

Modeling somatic evolution in tumorigenesis

Ryan Gerety^{*,†}, Sabrina L. Spencer^{‡,†,||}, Kenneth J. Pienta[¶], Stephanie Forrest^{*,§}

December 28, 2005

*Department of Computer Science, University of New Mexico, Albuquerque, NM 87131-0001

‡Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

¶Department of Medical Oncology, University of Michigan School of Medicine, Ann Arbor, MI 48109

§Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

† These authors contributed equally to this work.

|| To whom correspondence should be addressed. E-mail: spencers@mit.edu.

Abbreviations: SG, self-sufficiency in growth signals; IA, insensitivity to anti-growth signals; EA, evasion of apoptosis; LR, limitless replicative potential; SA, sustained angiogenesis; GI, genetic instability.

Abstract

Tumorigenesis in humans is thought to be a multistep process where certain mutations confer a selective advantage, allowing lineages derived from the mutated cell to out-compete other cells. This paper analyzes the computational implications of cancer progression presented by Hanahan and Weinberg in *The Hallmarks of Cancer*. We model the complexities of tumor progression as a small set of underlying rules that govern the transformation of normal cells to tumor cells. The rules are implemented in a stochastic multistep simulation model. The model predicts that (i) early-onset cancers proceed through a different sequence of mutation acquisition than late-onset cancers, (ii) tumor heterogeneity varies with the predominance of *genetic instability*, tumor category, and selective pressures during tumorigenesis, (iii) there exists an optimal initial telomere length which lowers cancer incidence and raises time of cancer onset, and (iv) the ability to initiate angiogenesis is an important stage-setting mutation, which is often exploited by other cells. The model offers insight into how the sequence of mutations acquired affects the timing and cellular make-up of the resulting tumor and how the cellular-level population dynamics drive neoplastic evolution.

1 Introduction

Cancer progression is a form of somatic evolution in which certain mutations give a cell a selective proliferation advantage [1]. Evidence strongly supports mutation as one of the dominant factors in setting rate-limiting steps in tumor progression, resulting in variation in the timing of progression between tumors [2]. Tumorigenesis is thought to require four to six stochastic rate-limiting mutation events to occur in the lineage of one cell [3, 4, 5]. Hanahan and Weinberg suggest that six cellular alterations, or

hallmarks, collectively drive a population of normal cells to become a cancer [6]. The six hallmarks are (i) self-sufficiency in growth signals (SG), (ii) insensitivity to anti-growth signals (IA), (iii) evasion of apoptosis (EA), (iv) limitless replicative potential (LR), (v) sustained angiogenesis (SA), and (vi) tissue invasion and metastasis. A specific genetic mutation may contribute only partially to the acquisition of a single hallmark capability, whereas in other cases, a single mutation may confer several hallmark capabilities at once. Genetic instability (GI) is defined as an “enabling characteristic” that facilitates the acquisition of other mutations due to defects in DNA repair signaling. These hallmarks form a candidate set of rules that underlie the transformation of a normal tissue to a cancerous one. The quantitative ramifications of these rules are explored in this paper, and lead to a number of interesting phenomena and hypotheses.

We model a simplified view of cancer progression using a stochastic model of tumorigenesis based on the hallmarks. The complexity of cancer cannot be understood by considering individual mutations independent of their interactions. Rather, the effect of a mutation often depends on other mutations within the same cell, on other mutant cells within the tumor, and on the tumor microenvironment. Our model allows us to study the evolutionary dynamics of early mutations, which generally go undetected in clinical settings, and thereby study the initial forces that drive neoplastic evolution.

2 Materials and Methods

This paper extends earlier work using the hallmarks of cancer to study cancer progression [7, 8]. Here, we model the hallmarks phenotypically (one mutation, one hallmark) in a three-dimensional agent-based model [8]. The computational grid contains a maximum of 10^6 cells initialized with a single normal cell and a single blood source. Successive cell divisions occur with a frequency described by a uniform distribution ranging between 5 and 10 time steps. We define the tissue such that angiogenic and growth factor constraints will prevent normal tissue from exceeding 8.6×10^4 cells. We define the onset of “cancer” to occur when the tissue grows to occupy 90% of the grid, or 9×10^5 cells, at which point the run terminates. In sections 3.1 and 3.2, we performed 1000 runs of the model with the default parameter settings described below, of which 986 end in cancer before 50,000 timesteps.

Normal cells require mitogenic growth signals to proliferate; cancer cells generate many of their own growth signals and are less dependent on exogenous stimulation [6]. In the model, normal cells divide only if they are within a pre-defined spatial boundary that represents a growth factor concentration threshold. The acquired mutation *self-sufficiency in growth signals* (SG) allows cells to proliferate regardless of the concentration of growth factor.

Contact inhibition is a type of anti-growth signal which stops overcrowding by preventing cell division

if a cell is surrounded by many other cells. In the model, each cell has up to 26 neighbors (neighbors need only make contact at a corner). Normal cells do not divide if all 26 positions are occupied by other cells. In the model, mutation in *insensitivity to anti-growth signals* (IA) allows cells to divide even when there is no space for the daughter cell. The daughter cell competes for survival with a randomly chosen neighbor with a $1/g$ chance of success. We selected g to be 5.

Premalignant and malignant cell populations undergo apoptosis [6], often in response to signals such as DNA damage and oncogene overexpression [9]. In the model, cells are checked for the presence of mutations each cell division, and if mutation is detected, the cell is eliminated. The probability of detecting genetic damage in the cell is n/e , where $n=0,\dots,6$ is the number of mutations carried by the cell. An informal sensitivity analysis led us to chose e to be 20, such that this parameter plays an important role but does not dominate tumorigenesis. The *evasion of apoptosis* mutation (EA) allows a cell with mutations to avoid mutation detection and the subsequent death that would normally occur.

In cell culture, telomere shortening limits normal human cells to 25-70 doublings [6, 10], a feature that curbs tumor growth. However, unlike normal tissues, 80% of human cancers show expression and activity of telomerase, an enzyme that is able to lengthen telomeres, conferring limitless replicative potential [11]. In the model, the initial cell is created with telomeres of length t , where t is 50 unless otherwise specified. To simplify the model, we do not consider the process of senescence, which usually occurs after cells have undergone a number of doublings. Under certain conditions, senescence can be circumvented, enabling cells to continue multiplying until they enter “crisis,” a state characterized by massive cell death [6]. When telomere length reaches zero in the model, we consider the cell to be in crisis, and it dies. The mutation *limitless replicative potential* (LR) allows the cell to divide indefinitely.

Cells cannot survive at distances of more than $100\mu m$ from a blood supply, which limits the size of tumors to about 10^6 cells without angiogenesis [12]. However, cancer cells have the ability to signal for angiogenesis when the concentration of nutrients and oxygen is low. In the model, lack of oncogenic angiogenesis limits the tissue to 8.6×10^4 cells; the mutation *sustained angiogenesis* (SA) confers the ability to signal for growth of new blood vessels.

The spontaneous mutation rate in human cells is estimated to be in the range of 10^{-7} to 10^{-6} mutations/gene/cell division [13]. The loss of DNA repair signaling can increase the mutation rate by a factor ranging from 10^1 to 10^4 [14]. In the model, cells that gain the mutation *genetic instability* (GI) have their base mutation rate, m , scaled up by i . We chose to use midrange values, defining i as 10^3 and m as $1/5.5 \times 10^6$.

We do not model tissue invasion and metastasis, as the model represents a single tissue. Additionally, there is a $1/a$ probability of random cell death due to non-specific causes, where a is 1000. The source

code that implements the model is available online (<http://cs.unm.edu/~forrest/software/cancersim>).

3 Results

3.1 Pathways to cancer vary with the timing of cancer onset

The standard perspective on tumorigenesis suggests that cells derived from the same lineage acquire multiple mutations in discrete steps [6]. We define the sequence of mutation acquisition as a “pathway” to cancer. With six different mutations possible and a variable number of mutations within each cell, there are $6! \sum_{i=0}^5 \frac{1}{i!} = 1956$ distinct sequences of mutation history. We are interested in the pathways cells take to become a tumor, and how particular pathways affect the dynamics of tumorigenesis. It is currently believed that it is not simply the accumulation of mutations, but also the order in which they are acquired, that determines the timing and extent of tumorigenesis [6, 15]. Our model allows us to study the sequence in which mutations are accumulated. We find that the pathways leading to early onset cancer (arising early in the life of the simulated tissue) are different from the pathways leading to late onset cancer.

The importance of genetic instability in cancer progression is a controversial issue in cancer biology. Loeb and others have argued that early acquisition of a mutator phenotype is necessary for tumorigenesis [16, 17], while Tomlinson, Bodmer, and others believe that increased cell division provides sufficient opportunities for mutation accumulation [18, 19, 20]. Our model proposes a compromise between these two viewpoints.

We find that *genetic instability* is often the causative mutation in earlier onset tumors, while *limitless replicative potential* is often the causative mutation in later onset tumors (Figure 1, Figure 2A). Figure 2A shows that very early tumors (before 2000 time steps) frequently result from acquiring *genetic instability* as the first mutation. This can also be seen in Figure 1, where the majority of the tumors that developed in the 0-5K timeblock have *genetic instability* as their first mutation. This is consistent with the fact that many early onset cancers are the result of inherited mutations in genes affecting the stability of the genome. These include xeroderma pigmentosum, ataxia telangiectasia, Nijmegen breakage syndrome and Bloom syndrome [20].

As indicated by Figure 1, a tumor is unlikely to form early (prior to 5000 timesteps) without either the increase in mutation rate associated with *genetic instability* or a rapid proliferation of cells. For example, in the rare chance that a cell on the tissue boundary acquires *self-sufficiency in growth signals*, the subsequent rapid clonal expansion will provide ample opportunity for further mutations to be acquired. This is illustrated by the prevalence of *self-sufficiency in growth signals* as a first mutation in early

tumors.

Later onset tumors in the model, however, are initiated by the acquisition of *limitless replicative potential* followed by either *insensitivity to anti-growth signals* or *evasion of apoptosis* (Figures 1 and 2). For later onset tumors, the many rounds of cell division offer sufficient opportunity for mutation accumulation. However, by limiting the life-span of the cell, telomere shortening prevents acquisition of a sufficient number of mutations. Without *limitless replicative potential*, mutations can be accumulated for only a limited number of cell divisions. The acquisition of *limitless replicative potential* immortalizes cells, allowing them to maintain a mutational line over the lifetime of the tissue. Clinical evidence supports our finding of limitless replicative potential as an early event in tumor development, demonstrated by the presence of telomerase activity in many precancerous lesions (reviewed in [21]). *Evasion of apoptosis* remains an important mutation to gain early in the pathway, regardless of when the tumor arises, as it prevents detection of acquired mutations. Conversely, *genetic instability* becomes a relatively rare mutation as time progresses (Figure 2B).

The model underscores the importance of accumulating the necessary mutations in a *single* cell. Due to the low probability of acquiring multiple mutations in an individual cell, *genetic instability*, which increases the mutation rate, and *limitless replicative potential*, which allows cells to live long enough to acquire multiple mutations, were either the first or second mutation in the most common pathways of all but three of the 986 observed tumors (data shown in supporting information). Thus, the model suggests the following compromise in the mutator phenotype debate: *genetic instability* plays a key role in early-onset tumors, while for later tumors, *limitless replicative potential* provides sufficient opportunity to acquire an equal number of mutations by allowing an increased number of cell divisions.

3.2 Heterogeneity varies with the predominance of *genetic instability*, tumor category, and selective pressures during tumorigenesis

Tumor heterogeneity has important consequences for the progression and treatment of cancer [22]. Our model makes three predictions: (i) tumor heterogeneity is strongly correlated with the predominance of *genetic instability* within the tumor; (ii) tumors with distinct causative mutation pathways exhibit distinct levels of heterogeneity; (iii) during tumor progression there is a marked increase in heterogeneity followed by a less dramatic decline.

Although there are several possible measures of heterogeneity within a tumor, we define tumor heterogeneity to be the average pathway distance between all cells of a given tumor. Pathway distance is measured as the number of steps backward through a pathway necessary to reach a common mutation. For instance, a cell with pathway *LR EA GI SG* and a cell with *LR EA IA SG* would have a distance of

2. By choosing the pathway as the level of analysis, rather than the set of mutations disregarding order, we capture the process of mutation acquisition.

Our results suggest that tumors initiated by *genetic instability* early in their pathways are more heterogeneous than other tumors ($p < 0.001$, $R^2 = .43$, ordinary least-squares regression). These results are consistent with evidence that points to *genetic instability* as a source of heterogeneity [23]. Once the predominance of *genetic instability* is accounted for, the time of cancer onset is not a significant determinant of heterogeneity.

We find that the five most common tumor categories, as defined by their most common initial two mutations, have statistically distinct levels of heterogeneity ($p < 0.01$, t-test). We find that only *LR IA* and *SG LR* are not significantly different. The boxplot in Figure 3B provides the distributions of heterogeneity among the various tumor groups. Interestingly, *GI LR* and *LR GI* tumors have the highest levels of heterogeneity. The increased level of heterogeneity is significant ($p < 0.001$, OLS) even when accounting for the early occurrence of *genetic instability*.

During tumorigenesis, there is a marked increase in intra-tumor heterogeneity. In the majority of tumors, the level of heterogeneity peaks, declines, and then rises again as the tissue becomes a tumor. Figure 3A illustrates this pattern within three tumors. This finding suggests that rapid proliferation allows for increased diversity upon which selection can act to choose the fittest cells, resulting in a decrease in heterogeneity.

Finally, we find that earlier onset tumors are more diverse than tumors arising later, clearly illustrated in Figure 1. Later tumors are frequently dominated by the same first two mutations, while earlier tumors contain a larger variety of first two mutations.

3.3 Initial telomere length affects cancer onset time and incidence

Limited replication protects against indefinite proliferation of mutant cell lines. Lengthened telomeres can promote tumorigenesis by allowing for the cell divisions necessary to develop a tumor. However, short, dysfunctional telomeres also promote tumorigenesis by creating chromosomal instability [21].

Analogously, we find that both long (90-170 units) and short (25-40 units) telomere lengths lead to higher incidence and earlier time of cancer onset than do intermediate lengths (Figure 4). In addition, there exists a trade-off between low early incidence and higher late incidence of cancer. The initial tumorigenic effects of long telomere length are due to the large replicative potential of the cells, but as the cells continue to divide, their telomeres reach an intermediate length where there is lower cancer incidence. After more cell divisions, the cells' telomeres become very short, causing cell death. This increases cancer incidence in the model because additional mutations can occur during the cell divisions

that replace the exhausted cells.

The initial telomere length that minimizes incidence at any particular point is a function of the tissue’s age. Figure 4B reveals that, with a life-span of 21900 timesteps, a telomere length of 55 corresponds to the lowest incidence of cancer, and a length of 50 produces the latest mean onset time. Thus, there exists in the model an optimal range of initial telomere lengths, which lowers cancer incidence and raises time of cancer onset. This raises the possibility that evolution plays a role in optimizing telomere length.

3.4 Angiogenesis sets the stage for tumor growth

The makeup of the final tumor does not necessarily reflect the complete evolutionary dynamics by which it was produced. *Sustained angiogenesis* is an example of a stage-setting mutation that is essential for tumorigenesis, but which can be exploited by other cells. Once the blood supply is established, angiogenic cells without other selective advantages can experience limited clonal expansion or be driven to extinction. Limited clonal expansion is caused by competition from other cells proliferating into the the newly vascularized region, often preventing additional angiogenesis.

In some cases, the clonal expansion of cells with *sustained angiogenesis* is not only impeded, but the population completely recedes (Figure 5A). Here, the initial proliferation of a population of *LR SA* cells provides for the expansion of a population of *LR* cells and then declines. A second rise in this population of *LR SA* cells then provides for the proliferation of *LR IA* cells, which “free ride” on the *LR SA* population, quickly driving the population of angiogenic cells to extinction. The decline of this *LR SA* population considerably slows tumorigenesis.

Figure 5B illustrates the formation of a new evolutionary niche by a population with *LR IA EA SA*, allowing neighboring *LR IA EA* cells to proliferate as well. The population of *LR IA EA* cells initially expands as a blood supply becomes available. However, as angiogenesis is completed within the region, the selective advantage of the mutation disappears and the angiogenic cells are constrained by other cells. In this situation, the *LR IA EA SA* cells creating the blood supply have three additional mutations, which allow them to compete more successfully than the *LR SA* cells did in Figure 5A. Thus, the angiogenic population is able to maintain itself rather than be driven to extinction. Due to the fact that a few cells with a *sustained angiogenesis* mutation provide a collective benefit to all cells nearby, and that once a blood supply is established the angiogenic cells lose their selective advantage, this mutation rarely occurs in the pathways composing the final tumor (Figure 2).

3.5 Sample simulation run

Due to the stochastic nature of the model, each run displays different cell population dynamics. Here, we describe the dynamics of a representative run for the parameters described in the methods section.

Several populations of *insensitivity to anti-growth signals* cells independently arise and die out in the early stages of the run (Figure 6A). At timestep 8888, a cell acquires *limitless replicative potential* and begins to spread (Figure 6A). This is the first cell lineage to arise that remains in the tissue until the end of the run, and thus can be considered a “causative” tumorigenic mutation. At timestep 20176, another cell independently develops *limitless replicative potential* and begins to spread. Snapshots of the tumor at timestep 20780 show these two separate colonies (Figure 7A). Shortly after timestep 23420, four populations of cells with *LR EA* emerge within the first population of *limitless replicative potential* cells. The slight fitness advantage conferred by the mutation allows for a slow clonal expansion and a decline in the parent *LR* population (Figure 6B). During this expansion, a cell with *limitless replicative potential* gains *insensitivity to anti-growth signals* and begins to proliferate. Although the *insensitivity to anti-growth signals* mutation alone was previously unsuccessful, the combination of *limitless replicative potential* and *insensitivity to anti-growth signals* allows these cells to proliferate rapidly. Quickly, the population with *LR IA* begins to dominate other cells. This expansion allows several *LR IA* cells to independently gain *evasion of apoptosis*, becoming *LR IA EA* cells. Ultimately, this population takes over the normal tissue and replaces the previous mutant populations (Figure 6C). Many cells acquire additional mutations, but they remain limited in population size.

The population of cells expands beyond the normal tissue boundary when one of the cells gains *sustained angiogenesis* and signals for vascularization outside the tissue’s natural extent, allowing the population to proliferate in a previously unoccupied region. The *LR IA EA* cells exploit the angiogenesis, successfully competing for the new habitable region. Clonal expansion is limited until a cell near the bounded region acquires *self-sufficiency in growth signals*. When this occurs, the population of *LR IA EA SG* cells expands (Figure 6C), and once this population gains *sustained angiogenesis*, the tumor grows rapidly in size until it reaches 9×10^5 cells at timestep 29272, and the run ends.

Figure 3A shows the increased diversity in the last stages of tumor progression. The pathway heterogeneity within the tissue gradually increases until timestep 25000 (Figure 3A), at which point there is a dramatic increase in heterogeneity corresponding to the beginning of a gradual increase in the number of mutant cells (Figure 8[1]). During tumorigenesis, the tissue experiences a number of rate-limiting steps, shown in Figure 8. Each passage through a bottleneck corresponds to acquisition of a mutation in the sample run. The tumor initially grows within the normal tissue’s natural extent after acquisition of *insensitivity to anti-growth signals* [1]. The acquisition of *sustained angiogenesis* allows for

rapid expansion beyond this region [2]. This growth is limited until a population with *self-sufficiency in growth signals* clonally expands [3], which is in turn limited until further *sustained angiogenesis* mutations occur [4]. In general, *genetic instability* and *limitless replicative potential* are frequently the causative mutations in early and late onset cancers respectively, but *sustained angiogenesis* and *self-sufficiency in growth signals* are key for providing transitions through proliferation bottlenecks. Targeted therapies in cancer treatment are based on identifying molecular bottlenecks and exploiting them for therapeutic intervention. For example, the vascular endothelial growth factor (VEGF) inhibitor bevacizumab is an antibody that binds to the growth factor and prevents it from binding to its receptor. Bevacizumab is approved for use in colon cancer and is a prototypical example of a targeted therapy based on a tumorigenesis bottleneck [24].

4 Discussion

The key findings from this model are: (i) Early-onset cancers proceed through a different sequence of mutation acquisition than late-onset cancers. Specifically, *genetic instability* is the most common first mutation in early-onset cancers whereas *limitless replicative potential* is the most common first mutation in late-onset cancers. (ii) Heterogeneity varies with tumor category, the predominance of *genetic instability* and the selective pressures during tumorigenesis. (iii) There exists an optimal initial telomere length which lowers cancer incidence and raises the time of cancer onset. (iv) The ability to initiate angiogenesis is an important stage-setting mutation, which is often exploited by other cells and is therefore infrequently present in final tumors.

This model presents a first step toward tracking early precancerous mutations computationally. Early events responsible for neoplastic progression are difficult to investigate experimentally for the very reason that they have not yet been detected due to lack of symptoms. The main limitation of our model is the difficulty of experimental validation, which would be necessary to demonstrate its clinical applicability. A thorough testing of the model would require periodically testing single cells for the presence of mutations in the hallmark categories, beginning before an animal has developed a clinically detectable cancer. Initiating testing for mutations once the animal has a palpable tumor ignores early mutation dynamics that, according to this model, are important for determining the timing and cellular makeup of the tumor that develops. In addition, the current technology of sequencing populations of tumor cells for various mutations at various time points does not provide explicit support of the multistep model of tumorigenesis, which requires multiple mutations to be found in a single cell.

The work presented here results from a number of simplifying assumptions, and thus represents a model rather than a complete description of tumorigenesis. The main assumptions relate to tissue

architecture and cellular interactions. For instance, the modeling of anti-growth signals is limited to contact inhibition. Further, we assume that all mutations fit into one and only one of the six hallmarks, whereas in reality, p53 is known to be involved in cell cycle inhibition, apoptosis, genetic stability and inhibition of blood-vessel formation [25]. Additionally, while we have represented mutations as a binary switch, many of these mutations, especially genetic instability, would be more accurately represented on a continuum. Similarly, we overlook the process of senescence as an intermediate step before cellular crisis. Future work could address these issues and could also extend the model to include cellular signal transduction pathways, invasion and metastasis, and cancer stem cells.

While simplifications are inevitable in a theoretical model, this work nevertheless reveals important consequences of the well-established qualitative concepts proposed by Hanahan and Weinberg [6]. This type of model can reveal unexpected nonlinear interactions, such as those we find among the six hallmarks. The model provides novel insight into the early dynamics of neoplasia currently inaccessible to experimental investigation and thus serves as an encouraging tool for hypothesis generation.

References

1. Cahill DP, Kinzler KW, Vogelstein B, Lengauer C (1999) Genetic instability and darwinian selection in tumours. *Trends Cell Biology* 9:M57–M60.
2. Frank S, Nowak M (2004) Problems of somatic mutation and cancer. *BioEssays* 26:291–299.
3. Nowell PC (1976) The clonal evolution of tumor cell populations. *Science* 194:23–28.
4. Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* 8:1–12.
5. Renan MJ (1993) How many mutations are required for tumorigenesis? implications from human cancer data. *Molecular Carcinogenesis* 7:139–146.
6. Hanahan D, Weinberg R (2000) The hallmarks of cancer. *Cell* 100:57–70.
7. Spencer S, Berryman M, Garcia J, Abbott D (2004) An ordinary differential equation model for the multistep transformation to cancer. *Journal of Theoretical Biology* 231:515–524.
8. Abbott R, Forrest S, Pienta K Simulating the hallmarks of cancer. *Journal of Artificial Life* In press.
9. Nilsson J, Cleveland J (2003) Myc pathways provoking cell suicide and cancer. *Oncogene* 22:9007–9021.
10. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, et al. (2002) *Molecular Biology of the Cell*. Garland Science.
11. Djojotubroto M, Choi Y, Lee H, Rudolph K (2003) Telomeres and telomerase in aging, regeneration and cancer. *Molecules and Cells* 15:164–175.
12. Folkman J (1990) What is the evidence that tumors are angiogenesis dependent? *Journal of the National Cancer Institute* 82:4–6.
13. Jackson AL, Loeb LA (1998) The mutation rate and cancer. *Genetics* 148:1483–1490.
14. Tomlinson IPM, Novelli MR, Bodmer WF (1996) The mutation rate and cancer. *PNAS* 93:14800–14803.
15. Kinzler K, Vogelstein B (1996) Lessons from hereditary colorectal cancer. *Cell* 87:159–170.
16. Loeb LA (1991) Mutator phenotype may be required for multistage carcinogenesis. *Cancer Research* 51:3075–3079.
17. Rajagopalan H, Nowak MA, Vogelstein B, Lengauer C (2003) The significance of unstable chromosomes in colorectal cancer. *Nature Reviews Cancer* 3:695–701.
18. Tomlinson IPM, Bodmer WF (1995) Failure of programmed cell death and differentiation as causes of tumors: some simple mathematical models. *PNAS* 92:11130–11134.
19. Tomlinson I, Bodmer W (1999) Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nature Medicine* 5:11–12.
20. Sieber OM, Heinimann K, Tomlinson IPM (2003) Genomic instability – the engine of tumorigenesis. *Nature Reviews Cancer* 3:701–708.
21. Hahn WC (2003) Role of telomeres and telomerase in the pathogenesis of human cancer. *Journal of Clinical Oncology* 21:2034–2043.
22. Shah RB, Mehra R, Chinnaiyan AM, Shen R, Ghosh D, et al. (2004) Androgen-independent prostate cancer is a heterogeneous group of diseases: Lessons from a rapid autopsy program. *Cancer Research* 64:9209–9216.
23. Misra A, Chattopadhyay P, Dinda AK, Sarkar C, Mahapatra AK, et al. (2000) Extensive intra-tumor heterogeneity in primary human glial tumors as a result of locus non-specific genomic alterations. *Journal of Neuro-Oncology* 48:1–12.
24. Ellis LM (2005) Bevacizumab. *Nat Rev Drug Discov Suppl*:S8-9.
25. Vogelstein B, Lane D, Levine A (2000) Surfing the p53 network. *Nature* 408:307–310.

26. Solé RV, Deisboeck TS (2004) An error catastrophe in cancer? *Journal of Theoretical Biology* 228:47–54.
27. Barrett M, Sanchez C, Prevo LJ, Galipeau P, Wong D, et al. (1999) Evolution of neoplastic cell lineages in barretts esophagus. *Nature Genetics* 22:106–109.

5 Figures

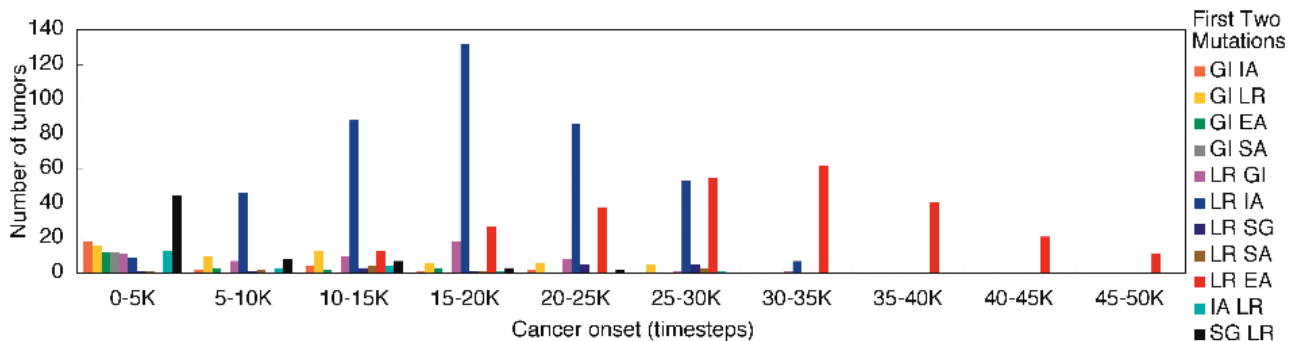


Figure 1: The timing of cancer onset is correlated with the most common first two mutations of a tumor. In the 0-5K timeblock, *GI* is the most common first mutation. The prevalence of *SG LR*, *GI IA*, *GI EA*, *IA LR* and *GI SA* sharply decreases after 5000 timesteps, while *GI LR* and *LR GI* remain roughly constant for several time blocks. The sequences *LR IA* and *LR EA* are distributed around 15-20K and 30-35K timesteps respectively. Pathways not shown occur in less than 1% of the tumors.

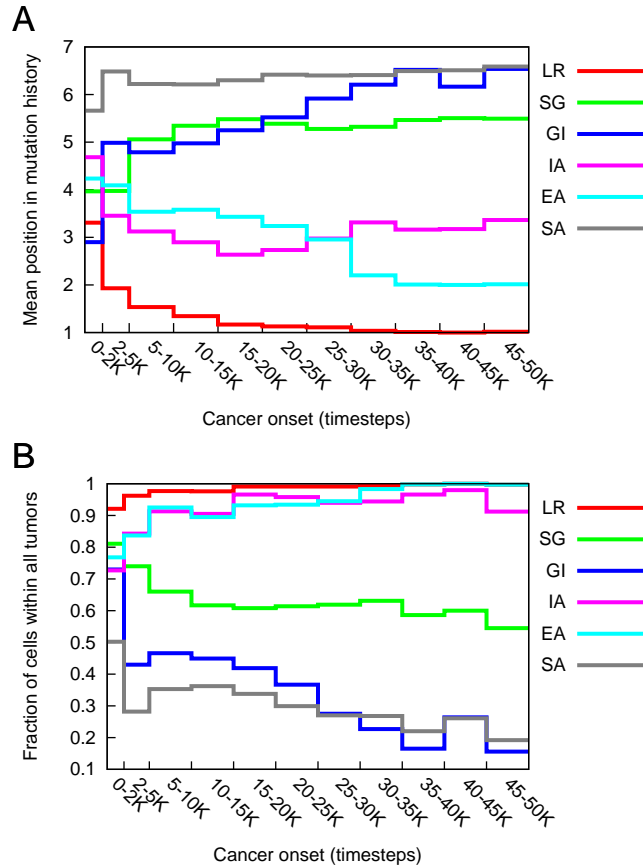


Figure 2: Pathways to cancer vary with the timing of cancer onset. (A) Average position of the mutations in the pathways present in the 986 runs that terminated in cancer, for tumors acquired in the time blocks specified on the x -axis. A position of one indicates that it was the first mutation of a pathway. A position of seven indicates that a given mutation did not appear in the pathway. For cancers arising before 2000 timesteps, *genetic instability* was frequently the first mutation, as it has the lowest mean position. As time progresses and telomere length shortens, *limitless replication* becomes the first mutation in all pathways. Standard deviation and sample size data are included in the supporting information. (B) Fraction of cells carrying a given mutation for all tumors acquired in the time blocks specified on the x -axis. Cells with mutations in *limitless replicative potential*, *insensitivity to anti-growth signals*, and *evasion of apoptosis* make up the majority of the tumors. The frequency of *genetic instability* drops substantially as time of cancer onset increases, whereas the frequency of *sustained angiogenesis* remains low for all time blocks.

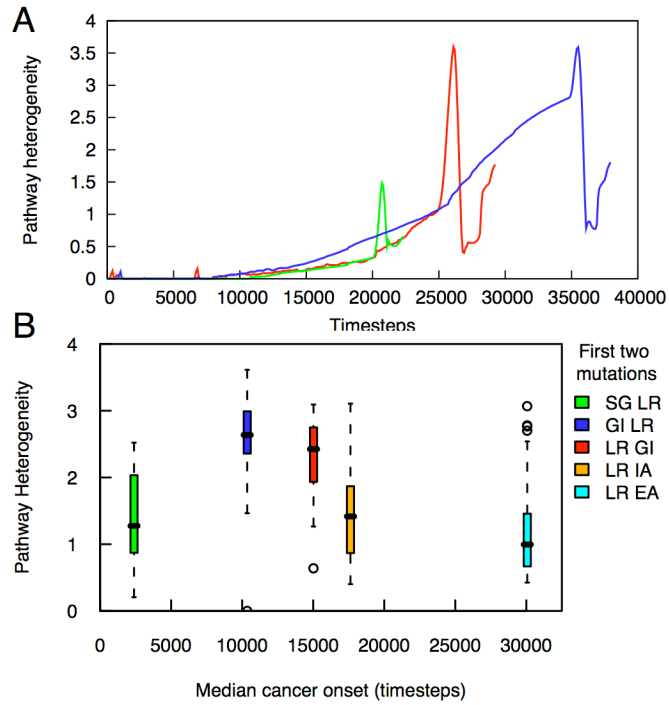


Figure 3: (A) Example dynamics of heterogeneity during the development of three tumors. Each line corresponds to the development of one tumor; the red curve is the tumor discussed in section 3.5. As the tumor begins to form there is typically a steady increase in the degree of tissue heterogeneity followed by a notable increase, an equally sudden decrease, and then often another increase as the tissue reaches 9×10^5 cells. (B) The five most common tumor categories, defined by the most common first two mutations, vary in their pathway heterogeneity.

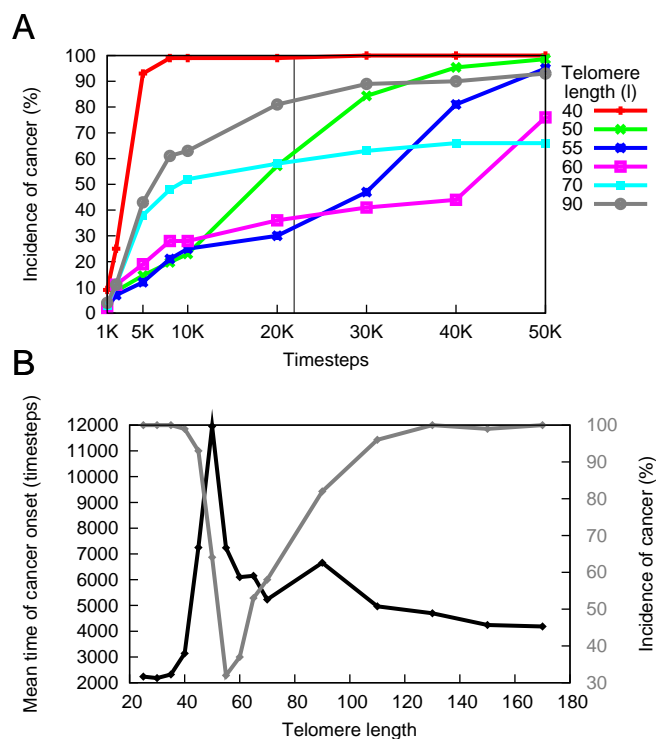


Figure 4: Initial telomere length affects incidence and onset of cancer. (A) Initial telomere length affects the pattern of incidence across time. Each point in time on the x -axis represents the cumulative incidence of cancers that arose before that time. The initial telomere length governs the tradeoff between the incidence of early and late cancer onset. Short (40 units) and long (90 units) telomeres produce an earlier, higher incidence of cancer than do telomeres of intermediate length. (B) Mean onset time and incidence for cancers acquired before 21900 time steps as a function of initial telomere length. This subfigure represents a snap shot at time $t = 21900$, indicated by the grey line in (A). Note that the gray line corresponds to the secondary y -axis.

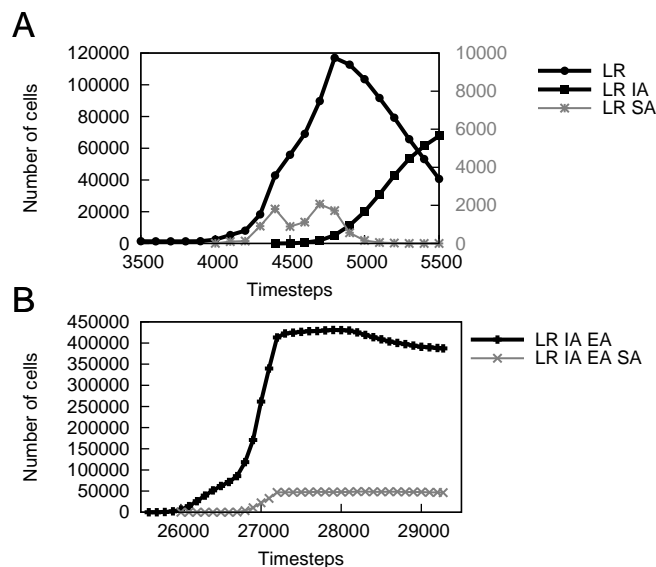


Figure 5: *Sustained angiogenesis* creates a niche for other cell populations. (A) A population of *LR SA* cells allows the population of *LR* cells to proliferate. Eventually, the *LR IA* cells replace the *LR SA* cells, temporarily preventing the development of new vasculature. (B) In this case, the cells creating the blood supply have additional mutations, allowing the *LR IA EA SA* population to plateau rather than decline.

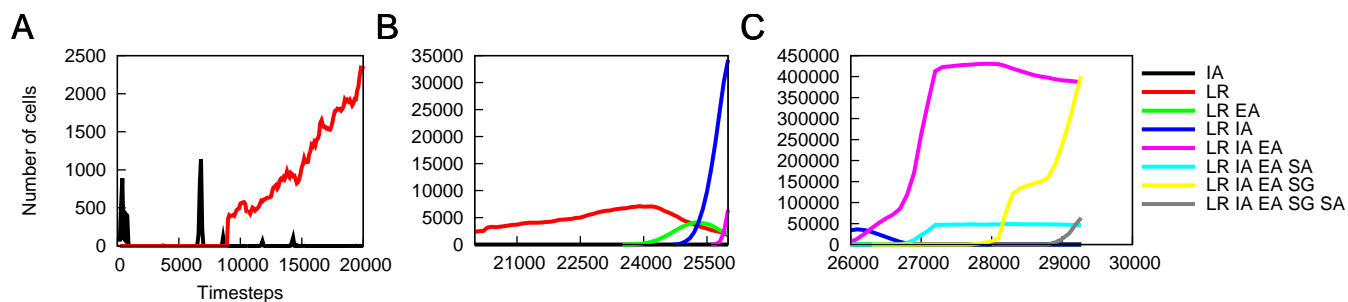


Figure 6: Growth dynamics of mutant cells in three graphs corresponding to three different time periods. (A) Populations with *IA* rise and fall, and cells with *LR* emerge. (B) Around timestep 24000, a cell with *LR EA* undergoes clonal expansion, resulting in a decline of the parent *LR* population. Near timestep 25000, cells with *LR IA* begin to out-compete the *LR EA* population. (C) At timestep 26000, cells with *LR IA EA* expand while the *LR IA* population declines. The emergence of a clonal population with *LR IR EA SA* provides the angiogenesis for the *LR IA EA* population to expand rapidly. At timestep 28000, *LR IA EA SG* cells begin to double the size of the tumor, aided by the *LR IA EA SG SA* cells.

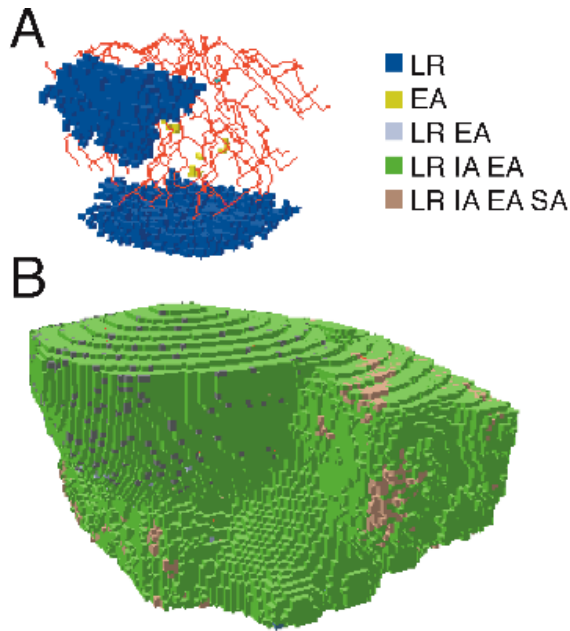


Figure 7: The tissue at various points during the run. (A) The tissue at timestep 20780. Normal cells have been removed to reveal two clonal populations with *LR*. In (B), the tumor at timestep 26982 has grown beyond the normal tissue extent through the acquisition of *SA*.

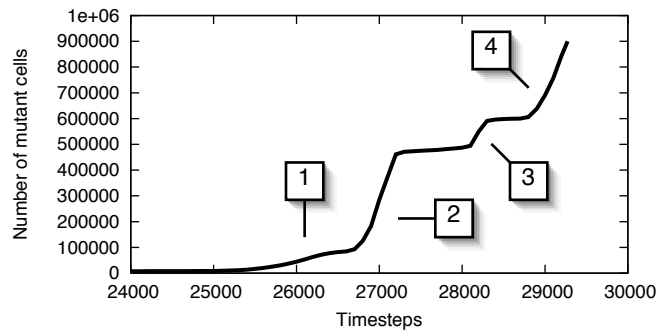


Figure 8: Cell proliferation pattern in mutant cells. The boxed numbers indicate clonal expansions that result from overcoming proliferation bottlenecks. The first expansion occurs with the acquisition of *insensitivity to anti-growth signals*. The second expansion is the result of a *sustained angiogenesis* mutation. The third occurs with acquisition *self-sufficiency in growth signals*, and the final expansion occurs with the acquisition of *sustained angiogenesis*.

6 Supporting Information

Table 1: The number of tumors as categorized by the most common first two mutations. Bold indicates tumors that do not have *LR* or *GI* as a first or second mutation.

Most common first two mutations	Number of tumors
LR IA	421
LR EA	268
SG LR	65
GI LR	56
LR GI	56
GI IA	27
IA LR	22
GI EA	20
LR SG	16
GI SA	12
LR SA	11
GI SG	4
SG GI	3
EA IA	1
IA EA	1
IA GI	1
SA LR	1
SG EA	1

Table 2: Mean, standard deviation, and sample size for data in Figure 2A

Timeblock	Average position (std)						# of tumors
	LR	SG	GI	IA	EA	SA	
0-2K	3.31 (0.38)	3.97 (0.46)	2.90 (0.49)	4.68 (0.59)	4.23 (0.56)	5.66 (0.67)	86
2-5K	1.93 (0.24)	3.97 (0.46)	4.99 (0.71)	3.46 (0.44)	4.09 (0.49)	6.48 (0.78)	60
5K-10K	1.54 (0.18)	5.06 (0.60)	4.79 (0.71)	3.12 (0.39)	3.54 (0.42)	6.22 (0.77)	84
10K-15K	1.34 (0.17)	5.35 (0.64)	4.97 (0.68)	2.89 (0.36)	3.58 (0.43)	6.21 (0.75)	148
15K-20K	1.17 (0.15)	5.48 (0.68)	5.25 (0.72)	2.64 (0.32)	3.43 (0.41)	6.30 (0.78)	194
20K-25K	1.13 (0.15)	5.39 (0.69)	5.52 (0.76)	2.73 (0.35)	3.24 (0.40)	6.42 (0.82)	147
25K-30K	1.11 (0.14)	5.28 (0.69)	5.92 (0.85)	2.97 (0.41)	2.96 (0.41)	6.40 (0.88)	124
30K-35K	1.04 (0.15)	5.32 (0.76)	6.21 (0.96)	3.31 (0.48)	2.20 (0.32)	6.41 (0.97)	70
35K-4K	1.01 (0.16)	5.47 (0.86)	6.52 (1.06)	3.16 (0.51)	2.01 (0.31)	6.49 (1.06)	41
40K-45K	1.00 (0.15)	5.51 (0.86)	6.16 (1.02)	3.17 (0.49)	2.00 (0.31)	6.51 (1.05)	21
45K-50K	1.02 (0.14)	5.49 (0.80)	6.54 (0.93)	3.36 (0.47)	2.01 (0.28)	6.59 (0.95)	11