Modeling the Role of Neutral and Selective Mutations in Cancer

C. C. Maley[1] and **S. Forrest**[1,2]

[1]Department of Computer Science, UNM, Albuquerque, NM 87131
[2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

Abstract

The transformation of normal cells into cancerous cells is an evolutionary process. Populations of precancerous cells reproduce, mutate, and compete for resources. Some of these mutations eventually lead to cancer. We calculate the probability of developing cancer under a set of simplifying assumptions and then elaborate these calculations, culminating in a simple simulation of the cell dynamics. The agent-based model allows us to examine the interactions of neutral and selective mutations, as well as mutations that raise the mutation rate for the entire cell. The simulations suggest that there must be at least two selectively neutral mutations necessary for the development of cancer and that preventive treatments will be most effective when they increase this number.

Cancer

Cancer is an evolutionary problem. This is the basis for both its virulence and our difficulties in treating it. The dynamics of cancer cells demonstrate the sufficient conditions for natural selection: heritable variation in the population and differential reproduction based on that variation. The variation in the population of precancerous cells (Fujii et al., 1996; Barrett et al., 1999) arises from the normal process of somatic mutations as well as the dramatic rise in mutation rates that is characteristic of the progression to cancer (Paulovich et al., 1997). Differential reproduction of the mutants is accomplished through phenomena such as the subversion of check points in the cell cycles of the mutants (Sherr, 1996). Nowell, 1976, argued for the importance of evolution in cancer more than two decades ago. Any mutations that redirect more of the body's resources to the cancer cells will be selected. This includes the invasion of new tissues and metastasis. The fact that the population of cells includes significant heterogeneity means that most treatments will not eradicate all the cells, leaving some resistant cells. Furthermore, since each patient's cells evolve through an independent set of mutations and selective environments, the resulting population of cancer cells in each patient is likely to be unique. This suggests that general treatments that will work for all, or even most, patients will be difficult to find. The fact that evolution within a tumor works against us in cancer means that not only is cancer an evolutionary problem, but that it will only be solved as an evolutionary problem.

Artificial life provides approaches that are ideal for addressing such evolutionary problems. The field of artificial life has grown up around evolutionary theory (Collins and Jefferson, 1992; Maley, 1998; Levin et al., 1997), and for good reason. When we try to represent heterogeneous populations of individuals interacting in a spatially structured environment, it is difficult to represent and analyze such systems with tractable mathematics. Computational models can help to extend analytical theory to the dynamics of systems with heterogeneous populations that are interacting and evolving. Computational models can help to test the simplifications necessary to reduce the biological system to a mathematically tractable formulation. At its best, artificial life models applied to theoretical biology lead to testable hypotheses.

This paper extends an analytical model of the risk of developing cancer and derives testable hypotheses about the genetic nature of the development of cancer from these models. We focus on a type of esophageal cancer known as esophageal adenocarcinoma, and its precancerous state, which is known as Barrett's esophagus (Reid, 1991; Neshat et al., 1994; Barrett et al., 1999).

Estimating Cancer Risks

Two dominant characteristics of cancer cells are their genetic instability (Lengauer et al., 1998) and uncontrolled proliferation (Kastan, 1997). The most commonly mutated tumor suppressor gene across all cancers is p53 (Smith and Fornace, 1995). The loss of this gene results in genetic instabilities (increased mutation rate), often with the loss or duplication of entire chromosomes (Smith and Fornace, 1995). The appearance of such aneuploid cells in Barrett's Esophagus is one of our most reliable indicators of a poor prognosis (Neshat et al., 1994). In contrast, p16 (a.k.a. CDKN2A and INK4a) is a gene thought to be responsible for shifting a cell from

a proliferative state to a quiescent state (G0) (Sherr, 1996). Loss of a p16 allele is associated with the spread of cells with that mutation throughout the Barrett's region. But, at least in Barrett's Esophagus, mutations in both p53 and p16 are not sufficient to cause cancer (Barrett et al., 1999). How many other genes are involved and what are their roles?

There is a body of mathematical modeling work which argues that the development of cancer is best understood as a sequence of two or more stages (Moolgavkar and Luebeck, 1990; Moolgavkar, 1999; Little, 1995; Luebeck and Moolgavkar, 1994; Sherman and Portier, 1996). The two stages might be called "precancerous" and "malignant." The two-stage model involves at least 6 rate parameters: the rate of cells changing from a normal state to the precancerous state, the rate of reproduction of precancerous cells, the rate of loss of precancerous cells, the rate of cells changing from the precancerous to the malignant state, and the rates of reproduction and loss of the malignant cells. These parameters appear to be sufficient to fit the model to most epidemiological data on the incidence of cancer. Moolgavkar, 1999 argues "without ancillary biological information there is little point to fitting models postulating more than two stages to tumor incidence data." It has been shown that models which fail to include the stochastic birth and death dynamics of cells in the stages give different results than those models which do include those dynamics (Luebeck and Moolgavkar, 1994). These stage models, also promoted by experimentalists (Fearon and Vogelstein, 1990), abstract away the evolutionary dynamics of cancer. Progression to cancer is seen as a progression through a linear sequence of stages, rather than a diversification into a phylogeny of cell lines. There are no interactions between cells in these models, such as competition for resources.

Theoretical work could potentially help guide research into this fundamental area of cancer genetics. For example, we could ask, if cancer requires 2 (or more) selective mutations in genes such as p16, what is the chance of developing cancer? Or, if a mutation in a gene such as p53 boosts the mutation rate, how would this affect the probability of getting cancer? Since we have good epidemiological data on the probability of getting cancer, we can then make guesses as to the number and kind of mutations that are necessary for its development. We will begin with some simple analytical calculations and incrementally elaborate them until we are forced to move to a simulation-based model of the evolution of cancer.

Loeb's Paradox

In 1991 Loeb formulated the following paradoxically calculation for the incidence of cancer. From the literature on human cell cultures he takes a per base pair, per cell division mutation rate of 10^{-10} (Oller et al., 1989; Monnat Jr., 1989; Fukuchi et al., 1989; Seshadri et al., 1987). He estimates that there are approximately 10^{16} cell divisions in a human lifetime. Finally, there are on the order of 10^9 base pairs in the human genome. Putting this together, we should expect $10^{-10} \times 10^{16} \times 10^9 = 10^{15}$ mutations in our cells during a human lifetime. If we are interested in the incidence of cells with two mutations at any loci, then this should occur $10^{-10} \times 10^{-10} \times 10^{16} \times 10^9 = 10^5$ times in a human lifetime. However, if a genetic disease requires 3 mutations to occur in the same cell, this should happen only once in 10^5 people. The chance of incurring 4 mutations is astronomically small. If these mutations must occur in specific loci, such as the coding regions of tumor suppressor genes and oncogenes, then the probability of developing cancer would be even smaller. Yet we believe that cancer requires a whole series of mutations (Armitage and Doll, 1954; Renan, 1993; Stein, 1991; Renan, 1993), and cancer is a frequent event during human lifespans.

Mutator Phenotype

One explanation for this paradox, offered in Loeb, 1998, is the idea of a "mutator" phenotype. Loeb's calculation changes if an early mutation, perhaps in p53, increases the mutation rate in the rest of the cell. Let us assume that the first event in this progression is a mutation that raises the mutation rate by c_m. Let μ be the mutation rate per locus per cell generation, k_m the number of critical genes necessary and sufficient to cause cancer, l_c the number of loci in a critical gene vulnerable to a cancer causing mutation, and let n_b be the number of cells in a human lifetime. To be generous, we will estimate that there are 100 different genes which, if they mutated, might raise the mutation rate. The expected number of cells that will independently develop cancer should be:

$$E[\text{Tumors}] = n_b \left[1 - (1-\mu)^{l_c 100}\right] \left[1 - (1 - c_m\mu)^{l_c}\right]^{k_m} \tag{1}$$

Where $(1-\mu)^{l_c 100}$ is the chance that a cell avoids a mutation in all $l_c 100$ loci that would produce the mutator phenotype. Thus $1 - (1-\mu)^{l_c 100}$ is the probability that a cell has a mutation in at least one of the 100 genes that lead to the mutator phenotype. Here $c_m\mu$ is the increased mutation rate. Loeb estimated $n_b = 10^{16}$ and $\mu = 10^{-10}$. There are approximately 10^3 loci in a human gene at which point a deletion, insertion, or substitution is likely to affect the polypeptide which that gene encodes. So we will consider $l_c = 10^3$. Comparison of normal and malignant cell cultures has estimated a change in mutation rate due to malignancy of 1 to 3 orders of magnitude (Seshadri et al., 1987). If we assume that cancer requires the initial mutation in the mutator gene and then 3 more mutations, a total number of mutations that was astronomically unlikely in

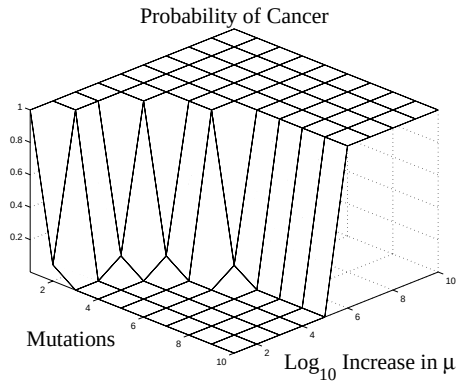

Figure 1: The expected number of cancerous cells that will develop during a person's lifetime. Two parameters are examined. The first parameter c_m is the increase in the mutation rate μ due to an initial mutation creating a mutator phenotype. This was calculated over the range of 10^1 to 10^{10}. The second parameter k_m is the number of mutations that are necessary and sufficient to cause cancer once the mutator phenotype has appeared, from 1 to 10. The expected number of cancerous cells has been truncated at 1.

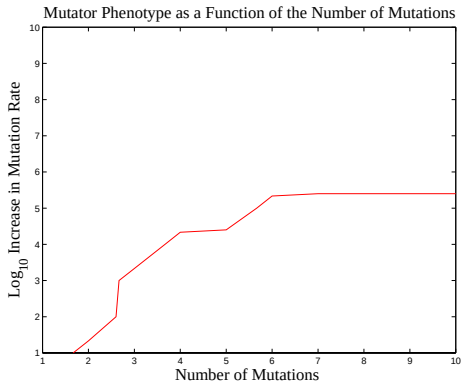

Figure 2: The predicted relationship between the increase in mutation rate of a "mutator phenotype" versus the number of mutations necessary to cause cancer after the appearance of the mutator phenotype. This an isocline calculated from Figure 1. This figure assumes a 0.4 probability of developing cancer during a lifetime. If the development of cancer requires many mutations, then the mutator phenotype would have to raise the mutation rate by at least 5 orders of magnitude.

Loeb's original estimation, and we assume that the mutator phenotype increases the mutation rate by 3 orders of magnitude, $c_m = 10^3$, then cancer should develop in $10^{16}[1-(1-\mu)^{10^3 10^2}](1-(1-10^{-10}10^3)^{10^3})^3 \approx 0.1$ cells in a human's lifetime. Figure 1 shows the log_{10} expected number of cancer cells dependent on k_m the number of mutations required and c_m the increase in the mutation rate due to the mutator phenotype. We have truncated the data at an expected single tumor because we are interested in the probability of developing cancer at least once.

Figure 1 shows that there is only a narrow window of mutation rate and number of sufficient mutations to develop cancer that result in realistic probabilities for developing cancer. In the United States, the chance of developing cancer during one's entire lifetime is approximately 40% (Ries et al., 1998). Figure 2 shows a view of the isocline where the probability of developing cancer is 40%. From this we can predict the relationship between the change in the mutation rate due to the emergence of the mutator phenotype and the number of mutations that are sufficient to cause cancer. For example, Figure 2 suggests that if the development of cancer requires 6 or more mutations after the initial rise in the mutation rate, then that initial increase must raise the mutation rate by at least 5 orders of magnitude.

Clonal Expansion

Loeb notes that Nowell, 1976, proposes another solution to his paradox. Some mutations can have selective effects and so increase the population of cells with that mutation (Nowell, 1976). We can elaborate Loeb's calculations with the assumption that the necessary mutations along the progression to cancer all have selective effects. Thus, if a cell incurs such a mutation, it will increase in frequency to some number n_t which is approximately equal to the number of cells in a tumor. Again μ is the mutation rate, k_m the number of critical genes, l_c the number of loci in a critical gene vulnerable to a cancer causing mutation, and n_b is the number of cells in a human lifetime. We will assume that the mutations can occur in any order.

The chance of the first mutation occurring is 1 minus the chance that it doesn't occur:

$$Pr[\text{first mutation}] = 1-(1-\mu)^{l_c k_m n_b} \qquad (2)$$

This will cause the cell with that mutation to expand to n_t cells. ¿From then on, each new mutation has n_t chances of occurring in a background of cells carrying all the previous mutations. The probability that the remaining $k_m - 1$ mutations occur is then:

$$Pr[\text{other mutations}] = \left[1-(1-\mu)^{l_c n_t}\right]^{k_m - 1} \qquad (3)$$

Let us make some reasonable assumptions for the values of l_c, n_t and k_m. To estimate n_t we will consider Barrett's Esophagus, a precancerous condition of the esophagus studied by the Reid lab at the Fred Hutchinson Cancer Research Center. Biopsies collected from the neoplastic tissue of the patients typically include 10^6 cells in a 2mm by 5mm section of epithelium. The entire Barrett's region averages approximately a surface area of 50mm by 60mm, or 10 biopsies by 30 biopsies. So

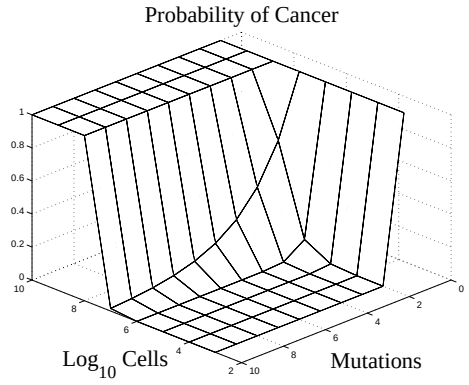

Figure 3: The probability of developing cancer during a person's lifetime. Two parameters are examined. The cell population size to which a selected mutant grows has been calculated over the range of 10^3 to 10^{10}. The second parameter is the number of selected mutations that are necessary and sufficient to cause cancer, from 1 to 10. These calculations estimate that if the selected population size is below 10^6 there is little chance of developing cancer. If it is 10^8 or above, a person is guaranteed to develop cancer during their lifetime.

the entire surface area can be sectioned into 300 biopsies of 10^6 cells, for a total of 3×10^8 cells. Since mutant clones are often observed to have expanded over the entire Barrett's region of a patient, it seems reasonable to set $n_t = 10^8$. Let us consider the case where $k_m = 4$ mutations are necessary to cause cancer. Recall that Loeb calculates the chance of 4 mutations occurring in the same cell to be astronomically small. Then,

$$Pr[\text{first mutation}] = 1 - (1 - 10^{-10})^{4 \times 10^{19}} \approx 1 \quad (4)$$

This number is so close to 1 that most computers cannot represent it as anything other than 1. So many cells are generated in a human lifetime that there are probably many cells that carry a mutation at any given locus. The interesting dynamics lie in the sequence of mutations that follow the first one:

$$Pr[\text{other mutations}] = \left[1 - (1 - 10^{-10})^{10^{11}}\right]^3 = 0.99986 \quad (5)$$

Given our assumptions, we estimate that 4 specific selected mutations are almost certain to occur in the lifetime of an individual. Of course, our estimates may be off. Figure 3 shows the probability of suffering cancer as a function of the number of cells to which selected mutant expands (n_t) and the number of selective mutations necessary and sufficient to cause cancer (k_m).

Figure 3 shows a precipitous drop in the probability of experiencing cancer as we reduce our estimate of the number of cells in a tumor from 10^8 to 10^6. The SEER report from the National Cancer Institute (Ries et al., 1998) estimates the lifetime probability of being diagnosed with cancer in the US is 45% for men and 38% for women (for all races and cancer sites combined). To match this estimate, our rough calculations suggest that in general cancers would require 3 selected mutations and those mutant clones would tend to spread to populations of 10^7 cells. Of course, this is an extremely simplified model of the incidence of cancer. We have not accounted for any environmental effects, genetic predispositions, or indeed any mutations that are necessary but do not spread through selection. Nevertheless, the elaboration to Loeb's calculations shows that Nowell's insight resolves the paradox. We develop cancer because the cells in the neoplastic tissues are evolving.

Both Mutator and Selective Effects

The two elaborations of Loeb's calculations consider selective and mutator mutations separately. A more realistic view of the development of cancer would likely consider both selective mutations and mutations that raise the mutation rate, and their interactions. In addition, there may be "neutral" mutations which have no effect on cell proliferation rates or mutation rates.

Consider the case in which a mutator or neutral mutation arises in a cell of the tumor. There is no reason to believe that this mutation would spread rapidly in the tumor. Without a selective advantage, such a mutation would be unlikely to grow to dominate the entire tumor. Meanwhile, if a selective mutation occurred in a cell which lacked the mutator or neutral mutation, the selective mutation would tend to expand throughout the tumor and thereby displace the mutant population with the mutator or neutral mutation. Thus, it is important to keep track of both the cells with the mutator or neutral mutations, as well as the cells that are free of those mutations but may yet suffer selective mutations. Each subpopulation can be characterized by the number of selective, mutator, and neutral mutations it has suffered, along with its population size. A set of difference equations can describe the growth dynamics of these subpopulations, as well as mutations that move cells from one subpopulation to another. But what growth dynamics should we use? The fundamental dynamic of biological reproduction is exponential. Is this a reasonable representation of tumor dynamics in humans?

In the esophagus, as in most of the digestive tract, cells along the lining (epithelium) are constantly being sloughed off and destroyed. These losses are replenished by the division of cells in the lining. In the case of Barrett's Esophagus, these cells are precancerous and hyperproliferative. The estimated turnover time is about once a week (Madara, 1995). As the cells are spatially structured as a two-dimensional layer (lining of a cylinder), there are severe spatial constraints restricting exponential growth. Further, cell division (mitosis) is a local

process, and so most new cells must compete for space with their immediate ancestors. The easiest way to represent a heterogeneous population of cells growing in a two-dimensional environment is with a two-dimensional model resembling a cellular automaton.

The Model We represented the the states of all the precancerous cells in the lining of the Barrett's region of an esophagus. We instantiated this as a two-dimensional discrete-event simulation in the shape of a column with "wrap-around" boundaries on the left and right sides, but not on the top and bottom. The state of a cell in this grid has four components: the number of selective mutations it has suffered (0-4), the number of neutral mutations it has suffered (0-4), whether or not it has suffered a mutation that increases its mutation rate (a "mutator" mutation), and its age (0-16). The population of cells is updated serially in a time step which represents approximately half a day. The time until the next reproduction (mitotic) event for each cell is drawn from a normal probability distribution with a mean of 8 times steps and a standard deviation of 2 time steps. Each selective mutation has the effect of doubling the replication rate of the cell. Thus a cell that has incurred 2 selective mutations reproduces 4 times as fast as a normal cell. When a cell divides, the new cell has a 50% chance of displacing one of the 9 cells, selected with uniform probability, in the 3 by 3 cell neighborhood centered on the parental cell. A run of the model began with all cells at age 0 with no mutations. With each time step representing 12 hours, we ran the model for 54,000 time steps (approximately 74 years), or a human lifetime. This put practical limitations on the number of cells we could model, with a maximum of 256 by 256 (65,536) cells. In the future we hope to model more realistic tumor sizes with approximately 10^8 cells.

We model the mutation rate as a Bernoulli process. The probability of a cell changing state is

$$Pr[\text{mutation}] = 1 - (1-\mu)^{(S+N+M)l_c n_p} = P \qquad (6)$$

Where $\mu = 10^{-10}$ is the mutation rate per base pair per cell generation, S, N, and M are the numbers of selective, neutral, and mutator genes sufficient and necessary to cause cancer if mutated, $l_c = 10^3$ is the number of critical base pairs (loci) in each gene at which a mutation could have a carcinogenic effect. In most cases, we assume that these mutations "knock out" the gene by either turning it off or destroying the functional effects of the normal protein produced by the unmutated gene. The last parameter, $n_p = 2$, is the number of independent pathways to cancer. This is an estimate of the number of genes in which a mutation will have the same carcinogenic effect. If a cell had at least one mutator gene mutated then μ increased by 10^3. This parameter for the increase in the mutation rate was called c_m in our earlier calculations. We primarily experimented with parameters S, N, and M, with some exploration of μ, l_c and the degree of increase in μ due to the mutator phenotype. A cell was called malignant if it had S selective mutations and N neutral mutations. We assumed that the mutator phenotype was not necessary for malignancy but only played a facilitating role through the increase in the mutation rate of the selective and neutral genes.

A Bernoulli process can be simulated by calculating the interarrival time for the next success. That is, instead of flipping a biased coin with probability of success P for each trial of the Bernoulli process, we can ask when the next success will happen. The probability mass function for the interarrival time k, the number of trials up to and including the next success, of a Bernoulli process is the geometric distribution:

$$Pr[k] = P(1-P)^{k-1} \qquad (7)$$

for $k = 1, 2, \ldots$. The expected value of k is $E[k] = 1/P$. When P is very small, as it is for most mutation rates, this function very gradually drops off. In this case, for the purposes of efficiency, it is reasonable to approximate $Pr[k]$ as a uniform distribution from 0 to $2/P$, which has the same expected value $E[k] = 1/P$ although a smaller variance. We calculated this with a single call to the pseudorandom number generator, using a version of Knuth's subtractive method (Knuth, 1981, pp. 171–172) to generate the pseudorandom numbers. We assume that the processes of DNA synthesis and cell division are the primary causes of mutations. Thus, in our model mutations only occur at cell division (Paulovich et al., 1997; Zheng et al., 1993). Mutations have an equal probability of occurring in the new or parental cell.

At the end of a run we measured the proportion of cells that suffered enough mutations to cause cancer (S and N). We ran the model at least 50 times for each parameter setting. A grey-scale picture of the model in the midst of a run is shown in Figure 4.

Results A run of the model was considered to have led to cancer if the final population had at least 1 cell with number of mutations required for malignancy (S and N). Figures 5 and 6 show the resulting probability of developing cancer as a function of the number of selective mutations S and neutral mutations N necessary and sufficient for developing cancer. Figure 5 shows the probabilities when there is no mutator gene to raise the background mutation rate. Figure 6 shows the results of the same parameter configurations when there is a mutator gene that may also mutate and thereby raise the mutation rate by 3 orders of magnitude.

Figure 7 is an extraction of a single curve from Figures 5 and 6 where $N = 1$. Figure 7 also shows the 90% confidence intervals around these curves calculated by treating the probability of developing cancer as a Bernoulli process. When there is no mutator gene in the

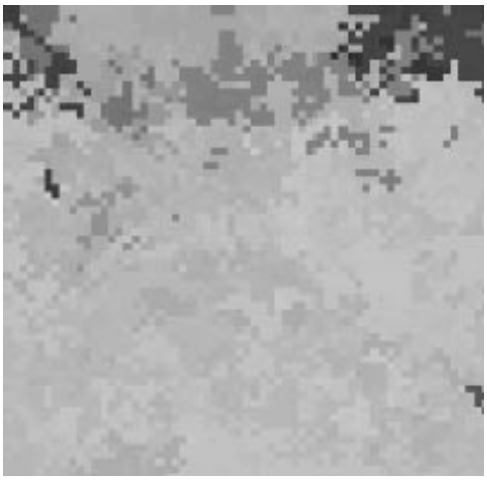

Figure 4: A view of the model running. The cells are color coded by lineage. The lighter grey lineages share an ancestor that suffered a selective mutation. This mutant clone is in the process of sweeping through the entire tissue.

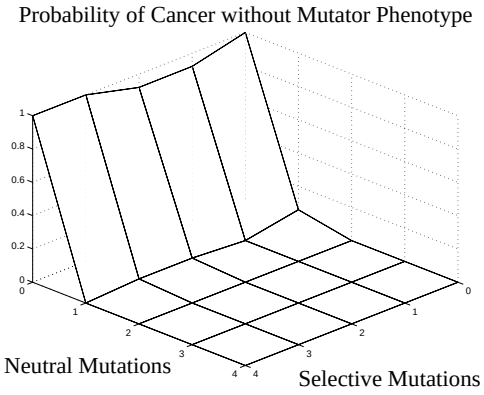

Figure 5: A plot of the probability of developing cancer as a function of the number of selective and neutral mutations necessary and sufficient to cause the disease. These probabilities have been calculated in the absences of a mutator gene. The probabilities are dominated by the number of neutral mutations that are necessary. The probabilities were calculated by at least 50 runs of the agent-based model with only 4096 cells.

system, the probability of developing cancer decreases with the number of selective mutations that are required. This seems reasonable in light of our earlier calculations. However, in the presence of a mutator gene that can raise the mutation rate at any time, the probability of developing cancer actually increases with the number of necessary selective mutations. This is because a selective mutation generates a large number of new cells, each cell representing a potential new mutation. When the mutation rate is high enough, this forms a positive feedback system in which one selective mutation generates the next selective mutation and so on until the system reaches malignancy.

Our explorations of other parameters in the system all show a relationship between the parameters and the probability of developing cancer that is either linear or sub-linear. In all of these cases we assume that 1 neutral and 2 selective mutations is necessary and sufficient for the development of cancer. .The exponents for these relationships were derived from the slope of the line that was fit to the log transformation of the data. It should be noted that in all cases the line was fit with only 3 or 4 data points, and so the results should be taken only as a qualitative indication of the dynamics of the system. Figures 8 and 10 are log-log plots of the relationship of the parameter to the probability of developing cancer. Figures 9 and 11 had to be plotted as a log-linear plots due to the calculated 0 probability of developing cancer in some instances. Figure 8 shows that the probability of developing cancer increases as a square root (in the presence of a mutator gene) or linear function (in the absence of a mutator gene) of the number of cells produced by a selective mutation. Figure 9 shows that the probability of developing cancer increases in proportion to the square root of the mutation rate. Figure 10 shows that this probability also increases in proportion to cube root of the change in mutation rate caused by a mutation in the mutator gene, i.e., the difference between the normal and the mutator phenotype. Finally, Figure 11 shows that the probability of developing cancer increases roughly in proportion to the number of base pairs in the genes at which a mutation can have a carcinogenic effect. Of course, since probabilities are bounded at 0 and 1, these relationships may break down as they near those boundaries.

Discussion

Other researchers have studied the relative merits of the two solutions to Loeb's paradox (Tomlinson and Bodmer, 1995; Tomlinson et al., 1996). Tomlinson et al., 1996 concluded that selective mutations alone are sufficient to explain the mutations observed in cancer. In their investigation of the mutator phenotype, they investigated the case where either 2 or 6 neutral mutations were necessary to cause cancer (Tomlinson et al., 1996). They assumed the mutator phenotype raised the mutation rate from 10^{-8} to 10^{-4}. They found that in the case of requiring 2 neutral mutations, cancer often developed before the mutator phenotype appeared, but with 6 required mutations, the mutator phenotype would appear before cancer. They argue that the importance of a mutator cell will be wiped out if any of the other mutations have selective effects. Our results do not support this. The presence of a few selective mutations amongst many neutral mutations has little effect. However, the

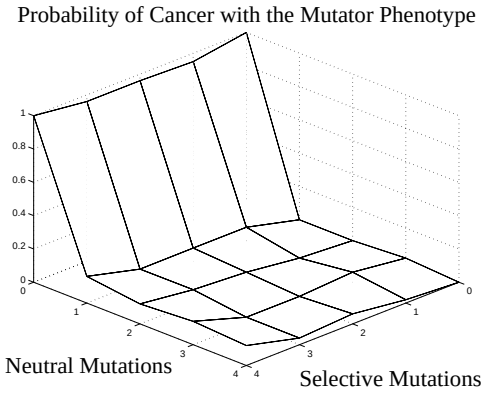

Figure 6: A plot of the probability of developing cancer as a function of the number of selective and neutral mutations necessary and sufficient to cause the disease. These probabilities have been calculated in the presence of a mutator gene that raises the background mutation rate from 10^{-10} to 10^{-7} when it is mutated. The mutator gene has the same probability of mutation as the other genes, and thus the background mutation rate may change at any time during the run of the model. This contrasts with Figure 1 in which we assumed that the mutator gene had been mutated before the other genes. Note that the probability of developing cancer rises with the number of selective mutations involved irrespective of the number of necessary neutral mutations.

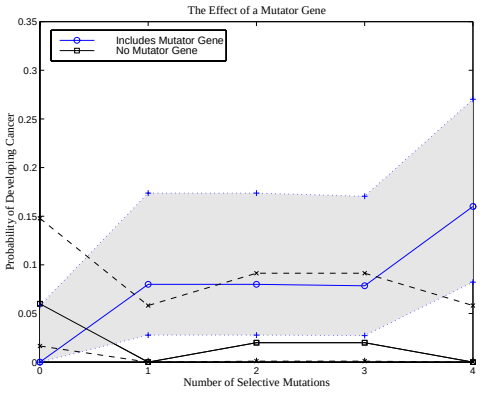

Figure 7: The interaction of mutator and selective genes. The solid lines show the probability of developing cancer as a function of the number of necessary selective mutations. In all cases 1 neutral mutation was required. Both solid lines are surrounded by their 90% confidence intervals shown in dotted lines for the mutator case and dashed lines for the case without a mutator gene. The confidence interval for the mutator case is shaded in grey. There is a synergy between large numbers of necessary selective genes and the mutator gene. In the presence of a mutator gene, the probability of developing cancer actually increases with the number of necessary selective genes. In this case the process of developing cancer has a sort of positive feedback effect that quickly generates malignant cells. In the absence of a mutator gene the probability of developing cancer goes down with the number of necessary selective genes.

combination of selective and mutator mutations dramatically increases the probability of developing cancer, as is shown in Figure 7.

An important aspect of both the analysis of selective and mutator mutations in cancer is that the parameters of the predictions are observable and thus the predictions are experimentally testable. Data is becoming available on the population sizes of cells with selective mutations, and it is becoming feasible to measure the mutation rate in cells with mutator phenotypes, perhaps through the loss of p53. Similarly, it should be possible to derive accurate measurements of the number of critical loci in any given gene relevant to the development of cancer. In the model we assumed this number was about 10^3 for all genes, an estimate that could be improved significantly. In the foreseeable future we will be able to reduce the ranges of the significant parameters in the model when information about the number and kinds of mutations that are sufficient for the development of cancer is determined.

Our simulation of the development of cancer is only a toy model and as such it avoids many of the known complexities of the biological system. We have implicitly assumed that each mutation is independent of the others, and so can occur in any order. Further, we have not explicitly represented the phenomenon of dominance in which a recessive phenotype might require two mutations before it appeared. However, this could be represented by the combination of a neutral mutation, which occurs first, and a selective mutation, which would follow the neutral mutation. We have also ignored the effects of cell senescence. Most cells stop dividing after some number of divisions have shortened the telomeres to the point where they no longer protect the ends of the chromosomes.

Only one type of selective effect has been modeled. However, mutations can have strong selective effects without changing the generation time of a cell. Mutants that tend to compete successfully for space, either by displacing their neighbors or by resisting displacement by future competitors, would also spread in the population. There are probably a variety of other genetic innovations that would have beneficial phenotypic effects. Most of these could be represented and explored in an elaborated model.

Our model of the mutator phenotype is probably inappropriate. We have modeled the mutator phenotype as a dramatic boost in the background mutation rate. This assumes that mutations occur independently throughout

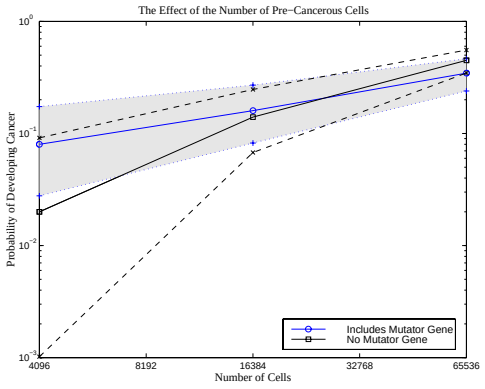

Figure 8: The effect of running the simulation with more cells. The values on the horizontal axis represent the number of cells that is produced through the clonal expansion of a selective mutation, or roughly the number of cells in the precancerous tissue. Again the 90% confidence intervals are plotted around each solid line and the interval with the mutator gene is shaded in grey. There is some indication that as the number of cells in the system rises, the effect of the mutator gene diminishes, but the confidence intervals generally overlap and so little of significance can be asserted. The slope of the best fit line for the mutator case is 0.5, indicating that the probability of developing cancer is proportional to the square root of the number of cells in a tumor. The slope for the non-mutator case is 1.1, indicating that in the absence of a mutator gene, the probability of developing cancer rises in proportion to the number of cells in a tumor.

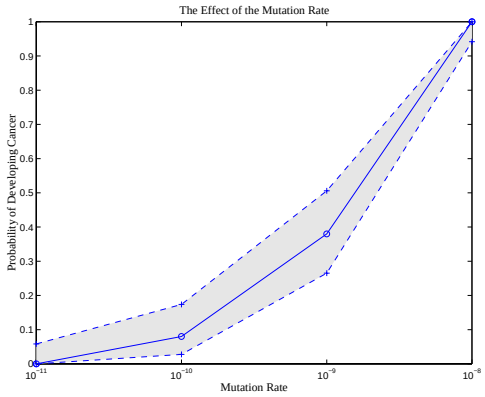

Figure 9: A log-linear plot with %90 confidence intervals of the effect of changing the background mutation rate μ. The increase in the chance of developing cancer is roughly proportional to the increase in the mutation rate. If we fit a line to the log-transform of the axes, ignoring the 0 value, the probability of developing cancer is proportional to the square root of the mutation rate (the slope = 0.5).

the genome. However, our archetypal candidate for a mutator gene, p53, seems to cause the loss (and gain) of whole chromosomes as well as prevent the repair of damaged DNA. In the case of chromosome loss, mutations in genes are not independent and tend to occur in massive clusters. Furthermore, we have not modeled the effects of deleterious mutations. We would expect an increase in the background mutation rate to also increase the frequency of deleterious mutations, which would result in a selective disadvantage, and sometimes fatal, effect on the host cell.

Finally, we have completely ignored the immune response. We know that the human immune system sometimes attacks precancerous and cancerous cells (Jantscheff et al., 1999), but the details of these dynamics are still unknown. The immune system would clearly have selective effects on the populations of cells. The immune system could lower the probability of developing cancer relative to our estimates.

The simplifications of our models and our ignorance of realistic parameter values prevent us from making highly focused experimental predictions. However, the qualitative behaviors of the models do lead to two predictions:

Prediction 1 *The development of cancer requires at least 2 selectively neutral mutations.*

Our model of 2^{16} cells with 1 neutral and 2 selective mutations sufficient for developing cancer, in the presence of a mutator gene, led to a cancer incidence of 35%. With a more realistic number of cells in a tumor, perhaps 10^8, the simulated incidence of cancer would be unrealistically high. Requiring more selective mutations only makes the incidence of cancer higher. Thus, cancer must require more selectively neutral mutations.

Prediction 2 *The development of cancer involves a number of neutral mutations that is within the same order of magnitude as the number of selective mutations.*

The clonal expansion of a selective mutant produces a large population of mutant cells and involves a large number of cell divisions in which new mutations may arise. The chance of a neutral mutation is greatly enhanced if it follows a selective mutation. However, if more neutral than selective mutations are required for the development of cancer, then the neutral mutations form bottlenecks in path to cancer and make malignancy more unlikely. This was at the heart of Loeb's paradox. On the other hand, if few neutral mutations but many selective mutations are required, and there exist mutator genes in the system, then the mutator genes and the selective genes form a positive feedback system that accelerates the system towards cancer. Since mutations in p53 are common across most forms of cancer (Smith and Fornace, 1995), it is reasonable to suppose that there is a mutator gene in the system. In this case, with few necessary neutral mutations, the probability of developing

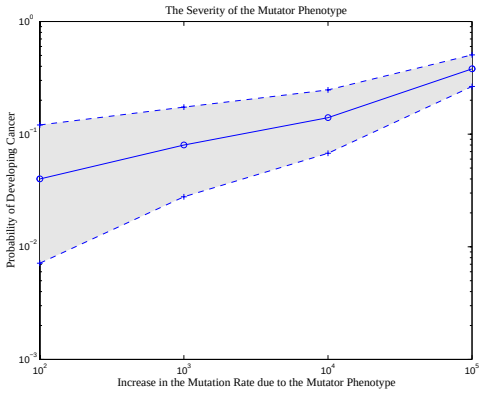

Figure 10: A log-log plot with %90 confidence intervals of the results from adjusting the effect of the mutator phenotype. The horizontal axis shows the change in the background mutation rate caused by a mutation in the mutator gene. The probability of developing cancer increases in proportion to the cube root (exponent = 0.3) of the change in the mutation rate due to the mutator phenotype.

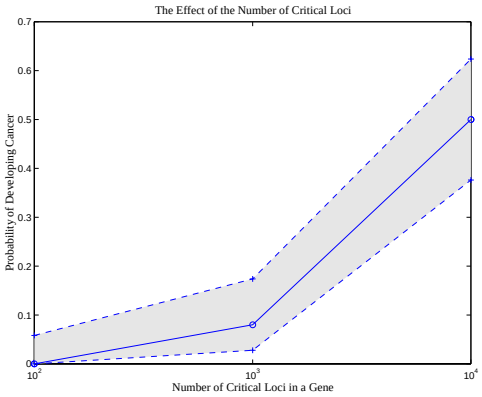

Figure 11: A log-linear plot with %90 confidence intervals of the effect of varying the assumed number of base-pairs or loci in a gene at which a mutation could have a carcinogenic effect. In the case of a tumor suppressor gene, this would correspond to the number of different mutations that could knock out the gene. It is difficult to fit a line to the log transform of the data since 1 of the 3 data points is 0. If we guess that the probability of getting cancer when each gene has 10^2 critical loci is between 0.01 and 0 (we only ran the model 50 times so we lack the resolution to distinguish probabilities this low), and replace that 0 value with 0.005, the slope of the line is 1. Thus, the probability of developing cancer is roughly proportional to this number of critical loci in a gene.

cancer is too high to be realistic. Thus, the number of necessary neutral mutations must be close to the number of selective mutations. Our guess for the meaning of "close" is the same order of magnitude.

What insights might we derive from these results for the treatment or prevention of cancer? All of the analyses suggest that neutral mutations are the bottleneck in the development of cancer. This implies that an effective prevention program would be one in which would add at least one additional neutral mutation to the set of necessary mutations for the development of cancer. In other words, we should try to add bottlenecks to the development of cancer. This might, for example, be achieved by treatments for which the precancerous cells would have to generate recessive mutations in order to escape the treatment and to progress on towards cancer. If the susceptible phenotype is completely dominant, then a mutation in one of the two alleles of a homozygous dominant cell will have no phenotypic effect and will thus be selectively neutral. Similarly, cocktails of multiple drugs (Hughes and Frenkel, 1997) that require mutations at multiple sites in order to develop resistance to all of the drugs in the cocktail might be particularly effective.

The authors gratefully acknowledge the support of the National Science Foundation (grants IRI-9711199 and CDA-9503064), the Office of Naval Research (grant N00014-99-1-0417), and the Intel Corporation. Thanks to Terry Jones for coding Knuth's pseudorandom number generator, Keith Wiley for writing the GUI (and capturing it in the act), the UNM adaptive group and two anonymous referees for their comments. A special thanks to Miller Maley for his advice on modeling probabilistic processes. We would also like to thank the Brian Reid lab at the Fred Hutchinson Cancer Research Center for their welcome, enthusiasm, and generous collaboration. Without their support, this paper could not have been written.

References

Armitage, P. and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*, 8:1–12.

Barrett, M. T., Sanchez, C. A., Prevo, L. J., Wong, D. J., Galipeau, P. C., Paulson, T. G., Rabinovitch, P. S., and Reid, B. J. (1999). Evolution of neoplastic cell lineages in Barrett oesophagus. *Nature Genetics*, 22:106–109.

Collins, R. and Jefferson, D. (1992). The evolution of sexual selection and female choice. In Varela, F. J. and Bourgine, P., editors, *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, pages 327–336, Cambridge, MA. MIT Press.

Fearon, E. R. and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61:759–767.

Fujii, H., Marsh, C., Cairns, P., Sidransky, D., and

Gabrielson, E. (1996). Genetic divergence in the clonal evolution of breast cancer. *Cancer Research*, 56:1493–1497.

Fukuchi, K., Martin, G. M., and Monnat Jr., R. J. (1989). Mutator phenotype of werner syndrome is characterized by extensive deletions. *Proceedings of the National Academy of Sciences USA*, 86:5893–5897.

Hughes, R. S. and Frenkel, E. P. (1997). The role of chemotherapy in head and neck cancer. *American Journal of Clinical Oncology*, 20:449–461.

Jantscheff, P., Herrmann, R., and Rochlitz, C. (1999). Cancer gene and immunotherapy: Recent developments. *Medical Oncology*, 16:78–85.

Kastan, M. B. (1997). Molecular biology of cancer: The cell cycle. In DeVita, Jr., V. T., Hellman, S., and Rosenberg, S. A., editors, *Cancer: Principles and Practice of Oncology, Fifth Edition*, pages 121–134. Lippincott-Raven Publishers, Philadelphia, PA.

Knuth, D. E. (1981). *The Art of Computer Programming, vol. 2.* Addison-Wesley, Reading, MA.

Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature*, 396:643–649.

Levin, S. A., Grenfell, B., Hastings, A., and Perelson, A. S. (1997). Mathematical and computational challenges in population biology and ecosystems science. *Science*, 275:334–343.

Little, M. P. (1995). Are two mutations sufficient to cause cancer? *Biometrics*, 51:1278–1291.

Loeb, L. A. (1998). Cancer cells exhibit a mutator phenotype. *Advances in Cancer Research*, 72:25–56.

Luebeck, E. G. and Moolgavkar, S. H. (1994). Simulating the process of malignant transformation. *Mathematical Biosciences*, 123:127–146.

Madara, J. L. (1995). Epithelia: Biologic principles of organization. In Yamada, T., editor, *Textbook of Gastroenterology, second edition*, pages 237–257. J. B. Lippincott Co., Philadelphia.

Maley, C. C. (1998). *The Evolution of Biodiversity: A Simulation Approach*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Monnat Jr., R. J. (1989). Molecular analysis of spontaneous hypoxanthine phosphoribosyltransferase, mutations in thioguanine-resistant HL-60 human leukemia cells. *Cancer Research*, 49:81–87.

Moolgavkar, S. H. (1999). Stochastic models for estimation and prediction of cancer risk. In Barnett, V., Stein, A., and Turkman, K. F., editors, *Statistics for the Environment 4: Pollution Assessment and Control*, pages 237–257. John Wiley and Sons, Ltd., New York, NY.

Moolgavkar, S. H. and Luebeck, E. G. (1990). Two-event model for carcinogenesis: Biological, mathematical, and statistical considerations. *Risk Analysis*, 10:323–341.

Neshat, K., Sanchez, C. A., Galipeau, P. C., Cowan, D. S., Ramel, S., Levine, D. S., and Reid, B. J. (1994). Barrett's Esophagus: A model of human neoplastic progression. *Cold Spring Harbor Symposia on Quantitative Biology*, 59:577–583.

Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*, 194:23–28.

Oller, A. R., Rastogi, P., Morgenthaler, S., and Thilly, W. G. (1989). A statistical model to estimate variance in long term-low dose mutation assays: Testing of the model in a human lymphoblastoid mutation assay. *Mutation Research*, 216:149–161.

Paulovich, A. G., Toczyski, D. P., and Hartwell, L. H. (1997). When checkpoints fail. *Cell*, 88:315–321.

Reid, B. J. (1991). Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology Clinics of North America*, 20:817–834.

Renan, M. J. (1993). How many mutations are required for tumorigenesis? Implications from human cancer data. *Molecular Carcinogenetics*, 7:139–146.

Ries, L. A. G., Kosary, C. L., Hankey, B. F., Miller, B. A., and Edwards, B. K., editors (1998). *SEER Cancer Statistics Review, 1973-1995.* National Cancer Institute, Bethesda, MD.

Seshadri, R., Kutlaca, R. J., Trainor, K., Matthews, C., and Morley, A. A. (1987). Mutation rate of normal and malignant human lymphocytes. *Cancer Research*, 47:407–409.

Sherman, C. D. and Portier, C. J. (1996). Stochastic simulation of a multistage model of carcinogenesis. *Mathematical Biosciences*, 134:35–50.

Sherr, C. J. (1996). Cancer cell cycles. *Science*, 274:1672–1677.

Smith, M. L. and Fornace, A. J. (1995). Genomic instability and the role of p53 mutations in cancer cells. *Current Opinion in Oncology*, 7:69–75.

Stein, W. D. (1991). Analysis of cancer incidence data on the basis of multistage and clonal growth models. *Advances in Cancer Research*, 56:161–213.

Tomlinson, I. P. M. and Bodmer, W. F. (1995). Failure of programmed cell death and differentiation as causes of tumors: Some simple mathematical models. *Proceedings of the National Academy of Sciences of the United States of America*, 92:11130–11134.

Tomlinson, I. P. M., Novelli, M. R., and Bodmer, W. F. (1996). The mutation rate and cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 93:14800–14803.

Zheng, C.-J., Byers, B., and Moolgavkar, S. H. (1993). Allelic instability in mitosis: A unified model for dominant disorders. *Proceedings of the National Academy of Sciences USA*, 90:10178–10182.