CancerSim: A Computer-Based Simulation of Hanahan and Weinberg's Hallmarks Of Cancer

by

Robert Abbott

B.S., Brigham Young University, 1998

THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Computer Science

The University of New Mexico

Albuquerque, New Mexico

April, 2002

CancerSim: A Computer-Based Simulation of Hanahan and Weinberg's Hallmarks Of Cancer

by

Robert Abbott

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Computer Science

The University of New Mexico

Albuquerque, New Mexico

April, 2002

CancerSim: A Computer-Based Simulation of Hanahan and Weinberg's Hallmarks Of Cancer

by

Robert Abbott

B.S., Brigham Young University, 1998

M.S., Computer Science, University of New Mexico, 2002

Abstract

Hanahan and Weinberg's important paper "The Hallmarks of Cancer" is a distillation of the immense body of cancer research into a small number of underlying principles. The paper poses six cellular alterations as essential to malignant growth. Cancerous cells defy the normal limits of mitosis, ignore growth-inhibition signals, escape dependence on external growth stimulation, cause extra vascularization, become mobile and invasive, and disable the safety mechanisms that normally detect mutation and trigger apoptosis. Genetic instability is an additional factor which may be necessary to account for the high incidence of cancer in its various forms.

The minimal model of cancer proffered by the Hallmarks paper is realized in CancerSim, a computer-based simulation. By implementing the Hallmarks, we seek

to determine whether the complex phenomena of cancer can indeed arise from just a handful basic principles.

To the extent that cancer is successfully simulated, we can examine the dynamics of a developing tumor and alter its progression by tuning the model parameters. CancerSim has both command-line and graphical user interfaces support both interactive investigation and off-line statistical analysis of the model.

Contents

List of Figures

List of Tables

Chapter 1

Cancer

Each year, there are 6 million deaths from cancer, and 10 million new cases diagnosed. Over the next 25 years, as deaths due to infection decrease, the number of diagnosis of cancer is expected to double to 20 million[6]. Over the last 30 years, the billions spent on cancer research and treatment have lead to better care overall, but breakthroughs only in the case of a few types of cancer.

1.1 Hanahan and Weinberg's Hallmarks of Cancer

Hanahan and Weinberg's important paper "The Hallmarks of Cancer" poses six cellular alterations as essential to malignant growth. These six hallmarks are believed to be common to most if not all human tumors. Genome instability is an additional factor which may be necessary to explain the high incidence of cancer.

1.2 CancerSim

To explore Hanahan and Weinberg's hallmarks of cancer, a computer-based model was created. The program is named "CancerSim." It allows the user to explore the phenomenon of Cancer by simulating and visualizing cell growth and genetic mutation. Users can see how the hallmarks interact, and affect the simulation by changing the model parameters. The affect of each mutation can be studied, as can the order in which the mutations are most likely to take effect. CancerSim may help assign estimates of relative importance to the hallmarks, providing focus for further research.

CancerSim does not simulate every nuance of cancer in its various forms. Instead, the goal is to determine whether the hallmarks posited by Hanahan and Weinberg can account for the appearance and behavior of cancer. Cellular mechanisms are simulated only in abstracted forms. The "genotype" of a cell is simply a boolean vector whose values enable or disable certain behaviors. Table 1.1 shows the information maintained for each cell. This, of course, is a vast simplification of reality. Studies based on actual detection rates place the number of mutations in cancerous cells at more than 10,000, and mathematical models suggest that the actual number of mutations is probably 10^{12} or more[8]. However, the vast majority of these probably have no effect on the progression of the malignant growth.

Generally, it is also not feasible to simulate realistic parameter values. In CancerSim, simulations of more than a million cells or so become painfully slow. This falls far short of the human body, with its approximately $5 * 10^{13}$ cells. Compromising on this parameter requires adjustment to all the others. Use of a realistic mutation rate, perhaps $5 * 10^{-9}$ per nucleotide per generation[8], would prevent tumor development because of the "small" number of cells simulated. Therefore, only mutation rates down to about 10^{-6} are of interest in CancerSim. Hopefully, scaling the parameters

Name	Datatype	Notes
Genetic Instability	boolean	
Ignore Growth Inhibit	boolean	
Evade Apoptosis	boolean	
Limitless Replication	boolean	
Sustained Angiogenesis	boolean	
Self Growth	boolean	
Telomere Length	integer	Replicative potential of this cell
Mutation Pathway	Genotype_History	Records the order in which mutations were obtained
Capillary	boolean	Whether there is a capillary at the cell's location
Nutrient	real-valued	Concentration of oxygen etc. at the cell's location
Sequence Number	integer	Tag to distinguish between cells reusing the same grid location

Table 1.1: CancerSim Cell State

in this way does not qualitatively change the dynamics of the system.

Chapter 2

The Hallmarks In CancerSim

CancerSim models cellular growth on a three-dimensional Cartesian grid. Each grid location may be empty or occupied by either a cell, a capillary, or both. The number of cells simulated is typically on the order of 1 million. The progression of the simulation is driven by discrete instantaneous events. Each event represents an action taken by a cell, such as mitosis.

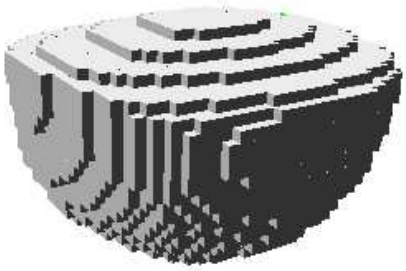

Figure 2.1: Healthy Tissue

The simulation begins with the creation of a single cell having no mutations. The

first event scheduled is the initial mitosis of this cell. When simulated, this mitosis prompts the scheduling of mitosis for both daughter cells, and so on. The simulation concludes when all the cells die (from telomere shortening), when cancer develops, or after an arbitrary number of timesteps. The criteria for "cancer" is the occupation of 90% of the computational grid by living cells. It is not possible for normal cells to reach this limit.

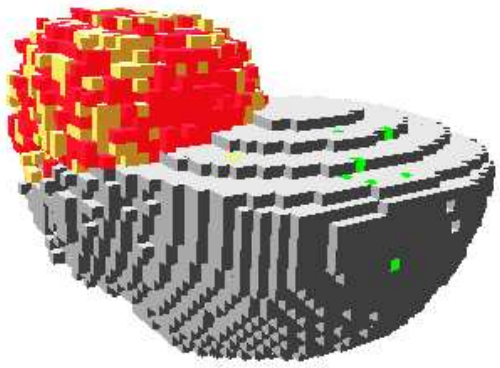

Figure 2.2: Cancerous Tissue

To detail the behavior of CancerSim, we examine each of the hallmarks of cancer along with its manifestation in the simulation.

2.1 Self-Sufficiency in Growth Signals

Normal cells proliferate only when stimulated by growth signals from other cells. The signal chemicals are bound by transmembrane receptors, prompting cell growth and division. In cancerous cells, these receptors are over-expressed, causing hyper-sensitivity to otherwise ambient levels of growth signal.

In CancerSim, cells undergo mitosis only if they are within a predefined boundary.

This boundary represents a threshold in the concentration of growth factor, beyond which growth signals are too faint to prompt mitosis. Cells may escape the tissue's natural extents through the "Growth Signal" mutation. This mutation causes cells to proliferate regardless of the concentration of growth factor.

2.2 Insensitivity to Growth-Inhibitor Signals

Normal cells eventually enter a post-mitotic, differentiated state. By ignoring anti-growth signals, cancer cells forgo full maturity and normal functioning, thus maintaining replicative potential.

CancerSim models contact inhibition, which is a type of growth inhibition. Contact inhibition prevents overcrowding by arresting the growth and division of cells already in contact with many other cells. In the three-dimensional grid space of CancerSim, each cell has up to 26 neighbors (to be considered to be neighbors, cells need only make contact at a corner). Cells forgo mitosis if all of the neighboring grid locations are already occupied by living cells. The "Ignore Growth Inhibit" mutations allows cells to divide even when there is no empty location for the new daughter cell. In this case, the daughter cell competes for survival with a randomly chosen neighbor with a $1/g$ likelihood of success, where g is a tunable parameter.

2.3 Evasion of Apoptosis

Cells have mechanisms to effect apoptosis (self-destruction). Normal cells express receptors for apoptosis signals. Such signals are sent by intra-cellular sensors which detect DNA damage, signaling imbalance, survival factor insufficiency, or hypoxia. Oncogene over-expression is among the factors that trigger early apoptosis. Tumor

cells of lack some of the surface features present on normal cells, and express some features not normally present on mature cells[7]. The immune system detects and responds to these cells as non-self. In this way, the body is protected from malfunctioning cells. To become cancerous, cells must disable either the effectors of mitosis or the receptors that trigger them. Over 50% of human cancers contain a mutated, nonfunctional P53 tumor suppressor gene. This prevents manufacture of P53 protein - an necessary component in some pathways of DNA damage detection. (P53 mutation is not the only way to evade apoptosis triggered by DNA damage, but is probably the most important.)

Before undergoing mitosis, each cell in CancerSim is checked for genetic damage. If mutation is detected, the cell undergoes apoptosis instead of mitosis. The probability of detection of damage in the cell is n/e, where n is the number of mutations carried by the cell and e is a tunable parameter. Thus, the probability of apoptosis due to genetic damage is proportional to the number of mutations. The "Evade Apoptosis" mutation causes cells to ignore apoptosis signals. This removes the selective disadvantage of genetic damage.

2.4 Limitless Replicative Potential

Cells normally become senescent after a fixed number of divisions. In culture, various human cells are limited to 60-70 divisions, though most die before exhausting this potential. This limitation precludes the formation of macroscopic tumors. Experiments show, however, that cells occasionally acquire limitless replicative potential. Such cells are said to be "immortal."

The mechanism of limited replicative potential lies in the inability of DNA polymerase to completely replicate chromosomes. The ends of chromosomes are called "telomeres." With the duplication of each chromosome, 50-100 base pairs at the end

of each telomere are lost. Eventually, the DNA of the chromosome is exposed. The unprotected DNA fuses together. Once the cell's genetic material is in this deformed configuration, the death of the cell is imminent.

In cancerous cells, the telomeres are extended to compensate for what is lost by DNA polymerase. Usually this is accomplished by overproducing telomerase enzyme, which appends hexanucleotide molecules to the telomeres. Alternately, telomeres may be lengthened by homologous recombination. This is usually prevented by mismatch repair machinery, a defect in which is one form of genetic instability[1].

Telomere shortening is simulated by CancerSim. The initial cell is created with telomeres of a specified length. With each mitosis, the telomeres are shortened by one unit length. When the telomere length falls to 0, the cell dies. The "Limitless Replicative Potential" mutation stops telomere shortening. Cells with this mutation can undergo mitosis any number of times.

2.5 Sustained Angiogenesis

Cells cannot survive at distances of more than about 100um from blood supply. Despite this fact, cells are normally unable to induce new capillary growth. Without such growth, tumors can only grow to about 0.5mm[2]. The importance of angiogenesis to tumorigenesis has been studied since the early 1970's and has resulted in clinical trials of several angiogenesis inhibitors. Because cancerous cells are genetically unstable, they are highly adaptive and quickly develop resistance to chemotherapy. The stromal and endothelial elements of the tumor are not in themselves malignant, and may be more susceptive to sustained treatment[4]. Such treatments alone only suppresses (rather than eliminating) tumors. Another problem is that angiogenesis is not specific to tumors; it is a normal part of homeostasis[2].

Angiogenesis is regulated by a variety of signals; over two dozen inducers (primarily bFGF and vascular endothelial growth factor, or VEGF) and a similar number of inhibitors are known. Some of the inducers are not specific to endothelial cells; they promote growth in other cells as well[2]. Cancerous cells may over-produce inducers, may lack inhibitors, or both. The "angiogenic switch" is the threshold in the ratio of inducers to inhibitors sufficient to trigger angiogenesis.

In CancerSim, angiogenesis is inhibited outside of a predefined area. This area overlaps with the area in which growth factor is concentrated enough to prompt mitosis, but neither area is completely within the other. Vasculature outside this area can be developed only in response to signals from cells with the "Sustained Angiogenesis" mutation. Cells signal for angiogenesis only when the distance to the nearest capillary is too far. This is a simple model of hypoxia, "the most potent external stimulus of angiogenic factor"[2].

2.6 Tissue Invasion and Metastasis

Metastasis - the spreading of cancer in the body - is the final stage of cancerous development. 90% of cancer fatalities are due to growths other than the original tumor. Localized cancer has a limited impact, while metastatic cancer takes root wherever possible within the body.

To circulate through the body, cells must decouple themselves from surrounding tissue. Cell adhesion molecules and integrins are disrupted by mutation, transcriptional repression, or proteolysis. Once freed from their moorings, cancer cells escape the influence of regulatory signals. Cancer cells also increase protease production, perhaps to facilitate the destruction inherent in the invasion of tissue.

CancerSim does not simulate metastasis, because only a localized area is simu-

lated.

2.7 Genome Instability

Genetic transcription is a very accurate process. It seems unlikely that all the mutations preconditioning cancer should occur during a human lifetime, yet cancer is a common occurrence. Cytotoxic carcinogens cause genetic instability by altering the cellular environment. This selects cells for the ability to quickly adapt, which requires genetic instability[1]. In an environment absent of carcinogens, cells are already well adapted. Since most mutations are disadvantageous, genetic stability is beneficial.

There are various sensing and repair enzymes in the body to detect and repair different kinds genetic damage. This indicates the possibility for various specific types of genetic instability[1].

In CancerSim, genetic instability is a single mutation which causes other genes to be copied with increased probability of error. For such cells, the base mutation rate is scaled by a factor of i, which is a tunable parameter.

2.8 Summary Of CancerSim Settings

Each setting can be specified either by command-line parameter switches (Table 2.8) or from the 'Settings' dialog box. Switches are used in this way:

```
GUI -n1000000 -r123 -t30
```

Setting Name	Switch	Description
Grid Size	-n	The computational grid contains n cell locations
Mutation Rate	-m	Each gene is copied with a $\frac{1}{m}$ chance of mutation
Random Apoptosis	-a	Each cell cycle exposes every cell to a $\frac{1}{a}$ chance of death from non-specific causes
Telomere Length	-t	Telomere shortening limits the initial cell to t replications
Evade Apoptosis	-e	A cell with n mutations has an extra $\frac{n}{e}$ likelihood of dying each cell cycle, unless one of its mutations is "Evade Apoptosis"
Genetic Instability	-i	Increase the likelihood of mutation by a factor of i for cells with this mutation
Ignore Growth Inhibit	-g	Cells with this mutation have a $\frac{1}{g}$ chance of killing off neighbor to make room for mitosis
Random Seed	-r	Seed value for the pseudorandom number generator

Table 2.1: CancerSim Settings

Chapter 3

Simulation Issues

3.1 Event Model

CancerSim's computational model of space is a three-dimensional Cartesian grid. This grid divides space into cubes, each defining a cell location. The simulation can be thought of as visiting each cube at each timestep, updating, creating, and destroying cells. In practice, this approach would be inefficient; most cubes are not occupied by cells, especially in the simulations' early stages. For instance, with a grid size of 125,000 cubes, no more than about 16,000 can be occupied until several genetic mutations have occurred. And even where there are living cells, they do not usually need to be updated at every timestep.

For this reason, CancerSim proceeds in an alternate fashion. Each future event is represented by a data structure. The set of all future events is stored in a priority queue, ordered on event time. This way, the grid need not be traversed. However, the priority queue of events must be maintained, consuming some memory and requiring some extra processing.

3.2 Random Number Generation

Random number generation is an important issue in CancerSim. In fact, almost all the parameters to the model are to influence the frequency of random events. In the simulation, all random behavior is driven by a random number generator, which produces a sequence of random numbers. In the same way that dice influence the progression of a board game, these random numbers are interpreted by CancerSim to determine, for instance, whether genetic mutation will occur during a given cellular mitosis.

The key fact of computer generated random numbers is that they are actually not random at all. This extends from the deterministic nature of computers themselves; they are designed to carry out instructions in a predictable way. Hardware dedicated to the random number generation exists, based on random physical processes such as resistance noise[5] and radioactive decay. However, such hardware is not generally available. Worse, providing random input to a program renders its results non-repeatable.

For these reasons, uncertain events in CancerSim are driven by a *pseudo*-random number generator, or "PRNG", as is common in computer simulation. With each application, the PRNG returns a different, apparently random number. But because the internal state of the PRNG is finite in size, it has only a finite set of configurations. Once some state is entered for the second time, the PRNG is in a cycle and will generate the same sequence of "random" numbers indefinitely. The length of the cycle is the "period" of the PRNG. A good PRNG must have a long cycle.

CancerSim's PRNG is adapted from the GNU standard C library, which was in turn adapted from BSD (Berkeley Systems Distribution) UNIX. The formula used in this PRNG is $x_n = x_{n-1}^{31} + x_{n-1}^{3} + 1$, and x is a 112 bit number. This permits a maximum period of ($2^{112} \approx 10^{33}$), which is not of practical concern, but the

actual period depends on the start state, and is almost sure to be much shorter than this maximum. The least significant bit, in particular, has a very short period ($2^7 - 1 = 127$), but entropy from lower-order bits propagates to higher-order bits. Of the 112 internal bits, only the 31 most significant are output in each pseudo-random number.

The start state of a PRNG is created from a "seed" value provided by the user. For each seed, the PRNG generates a different sequence of numbers. The user must be allowed to specify this value to make runs variable, yet repeatable. In CancerSim, the user influences the random number seed through the parameter r. Rather than directly specifying the random seed, r is combined with the other starting parameters using a hash function to determine the random seed. This way, pseudo-random number sequences are re-used only between runs with identical parameter sets.

The key random events in CancerSim are genetic mutations. The most straightforward way to test for such an event is to generate a random number in the interval $[0, 1)$ and call it a "success" if the random number is less than p, the probability of the event. Genetic mutations, however, are very rare. Instead of generating a random number each time the unlikely event *might* occur, we instead generate a number which represents the number of failures before the next success. This value initializes a counter, which is decremented each time the event might occur. When the count falls to 0, the event occurs, and the number of trials until the next success is recalculated. The number of trials between each success in such an event is given by the geometric distribution. Knuth[5] gives a very convenient formula for generating a number from the geometric distribution. Here, the rand function is expected to return a number in the interval $[0, 1)$ from the uniform distribution.

$$n = \lceil \frac{log(rand())}{log(1-\frac{1}{p})} \rceil$$

3.3 Command Line Interface

CancerSim offers two user interfaces: graphical (GUI) and command-line (CLI).

The command-line interface is for non-interactive use. It is normally run with a set of parameters to influence the simulation, and its output is typically redirected to a file. At intervals of 100 timesteps, information about the state of the simulation is output for later analysis. It is either run from a controlling program or using a shell, i.e.:

```
CLI -n27000 -i100 -g10 -t100 -m1000000 -e20 -a1000 -r1 > output
```

3.4 Graphical User Interface

The graphical user interface (Figure 3.1) is for interactive use. The GUI can be run from a command-line allowing starting parameters to be specified, just as with the CLI. This is a convenient way to initialize the simulation when the parameters are known in advance.

3.4.1 Commands

The left portion of the CancerSim GUI window is a column of command buttons.

The "Run" button is pressed to begin the simulation and again to pause it.

The "Single Step" button takes the simulation through one timestep. This does not imply that only one event is processed; generally, many events occur on each timestep.

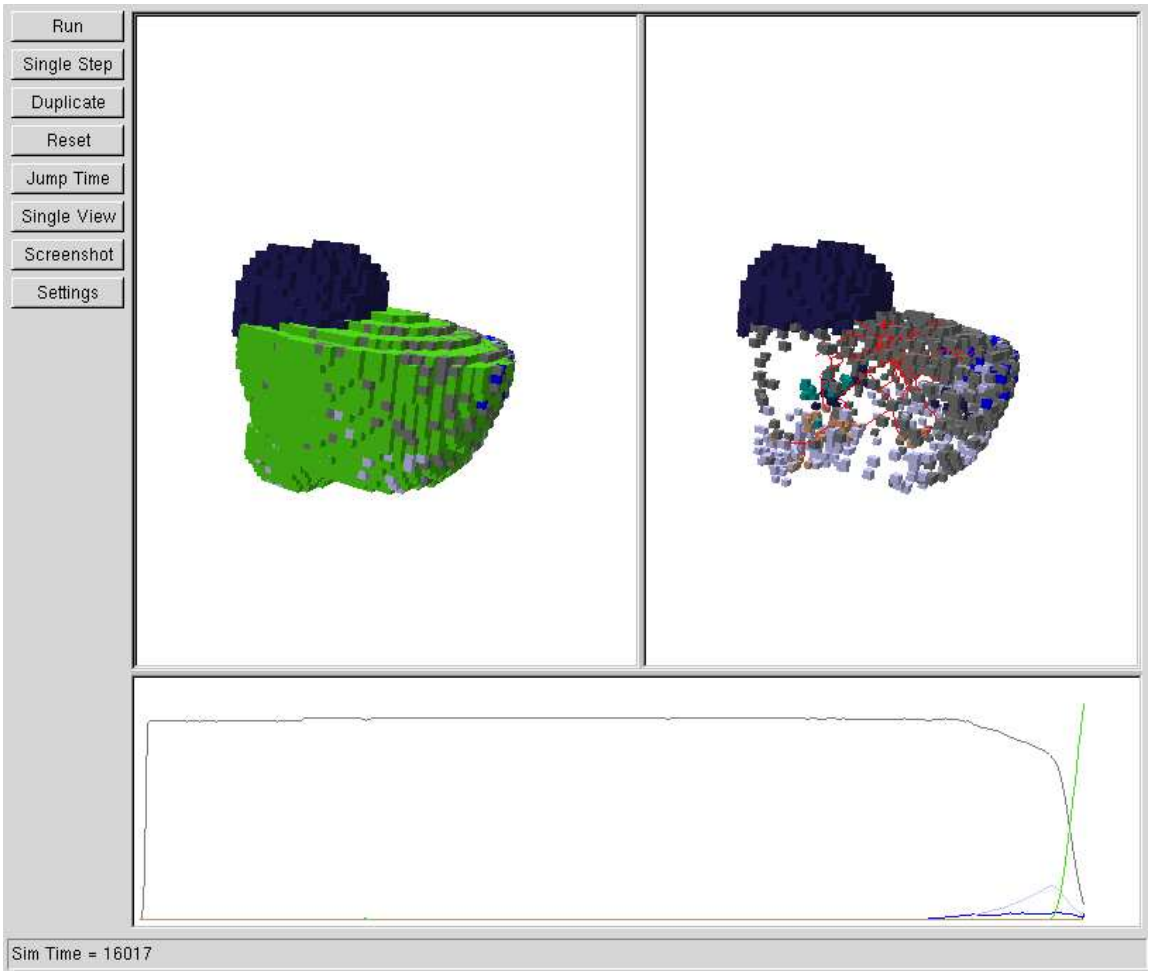

Figure 3.1: CancerSim GUI

The "Duplicate" button creates a new CancerSim window. The simulation state of the new window is identical to the original, but proceeds forward independently. This functionality is used either to create a snapshot of a simulation at a certain timestep or to test the effect of a parameter by altering it in a duplicate window and then continuing the simulation.

The "Reset" button resets CancerSim to its starting configuration of a single cell. This button brings up a small dialog box requiring confirmation, since resetting the

simulation discards its previous state. In this dialog box is the grid size specification. The maximum number of cells, n, is specified here. The CancerSim grid is always cubic. The side length of the cube is $\lfloor n^{\frac{1}{3}} \rfloor$, so the n specified is reduced to the next lowest perfect cube if necessary.

The "Jump Time" button brings up a dialog for the specification of a new sim time. If the user confirms this action, the simulation is "jumped" (without the normal 3-D view updates) to the specified time, just as if the user had run the simulation until the specified time and pressed the "Stop" button. If the new sim time is in the past, the result is as if the user had stopped the simulation at that time.

The "Dual View" button divides the 3-D view into two panes. Both panes show the tissue from the same angle. However, the two views may have different cell visibilities. (See below for more information on cell visibility).

The "Screenshot" button allows the user to save an image of the 3-D view in Microsoft's "Bitmap" (.bmp) format. This format is simple and widely supported in imaging and word processing software.

The "Settings" button summons a dialog for the editing of all simulation parameters (except grid size; see "Reset," above). These parameters are defined near the beginning of this chapter. The initial values of these parameters can be specified from the command line.

3.4.2 3-D View

The 3-D view initially shows just a dot representing the first cell in the tissue. As the simulation is run, the 3-D view is updated every 100 timesteps, or whenever the simulation is paused.

In the 3-D view, each cell is rendered as a small cube. The cube is colored

according to the genotype of the cell it represents. The genotype of any cell can be displayed by left-clicking on the cell. This reports the cell's genotype in the upper-lefthand corner of the 3-D view, and outlines the cell in yellow. The mutations listed for the cell are listed in the same order as they were obtained by the cell's ancestors.

To view the tissue from a different angle, the user may left-click on a cell, then move the mouse while still holding the left button down.

To see inside the tissue, or simply to declutter the display, the user can hide all cells of any genotype by right-clicking on a cell of that genotype. When cells are hidden, the vasculature of the tissue becomes visible (as branching structure drawn in red). To make all cells visible, right-click anywhere but on a cell.

3.4.3 History View

The 2-D graph below the 3-D view shows information about the history of the simulation. It plots the number of cells of each genotype against time. In this view, cell populations are not differentiated by the order in which their ancestors acquired mutations. As time passes (by running the simulation) the graph becomes more compressed.

When the user left-clicks within the History View, the coordinate on the curve nearest the mouse cursor is displayed, e.g: "$t = 87, n = 1501$." Here, t is the sim time, and n is the number of cells of that genotype living at that time. When a point of interest is located on this curve, the simulation reverted using the "Jump Time" button (see above).

3.4.4 CancerSim Review

Running CancerSim to completion can take a significant amount of time, depending mainly on the grid size and the speed of tumor progression. This makes it difficult and time-consuming to interactively compare the results of several runs, or to quickly locate a set of parameters that will generate interesting behavior.

To address this issue, a coarse parameter sweep of CancerSim has been performed (comprising 4374 runs and over 1.3 *trillion* cell events in total). Data from the output of these runs can be quickly visualized using the CancerSim Review GUI (Figure 3.2). The tradeoff of using this ready-made data is that only a few settings of each parameter are available. The available settings are shown in Table 3.1.

The graph displayed by CancerSim review displays 3 curves. Each curve corresponds to a run of CancerSim. The 3 runs vary only by random seed; all other parameters are identical. Each curve shows how the number of cells varied over time during the course of a run. If the curve reaches the top of the graph, the run ended with cancer.

The CancerSim Review Settings Window selects the information to be displayed on the graph. First is a drop-down box designating the type of cell to be counted. By default, all cells are counted. Any of the mutations can be selected, creating a graph of the prevalence of that mutation. The radio buttons in the Settings Window select runs with various parameters. At the bottom of the settings window is a button entitled "CancerSim." Pressing this button launches the CancerSim GUI preloaded with the parameters from CancerSim Review. The user can then use the GUI to visualize the run summarized by CancerSim Review. Naturally, the parameters may be fine-tuned (or changed completely) once the GUI is launched; in this case the interactive simulation will not correspond to the plot in CancerSim Review.

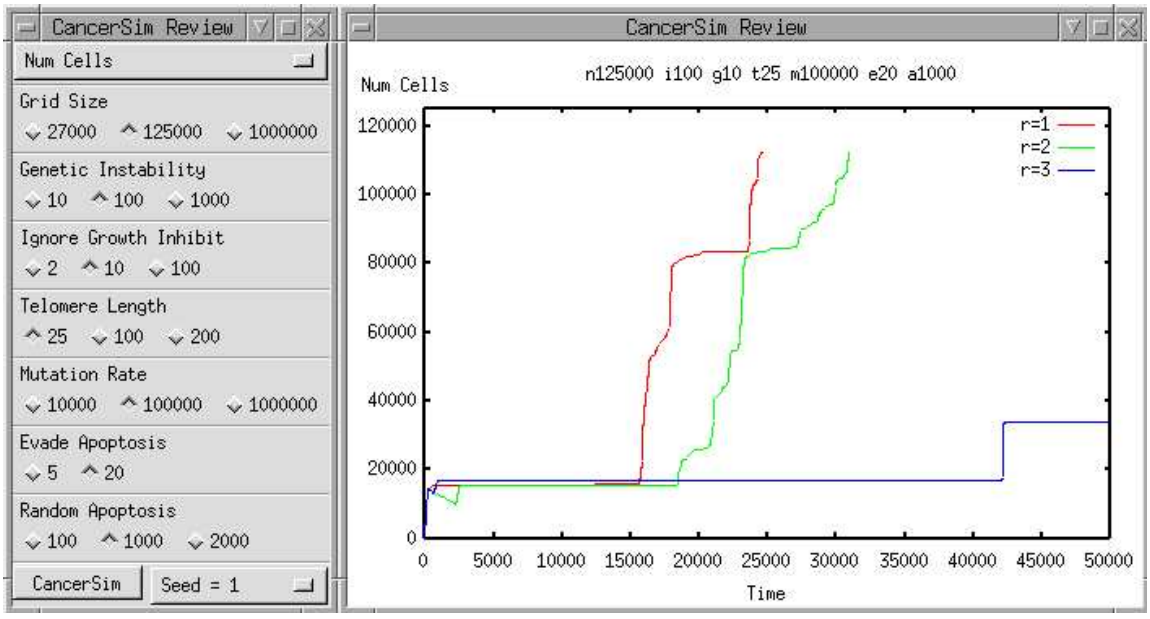

Figure 3.2: CancerSim Review GUI

Setting	Values
Grid Size	27000, 125000, 1000000
Mutation Rate	10^4, 10^5, 10^6
Random Seed	1, 2, 3
Random Apoptosis	100, 1000, 2000
Telomere Length	25, 100, 200
Evade Apoptosis	5, 20
Genetic Instability	10, 10^2, 10^3
Ignore Growth Inhibit	2, 10, 100

Table 3.1: CancerSim Review GUI Parameter Settings

Chapter 4

Results

The primary "result" of CancerSim is the degree to which its simple implementation of Hanahan and Weinberg's Hallmarks of Cancer realistically models cancer. This assessment can be made only by medical experts. It is bound to be subjective, and will depend on the user's interests. If nothing else, CancerSim hopefully provides a frame of reference for more precise discussion of the Hallmarks and a useful starting point for higher fidelity simulations of specific phenomena.

Additionally, the behavior of CancerSim itself can be studied. However, the significance of these results hinges on the larger question of whether CancerSim resembles reality at all.

In the spirit of optimism, but with these significant disclaimers never out of mind, we tender some analysis of CancerSim's output.

For each parameter, CancerSim Review was used to locate a region of interest. The parameter was stepped through this region, varying only the parameter and the random seed. These single-parameter sweeps are intended to indicate the parameter's general effect, though in fact the parameter might have a different effect with another

set of fixed parameters. Finally, 100 runs were performed varying only the random seed in order to examine the stability of model. Using this data, studies of the "pathways to cancer" and "longevity" were performed.

4.1 Pathways To Cancer

An issue raised by the Hallmarks paper is the sequence of the changes leading to cancer. The authors believe that "virtually all cancers must acquire the same six hallmark capabilities [3]." Does this imply that the progression from normalcy to cancer is inflexible and predictable? Not according to Hanahan and Weinberg. Instead, the hallmark behaviors may have varied causes, and be obtained in varying orders. CancerSim records the history of the genotype for each cell. This is simply the order in which the cell's mutations were acquired by its ancestors, or its genetic "pathway." Only the current genotype of the cell affects its behavior, but the pathway is recorded for the purposes of this analysis.

In CancerSim, as in vivo, the tissue is practically never homogeneous by the time of death; it contains cells with various genotypes. Moreover, cells with identical genotypes may have obtained them by different pathways. In fact, even if the tissue consists *only* of cells with all 6 of the simulated hallmarks, up to $6! = 720$ different pathways may be observed. In such a tissue, no cell has any selective advantage, so the variety of pathways is unlikely to diminish no matter how long the simulation is run. Finally, cells having taken the same pathway are *probably* but *not necessarily* related (other than through the original cell); it may be coincidence that two distinct cell lines receive mutations in the same sequence.

We call a pathway "predominant" if it is shared by more than 50% of the cells in the tissue at the time of death. By this definition, a tissue might not have any predominant pathway. Figure 4.1 shows how the probability of finishing a simulation

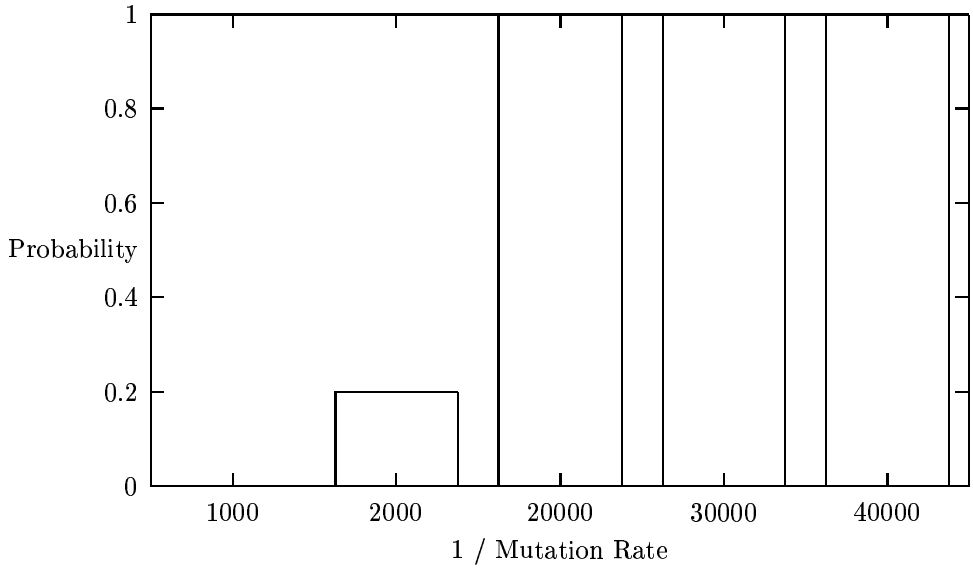

Figure 4.1: Likelihood Of Predominant Pathway Vs. Mutation Rate

with any predominant genotype history varies according to the base mutation rate, taking the average over all other parameters. Only runs ending with cancer are factored in.

When the mutation rate is low, the simulation proceeds in a rather orderly manner. First, some mutation occurs. If the mutation conveys a selective advantage, descendants of the mutant cell multiply until the selective advantage is fully exploited. If the cell has all the necessary mutations, this means filling all available space, causing the simulation to end. Otherwise, the simulation proceeds until the next mutation. Thus, low mutation rates raise the probability of some pathway predominating.

When the mutation rate is very high, mutations are not given time to play out independently. Descendants of a mutant cell are likely either to be hampered by other more degenerate cells, or else develop their own mutations.

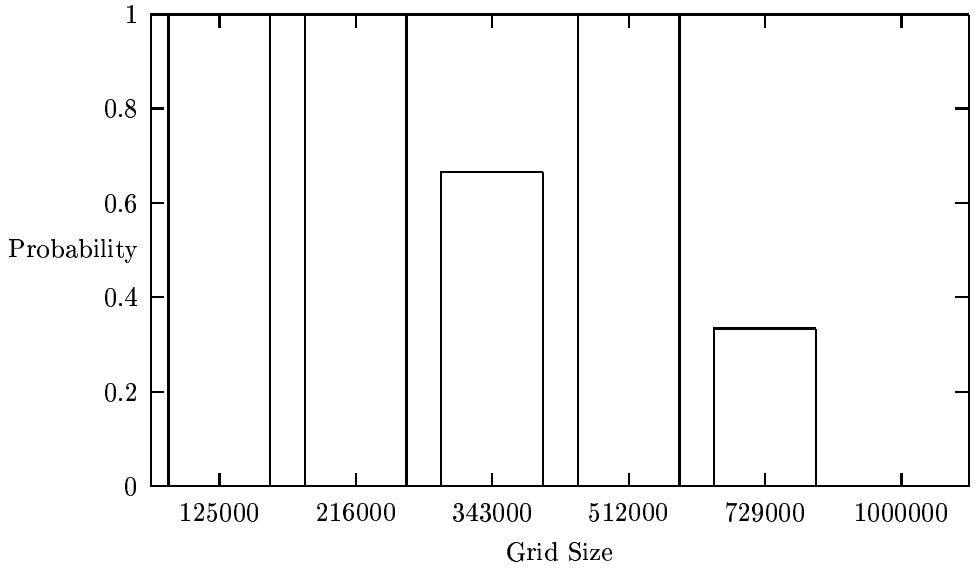

Figure 4.2: Likelihood Of Predominant Pathway Vs. Grid Size

Figure 4.2 indicates that running on a larger computational grid also decreases the likelihood of some pathway predominating. This may be because it takes longer for each advantageous mutation to sweep the larger tissue, which allows more time for additional mutations. Thus the effect is similar to the case in which the grid is smaller, but the mutation rate is higher.

Having examined the likelihood that some pathway will predominate, we can ask which pathways are the most likely to do so. In the random seed parameter sweep, cancer arose 96 of 100 runs. Of those 96 runs, 90 completed with some predominant pathway. Of these, all but one were permutations of the same 4 mutations: limitless replication, evasion of apoptosis, ignore growth inhibit, and self growth. The single exception had genetic instability as well. In all, only 7 different pathways predominated, with the top 2 accounting for over 50% of the runs. Clearly, in CancerSim, all pathways to cancer are not equally likely.

Surprisingly, the Sustained Angiogenesis mutation appears in only one pathway.

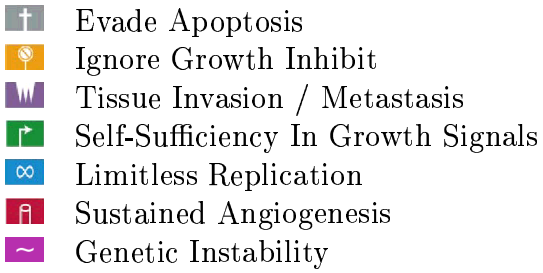

Figure 4.3: Pathways of Cancer Legend

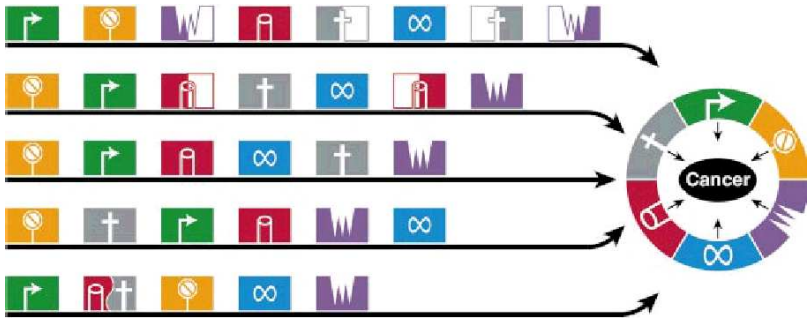

Figure 4.4: Pathways of Cancer (Reproduced From Hanahan [3])

Figure 4.5: Pathways of CancerSim. The number of occurrences of each pathway is indicated.

This does *not* mean that Sustained Angiogenesis was unimportant in these runs, only that the predominant genotype at the end of the run did not have this mutation.

This is probably inaccurate.

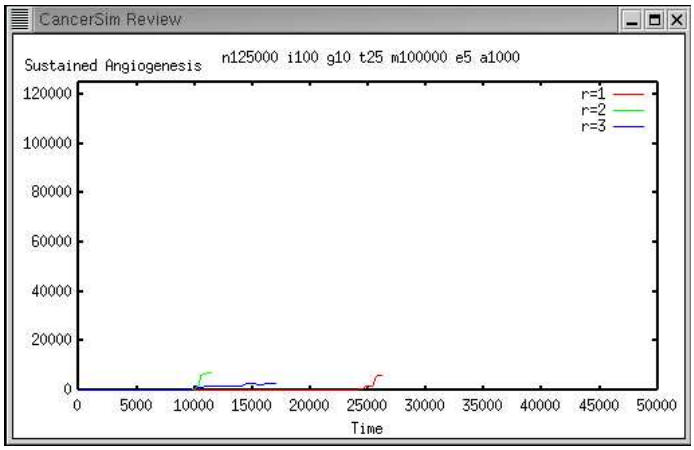

Figure 4.6: Consistent Low-Level Angiogenesis

Figure 4.6 shows the population of angiogenic cells during a typical simulation that resulted in cancer. In each case, there is a significant number of angiogenic cells. However, the number of such cells is small, and not part of the predominant pathway. Still, these cells cause the angiogenesis necessary to support the growth of the tumor. Ultimately, "Cause Angiogenesis" is a rare mutation because its effects can be shared.

Evade Apoptosis, on the other hand, appears in every pathway. However, it never occurs first in the sequence. This makes sense, since it conveys no selective advantage until the cell has at least one other mutation.

Only two pathways involve genetic instability. This does not agree with studies indicating that genetic instability is an important hallmark of cancer, and that it appears early in carcinogenesis[1]. It seems that in CancerSim, genetic instability is is a very weakly selective mutation.

Limitless Replication usually occurs first. This is a bit surprising since this data was collected from the random seed parameter sweep, which used an initial telomere

length of 100, and 2^{100} is about 10^{30} - an enormous number of cells. This means that the *average* cell is under little pressure to escape apoptosis due to telomere shortening. However, to benefit from highly selective mutations and sweep the tissue, precancerous cells must be able to replicate many times.

The "Ignore Growth Inhibit" mutation is consistently near the end of the pathway. It may be that other mutations are more selective until most of the empty space (initially inaccessible due to insufficient growth factor and vascularization) is occupied.

4.2 Longevity

We define longevity as the amount of sim time taken for the tissue to develop cancer, or the arbitrary cutoff of 50,000 time steps. If the tissue dies before the maximum number of time steps (due to telomere shortening), the longevity is simply the maximum number of time steps; that is, only cancer decreases longevity. Each data point in the longevity plot represents the average longevity of the tissue with a fixed set of parameters (other than random seed) over some number of trials. The number of trials is indicated on the plot.

4.2.1 Grid Size

In CancerSim, the natural extent of the tissue (as defined by growth factor concentration and vascularization) is proportional to the grid size. The larger the grid, the more cells in the tissue. This seems to generally increase the prevalence of cancer (Figure 4.7). In the model, more cells simply provide more opportunity for mutation and the onset of cancer.

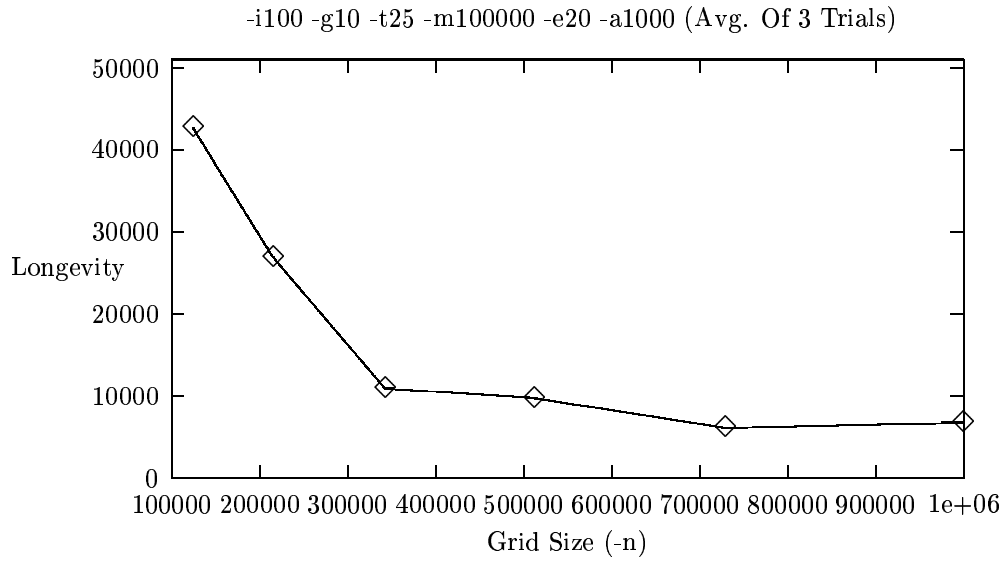

Figure 4.7: Longevity Vs. Grid Size

This leads to a hypothesis about cancer in large vs. small organisms. An occurrence of cancer anywhere in the body will metastasize and invade other tissue. Ultimately, it will kill the organism. In effect, each cell in a multicellular organism puts the others at risk. But because cancer is not contagious, the risk is limited to the individual. Thus, a given number of cells will be more resistant to cancer when divided among a larger number of individuals. The joint liability of multicellularity seems to favor smaller organisms and exert more evolutionary pressure on large ones. For this reason, organisms with more cells (or more cellular turnover) probably have more highly evolved cancer defense mechanisms.

One early observation of cancer research was the relatively high resistance of human cells to carcinogenesis when compared to rodent cells. Subsequent work has elucidated the relative sophistication of the human mechanisms of cell growth, programmed apoptosis, and differentiated function; all of these counteract oncogenes[9]. It seems likely that this relation holds generally for large animals relative to smaller

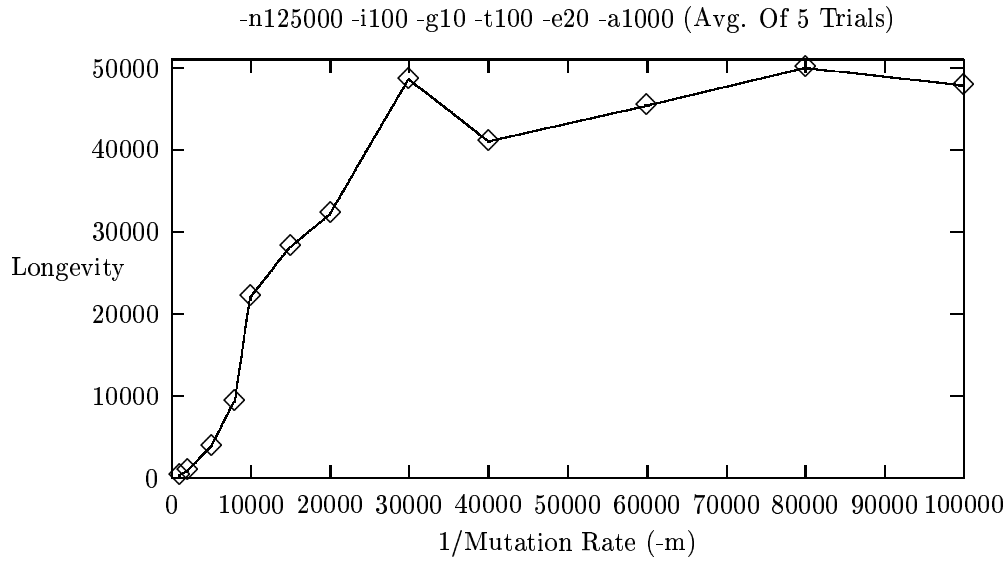

Figure 4.8: Longevity Vs. Mutation Rate

ones.

4.2.2 Mutation Rate

CancerSim is very sensitive to the mutation rate parameter (Figure 4.8). This confirms all indications that a mutagenic environment is likely to cause cancer.

4.2.3 Genetic Instability

Genetic instability has some effect, though it is not very pronounced (Figure 4.9). Considering the dramatic effect of the base mutation rate (see Figure 4.8, this deserves examination.

Unlike the other mutations, genetic instability is of no direct benefit to the cell because it does not increase the likelihood of survival or of reproduction. It does make

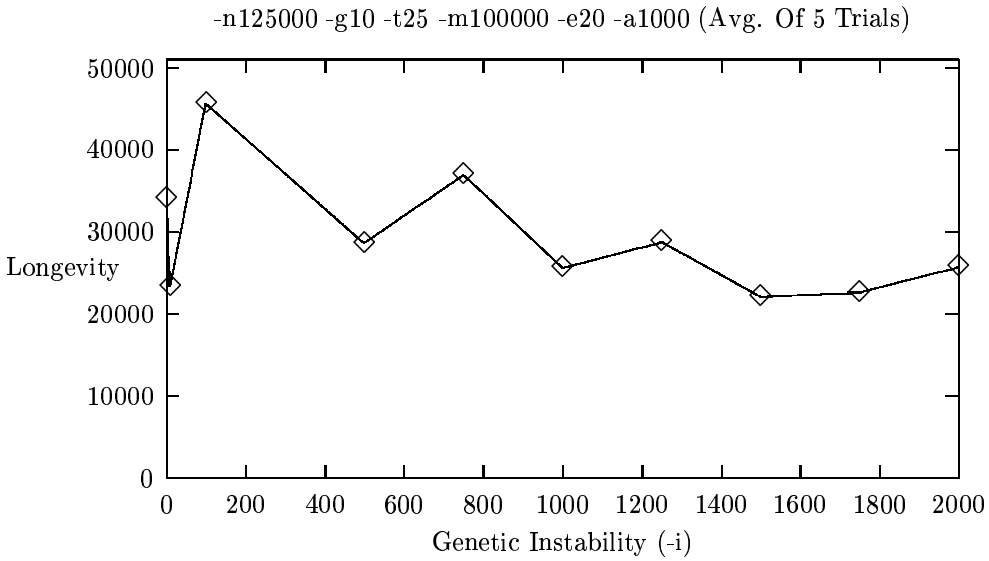

Figure 4.9: Longevity Vs. Genetic Instability

future mutations more likely, but apparently this is only enough to decrease longevity slightly. This model of genetic instability is called the "associated-selection model" [1]. One alternative, the "direct-selection model," suggests that genetic instability must itself provide some growth advantage.

Other models of genetic instability would have a much larger effect. To simulate environmental effects, all cells might obtain genetic instability together. This would be equivalent to raising the base mutation rate. This might better model the effect of carcinogens, whose effect is never localized to just a single cell.

4.2.4 Random Apoptosis

In CancerSim, the only effect of cell aging is telomere shortening, which is an artifact of mitosis. Under this model, a cell that never reproduces (because of contact inhibition) never ages, and will never die. Given sufficient initial telomere length,

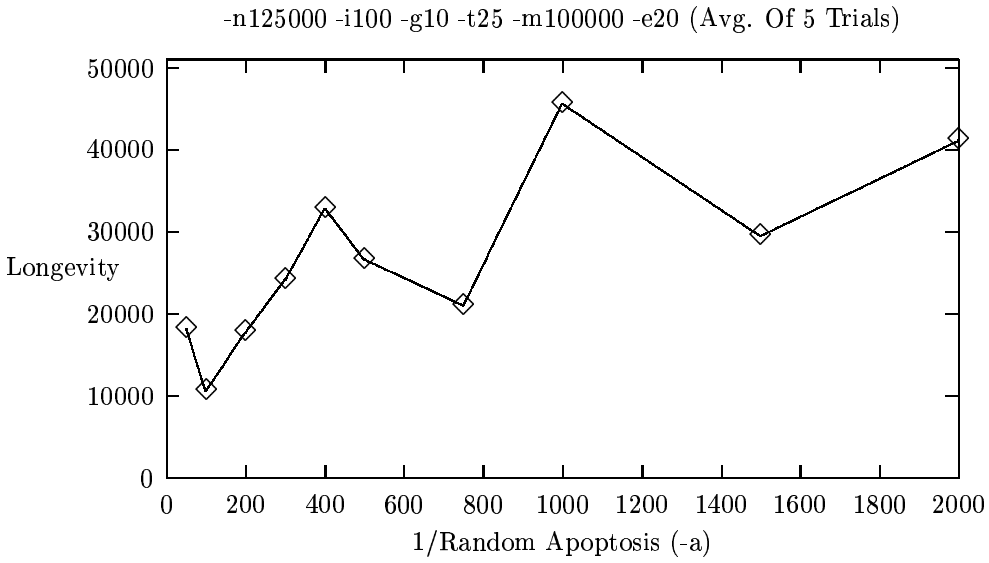

Figure 4.10: Longevity Vs. Random Apoptosis

the tissue quickly fills all available space. Cell-cell contact then inhibits all mitosis, precluding genetic mutation. The simulation is at an impasse.

To ensure the progress of the simulation, an element of random apoptosis is added. Each cell cycle, each cell is subjected to some risk of death. This might be due to mechanical, chemical, or radiological damage, aging, or the immune system.

High rates of random apoptosis promote cancer (Figure 4.10). Increased cell turnover provides more opportunity for mitosis, and therefore for mutation[8].

4.2.5 Telomere Length

The length of telomeres (t) in the initial cell places a limit of 2^t on the size of the cell population. This limit clearly manifests itself for extremely small t. Figure 4.11 shows the total number of cells created during runs with very small values of t.

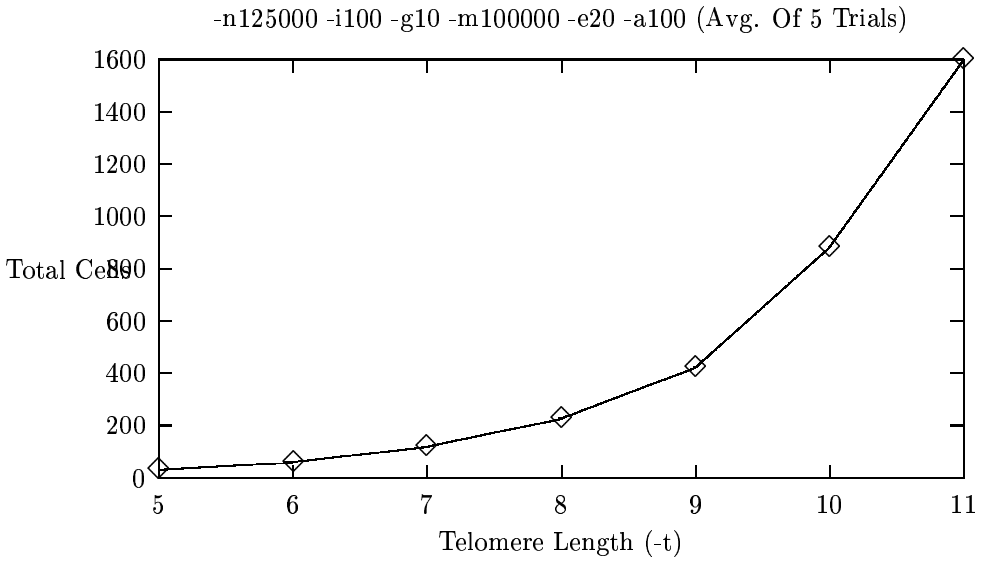

Figure 4.11: Cell Count Vs. (Short) Telomere Length

For any reasonable telomere length, though, 2^t is very large. Typically, at least one cell will acquire the "Limitless Replicative Potential" mutation. Thus, unless t is very small, the number of cells created doesn't noticeably correspond to the theoretical limit.

For moderate values of t (approximately 50 to 150 in Figure 4.12), telomere shortening appears to increase the risk of cancer. Even though there is ample replicative potential to fill available space, the increase in cell turnover due to telomere exhaustion provides more opportunity for mutation.

4.2.6 Evade Apoptosis

The body's efficiency in eliminating defective cells has a significant impact on longevity (Figure 4.13). Mutant cells can sometimes be killed before proliferating. But if the cell is able to make enough redundant copies of itself before being destroyed, it is too

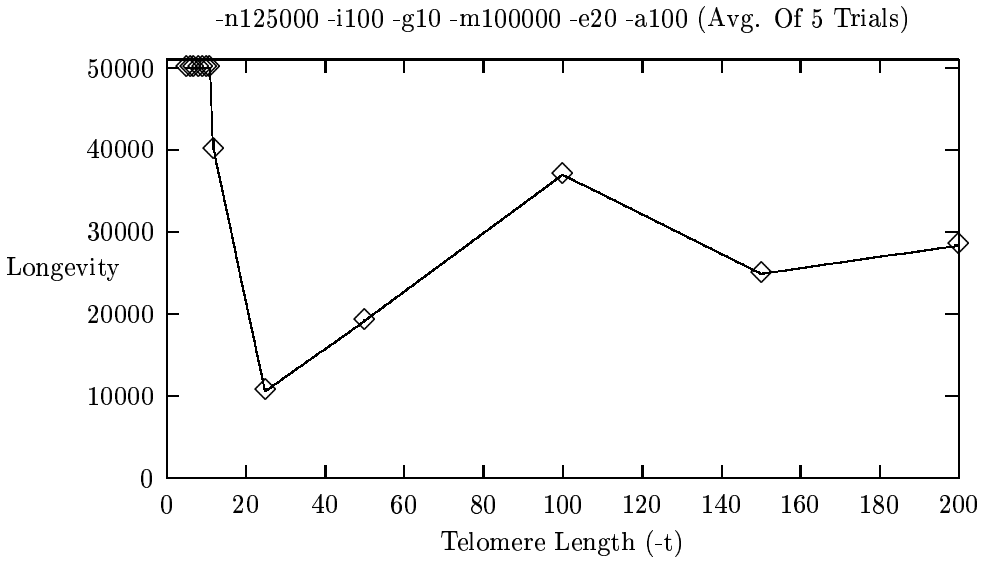

Figure 4.12: Longevity Vs. Telomere Length

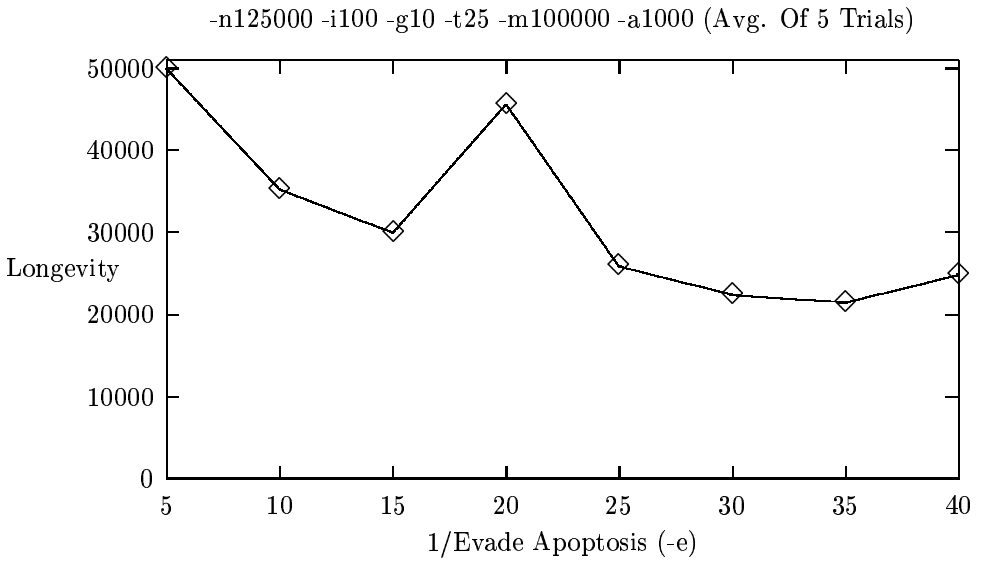

Figure 4.13: Longevity Vs. Apoptosis Evasion

late. Any mutant cells killed are quickly replaced, perhaps with additional mutations.

4.2.7 Ignore Growth Inhibit

In CancerSim, the tissue quickly reaches its natural extent. Without mutation, new cell growth is possible only when apoptosis frees up space. The "cause angiogenesis" and "self-growth" mutations can allow cells on the surface of the tissue to extend out into unused space beyond the natural extent of the tissue. However, most of the cells are interior to the tissue, not on its surface. The "ignore growth inhibit" mutation allows cells inside the tissue to reproduce even when there is no room available (resulting in the death of either the daughter cell or a previously existing neighbor cell). Because this mutation benefits every space-constrained cell, it is strongly selective, and sharply decreases longevity (Figure 4.14). However, it has little effect when the odds of displacing the incumbent cell are less than about 5 percent.

4.3 Observations

At several points during the development of CancerSim, seemingly small alterations to the mechanics of the simulation had large effects on the simulation's dynamics. For instance, the affect of the genetic instability phenotype was greatly magnified by applying it immediately instead of only to future mitoses. In another case, an error in the initialization of a random number generator obscured the effect of telomere shortening on tumorigenesis, even though the numbers generated were not regular in any noticeable way. Unfortunately (for the analyst), the dynamics of cancer probably *are* very nonlinear; in some cases, the difference between life and death might be the survival of a certain precancerous cell. This fact discourages quantitive analysis of

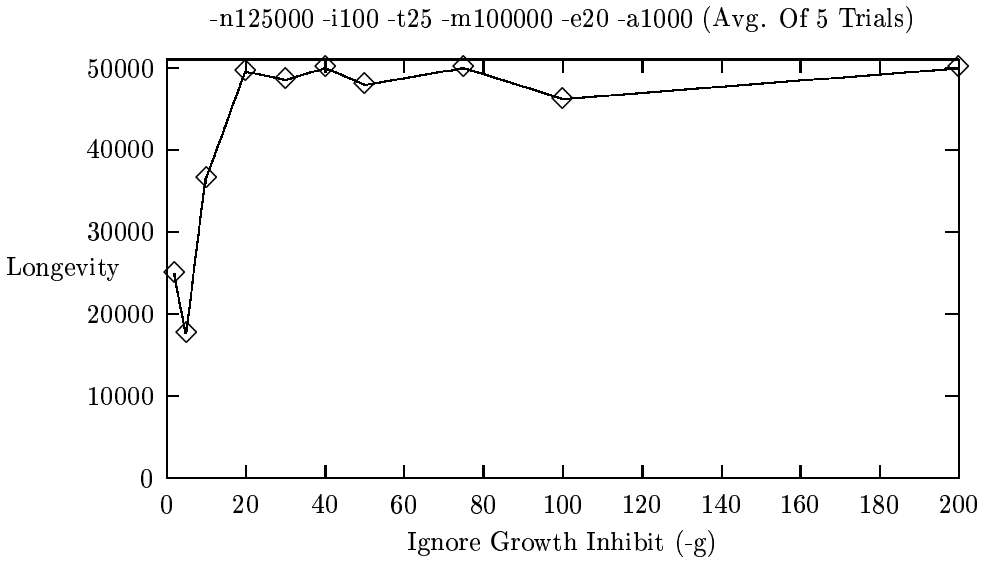

Figure 4.14: Longevity Vs. Ignore Growth Inhibit

cancer through simulation.

However, many interesting dynamics of CancerSim were preserved over various parameters and even modifications to the model.

4.3.1 The Mutation Rate Isn't Everything

The largest determinant of longevity in CancerSim is the number of mutant cells created. Predictably, cancer develops more quickly when the mutation rate his higher. However, the mutation rate is only one factor contributing to the total number of mutations. The size of the organism and the average lifespan of its cells must be factored in, too. These determine how often opportunities for mutation arise. This creates a dilemma in cancer treatment. On hand, the cancerous cells must be killed. On the other hand, trying to kill them may simply cull the weaker or less adaptable cell lines, leaving a more space and sustenance for the elites which remain.

4.3.2 Nice Guys Finish Last

Cancer is impossible wihout angiogenesis, yet the Cause Angiogenesis mutation is often not in the dominant genotype by the end of the simulation. This is because the beneficial effects of angiogenesis are shared with other, non-angiogenic cells. This may not be completely realistic, but it raises the question of why all cells aren't "cancerous," and instead cooperate with cells of different types. The mutations causing cancer are all selective at the cellular level, so cancer is inevitable given enough time. How did cells come to cooperate and differentiate? Could the forces behind the evolution of multicellularity be harnessed and enhanced in order to prevent cancer?

References

[1] Blagosklonny, M.V. "How carcinogens (or telomere dysfunction) induce genetic instability: associated-selection model" FEBS Letters. 2001; 506(3): 169-172

[2] Fidler, I., Kerbel, R., Ellis, L. "Biology of cancer: angiogenesis" In DeVita, V., Hellman, S., Rosenberg, S (eds): Cancer: Principles & Practice of Oncology. Philadelphia, Lippincott-Raven. 2001, pp 137-145

[3] Hanahan, D., Weinberg R.A. "The hallmarks of cancer" Cell. 2000 Jan 7; 100(1): 57-70

[4] Hidalgo, M. "Therapeutic angiogenesis inhibitors in the treatment of cancer" Advances in Internal Medicine. 2001; 47: 159-90

[5] Knuth, D. The Art of Computer Programming Volume 2: Seminumerical Algorithms. Addison-Wesley Pub Co. 1997.

[6] Sikora, K. "Developing a global strategy for cancer" European Journal of Cancer. 1999; 35(14); 1870-1877

[7] Thomson, D., Major, P., Shuster, J., Gold, P. "Tumor immunology" In Samter, M., Talmage, D., Frank, M., Austen, K., Claman, H. (eds): "Immunological Diseases" Boston/Toronto. Little, Brown and Company. 1988. pp 521-549

[8] Tomlinson, I; Sasieni P; Bodmer W. "How many mutations in a cancer?" American Journal of Pathology. 2002 Mar; 160(3): 755-8

[9] Yuspa, S.H. "Overview of carcinogenesis: past, present, and future" Carcinogenesis. 2000 Mar; 21(3); 341-4