2021](#)). A somewhat larger estimate is given in [McNeil et al., 2018](#) where it is noted that a 1 mm<sup>3</sup> adenoma contains on the order of  $5 \times 10^5 \text{ cells}$ , which translates into  $5 \times 10^8 \text{ cells}$  in an adenoma of size 10 mm<sup>3</sup>, or  $5 \times 10^4$  crypts. Therefore, we have explored a range of crypt expansion to sizes  $N_1 = 10^2$  and  $N_2 = 10^5$  crypts, see [Appendix 1—figure 7](#). Assuming, for type 6 crypts,  $\gamma_6 = 1.01 \text{ yr}^{-1}$  and the crypt death rate  $\delta_6 = 0.05 \text{ yr}^{-1}$ , we obtain that expansion from 1 type 6 crypt to  $N_1$  crypts will take  $\Delta T_1 = 4.79 \text{ years}$ , and expansion from 1 type 6 crypt to  $N_2$  crypts will take  $\Delta T_2 = 11.97 \text{ years}$  respectively.

Using these calculations, we obtain the deterministic prediction that expansion to  $10^2$  crypts will shift the incidence curve by  $\Delta T_1$  years, and expansion to  $10^5$  crypts will shift it by  $\Delta T_2$  years. Apart from this shift, the probability of crypt non-extinction has to be incorporated. If  $P(t)$  is the solution of [equation \(12\)](#) under system (LABEL:n1non-23), then we can approximate the probability  $P_i(t)$  that by time  $t$ ,  $N_i$  crypts of type 6 have been generated, by

$$P_i(t) = \left(1 - \frac{\delta}{\gamma_6}\right) P(t - \Delta T_i), \quad i = 1, 2. \quad (24)$$

The corresponding curves are marked as curves 1 and 2 in [Appendix 1—figure 7](#), see the orange and green solid lines respectively. To check this approximation, we have run Gillespie simulations that stopped as soon as  $N_1 = 10^2$  (orange symbols) and  $N_2 = 10^5$  (green symbols) crypts were generated. Again, very good agreement between the stochastic simulations and our deterministic approximation was observed.

Next, we will show that similar results are obtained when the model with an explicit type 6 crypt expansion phase is fit to the advanced adenoma incidence data. Fixing a crypt carrying capacity, we found the best fitting fission rates  $\gamma_3$  and  $\gamma_4$  for each value of  $r_1$ . The procedure was similar to that used in [Appendix 1—figure 5](#), except probabilities to have  $N_i$  type 6 crypts,  $P_i(t)$  ([equation \(24\)](#)) were used to fit the incidence data. Results are shown in [Appendix 1—figure 8](#). The plot of the fitting error again shows regions of low error (panel (a)). In those regions, nonzero values of crypt fission correspond to the best fit (see panels (b) and (c), where regions of low error are shaded red). As in [Appendix 1—figure 5\(b\)](#), the best fitting values of  $\gamma_4$  are lower than those of  $\gamma_3$ , and both are similar to those found in the literature, see [Table Appendix 1—table 2](#).

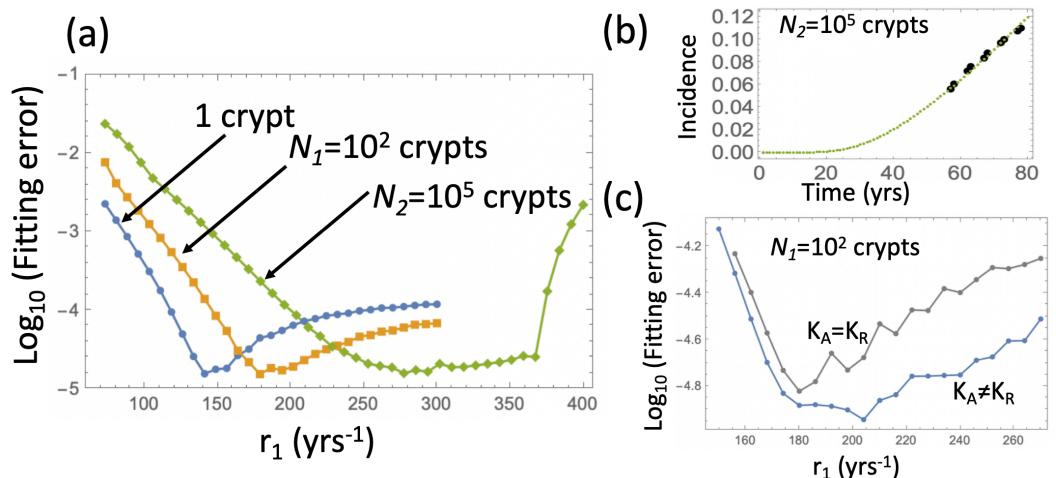


**Appendix 1—figure 8.** Fitting the adenoma incidence curve to the nonlinear model with an explicit expansion phase. Under a fixed  $K_A = K_R = 1000$ , for each value of  $r_1$  the best fitting values of  $\gamma_3$  and  $\gamma_4$  were found. (a) The [Appendix 1—figure 8 continued on next page](#)

## Appendix 1—figure 8 continued

fitting error is shown as a function of  $r_1$  for the original model (blue line, same as in **Appendix 1—figure 5(c)**); the model with expansion to  $N_1$  crypts (yellow curve), and the model with expansion to  $N_2$  crypts (green curve). (b) The best fitting values of crypt fission rates,  $\gamma_3$  and  $\gamma_4$ , for the model with expansion to  $N_1$  crypts. (b) The best fitting values of crypt fission rates,  $\gamma_3$  and  $\gamma_4$  for the model with expansion to  $N_2$  crypts. The rest of the parameters are as in **Appendix 1—figure 5**.

Therefore, to proceed, we will fix the crypt fission rates  $\gamma_3$  and  $\gamma_4$  to their values in Table **Appendix 1—table 2**, and fit the model to the advanced adenoma incidence curve to find the best rate  $r_1$  and crypt carrying capacities. This is a procedure similar to that in **Appendix 1—figure 6**, except type 6 crypt populations grow to a given size. The fitting error for the three cases (that is, growth to size 1, size  $10^2$ , and size  $10^5$  crypts) is shown in **Appendix 1—figure 9(a)** as a function of  $r_1$ . The fitting procedure was performed by using the ODE approximation (**equation (24)**), and the best fitting parameter set then used in a Gillespie simulation ( $5 \times 10^5$  simulations are used and the error bars are too small to see). **Appendix 1—figure 9(b)** shows the result of the Gillespie simulation corresponding to the best-fitting parameters for the case of expansion to  $10^5$  crypts. The ODE prediction and the fitted curve for the expansion to  $10^5$  crypts look identical.

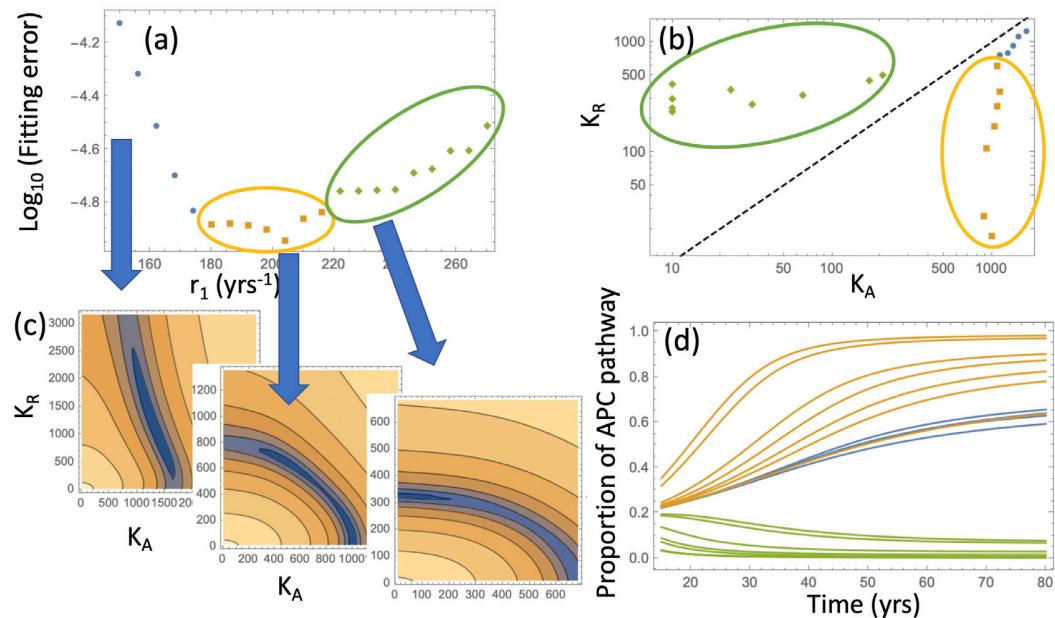


**Appendix 1—figure 9.** Fitting the adenoma incidence curve to the nonlinear model with an explicit expansion phase. Under fixed crypt fission rates, for each value of  $r_1$  the best fitting value of  $K_A = K_R$  was found. (a) The fitting error is shown as a function of  $r_1$  for the original model (blue line, same as in **Appendix 1—figure 6(b)**); the model with expansion to  $N_1$  crypts (yellow curve), and the model with expansion to  $N_2$  crypts (green curve). (b) The fitted incidence curve (Gillespie simulation) for the case of expansion to  $10^5$ . The ODE prediction and the fitted curve for the expansion to  $10^2$  crypts look identical. (c) The reduction in error resulting from relaxing the restriction  $K_A = K_R$ . The rest of the parameters are as in **Appendix 1—figure 6(a-c)**.

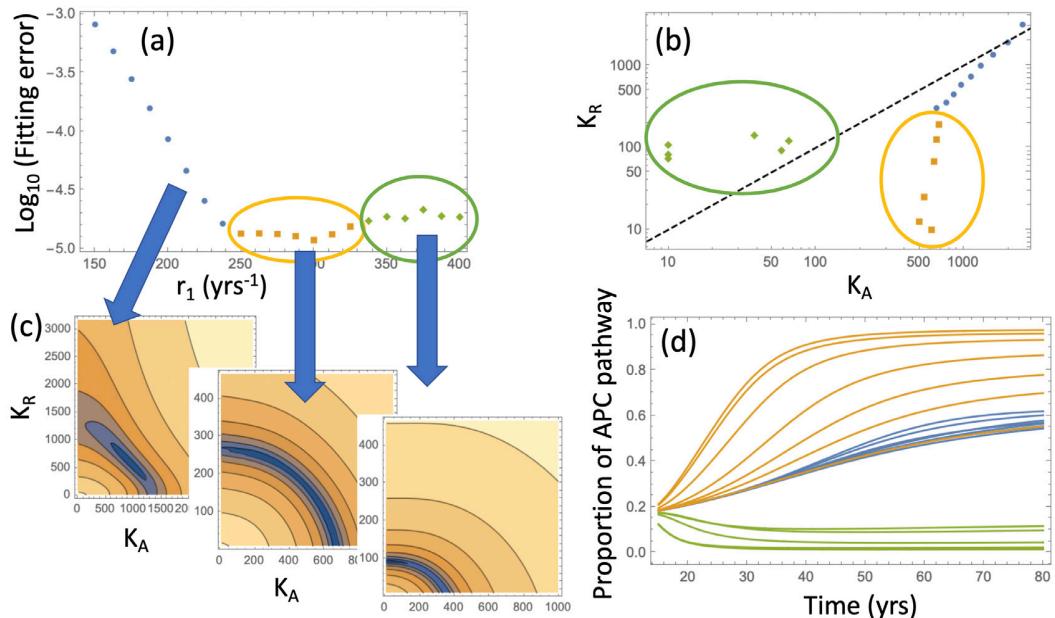
## Compartment dynamics and pathways to advanced adenoma

Relaxing the restriction  $K_A = K_R$  improves the fitting, as shown in **Appendix 1—figure 9(c)**. We will explore compartment dynamics in this more general setting, given that by AIC, the model with carrying capacities  $K_A$  and  $K_R$  separately fitted is more powerful. **Appendix 1—figure 10** shows the procedure of fitting in the case where type 6 crypts grow to size  $10^2$  before detection. For each value of  $r_1$ , the best fitting pair  $(K_A, K_R)$  was determined and the resulting error plotted (panel (a)). Panel (c) shows examples of heatplots of the error as a function of  $K_A$  and  $K_R$  (for a fixed  $r_1$  value). Depending on the size of the resulting error and the location of the minimum in the  $(K_A, K_R)$  space, three groups of fits can be distinguished. Of particular interest in the group marked by orange and green points. For the fits in the “orange” group, we have  $K_A > K_R$ , that is, the best fitting carrying capacity that characterizes type-3 crypts is larger than that for type-4 and 5 crypts. In contrast to that, the “green” group has  $K_A < K_R$ . This can be seen in **Appendix 1—figure 10(b)**, where the best fitting pairs  $(K_A, K_R)$  are shown for each  $r_1$  value. We reject the “green” group of fits because it corresponds to a larger carrying capacity of type-4 and 5 crypts compared to that of type-3 crypts, which is inconsistent with observations; in addition,

the fitting error is minimized for one of the fits in the “orange” group. The same procedure was also performed for the case where type-6 crypts grew to size  $10^5$  (see **Appendix 1—figure 11**), with very similar results.



**Appendix 1—figure 10.** Pathways to adenoma: Fitting the incidence curve to the nonlinear model with an explicit expansion phase. Under fixed crypt fission rates, for each value of  $r_1$  the best fitting pair ( $K_A, K_R$ ) was found. (a) The fitting error is shown as a function of  $r_1$ ; 3 groups of parameter combinations are marked by blue, orange, and green. (b) The best fitting pairs ( $K_A, K_R$ ) for each of the points in the three groups. (c) Heatplots of the fitting error as a function of ( $K_A, K_R$ ), for 3 different values of  $r_1$ , one from each group. (d) Proportion of the APC pathway (formula **Moolgavkar, 1978**) plotted as a function of time for different values of  $r_1$ , using the same color code for the three groups. Expansion to  $10^2$  crypts is assumed. The rest of the parameters are as in **Appendix 1—figure 6a-c**.



**Appendix 1—figure 11.** Same as **Appendix 1—figure 10**, except expansion to  $10^5$  crypts is assumed.

In order to determine the likelihood of type  $y_6$  produced from cells of type  $y_3$  or cells of type  $y_5$ , we looked at the probabilities of the two pathways,  $P_{APC}$  and  $P_{KRAS}$ , which stand for the probability to create adenoma by first inactivating the APC gene and then adding a gain-of-function mutation in KRAS gene, or by first activating KRAS and then inactivating the APC gene, see panel (a) of **Figure 3** of the main text. The two probabilities satisfy the following equations,

$$\dot{P}_{APC} = R_{36}n_3(1 - P_{APC}), \quad P_{APC}(0) = 0, \quad (25)$$

$$\dot{P}_{KRAS} = R_{56}n_5(1 - P_{KRAS}), \quad P_{KRAS}(0) = 0. \quad (26)$$

The proportion of adenomas that originated through the APC pathway is given by

$$P_{APC}/(P_{APC} + P_{KRAS}). \quad (27)$$

This function is plotted in **Appendix 1—figure 10(d)** for the parameters corresponding to both “orange” and “green” groups of fits. We can see that for parameter combinations from the orange group the APC-pathway is predominant, while for parameter combinations from the “green” group the KRAS pathway is predominant. The same qualitative results are observed in **Appendix 1—figure 11(d)**.

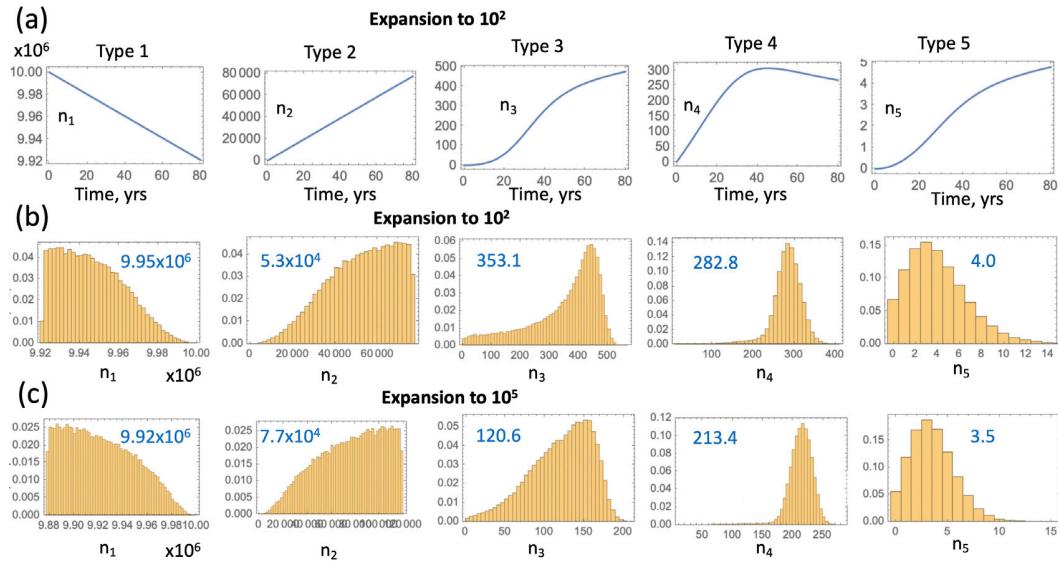
For further simulations, we used the two best fitting parameter combinations obtained from the fitting procedures in **Appendix 1—figure 10** and **Appendix 1—figure 11**. The parameter values obtained by means of these fitting procedure are summarized in Table **Appendix 1—table 3**.

**Appendix 1—table 3.** Parameters obtained from the fitting procedures in **Appendix 1—figure 10** and **Appendix 1—figure 11**.

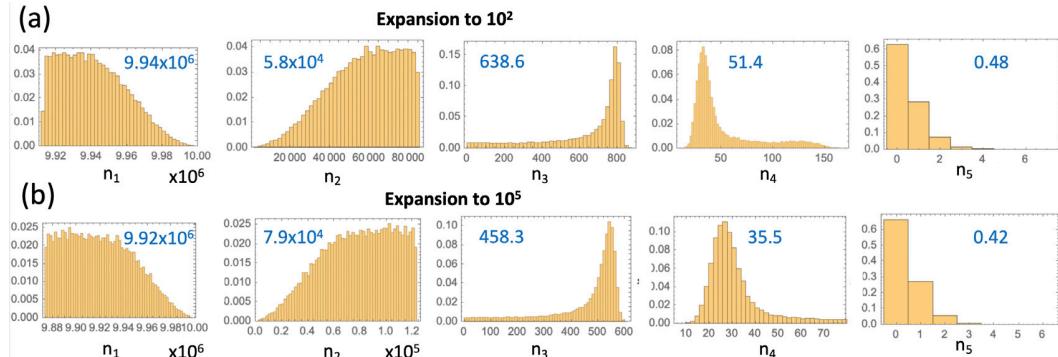
Parameter set #	$N_i$ (crypts)	$\Delta T_i$ (yrs)	$r_1$ (yrs-1)	$K_A$	$K_R$	Comment
1	$10^2$	4.79	204	1000	17	best fit
2	$10^2$	4.79	198	1039	171	2nd best fit
3	$10^5$	11.97	300	607	10	best fit
4	$10^5$	11.97	288	631	68	2nd best fit

System (LABEL:n1non-23, 12) does not only allow fitting the model to advanced adenoma incidence data, but also shows the prediction for the dynamics of the crypts of different mutational status leading up to the advanced adenoma formation, as shown in figure **Appendix 1—figure 12(a)**. Using the best fitting parameters of **Appendix 1—figure 6(b)**, we can visualize the mean trajectories for the numbers of crypts in each compartment,  $n_i$  for  $i = 1, \dots, 5$ . The stochastic Gillespie simulations provide additional information on the statistics of crypt numbers. **Appendix 1—figure 12(b)** shows the numerically obtained probability distributions for the crypt numbers in each compartment, at the time when the first type 6 crypt is created. These simulations of panel (b) correspond to the same parameter set as in panel (a); both assume expansion to  $10^2$  crypts and  $K_A = K_R$ . Panel (c) presents similar simulations but for the best fitting parameters where expansion to  $10^5$  crypts takes place. The mean crypt numbers are given in blue in each histogram; we can see that the orders of magnitude are similar for both cases. Similar results are obtained when the restriction  $K_A = K_R$  is dropped, see **Appendix 1—figure 13**; as expected, the size of the type-3 compartment is somewhat larger and the size of the type-5 compartment somewhat smaller than the corresponding values for the  $K_A = K_R$  case. Note that **Appendix 1—figure 12** and **Appendix 1—figure 13** do not contain information on pathways to advanced adenoma. In other words, shown are just the numbers of crypts of each type that are present in the entire simulated colon, at the time when the first crypt associated with the

advanced adenoma is generated, regardless of whether it was generated by a mutation of a type 3 or a type 5 crypt.



**Appendix 1—figure 12.** Compartment dynamics for the best fitting models with  $K_A = K_R$ . **(a)** The expected crypt numbers (using equations (LABEL:n1non-23)) for the 5 compartments are shown for the 2nd parameter set in table **Appendix 1—figure 9(c)** (expansion to  $10^2$  crypts). **(b)** For the same parameter values, probability distributions for the crypt numbers are shown at the time when the first type 6 crypt is created ( $5 \times 10^5$  simulations were run); mean crypt numbers are shown in blue. **(c)** Same as **(b)** but for the 3rd parameter set in table **Appendix 1—figure 9(c)** (expansion to  $10^5$  crypts).



**Appendix 1—figure 13.** Same as **Appendix 1—figure 12(b-c)**, but without the restriction  $K_A = K_R$ . **(a)** Parameter set #2 from Table **Appendix 1—table 3** was used. **(b)** Parameter set #4 from Table **Appendix 1—table 3** was used.

### Aspirin dosage and modeling aspirin's effect on the crypt dynamics

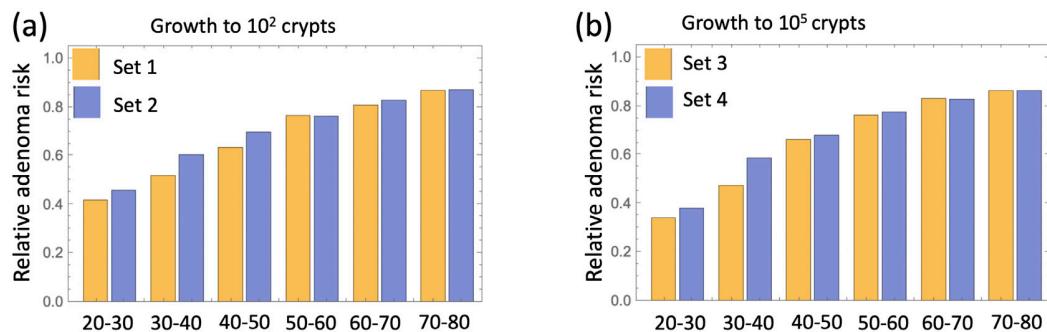
In mouse experiments performed in [Shimura et al., 2020](#) we used the doses of 15 mg/kg, 50 mg/kg, and 100 mg/kg aspirin each day. Using the conversion table in [Drew et al., 2016](#), and a mass of a human of 70 kg, we obtain the equivalent daily human doses of 85.1 mg/day, 283.5 mg/day, and 567 mg/day.

To compare this with the weekly intake of the participants in the study of [Chan et al., 2004](#), we note that a standard aspirin tablet contains 325 mg, such that 2 tablets a day (i.e. 14 tablets per week) is 650 mg/day. The patients were grouped by their intake into 0.51.5 tablets per week, 25 tablets per week, 614 tablets per week, and  $gt_{14}$  tablets per week. Therefore, the highest dose

administered in our mouse experiments (100 mg/kg aspirin per day) most closely matches the intake of the 614 tablets per week group, see **Table 1** of the main text.

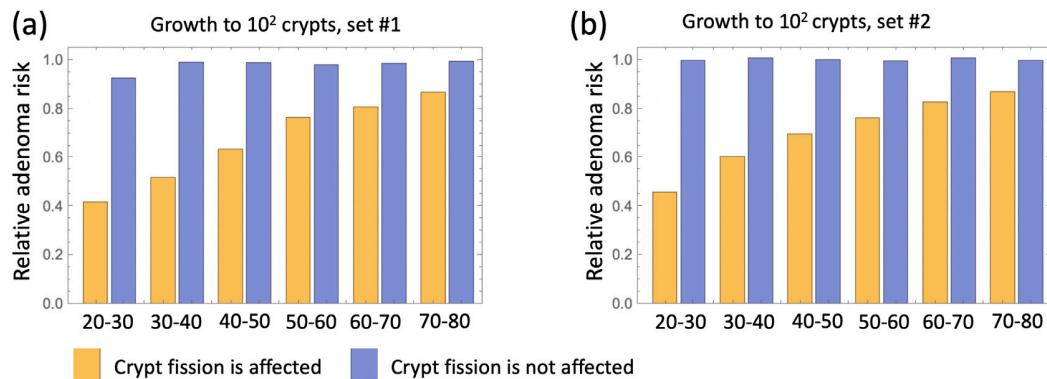
To explore the effect of aspirin on the incidence of advanced adenoma, we used Gillespie simulations of system (LABEL:n1non-23), also including a growth phase of type 6,  $\dot{n}_6 = R_{36}n_3 + R_{56}n_5 + (\gamma_6 - \delta_6)n_6$ .

In these simulations, aspirin-free parameter values were used for  $0 \leq t < T_{start}$ , and then during an interval of aspirin treatment  $T_{start} \leq t \leq T_{end}$ , parameter values modified by aspirin were used, see **Figure 4(b)** of the main text. As before, simulations were stopped at  $t = 80$  yrs or when the target number of type 6 crypts were generated, whichever happened first. Incidence curves and relative risk were constructed by processing a large number of such simulations.  $2.5 \times 10^5$  independent simulations per condition were used unless otherwise noted. The error bars for the age incidence curves or the relative risk are too small to be seen.

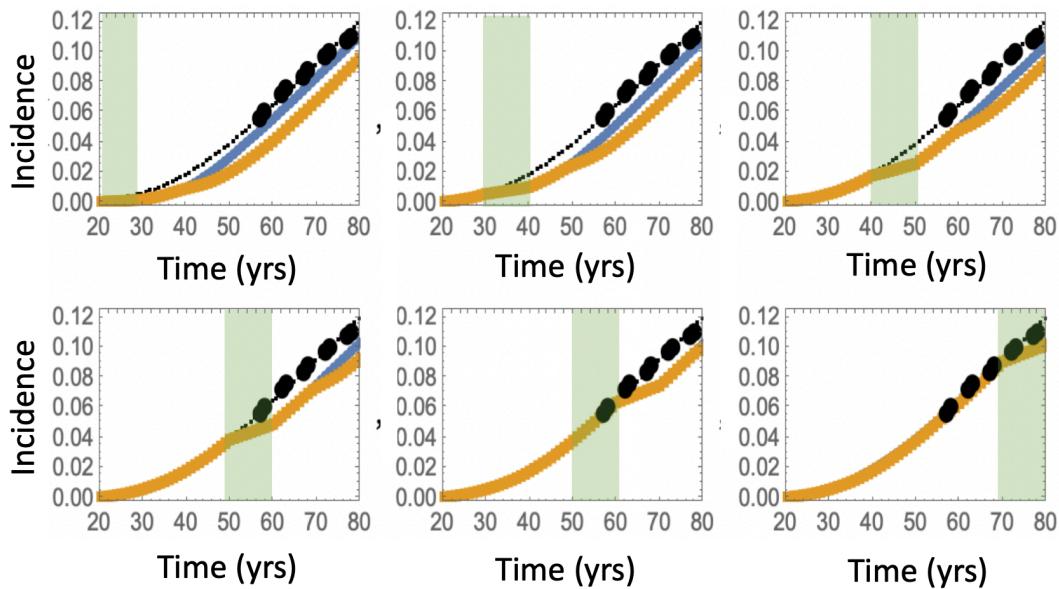


**Appendix 1—figure 14.** extends the results of main text **Figure 4(a)** and shows that relative adenoma risk predicted for each decade of treatment is very similar for all parameter sets (#1-4, Table **Appendix 1—table 3**).

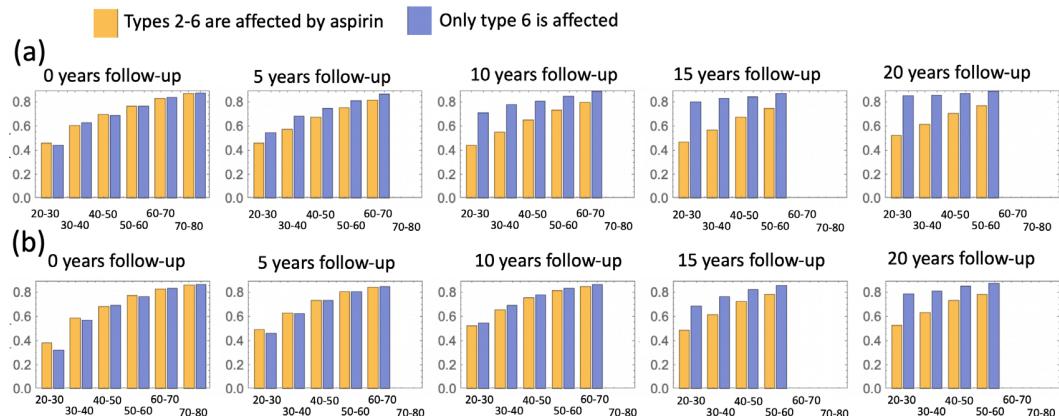
Figure **Appendix 1—figure 14**: Relative incidence of advanced adenoma, where (a) growth to  $10^2$  type 6 crypts is assumed and parameter sets #1 and #2 (Table **Appendix 1—table 3**) are compared, and (b) growth to  $10^5$  type 6 crypts is assumed and parameter sets #3 and #4 are compared. Aspirin affects type 2–6 cells, both through conversion rates and crypt fission/death. Treatment is applied for different decades (as marked under the pairs of bars), and the relative risk is evaluated at the end of the treatment decade.  $2.5 \times 10^5$  independent simulations are used for each condition.



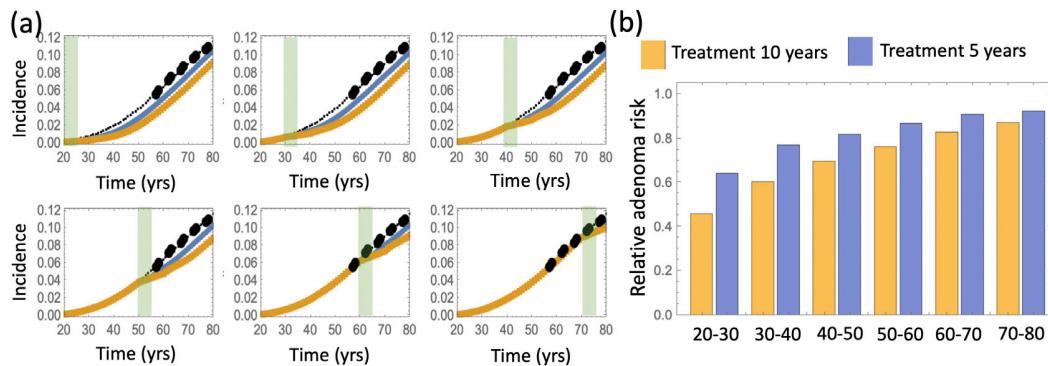
**Appendix 1—figure 15.** Relative incidence of advanced adenoma, where (a) parameter set #1 (Table **Appendix 1—table 3**) and (b) parameter sets #2 is used. In each panel, a comparison is presented between the case where aspirin affects crypt fission/death rates (yellow) and where it does not affect crypt fission/death rates (blue). Aspirin affects cellular kinetics through conversion rates in all cases; types 2–6 are affected.  $2.5 \times 10^5$  independent simulations are used for each condition.



**Appendix 1—figure 16.** Predicted advanced adenoma incidence in the absence of aspirin treatment (thin black lines are the fitted curves and black dots are incidence data); under aspirin treatment where the drug affects types 2–6 (yellow lines), under aspirin treatment where the drug affects type 6 only (blue lines). Each panel corresponds to aspirin treatment administered during one decade (20–30 years, 30–40 years, etc). The treatment period is shaded light green. It was assumed that type 6 crypts grow to  $10^5$ , and set #4 in Table **Appendix 1—table 3** was used.  $2.5 \times 10^5$  independent simulations are used for each condition.



**Appendix 1—figure 17.** A comparison of predicted relative advanced adenoma risk under the assumption that cell types 2–6 (yellow bars) and only type 6 cells (blue bars) are affected by aspirin. Aspirin is administered during different decades of patients' life (as marked under each pair of bars), and different panels correspond to different durations of the follow-up period. (a) Parameter set #2, and (b) parameter set #4 (Table **Appendix 1—table 3**) is used;  $2.5 \times 10^5$  independent simulations are used for each condition.



**Appendix 1—figure 18.** A comparison of predicted relative advanced adenoma risk under the assumption that aspirin treatment lasts 10 years (yellow) or 5 years (blue). (a) Predicted advanced adenoma incidence in the absence of aspirin treatment (thin black lines are the fitted curves and black dots are incidence data); the incidence curve for patients treated for 10 (yellow) and 5 (blue) years. Aspirin is administered during different decades of patients' life (the shorter, 5 year intervals, are marked green). (b) A comparison of predicted relative advanced adenoma risk under the assumption of treatment for 10 years (yellow bars) and 5 years (blue bars). All cell types 2–6 are affected by aspirin (both through conversion and fission/death). Parameter set #2 is used;  $2.5 \times 10^5$  independent simulations are used for each condition.

**Appendix 1—figure 15** explores the individual contributions of the two possible mechanisms by which the effect of aspirin could be manifested. The yellow bars correspond to the effect on both conversion rates and crypt fission/death rates, and the blue bars only include the effect on conversion rates. It is clear that the effect on conversion rates is not as strong compared to the effect on the crypt fission/death rates.

**Appendix 1—figure 16** is similar to main text **Figure 4(a)**, except it contains additional incidence functions that correspond to aspirin only affecting type 6 cells (blue lines). Not surprisingly, the resulting change in adenoma risk is smaller, that is, the blue line (effect on type 6 cells only) is closer to the thin black line (no-aspirin adenoma incidence) than the yellow line (where aspirin affects all types 2–6). It is interesting that the difference becomes larger with time. In **Appendix 1—figure 17** we change the length of the follow-up (see **Figure 4(b)** of the main text), and the difference between the two assumptions (crypts 2–6 affected vs only type 6 crypts are affected) increases as time goes by.

**Appendix 1—figure 18** shows the effect of a shorter aspirin duration time