Modeling the Effects of Dynamic Drug Treatment on the Evolution of Tumors toward Malignancy

Linda Doan, Matthew Pennington, Jane Wang, and Lin Wang

Abstract

Cancer is a clonal disease in which normal control over cell cycle in a population becomes disregulated. Left unchecked, a benign neoplasm may gain mutations that allow a subpopulation to prevail, expand, and become malignant, eventually destroying its host organism. Neoplasms are widely thought to be evolutionary systems in which Darwinian fitness governs the displacement of subpopulations and the prevalence of other subpopulations. Though neoplasms are often treated aggressively with drugs, there has been little study of the effects of treatment as it pertains to the evolution of cancer. To study how treatments may affect the evolution of a tumor, we formulate a benign-to-malignant evolutionary process within a space of 4^6 genotypes. The growth of the populations is modeled on a fitness landscape, and the fitness landscape is perturbed as a function of time to simulate the change in the environment caused by the drug. We separately treat the evolution of the cancer population as both a deterministic system and a stochastic system in order to broaden the spectrum of behaviors that we observe. First we consider how time-dependent perturbations to these systems (simulating dynamic drug treatments) affect the flux of the population towards malignancy. Secondly, we will consider how these perturbations may affect the population distribution by measuring what we call population entropy. We find that changes in the fitness landscape do lead to interesting and diverse behaviors in the system. It is also clear that relative mutation rates in the population play a big role in both the flux and the population entropy results in both models.

Keywords Cancer evolution, population entropy, cancer therapy, deterministic modeling, stochastic modeling

1. Introduction

Cancer is a disease in which cells with unregulated growth invade and harm the normal tissue of the host organism, eventually killing the host. Cancers typically begin as benign neoplasms, in which abnormal cells persist by dividing more than normal, or not dying off as normal. The neoplasm is usually contained in the host by various mechanisms. Left unchecked, the neoplasm can develop the ability to break out of containment and invade the surrounding tissue. This is when the benign neoplasm is said to transition to a malignant cancer. Although cancer is a clonal disease, it is characterized by a high mutation rate and thus leads to a genetically heterogenous population.

It has long been recognized that neoplasms are microcosms of evolution (Nowell 1976; Merlo 2006) where every tumor starts from healthy cells and becomes progressively more disregulated. As in classical evolution, populations are shaped by random mutations and selective pressures. For example, cancers that develop the ability to attract new blood vessels, to evade the immune system, or to turn on growth signals have a selective advantage over other subpopulations. The gain of a single advantage is not sufficient for malignancy. Rather, several of such cellular processes must become abnormal before a neoplasm can develop into a malignant tumor (Nowell 1976; Renan 1993, Stoler 1999). Once malignancy is reached, the cancer enters a different environment in which there are different selective pressures that govern the evolutionary process.

As mentioned previously, there develops in a tumor a heterogeneous population of cells that compete for space and resources. As in classical evolution, most genetic changes are either neutral or deleterious. However, heritable genetic changes do occur that allow clones to out-compete less fit subpopulations, making this an evolutionary system amenable to modeling. In order to treat this as an evolutionary process, the growth of the neoplastic population will be modeled on a fitness landscape. Use of fitness landscapes is complicated by the fact that interactions with environmental factors can change in numerous and unpredictable ways – especially within long time scales. However, in the context of cancer evolution, the model can reasonably include a starting point and an endpoint.

It is not in the interest of this paper to accurately portray the environment and the perturbation to the environment in the fitness landscapes. Rather, it is more interesting and immediately possible to address how a generic change in a landscape (generic drug treatment) can affect the evolution and population distribution of a neoplastic population. In order to make these generalizations, the model will be run over an ensemble of different fitness landscapes. In this model, a therapeutic intervention imposes a randomly generated fitness landscape onto the population. Two therapeutic interventions are thus represented by two different landscapes, and the frequency of the switch between the two drugs represented by the parameter ω.

While cancers have a relatively high mutation rate compared to normal cells, the probability that more than one non-neutral mutation occurs during one generation of a cell is extremely low. The chromosomal theory of cancer origin is not treated by this model, as these are thought to occur in specific cases (Loeb 1991; Duesberg 2005). This study seeks to treat the general topic of cancer in the context of classical evolutionary processes.

Modeling the entire genotype and the possible evolutionary changes in genotype space of a population is unrealistic given current computing abilities. In order to address the generic rules that govern cancer evolution, we have simplified the system such that the genomic state of the neoplasm is defined by 6 genes. Each gene can be defined by one of four discrete states, yielding 4096 genotypes in our space. Because mutation rates allow

only up to one gene to mutate in each generation, mutations are allowed only along single axes. Another mutation phenomenon that we incorporate into this model is that a large variety of different mutations can lead to gene breakage, whereas only very specific and precise mutations will fix a broken gene (Frank 2004). Thus, genes are more likely to break than to repair, and mutations accumulate.

In our model, the evolving cells are allowed to explore randomly generated landscapes within a finite multidimensional genomic space. The genomic space is represented as a network with asymmetric transition probabilities between different states. According to the current understanding of cancer evolution, mutation frequencies are estimated at 10^{-6} to 10^{-7} per locus per generation (Albertini 1990; Araten 2005). A key feature of this is that simultaneous mutations in two genes are rare, but big jumps along a single genetic axis are possible. This aspect of natural evolution is preserved in the model in this study.

Finally, the equations that govern the system allow for an entropy measurement of the system. In this study, the population entropy is interpreted as the breadth of the population over genotype space, or the efficacy with which cells explore the potential genome space. Increased genetic diversity positively correlates with progression of a tumor towards malignancy (Maley 2006). Currently, there is little knowledge in the field of cancer biology as to the effects of drug treatment on distribution entropy. The study will examine for the first time how population entropy is affected by dynamic drug treatment.

In this study, we separately treat the evolution of cancer as a deterministic system and as a stochastic system in order to better address various aspects of the evolution process. While the models must remain simple, they must address the salient features of cancer evolution and treatment as it is currently understood. Thus, the fitness of a subpopulation is a function of its genotype, and determines its proliferation rate, while the total size of the population is limited by uniform competition between the cells, as might be expected for space or nutrients. Drug treatment changes this landscape, and mutation between different genotypes that lie along a single genetic unit vector are connected by forward and backward linear hopping rates.

Understanding cancer progression as a microcosm of evolution can give us insight into the development, screening, and treatment of cancer. Therapies can be seen as changing the fitness landscape for a cancer. Because evolution is a dynamic process, perhaps the dynamics of therapy affect the evolution of cancer in an obvious way. Specifically, this study is aimed at understanding the general behaviors of how the dynamics of drug treatment can affect the rate at which a neoplasm becomes malignant, and how drug treatment dynamics can affect the population distribution.

2. Basic Model

We developed a simple model to test the effect of hypothetical dynamic drug treatments

on the evolution of premalignant neoplasms. Assuming universal logistic competition between all cells in a population, and expressing that population as a function of genotype *(p(g)*), we have, in general:

$$\partial_t p(g) = f \cdot p(g) - \alpha p(g) \int p(g')\,dg' + \int m(g',g) p(g')\,dg' - p(g) \int m(g,g')\,dg' \quad (1)$$

Here, *f* is a fitness defined as a function of genotype, α is a constant parameter, and *m(g,g')* represents a mutation rate: the rate that genotype g is converted into genotype *g'*. This partial integral-differential equation becomes significantly less intimidating if we discretize the genotype, which makes it a system of ODEs.

$$\partial_t p_g = f_g p_g - \alpha p_g \sum_{g'} p_{g'} + \sum_{g'} \left(m_{g',g} p_{g'} - m_{g,g'} p_g \right) \quad (2)$$

Here, p_g is the population at a particular genotype *g*, and f_g is the corresponding fitness. The matrix elements $m_{g,g'}$ represent the linear rate at which population at genotype *g* becomes genotype *g'*.

The customary division between genotype and phenotype in living things presents a difficulty, here, since we are trying to consider in some detail both mutations and a continuous fitness landscape. The conflict is essentially that mutations (in genotype space) occur in small increments, and the mechanism is fairly well known. Fitness landscapes, on the other hand, can only be considered continuous in phenotype space, where genes have graded activity. The mapping of genotype space (which is intrinsically discrete, and very high-dimensional) to phenotype space (which is lower dimensional, and more continuum-like) in any sort of detail is basically impossible for any single gene, as far as we know. To address this problem, we have to make two compromises. First, without detailed data on any particular genes and sequences, we must look for results that are general to the extent they are true for most possible or most likely situations. This means, essentially, testing an ensemble of fitness landscapes and mutation matrices. Second, this ensemble should be chosen to best preserve what is known about both fitness landscapes and mutation processes, though in detail it will look like neither. Due to computational constraints, it makes sense to limit ourselves to genome spaces that consist of 6 independent genes, each with 4 internal states.

In real tumors, the cellular mutation rate can be quite high, but is still such that the vast majority of genes do not experience a mutation between any two consecutive mitoses. For this reason, when a mutation occurs in a gene of interest, it is almost certainly the only mutation that has occurred in a gene of interest. Putting this property of genetic mutation into a mutation matrix involves setting to zero any elements that represent the interconversion of genotypes that differ along more than one genetic axis. For 3 genetic axes with 2 internal states for each gene, this connectivity is simply the edges of a cube. For more internal states, and higher dimensions, it is like the tiling of the volume with the appropriate hypercube. Furthermore, a single mutation in a single gene can have no

effect on the phenotype, an extreme effect, or anything in between. To represent this in a lower dimensional space, we allow connections of any length along a single genetic axis, though we make longer jumps less likely.

Lastly, for the mutation matrix, we make the assumption that a functional gene subject to random mutations will in all likelihood cease to function, whereas the converse is not true. This is based on the very high dimensionality of real genome-space, and the fact that a random walker in a high dimensional space is unlikely to ever return to its starting point. Essentially, we are saying that the neutral manifold of a functional gene should be considered a small, confined set in genotype space. Projecting this situation into the lower dimensional space related to phenotype used for simulations means introducing a systematic bias to the mutation rates, to result in a net drift toward "broken" genes.

For this study we are interested in probing any non-trivial behavior that comes from the *dynamics* of a fitness-affecting drug treatment. In equation (2), this means making f_g time-dependent. To get at this behavior specifically, and not the behavior specific to a hypothetical class of drugs, we decided to have the fitness landscapes representing the medicated and the unmedicated states be generated by identical random processes, that each one should have some smoothness in genotype space, and that they should have identical first and second moments. We make the time dependence explicit, and switch between pairs of landscapes with a frequency ω.

If one assumes that there is one parameter (the distribution width) that characterizes the fitness landscape and one parameter that characterizes the overall mutation rate, and rescales p and t appropriately, it becomes clear that the behavior of the system is governed by the ratio of these two distribution parameters. This suggests that we may have different results depending on whether the system is dominated by growth (and thus competition), by mutation, or neither. Furthermore, the choice of these parameters will set a spectrum of timescales intrinsic to the evolutionary system. We are introducing another, completely independent timescale by our choice of ω. Thus, a priori, we predict that there are at least six distinct parameter regimes for the model that exist generically:

$$\langle m \rangle < \sqrt{\langle f^2 \rangle} < \omega, \;\; \omega < \langle m \rangle < \sqrt{\langle f^2 \rangle}, \;\; \langle m \rangle < \omega < \sqrt{\langle f^2 \rangle}$$
$$\sqrt{\langle f^2 \rangle} < \langle m \rangle < \omega, \;\; \omega < \sqrt{\langle f^2 \rangle} < \langle m \rangle, \;\; \sqrt{\langle f^2 \rangle} < \omega < \langle m \rangle$$

It is also possible to talk about parameter regimes where these timescales are approximately equal, but due to the considerable complexity of the model's dynamics, it is difficult to locate these parameter regimes in advance, and they must be found by interpolation between known limits.

Now that we have built up a model that captures some aspects of the evolution of an isolated, internally-competing population in a controlled, dynamic environment, we need to think about what outputs of the model may be interesting. Classically, people concern themselves with the population of tumor cells, and focus on driving that population to extinction. We adopt a different tactic. The model is only designed to represent a benign,

self-limited growth, as evidenced by the competition term, and the total population shouldn't vary too dramatically. Clinically, what determines the quality of a treatment in a patient with this condition is the amount of time it takes for the tumor to progress (which happens with great certainty) to a non- self-limiting state by the accumulation of the appropriate mutations. For our purposes, we consider this to be the flux out of our model system by a mechanism other than cell-death, and add a perfect absorber at a specific genotype to provide this pathway. The rate of flux of population onto this absorber is a proxy for the rate at which cells become malignant, and is thus proportional to the inverse of the average time to progression in a simulation "patient."

3. Results

In this section, we research the basic model from two different views: the deterministic model and the agent-based model.

3.1 Deterministic model

To test our hypotheses about dynamical drug treatments and explore the model behavior in general, we integrated equation (2) many times. For each of 6 different sets of 2 fitness landscapes and one mutation landscape, we scaled the landscapes by each of 9 different relative rates determining $\sqrt{\langle f^2 \rangle} / \langle m \rangle$, and probed each of these parametrized models with 6 different dynamical treatment rates, ω. Each integration was performed three times for initial conditions corresponding to a steady state of each of the fitness landscapes used, as well as a steady state of the fitness landscape resulting from averaging the two primary landscapes. In addition to the raw population data, we calculated the total flux onto an absorber over time for each of these cases, and calculated the distribution entropy

$$S_d = \sum_g p(g) \log\left[p(g) \right]$$

as an index of how spread out the population distribution was over the possible genotypes, or how "actively" it might be considered to be evolving.

For such a simple system, the results are incredibly complicated. There is one observation that holds across all trials, however, and that is that the distribution entropy, which can be taken as a measure of how well the system is exploring genotype space, can go on long transients to higher-than-equilibrium values for a system that has an otherwise low mutation rate.

The following pictures show the distribution entropy of a population with different scaling between the mutation rate and the fitness.

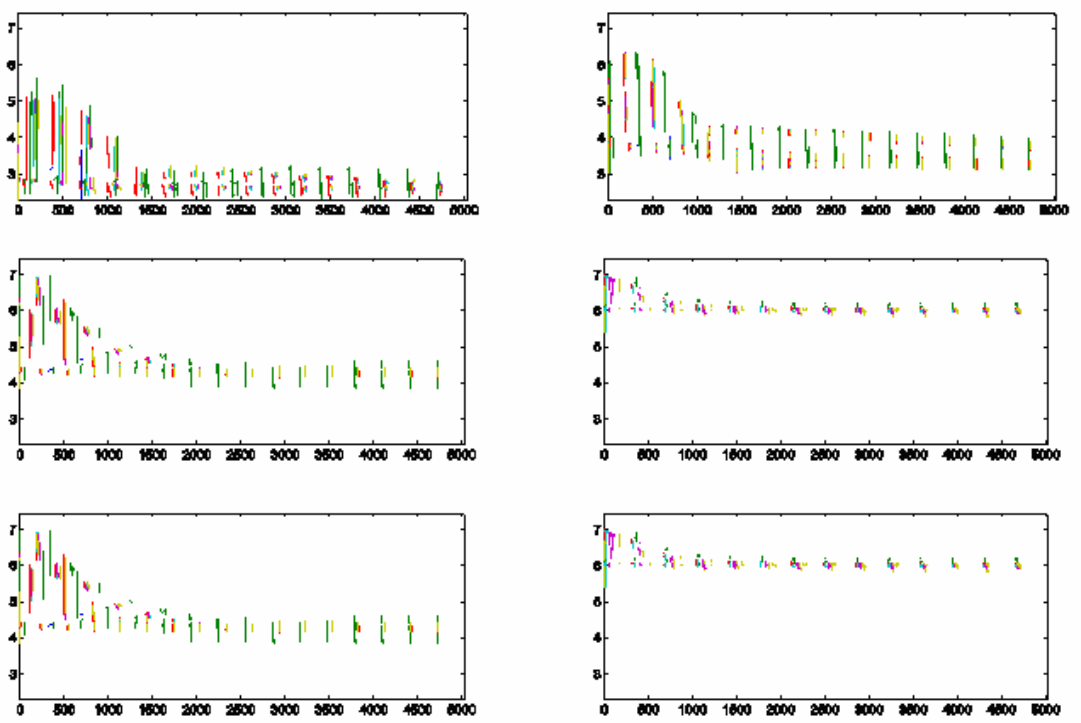

Figure 1: Distribution entropy of a population with different scaling between the mutation rate and the fitness

The top left axes contain results for a system where the mutation rate is very low. It gets progressively higher to the right and down, until it is distinctly in the high mutation rate limit in the lower right. Each graph contains three families of curves, with each family related to simulations started with distinct initial conditions. The individual curves within a family represent simulations with different timescales for the alternation of the two fitness landscapes. The large transient deviations to distribution entropy values characteristic of much higher mutation rates seen with the initiation of landscape dynamics observed at low mutation rates are very nearly the only universal behavior we have observed in this study.

The absorbed population for a pair of fitness landscapes with different relative mutation rates are illustrated in Figure 2. Mutation rates range from lowest (upper left) to highest, (lower right) reading left-to-right, then top-to-bottom. The blue lines in each group of curves represents the absorption for the unswitched landscapes starting with different initial conditions. The transiently high absorption observed in some of the low mutation rate traces is a fairly common, but not universal observation for this parameter regime. Conversely, the overall deviations from the unswitched behavior at higher mutation rates seem more predictable. The appearance of traces with different asymptotic absorption rates, as seen in the three graphs with the lowest mutation rates (compare the slope of the green traces in each of these graphs with the other-colored lines), is a common behavior for these parameter values, and appears to be confined to low mutation rates.

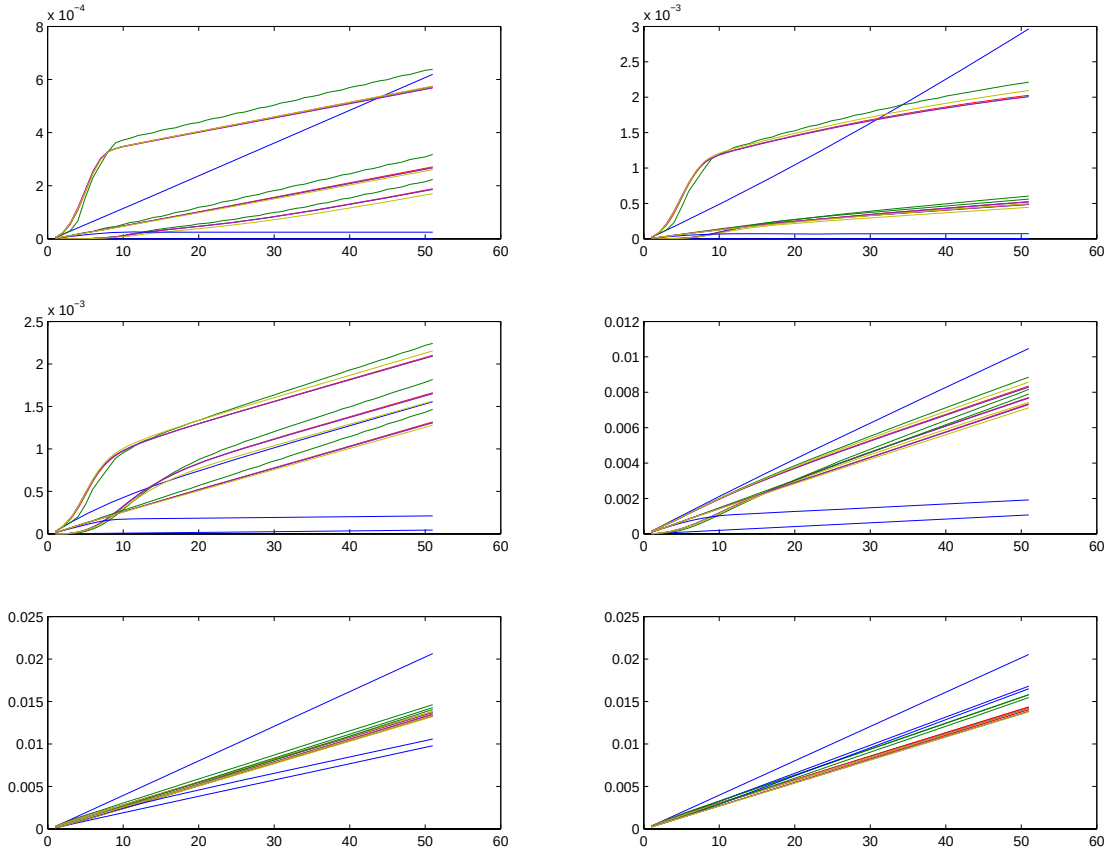

Figure 2: The absorbed population for a pair of fitness landscapes with different relative mutation rates

Beyond the unifying feature of genotype dispersal we have observed very diverse reactions of the evolving systems to fitness landscape dynamics. Ironically, it is the absence of certain behaviors that stands out. Namely, except in the low-mutation rate regime where it seems to have some bearing (along with initial conditions) on which of potentially several attractors a system finds, the particular rate of the dynamical treatment seems to have very little effect, merely that it *becomes dynamic* starting at a certain time seems sufficient to greatly modify the behavior in a stereotyped way (stereotyped for a certain instance of landscape and mutation matrix, that is).

The following few figures give an idea of the range of behaviors we have observed, and our preliminary attempts to generalize about the behavior of the system. Figure 3 contains plots of different simulations in the slow mutation limit.
Very diverse behavior is observed there. Multiple attractors for the population distribution limit cycle seem to be the rule. Initial transients representing anomalously high absorption were observed more than once, but it is difficult to generalize about this behavior with such a small data set. Figure 4 contains plots of different simulations with moderate mutation rates. The multiple limit sets characteristic of low mutation rates seem to have disappeared in this parameter regime. That said, the deviations of asymptotic absorption rate from the "limiting" cases representing non-dynamic treatment seem to have the capacity to be quite large. Figure 5 contains plots of different simulations in the

fast mutation limit. The limiting behavior of the dynamic treatment models seems to fall neatly within the limits set by the non-dynamic fitness landscapes (blue lines).

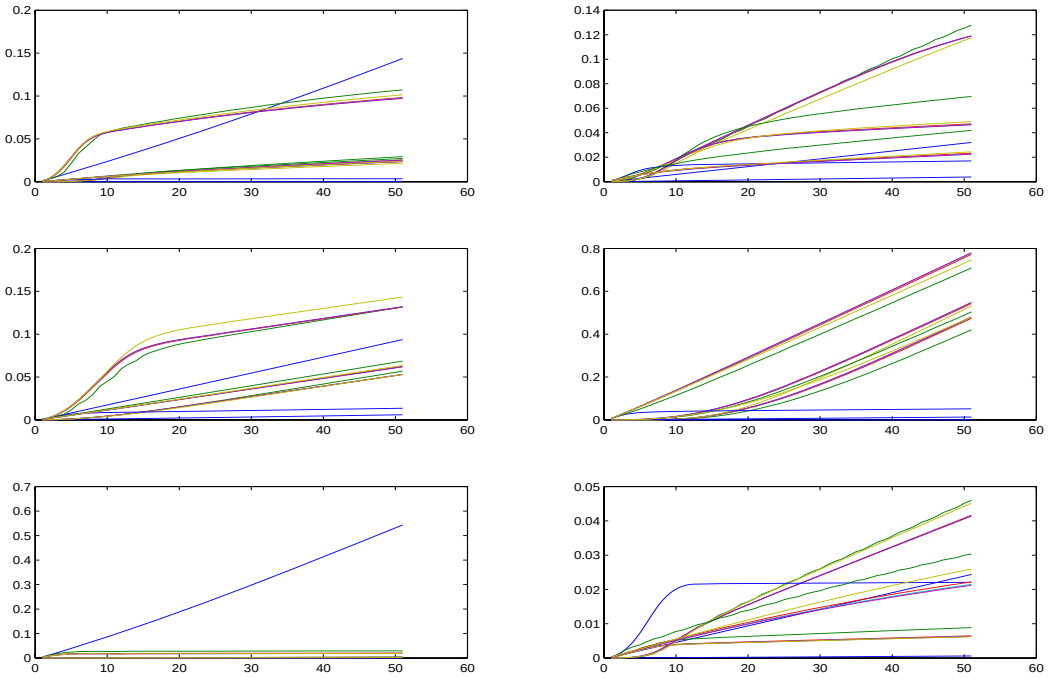

Figure 3: Plots of different simulations in the slow mutation limit

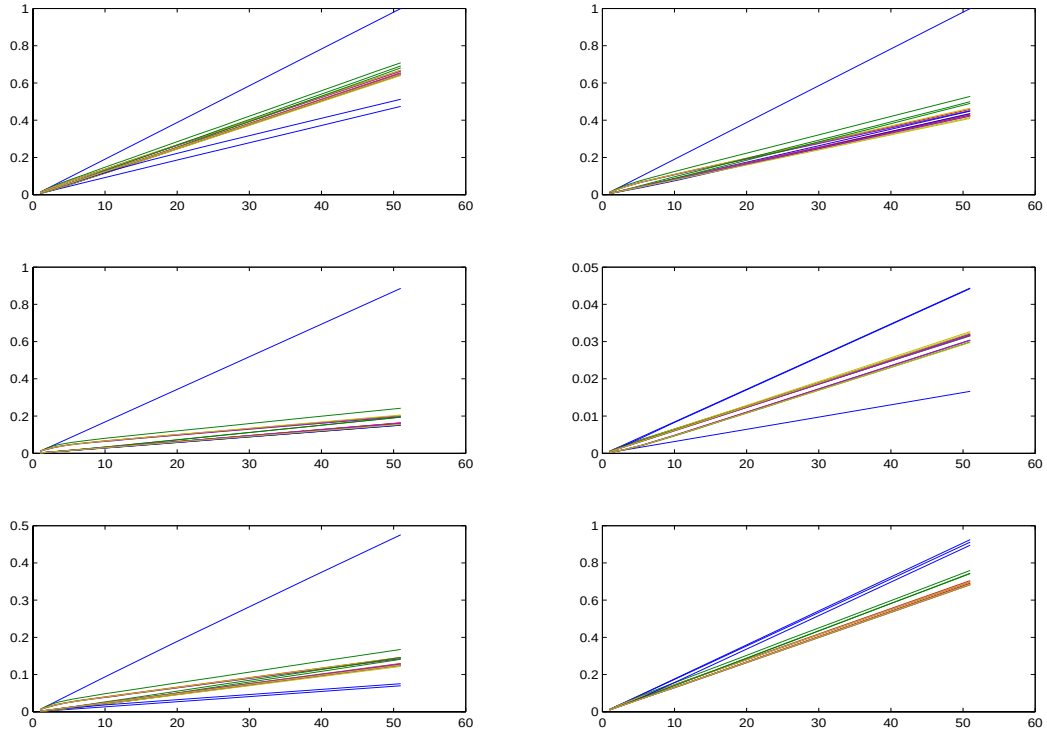

Figure 4: Plots of different simulations with moderate mutation rates

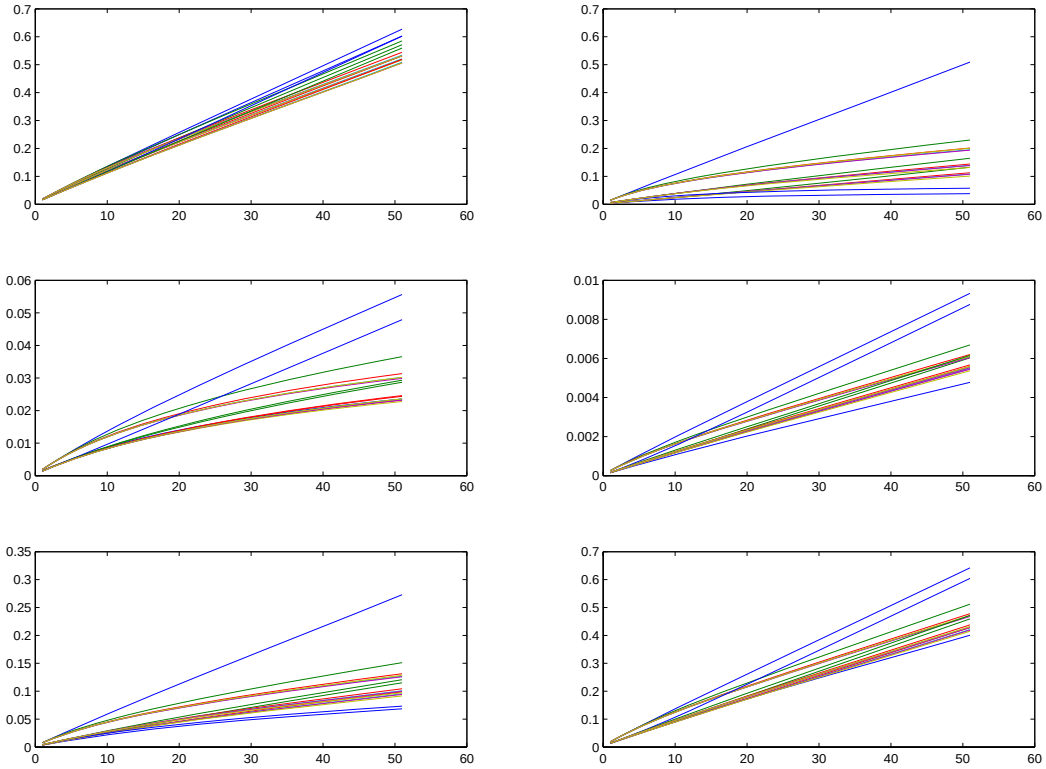

Figure 5: Plots of different simulations in the fast mutation limit

3.2 Agent-based stochastic model

In order to examine the problem from a different angle, we also implemented an agent-based model with finite autonomous cells interacting with the landscape. This allows investigation into the effects of stochasticity as well as discrete dynamics on the evolution of neoplastic populations. Using agent-based modeling also allows us to easily build in heterogeneous and dynamically changing mutation rates.

Individual cells track the fitness value of their present locations, which they combine with a competition term, proportional to the total population of cells, to determine the probability they will clone themselves, stay neutral, or die. They also are able to mutate a single gene at any one time step with a probability determined by a mutation rate parameter. We did not incorporate preferential genetic drift into this model for the sake of simplicity. Mutations can also only occur in small steps, not longer range ones as with the deterministic model. Mutation rates were made dynamic by coupling them to a single gene (g1) and its instantaneous value. Runs were executed with both dynamic mutation rates as well as unvarying ones in order to examine the effects of this.

The difference between the nondynamic mutation rate case versus the dynamic mutation rate case is striking, especially when looking at the number of cells which had been absorbed, i.e. reached the malignant state, over time. For the nondynamic mutation rate simulations, the number of absorbed cells over time steadily increases at a fairly constant

rate, consistent with the deterministic case. We cannot really discern a trend in the rate of cells reaching malignancy state as a function of frequency of landscape switching due to high variability between trials. Shown below (Figure 6) are two trial runs (trial 1 first row, trial 2 second row) which seem to depict almost exact opposite dependencies on frequency. Unfortunately proper statistics could not be run due to time and computational limitations. However we can discern a qualitative difference in the behavior between the dynamic and nondynamic mutation rate cases. For the dynamic mutation rate case, we do not see a constant linear absorption of cells into malignancy as we do with the nondynamic case. Instead we see that the trend is linear at first, until plateauing at a fairly constant value for the rest of the run.

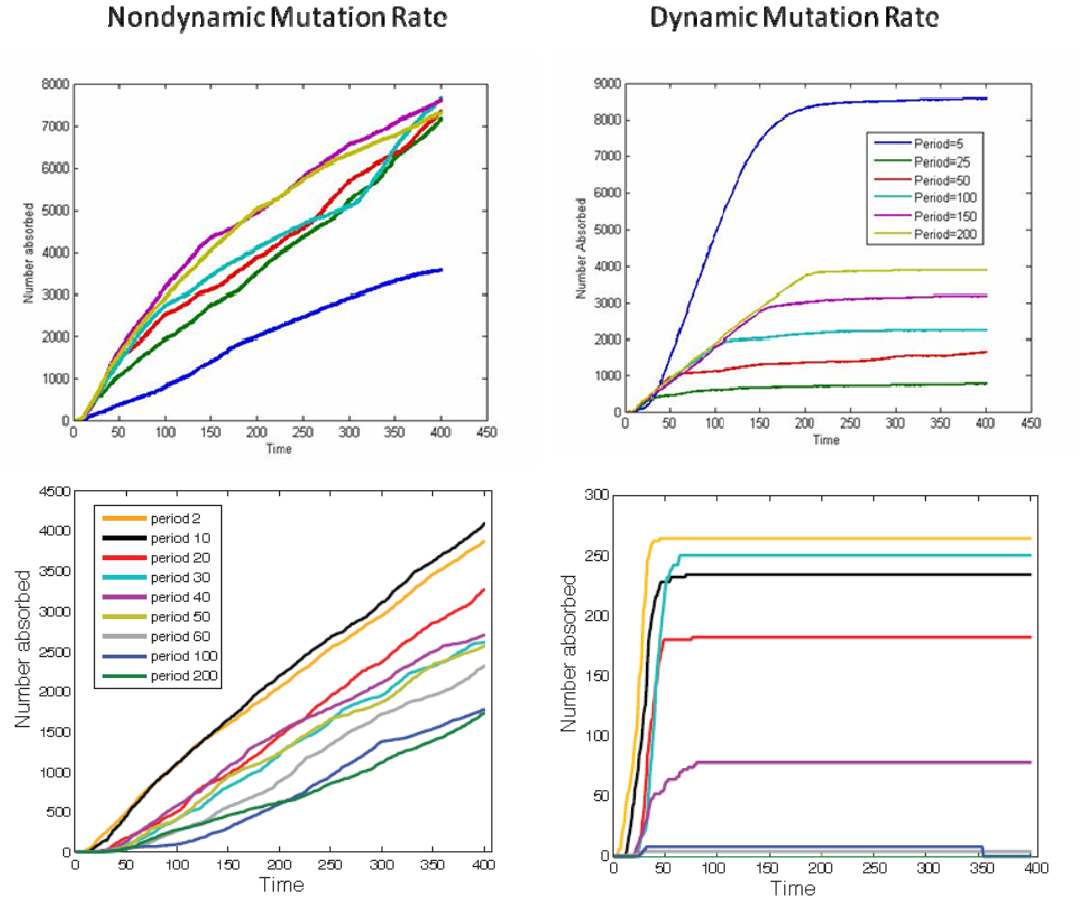

Figure 6

For the dynamic mutation case, we still see the general trend of more cells reaching malignancy state for higher frequencies of switching, but the number of cells over time abruptly plateaus after a short time. The absolute number of cells absorbed is also much smaller for the dynamic mutation rate case. It is unclear at this time what exactly is the reason behind this, although there is a very high level of heterogeneity among the mutation rates of cells at the end of the runs. Calculating the entropy of these two cases shows no significant difference (data not shown), meaning that a dynamic mutation rate isn't necessarily contributing to higher clustering of cells at high fitness landscape peaks, so it's unclear what exactly is the source of the low levels of absorption. Furthermore, the overall population of cells is both significantly higher stabilizes more quickly in the

dynamic mutation rate case.

From Figure 7, we can see that the dynamic mutation rate case promotes higher total population levels for all frequencies of switching between landscapes. While, Figure 8 and 9 illustrate that dynamic mutation rates also promote overall higher levels of average fitness for the cells, as well as lower standard deviations in this fitness. Additionally, from Figure 10, cells with dynamic mutation rates tend to occupy fewer total spaces in the fitness space than the nondynamic mutation rate case. This can be interpreted as more cells tending to cluster together at high fitness peaks when the mutation rate can be adjustable.

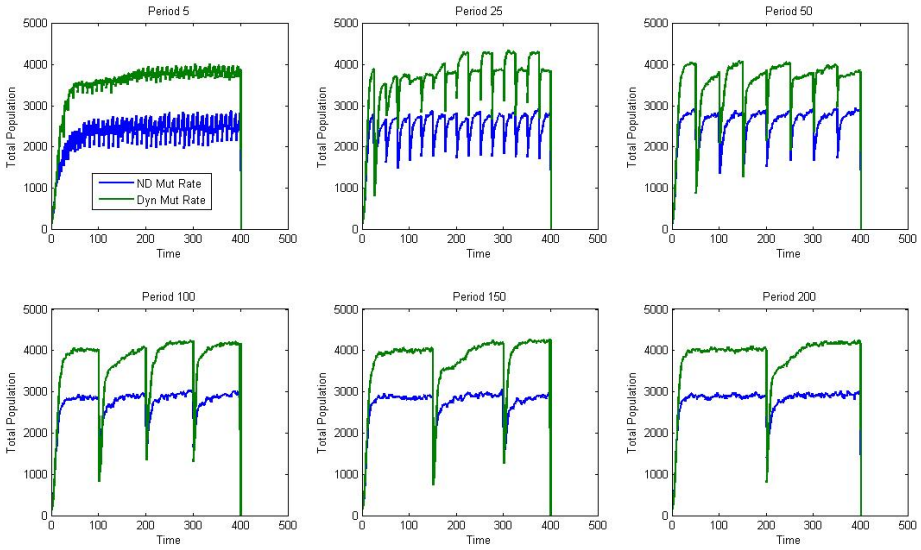

Figure 7: Total population levels

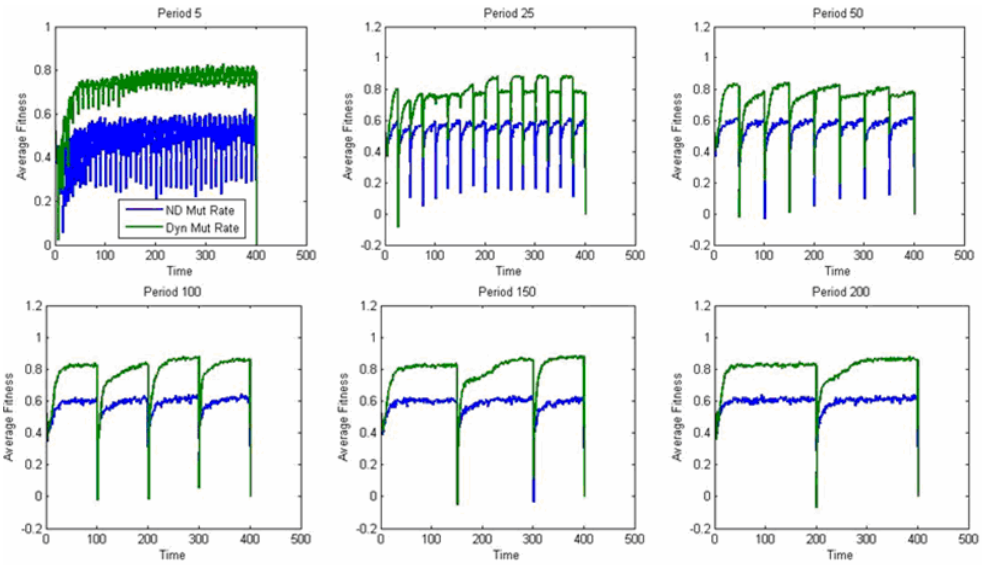

Figure 8: Average fitness value

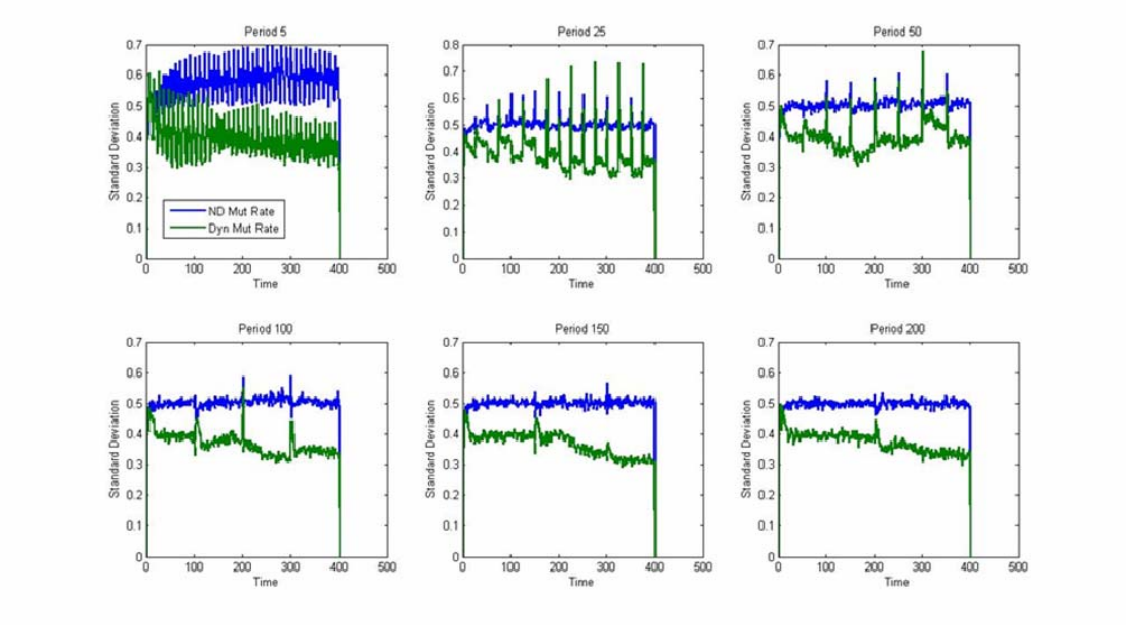

Figure 9: Standard deviations of fitness

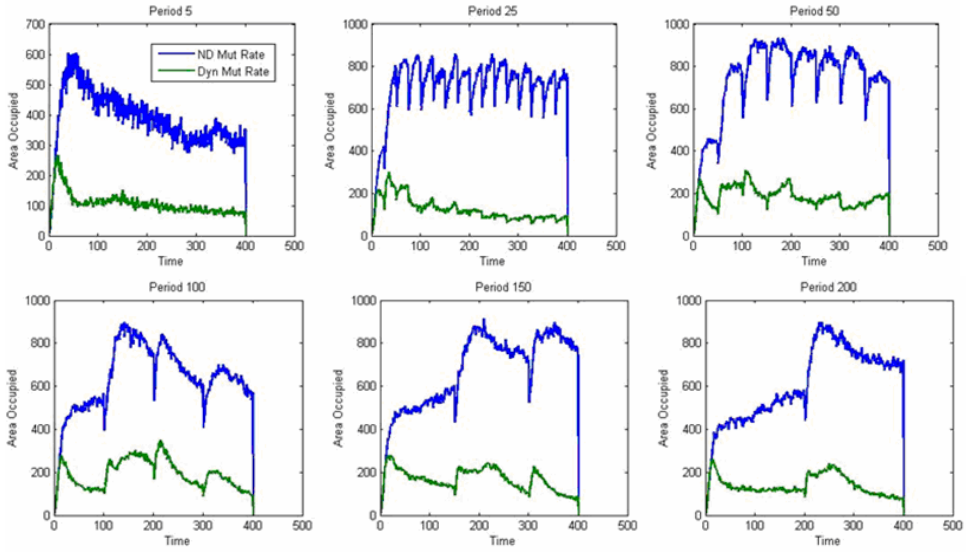

Figure 10: Dispersions of cells over landscape

An interesting result arose when we plotted the distribution of mutation rates at the end of a dynamic mutation rate run. From Figure 11, we see that for the higher frequencies, we get a fairly wide distribution of mutation rates, while for the lower frequencies we do not conserve higher mutation rates throughout the run. The initial mutation rate for all cells in this run was 0.3. This seems to suggest that a high frequency of landscape switching promotes the existence of a limited amount of cells with fairly high mutation rates, or possibly just a heterogeneous mix of mutation

rate

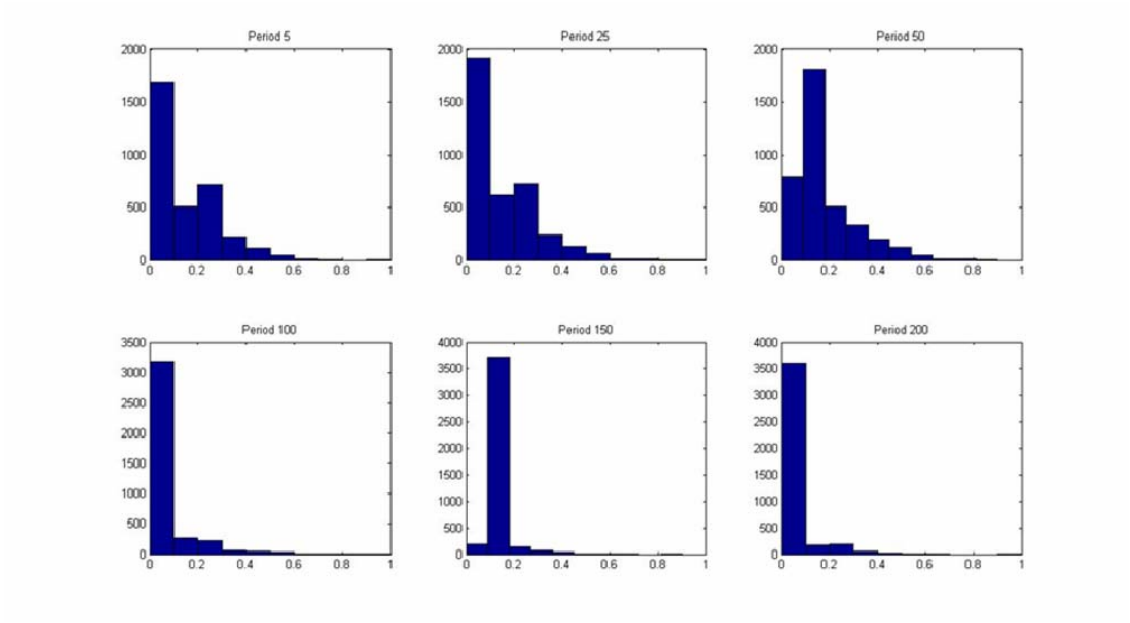

Figure 11: Distribution of mutation rates

The dynamics of this model are sensitive to both the competition term as well as the mutation rate. A too-high competition rate can lead to a population that dies out within a few generations. The mutation rate determines the probability that a given cell will mutate one of its genes at any one time step, and if this parameter is too low, a steady state system can be observed consisting of almost all cells being clustered exclusively at high fitness peaks. This usually results in no cells being absorbed due to lack of exploration of the gene space.

We can see that heterogeneity in mutation rates allows cells to survive switching landscapes by clustering on high fitness peaks. Unfortunately we were not able to discern a relationship between the frequency of landscape switching and the rate of cells reaching malignancy due to the high variability between runs and computational and time constraints. As the number of cells increases, the computational time needed to run the simulation increases linearly because each cell needs to be updated.

5. Discussions

To attempt to specific conclusions from this work would be unacceptably speculative. The one thing that is perfectly clear is that initiating large dynamic change to the fitness landscape of our simple model representing neoplasm evolution can lead to extreme changes in population dynamics, and very diverse behaviors. Going into this work, we predicted two main competing effects would come to play in the case of dynamic landscapes. First, we expected that since the peaks of maximum fitness would be periodically destroyed and relocated, the average population would be decreased by some amount, which would generally decrease the rate at which the neoplastic cells could convert to malignancy (be absorbed). Second, since the reduced population at peaks of high fitness would no longer be able to effectively suppress the population at points with

lower fitness, we would expect a compensatory expansion of population at low fitness, effectively broadening the distribution. We predicted that a broader population distribution would lead to more rapid conversion to malignancy, because the population would be increased at points that are more closely associated with the absorbing state (which is unrelated to the fitness structure). The question, as we originally framed it, involved exploring how these effects interacted in detail as a function of an easily controlled free parameter, ω.

Indeed, we observed both of these effects, as predicted. They do not appear, however, to explain the changes of flux through the system in any easy to summarize way. The question is opened, then, as to whether there is another behavior-dominating effect we haven't thought of controlling the system? Or have we simply been confused by the statistics of small sample sizes, and the two effects listed above are sufficient for understanding under at least some conditions? One phenomenon that seems potentially interesting is the behavior in response to dynamical treatment observed at low mutation rates. In addition to this being the physiologic limit of greatest interest, this case exhibits the greatest richness of behavior. The presence of multiple evolutionary attractors that can be accessed by dynamic treatment, and the similarity of the transitional states to higher-mutation rate simulations are both of potential clinical interest.

Juxtaposing the stochastic model next to the deterministic one provides some interesting insight into the effects of heterogeneity in mutation rates. In general, however, the two models cannot be directly compared because of a handful of significant modeling differences. It is especially difficult to relate the two parameter spaces to each other because one is associated with continuous variables while the other is purely stochastic and discrete. In addition, the addition of drift into the deterministic model precludes a direct analogy between the two. More exploration of this topic is required to be able to drawn meaningful conclusions from these observations.

Acknowledgements:

This collaboration was initiated at the 2008 China Complex Systems Summer School, co-sponsored by the Santa Fe Institute and the Institute of Theoretical Physics of the Chinese Academy of Sciences. The research and education programs of SFI are supported by core funding from the U.S. National Science Foundation and by gifts and grants from individuals, corporations, other foundations, and members of the Institute's Business Network for Complex Systems Research. SFI received direct support for the 2008 China CSSS from the US NSF under award OSIE-0623953.

References

1. Albertini, R.J., Nicklas, J.A., O'Neill, J.P., Robison S.H. *In vivo* somatic mutations in humans: measurement and analysis. *Annu. Rev. Genet.* **24**, 305-326 (1990).
2. Araten, D.J. *et al.* A quantitative measurement of the human somatic mutation rate. *Cancer Res.* **65**, 8111-8117 (2005).
3. Duesberg P, Li R, Fabarius A, Hehlmann R. The chromosomal basis of cancer. *Cell Oncol*. **27**, 293-318 (2005).
4. Frank, S.A. and Nowak, M.A. Problems of somatic mutation and cancer. *Bioessays*. **26**, 291-9 (2004).
5. Loeb, L.A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res*. **51**, 3075-3079 (1991).
6. Maley, C.C. *et al.* Genetic clonal diversity predicts prgression to esophageal adenocarcinoma. *Nature Genet.* **38**, 468-473 (2006).
7. Merlo, L.M., Pepper, J.W., Reid, B.J., Maley, C.C. Cancer as an evolutionary and ecological process. *Nature* **6**, 924-935 (2006).
8. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23-28 (1976).
9. Renan, M.J. How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol. Carcinogenesis* **7**, 139-146 (1993).
10. Stoler, D.L. *et al*. The onset and extent of genomic instability in sporadic colorectal tumor progresion. *Proc. Natl Acad. Sci. USA* **96**, 15121-15126 (1999).