Supplemental Materials and Methods

MODEL DESCRIPTION

These methods are presented in ODD protocol standard format for agent-based models (1) and the modeling was completed using NetLogo 4.0.4 (2).

Model overview

Purpose

This agent-based model was designed to explore the interactions of differentiation dynamics with the somatic evolution of the hallmarks of cancer during neoplastic progression. We set out to determine if there is there a canonical ordering to the acquisition of the hallmarks due to their effects on somatic evolution.

Entities, state variables, and scales

We explicitly model individual cells, each of which in a simulation has values for the state variables shown in Table S4.

Process overview and scheduling

Time is modeled with discrete steps, each of which is approximately one phase of the cell cycle. Thus, it takes 4 time steps to complete one cell cycle once a cell begins dividing. The flow chart in Fig. S4 outlines the decisions that a cell makes.

Model design concepts

Emergence

We are interested in the order of mutations acquired by the cells in simulations beginning with normal (no mutations) cells and ending in cancer. We observe that the number of cells in the simulation increases over time as individual cells acquire growth and survival mutations.

Adaptation

Cells in this model are not explicitly encoded with adaptive behavior. Instead, the set of mutations that they have increases the fitness of a cell by increasing its cell division rate, decreasing its death rate, or disrupting differentiation.

Sensing

While cells are not able to sense each other directly, they are able to indirectly sense their environment. In order to make division and death decisions, they sense whether there are sufficient nutrients or survival signals globally, which are not explicitly modeled. Instead, we use as a proxy a function of the actual number of cells in the model, and either the number of cells that can signal for angiogenesis or the expected number of cells, respectively.

Interaction

There are no direct interactions or communication between cells in the model. Cells compete for nutrients indirectly.

Stochasticity

Cell division, cell death, and mutation are all stochastic in this model. Each is constrained by parameterized possibilities.

Observation

Every 60 time steps (15 days), summary data are written to a file. These data include population size (total and at each differentiation stage), proportion of individual mutations, and the entire clonal population structure. Additionally, a new "clone" is created with each new mutation since we model new phenotypic mutations in the spirit of the infinite alleles model (3). The life history of each clone is recorded, including its set of mutations, the time step at which the mutations occurred, and its parent clone, allowing the explicit generation of cell lineages.

Model Details

Initialization

See Table S5 for the parameter values and their ranges used in the simulations. The number of differentiation stages is a parameter d to the model, and each cell i has a state variable that tracks its differentiation level at time t, $d_i(t)$. Stem cells have $d_i(t) = 0$, fully differentiated cells have $d_i(t) = d - 1$, and transient cells have $0 > d_i(t) > d-1$, where d is the total number of differentiation, or transient, stages possible. Simulations are initialized with $2^j s$ cells of each stage j for a total population size at time $t = 0$ of

$$n(0) = \sum_{j=0}^{d-1} 2^j s$$

The model is allowed to run for 50 time steps during which no new mutations are gained and telomeres are not decremented, in order for the population structure to reach

equilibrium before the experiments begin. All cells are initialized with the same parameter values (Table S5), and over time acquire differences in their state variables (Table S4). Parameters are converted from hours or days into time steps as necessary.

Submodels

The submodels below correspond to boxes in the flowchart (Fig. S4).

Each cell can die

The probability for apoptosis due to background mortality for a cell i is given by, $p_{i,a} = p_a / \alpha$, where p_a is the wild-type probability for apoptosis and α is the reduction in probability for apoptosis that occurs when a cell acquires the evasion of apoptosis mutation. The decision is made independently for each cell i if it dies by drawing a random number from a uniform distribution between 0 and 1 that is less than its probability for apoptosis, $p_{i,a}$.

When there are excess cells in the model, fully differentiated cells can die. Thus, when the actual number of cells in the simulation exceeds the target number of cells such that $n(t) > N$, where N is the target number of cells, cells with $d_i(t) = d\text{-}1$ are randomly removed from the simulation until $n(t) = N$ or there are no more fully differentiated cells.

Healthy cell populations have sufficient survival signals from their microenvironment. When the number of cells in the population increases over some threshold, the amount of inherent survival signals is insufficient to support the tissue and so apoptosis is triggered at higher rates. Cells that have acquired the evasion of apoptosis mutation do not die in

response insufficient survival signals. The probability that any cell dies due to insufficient survival signal is given by $p_{ss} = min(max(r_{ss} (n(t) - N), 1), 0)$. As before, a cell *i* dies if it draws a random number from a uniform distribution between 0 and 1 that is less than the probability for death due to insufficient survival signals, p_{ss}.

Each cell can divide

The actual cell division process takes 3 timesteps if the cell has the genomic instability mutation and 4 timesteps if not. Here, cells can move out of the resting state and into active cell division.

Cell division proceeds if there are enough nutrients to support the existing the number of cells, or $n(t) < k$, where k is the number of cells that can be supported by the tissue and can be found by $k = n_{max} + a\, n_{SA=T}$. Here, n_{max} is the maximum number of cells that can be supported by nutrients available to a crypt, a is the angiogenic tissue support rate, and $n_{SA=T}$ is the number of cells with the sustained angiogenesis mutation.

A fully differentiated cell *i*, $d_i(t) = d\text{-}1$, cannot divide. Each stem or transient cell i has a probability to divide given by

$$p_{i,d}(t) = \begin{cases} p_s + m_{i,IA}(t)p_{s,IA} + m_{i,SG}(t)p_{s,SG} & \text{if } d_i(t) = 0 \\ p_t + m_{i,IA}(t)p_{t,IA} + m_{i,SG}(t)p_{t,SG} & \text{if } 0 < d_i(t) < d-1 \end{cases}$$

where $m_{i,IA} = 1$ if cell *i* has the insensitivity to anti-growth mutation and 0 otherwise, $m_{i,SA} = 1$ if cell *i* has the self-sufficiency in growth signals mutation and 0 otherwise, $p_{s,IA} = p_{s,SG} = (1 - p_s) / 2$, and $p_{t,IA} = p_{t,SG} = (1-p_t) /2$. The proliferation rate increase associated with self-sufficiency of growth signals and insensitivity to antigrowth signals is unknown

and probably varies depending on the microenvironment and mechanism that generates those hallmarks. Both hallmarks increase proliferation. As a first approximation, we assume each raises the proliferation rate half way between the wild-type proliferation rate and the maximum possible proliferation rate (which is simply dividing at every time step, and so can vary with the time scaling of the model), with self-sufficiency in growth signals systematically benefitting from rounding approximations. In short, cells with mutations in the growth phenotypes will be more likely to divide than those without. A cell i divides if it draws a random number from a uniform distribution between 0 and 1 that is less than its probability for division, $p_{i,d}(t)$.

Copy DNA

While copying its DNA, a cell i acquires a mutation in one of its phenotypic characters if it draws a number from a uniform distribution between 0 and 1 that is less than $7\nu_{i,a}$, where 7 is the number of characters that can mutate and $\nu_{i,a}$ is the probability that cell i acquires a new mutation for a particular character per cell division, given by

$$\nu_{i,a}(t) = \begin{cases} \nu m & \text{if } m_{i,GA} \\ \nu & \text{otherwise.} \end{cases}$$

Here, the mutation rate for cell i increases when it has the genomic instability mutation by a factor of m. If a cell gets some new mutation, an integer is drawn between 1 and 7 to determine which trait is mutated. A new mutation is recorded for the cell if it does not already have a mutation for that phenotype. Reversions to wild-type do not occur if the same mutant trait is selected multiple times.

New mutations can be repaired

Repair of new mutations is implicitly implemented by modifying the mutation rate in the Copy DNA submodel. As such, cells with genomic instability have higher chances of retaining new mutations than those without it. In addition, cells with genomic instability skip this timestep and thus exit the cell cycle faster than those without genomic instability.

Finish cell division

Recall that the differentiation stage for a cell i at time t is stored in its state variable $d_i(t)$. Here, we determine the differentiation stage for daughter cells j and k. When a transient cell divides, both daughters become more differentiated, such that $d_j(t) = d_k(t) = d_i(t) + 1$. If the stem cell population size is less than the target stem cell population size s, then both daughter cells remain stem cells, such that $d_j(t) = d_k(t) = 0$. Otherwise, one daughter cell remains a stem cell, $d_j(t) = 0$, and the other differentiates, $d_k(t) = 1$. Cells with the loss of differentiation mutant phenotype, however, never differentiate upon division and $d_j(t) = d_k(t) = d_i(t)$ for all daughters.

The telomere length for a cell i at time t, $l_i(t)$, is decreased by 1 for transient amplifying cells and l_t / l_s for stem cells, which is the ratio of the number of times a transient cell can divide to the number of times a stem cell can divide. Cells with the limitless replicative potential mutation do not decrement their telomeres. All other state parameters are copied unmodified from parent i to daughter cells j and k.

DETERMINING THE NECESSARY AND SUFFICIENT CONDITIONS FOR CANCER PROGRESSION

Every possible combination of mutations was simulated and analyzed for cancer progression. For each condition, all of the cells were initialized to the same mutational state and no new mutations were acquired during the simulation. Data were recorded on the simulations for 30 years or until they progressed to cancer.

CALCULATING THE ENTROPY FOR A NEOPLASM AT CANCER

Entropy, or the Shannon Index (4), was used to summarize the diversity of a tumor and is given by

$$H = -\sum_{i \in P} p_i \log_2 p_i$$

where P is the set of all clones that existed at the end of the simulation and p_i is the proportion of each unique clone i.

Supplemental Figures

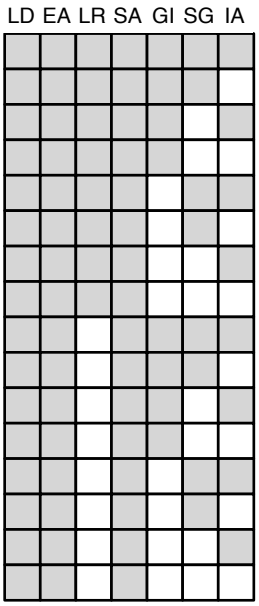

Fig. S1. The necessary and sufficient mutations for cancer are loss of differentiation (LD), evasion of apoptosis (EA), and sustained angiogens is (SA). All of the cells in a simulation were initialized with all possible mutations, and 10 simulations were run per combination. Rows represent the combinations of mutations that resulted in cancer in all 10 cases, and columns are the possible mutations. A gray mark for a mutation indicates that all cells in the simulations were initialized with that mutation. No other combination of simulations progressed to cancer. Abbreviations are as given in Table S1.

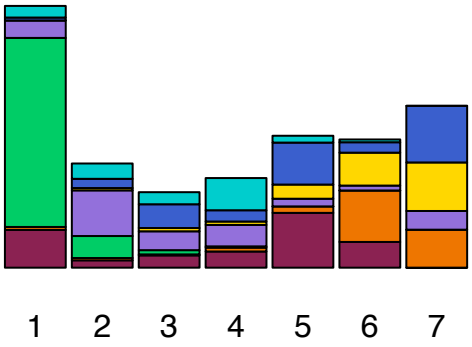

Fig. S2. Sequence logo plot (6) of the frequency at which mutations occur in various mutational positions. The height of the bars represent the information content in bits of the position, where higher bars correspond to larger departures from randomness. The frequency that a particular mutation was observed at any given position is proportional to its height within a bar. The first mutation is non-random and tends to be the loss of differentiation. Each mutation is represented by a different color as given in Fig. 3B.

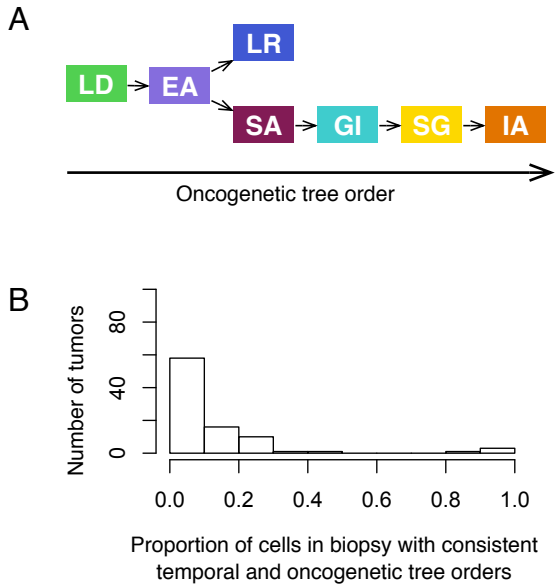

Fig. S3. The oncogenetic tree model for the simulations that progressed to cancer. (A) The maximum likelihood tree to fit the observed sets of mutations begins with the loss of differentiation followed by the evasion of apoptosis. The limitless replicative potential and sustained angiogenesis are independent events with approximately equal probability, and so are represented by a branch in the tree. Each mutation is represented by a different color given in Fig. 3B. (B) The proportion of cells in a simulation with temporal orders consistent with this oncogenetic tree model still tends to be low (mean = 11.5%, s.e.m. = 2.2%, n=90).

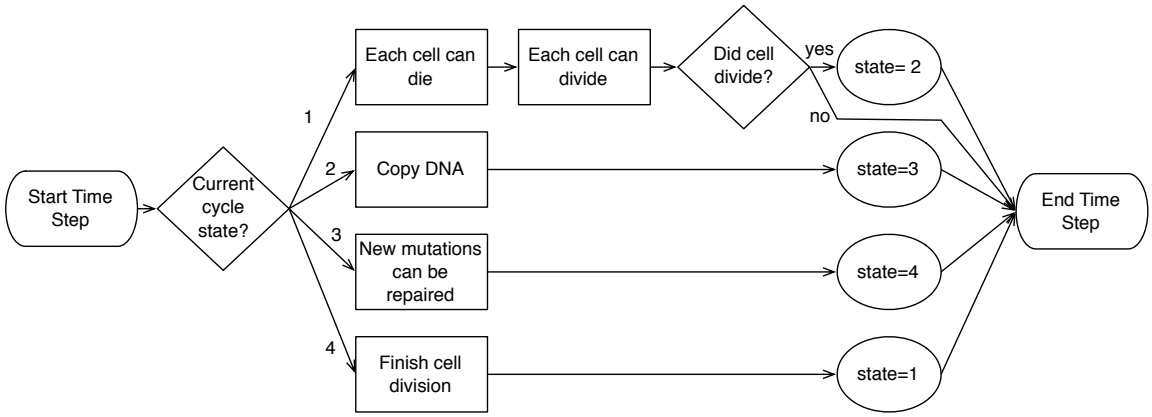

Fig. S4. The decision process that each cell follows at each time step during simulations. We limited our modeling to cell cycle processes necessary for cell division and death, genomic duplication, and mutation repair. In short, the current cycle state of a cell defines its next step. If a cell is in the resting stage (cycle state = 1), then it can die or divide. If a cell is in the S / G1 phase (cycle state = 2), then it copies its DNA. A cell in the G2 phase (cycle state = 3) can repair new mutations to its genome. A cell in the M phase (cycle state = 4) physically divides into two daughter cells.

Supplemental Tables

Table S1. Cancer phenotypes included in the computational model.

Abbrev	Mutant Phenotype	Phenotype
IA	Insensitivity to antigrowth signals	Mutation confers ability to divide in the presence of anti-growth signals
SG	Self-sufficiency in growth signals	Mutation confers ability to divide in absence of growth-signals
EA	Evasion of apoptosis	Mutation reduces a cell's response to apoptosis signals
LR	Limitless replicative potential	Mutation eliminates telomere loss during cell division
SA	Sustained angiogenesis	Mutation increases the carrying capacity of the tissue
LD	Loss of differentiation	Mutation removes a cell's ability to differentiate
GI	Genome instability	Mutation increases the probability that a cell will acquire a new mutation

Table S2. All cancer-outcome simulated neoplasms and their characteristics.

Time to cancer (years)	Entropy at cancer	Percent match with cross-sectional path	Largest clone		Total number of clones	Number of clones at cancer
			Temporal order	Percent		
0.2	0.10	0.00%	SA.LD.EA	99.10%	63	56
0.2	0.82	21.48%	LD.SA.EA	78.00%	92	77
0.2	0.47	0.00%	SA.LD.EA	92.07%	85	71
0.3	0.64	0.00%	LD.GI.SA.EA	94.06%	586	553
0.3	1.61	0.00%	LD.GI.SA.EA	76.09%	566	537
0.3	0.68	13.82%	LD.SA.EA	85.41%	82	74
0.3	0.68	5.26%	LD.GI.SA.EA.LR	91.83%	432	389
0.4	0.78	0.00%	SA.LD.EA	86.29%	81	68
0.4	0.40	0.00%	SA.LD.GI.SG.IA.EA	96.37%	310	280
0.4	0.58	1.17%	LD.SA.EA	91.08%	102	94
0.4	1.52	0.00%	LD.GI.SA.LR.EA	61.65%	395	337
0.4	0.04	0.00%	SA.LD.EA	99.72%	76	68
0.5	0.15	0.00%	LD.SA.EA	98.50%	94	82
0.6	0.99	0.03%	LD.LR.GI.SA.EA	87.16%	503	433
0.6	0.33	0.00%	SA.LD.EA	95.08%	64	58
0.7	3.55	0.00%	GI.SA.LD.EA	53.47%	456	418
0.7	0.08	0.00%	SA.EA.LD	99.32%	79	69
0.7	0.28	0.00%	SA.EA.LD	96.83%	71	67
0.7	2.50	0.00%	SA.LD.GI.EA	48.16%	469	452
0.7	0.99	0.00%	GI.SA.LD.EA.LR	88.32%	409	371
0.8	0.05	0.00%	SA.LD.LR.EA	99.64%	71	48

0.9	0.17	0.00%	SA.LD.SG.EA	98.33%	72	50
1.0	0.67	0.00%	SA.LD.LR.EA	87.68%	94	76
1.1	0.86	0.00%	SA.LD.EA.LR	76.36%	50	38
1.3	1.81	0.02%	LD.LR.GI.SA.EA	75.42%	383	350
1.4	2.23	3.19%	LD.GI.SA.EA.LR	63.40%	437	392
1.9	0.17	0.00%	EA.LR.SA.LD	98.29%	67	56
2.5	4.62	4.38%	LD.GI.SA.EA.LR	31.16%	638	446
4.2	0.80	0.00%	GI.LD.SG.IA.LR.SA.EA	90.45%	309	205
5.2	2.78	21.47%	LD.EA.SA.LR	31.00%	155	90
5.7	2.74	11.84%	LD.GI.EA.SA.SG	59.83%	447	402
5.9	2.74	4.51%	LD.GI.SA.EA.LR.SG	34.34%	439	332
6.0	1.17	28.46%	LD.SA.EA	69.20%	456	217
7.6	2.71	3.70%	LD.GI.SA.EA.LR.SG	34.63%	420	347
7.9	1.47	5.42%	LD.GI.SA.EA.LR	80.19%	441	370
8.6	2.54	0.22%	LD.EA.SA.GI.LR.IA	31.66%	390	314
9.4	1.34	2.12%	LD.GI.EA.LR.SA.SG	72.80%	413	301
9.6	1.96	24.25%	LD.GI.EA.SA.LR	56.70%	354	293
11.3	0.54	3.76%	LD.EA.GI.SA.SG.LR	94.61%	312	268
11.7	1.49	0.00%	IA.LD.GI.EA.SA.SG.LR	75.17%	309	181
11.8	2.18	25.73%	LD.EA.LR.GI.SA	57.40%	440	361
12.1	0.70	1.86%	LD.EA.SA.LR	87.26%	138	94
12.4	1.49	2.47%	LD.GI.EA.SA.LR	70.01%	510	406
12.4	0.74	2.05%	LD.EA.GI.SA.LR	92.70%	417	359
12.5	0.69	3.66%	LD.GI.EA.LR.SA	93.12%	452	390
13.4	2.68	14.22%	LD.EA.GI.SA.LR	45.46%	373	337

13.7	3.31	0.00%	GI.LD.EA.SG.LR.SA	42.41%	1029	557
13.8	3.02	2.13%	LD.LR.EA.GI.SA.IA	37.28%	295	239
14.7	1.43	9.11%	LD.EA.GI.SA.LR	81.33%	476	423
15.0	3.45	0.00%	LD.LR.SG.EA	23.77%	376	278
15.7	2.24	25.61%	LD.EA.GI.LR.SA	62.07%	437	341
15.7	0.21	1.87%	LD.EA.SA.LR	97.63%	129	74
16.8	2.34	18.97%	LD.EA.GI.SA.LR	38.58%	340	270
16.9	2.03	24.42%	LD.EA.GI.SA.LR	66.59%	388	312
17.0	1.49	24.81%	LD.EA.GI.SA.LR	71.17%	375	318
18.5	0.80	0.00%	EA.LD.IA.GI.SA.LR	90.78%	314	270
18.5	2.15	6.66%	LD.EA.LR.GI.SG.SA	63.75%	508	409
18.9	3.33	21.20%	LD.EA.LR.GI.SA	24.26%	425	345
19.3	2.37	16.40%	LD.EA.LR.GI.SA.IA	58.64%	335	277
19.4	2.97	0.00%	LD.LR.EA.GI.SA.IA	37.76%	411	339
20.8	2.72	0.00%	EA.LD.LR.GI.SA.IA	38.46%	321	256
21.7	1.87	22.01%	LD.EA.LR.GI.SA.SG	60.64%	357	280
21.8	2.82	18.26%	LD.EA.LR.GI.SA.IA	44.28%	386	305
21.9	3.05	18.24%	LD.EA.SA.GI.LR	40.92%	389	292
22.1	1.55	0.00%	LR.IA.LD.EA.SA.GI	68.88%	404	309
22.3	1.39	0.00%	LD.SG.EA.GI.SA.IA.LR	70.44%	190	125
23.9	1.90	17.61%	LD.EA.LR.GI.SA.SG	59.00%	311	261
23.9	2.51	0.00%	EA.LD.SA.LR	33.87%	339	212
24.0	2.92	16.93%	LD.EA.LR.GI.SA.SG	37.38%	354	277
24.9	2.32	0.01%	LD.EA.SG.GI.SA.IA	65.16%	366	266
25.2	2.53	21.85%	LD.EA.LR.GI.SA.IA	45.33%	351	276

25.3	1.55	18.02%	LD.EA.LR.GI.SA.IA	72.94%	330	262
26.1	2.16	6.95%	LD.EA.LR.GI.SA.IA	64.42%	438	334
26.3	2.54	0.00%	EA.LD.LR.GI.SA.SG	42.58%	350	256
26.6	2.56	19.55%	LD.EA.LR.GI.SA.IA	56.86%	333	248
26.9	2.30	9.71%	LD.EA.LR.GI.SA.IA	65.50%	397	314
27.2	2.22	18.72%	LD.EA.LR.GI.SA.IA	57.28%	354	271
27.2	0.48	0.00%	LD.EA.GI.IA.SG.SA.LR	94.45%	243	126
27.3	2.41	10.42%	LD.EA.LR.GI.SA.SG	64.67%	331	256
27.6	2.08	0.00%	LD.EA.GI.IA.LR.SA	62.55%	353	243
27.7	2.19	0.00%	LD.IA.LR.EA.GI.SA.SG	59.89%	142	61
27.8	2.08	10.39%	LD.EA.LR.GI.SG.SA	65.23%	351	276
28.2	1.82	18.44%	LD.EA.LR.GI.IA.SA.NU	65.48%	228	152
28.3	2.58	12.24%	LD.EA.LR.GI.SA.IA.NU	37.51%	250	171
28.9	2.95	0.46%	LD.LR.EA.SA.GI.IA	39.45%	371	285
28.9	1.97	19.37%	LD.EA.LR.GI.SA.SG.IA	56.70%	280	186
29.0	2.22	0.00%	LD.SG.EA.GI.SA.LR	54.46%	303	206
29.1	2.52	0.00%	LD.LR.EA.GI.SA.IA	54.80%	384	290
29.1	2.24	0.00%	EA.LD.LR.GI.SG.SA	59.16%	363	254
29.3	2.85	41.31%	LD.EA.LR.SA.GI.IA	44.04%	385	282

Table S3. Results for the parameter sweeps. One thousand simulations were run per condition under the default conditions defined in Table S5 and $\nu = 5.0 \times 10^{-4}$, unless otherwise indicated.

Parameters		Temporal and path match		Entropy		Largest clone		
Value	**Path Order**	**Mean**	**Std error**	**Mean**	**Std err**	**Mean**	**Std err**	**Cancer**
Number of divisions transient cell								
50	LD.EA.LR.GI.SA.SG.IA	8.40%	1.00%	4.2	0.2	41%	2%	9.5%
60	LD.EA.LR.GI.SA.SG.IA	7.50%	1.00%	4	0.1	41%	2%	8.5%
70	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
80	LD.EA.LR.GI.SG.IA.SA	4.30%	0.50%	4	0.2	42%	2%	10.8%
Number divisions somatic stem cell								
70	LD.EA.LR.GI.SA.SG.IA	6.60%	0.70%	3.8	0.2	41%	2%	9.6%
100	LD.EA.LR.GI.SA.SG.IA	7.50%	1.30%	4.3	0.2	44%	2%	11.2%
140	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
Angiogenic tissue support rate								
1.1	LD.EA.LR.GI.SA.SG.IA	2.70%	0.40%	4.1	0.2	43%	3%	5.4%
1.4	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
1.7	LD.EA.LR.GI.IA.SG.SA	7.00%	0.70%	4.3	0.1	36%	2%	15.3%
Survival support rate								
$5x10^2$	LD.EA.LR.GI.SA.SG.IA	4.00%	0.70%	4.1	0.2	41%	3%	6.8%
$5x10^4$	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
$5x10^6$	LD.EA.LR.GI.IA.SA.SG	5.50%	0.60%	4.4	0.1	38%	2%	10.1%
Default maximum number of cells								
2044	LD.EA.LR.GI.SG.IA.SA	5.00%	0.60%	3.6	0.1	47%	2%	12.4%
4088	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%

8176	LD.EA.LR.GI.SG.IA.SA	6.50%	0.80%	4.3	0.2	41%	2%	8.8%
Probability for stem cell division								
1/24	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
1/36	LD/EA.LR.GI.IA.SG.SA	4.10%	0.40%	4	0.1	40%	2%	11.1%
1/48	LD/EA.LR.GI.SA.SG.IA	5.20%	0.70%	4.2	0.1	40%	2%	9.5%
Number of differentiation stages								
4	LD.EA.LR.SG.IA.GI.SA	3.80%	0.60%	4	0.2	42%	2%	9.8%
5	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
6	LD/EA.LR.GI.IA.SA/SG	6.40%	0.50%	4	0.1	41%	2%	11.1%
Number of stem cells per crypt								
4	LD.EA.LR.GI.SA.IA/SG	4.30%	0.80%	4.4	0.2	39%	3%	5.5%
8	LD.EA.LR.GI.SG.SA.IA	4.80%	0.70%	3.7	0.2	48%	3%	8.2%
16	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
Mutation rate								
$1x10^{-6}$	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0%*
$1x10^{-5}$	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0%*
$1x10^{-4}$	LD.EA.LR.SA.GI.SG.IA	7.26%	1.01%	1.7	0.1	67%	2%	0.9%*
$5x10^{-4}$	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
Evasion of apoptosis mutation effect on probability for apoptosis								
2	LD.EA.GI.LR.IA/SG.SA	8.93%	0.70%	7.37	0.04	11.4%	0.4%	99.8%
10	LD.EA.LR.GI.SA.IA.SG	5.60%	0.70%	4.2	0.2	40%	2%	11.9%
100	LD.EA.SA.GI.LR.SG.IA	4.43%	2.39%	3.4	0.5	52%	8%	1.2%

*10,000 simulations were run.

Table S4. State variables for each cell in the model.

Parameter	**Description**
Differentiation stage [$d_i(t)$]	Differentiation stage of a cell *i* at time *t*, characterizes if cell *i* is a somatic stem cell or any stage of transient cell
Telomere length [$l_i(t)$]	Acts a counter to determine how many cell divisions remain for cell *i* at time *t*
Insensitivity to antigrowth signals mutation [$m_{i\text{-}IA}(t)$]	Presence of mutation in cell *i* at time *t* confers its ability to divide in the presence of antigrowth signals
Self-sufficiency in growth signals mutation [$m_{i\text{-}SG}(t)$]	Presence of mutation in cell *i* at time *t* confers its ability to divide in absence of growth signals
Evasion of apoptosis mutation [$m_{i\text{-}EA}(t)$	Presence of mutation in cell *i* at time *t* reduces its response to apoptosis signals
Limitless replicative potential mutation [$m_{i\text{-}LR}(t)$]	Presence of mutation in cell *i* at time *t* eliminates its telomere loss during cell division
Sustained angiogenesis mutation [$m_{i\text{-}SA}(t)$]	Presence of mutation in cell *i* at time *t* increases the carrying capacity of the tissue
Loss of differentiation mutation [$m_{i\text{-}LD}(t)$]	Presence of mutation in cell *i* at time *t* removes its ability to differentiate signals
Genetic instability mutation [$m_{i\text{-}GI}(t)$]	Presence of mutation in cell *i* at time *t* increases the probability that it will acquire a new mutation

Table S5. Parameter values

Parameter	**Description**	**Biological value**	**Model value (default value)**
Number divisions transient cell [l_t]	Number of time a transient cell can divide	50 - 80 (5)	50 - 80 (70)
Number divisions somatic stem cell [l_s]	Number of times a somatic stem cell can divide	70 - 145 (6, 7)	70 - 140 (140)
Angiogenic tissue support rate [a]	Additional number of cells supported in neoplasm supported per angiogenic cell	1.1 - 1.7 cells per angiogenic cell (8)	1.1 - 1.7 (1.4)
Survival support rate [r_{ss}]	Rate at which excess cells in the crypt are expected to survive	Unknown	$5.0\text{x}10^{-2}$ - $5.0\text{x}10^{-6}$ ($5.0\text{x}10^{-4}$)
Default maximum number of cells [n_{max}]	Number of cells supported by nutrients available to a crypt	Model-specific (2x target crypt size)	2044 - 8176 (4088)
Probability for stem cell division [p_s]	Probability that a stem cell divides	0.04167 - 0.020833 divisions per hour (9)	0.04167 - 0.020833 (0.020833)
Probability for transient cell division [p_t]	Probability that a transient cell divides	0.05882 divisions per hour (9)	0.05882 (0.05882)
Number of differentiation stages [d]	Number of differentiation, or transient, stages possible	4 - 6 (10, 11)	4-6 (5)
Number of stem cells per crypt [s]	Target number of stem cells per crypt	4 - 16 (10); 4-12 (11)	4 - 16 (16)
Mutation rate [v]	The chance of a cell acquiring a mutation each cell division	$1.6\text{x}10^{-8}$ - $1\text{x}10^{-3}$ mutations / cell division (12-15)	$5.0\text{x}10^{-4}$ - $1.0\text{x}10^{-6}$ ($1.0\text{x}10^{-4}$ mutations / cell

			division)
Mutator fold change [m]	The fold change in mutation rate that occurs when a cell acquires the genomic instability mutation	10 x greater (16)	10 x greater
Probability for apoptosis [p_a]	Probability that a given cell in a crypt undergoes apoptosis	0.0000452 dead / hour (17)	4.52×10^{-5} deaths / hour
EA mutation effect on probability for apoptosis [α]	The factor by which the EA mutation reduces the probability for apoptosis	Unknown	2 - 100 x lower (10 x lower)
Target population size of crypt [N]	Number of cells that typically exist in a crypt	2044 cells (9)	2044 cells

Supplemental References

1. Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, et al. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* 2006;**198**:115-126.
2. Wilensky U. NetLogo. http://ccl.northwestern.edu/netlogo/ *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL* 1999.
3. Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. *Genetics* 1964;**49**:725-38.
4. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal* 1948;**27**.
5. Shay JW, Wright WE. Hallmarks of telomeres in ageing research. *J Pathol* 2007;**211**:114-23.
6. Shepherd BE, Guttorp P, Lansdorp PM, Abkowitz JL. Estimating human hematopoietic stem cell kinetics using granulocyte telomere lengths. *Exp Hematol* 2004;**32**:1040-50.
7. Lansdorp PM. Telomere length and proliferation potential of hematopoietic stem cells. *J Cell Sci* 1995;**108 (Pt 1)**:1-6.
8. Deli G, Jin C-H, Mu R, Yang S, Liang Y, Chen D, et al. Immunohistochemical assessment of angiogenesis in hepatocellular carcinoma and surrounding cirrhotic liver tissues. *World J Gastroenterol* 2005;**11**:960-3.
9. Potten CS, Kellett M, Roberts SA, Rew DA, Wilson GD. Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut* 1992;**33**:71-8.
10. Potten C, Loeffler M. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development* 1990;**110**:1001--1020.
11. Nicolas P, Kim K-M, Shibata D, Tavaré S. The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Comput Biol* 2007;**3**:e28.
12. Araten DJ, Golde DW, Zhang RH, Thaler HT, Gargiulo L, Notaro R, et al. A quantitative measurement of the human somatic mutation rate. *Cancer Res* 2005;**65**:8111-7.
13. Bielas JH, Loeb LA. Quantification of random genomic mutations. *Nat Methods* 2005;**2**:285-90.
14. Siegmund KD, Marjoram P, Woo Y-J, Tavaré S, Shibata D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc Natl Acad Sci USA* 2009;**106**:4828-33.

15. Ushijima T, Watanabe N, Shimizu K, Miyamoto K, Sugimura T, Kaneda A. Decreased fidelity in replicating CpG methylation patterns in cancer cells. *Cancer Res* 2005;**65**:11-7.

16. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* 1993;**75**:1027-38.

17. Hall PA, Coates PJ, Ansari B, Hopwood D. Regulation of cell number in the mammalian gastrointestinal tract: the importance of apoptosis. *J Cell Sci* 1994;**107**:3569-77.