

Concepts in solid tumor evolution

Arend Sidow and Noah Spies

Departments of Pathology and of Genetics, Stanford University, Stanford, CA 94305, USA

Evolutionary mechanisms in cancer progression give tumors their individuality. Cancer evolution is different from organismal evolution, however, and we discuss where concepts from evolutionary genetics are useful or limited in facilitating an understanding of cancer. Based on these concepts we construct and apply the simplest plausible model of tumor growth and progression. Simulations using this simple model illustrate the importance of stochastic events early in tumorigenesis, highlight the dominance of exponential growth over linear growth and differentiation, and explain the clonal substructure of tumors.

Evolution seems so familiar

Everybody knows what evolution is. Right? One hundred and fifty-six years after the first modern treatment of evolution [1], and 53 years after discovery of ‘the’ molecular clock [2], most biologists have an intuitive feel for evolution. Evolution is a popular subject for science education, it is taught with passion in college and graduate courses, and mentioning it in scientific papers is *de rigueur* even if the study has nothing to do with evolution. After all, (i) every single biological process is the product of evolution and, therefore, (ii) nothing in biology makes sense except in the light of evolution [3].

Despite its popularity as a subject, however, evolution is surprisingly difficult to study, and every field that begins to embrace evolutionary thought and gather evolutionarily-relevant data is faced with the tension between studying something so seemingly familiar and producing actually meaningful and novel insight. This tension is now emerging in cancer, where technological advances are enabling the generation of sufficient data to address which general evolutionary principles apply, and how cancer-specific mechanisms extend them.

Cancer repeats, the tree of life does not

The most obvious difference between cancer and organismal evolution is that the substrate of the former is a somatic cell, whereas the latter is a germline process. Cancer is a dead end whether or not it kills its host, whereas germline-based organismal evolution forever marches on. Natural selection over the billions of generations since the last common ancestor of all life has produced an enormous range of organisms, but there is only one tree of life – life on earth is the realization of a single experiment that will never repeat. By contrast, each type of

human cancer is an experiment that is repeated over and over with similar initial conditions among the trials and a limited range of possible evolutionary trajectories, constrained by the specific molecular biology of the originating cell and by the human genome. A discussion of driver mutations is presented in [Box 1](#).

Thus, in cancer, evolution really does repeat itself, and that is its Achilles’ heel: recurrently mutated genes are used in screening and diagnosis, inform treatment choice, and are potential targets for drug discovery. It also means that particular evolutionary principles are likely operational, some of which are already discernible at this early stage of our understanding of cancer.

Tumors are clonal

Two lines of evidence from tumor genome-sequencing studies support the view that all tumor cells (either primary or metastatic) are derived from a single ancestral cell that acquired proliferative potential [4]. The first is that alternate allele frequencies of single nucleotide variants (SNVs), and the prevalence of larger genomic aberrations such as aneuploidies, are frequently consistent with all tumor cells carrying the changes [5,6]. If cells within a tumor carry several shared somatic changes, then they must derive from a single ancestral cell that harbored these changes. This is because the probability of two or more cells independently acquiring the same set of somatic mutations is infinitesimally small.

The second, broader line of evidence supporting the presence of subclones and the clonal origin of metastases comes from phylogenetic studies of distinct samples in the same patients [7–11]. In these studies, somatic mutations serve as unequivocal lineage markers to construct bifurcating trees. Branch points represent cell divisions that gave rise to two proliferating daughter lineages that eventually expand into subclones.

Cancer cells are not a mating population and their genomes do not recombine

Cancer cells arise by mitotic division and therefore do not exchange genetic material, neither with each other nor with the cells of the microenvironment. A comparison to organismal life, which forms the basis for our intuition about populations, provides an illustrative contrast. Organismal species, including bacteria, constantly exchange genetic variation through mating or other mechanisms. This has the paradoxical effect of maintaining variation but binding the genomes of all individuals together such that as a whole the species evolves in concert.

Cancer is entirely different. Once a cancer cell divides, the two daughter genomes are on a future path of absolute

Corresponding author: Sidow, A. (arend@stanford.edu).

0168-9525/

© 2015 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tig.2015.02.001>

Box 1. Identification of driver genes

The limited range of evolutionary trajectories is the basis for the identification of cancer driver genes, now accomplished by large-scale sequencing projects that identify recurrently mutated genes. In many cancers, mutations in a limited number of genes act as the most common drivers [5,33–38]. Gain-of-function changes in oncogenes (e.g., ERBB2, KRAS, or IDH1) tend to be caused by a small set of mutations that recur, but as a class these are rarer than loss-of-function mutations in tumor-suppressors (e.g., TP53, RB, or VHL). Because there are very many ways to inactivate a gene (loss-of-function point mutations, frameshifts, deletions, etc.), searches for driver mutations have been broadened to the gene level and do not require the exact same nucleotides to be recurrently mutated [32,39]. These searches have revealed that many cancers exhibit mechanistic commonalities, mostly along the lines of tissue of origin, but some cross those lines as well [40].

independence (Figure 1A). Without mating and recombination, mutations that occur in one lineage have no bearing on another lineage. The variants carried by a cell occurred in a specific order in its ancestral lineage only, back to the zygote (Figure 1B). Similarly, the context in which a new mutation arises is the totality of all previous mutations across the entire genome that occurred before it (Figure 1C). As a consequence, some driver mutations will only provide an advantageous phenotype if the cell previously acquired another driver that allows the new driver to have an effect. This explains why mutations in some pathways tend to occur early in tumor evolution, and others late [11–13].

Cell autonomy rules

Driver mutations act cell-autonomously, that is, they confer a growth advantage on the cell in which they arose. That is how they make themselves more frequent. A non-cell-autonomously acting mutation confers a growth advantage on cells that do not harbor it, and therefore cannot cause an increase of its own frequency in the cancer cell

population. The vast majority of cancer-causing mutations that have been functionally studied exhibit cell-autonomous activity via processes such as intracellular signaling, cell cycle control, or cell death. Note that mutations that cause a tumor cell to interact with its microenvironment, for example via VEGF (vascular endothelial growth factor) to stimulate vascularization, are considered to be cell-autonomous because they provide their growth benefit to the originating cells.

There are some exceptions to the cell-autonomy rule, however, such as in prostate cancer, where non-cell-autonomous signaling by androgens is permissive for tumor development. Androgen-deprivation therapy can lead to remission but, eventually, the cancer overcomes its dependence by cell-autonomous mechanisms [14]. Similarly, anti-EGFR (epidermal growth factor receptor) antibody therapies in colorectal cancer have been shown to be eventually rendered ineffective by preexisting (cell-autonomously acting) *KRAS* oncogene mutations [15]. In a novel mouse cancer model, non-cell-autonomous drivers lead to unstable tumor growth and can result in tumor collapse [16]. Finally, the accumulation of genomic change that in some solid tumors culminates in hypermutation and chromothripsis [17,18], as well as the accumulation of driver mutations in the ancestral lineages of tumor cells, are difficult to reconcile with non-cell-autonomous mechanisms. These examples illustrate that non-cell-autonomous mechanisms exist and play a role in cancer, but it is cell-autonomy that make cancer the disease of relentless cell proliferation.

Cell autonomy has profound implications for our understanding of the relationship between subclones within a tumor. Much cancer literature implies or asserts that clones compete with one another – as if cells within a clone behave concertedly and ‘gang up’ on all the cells of another clone (e.g., [19]). Concerted behavior, such as slowing of growth rates in larger tumors due to reduced nutrient

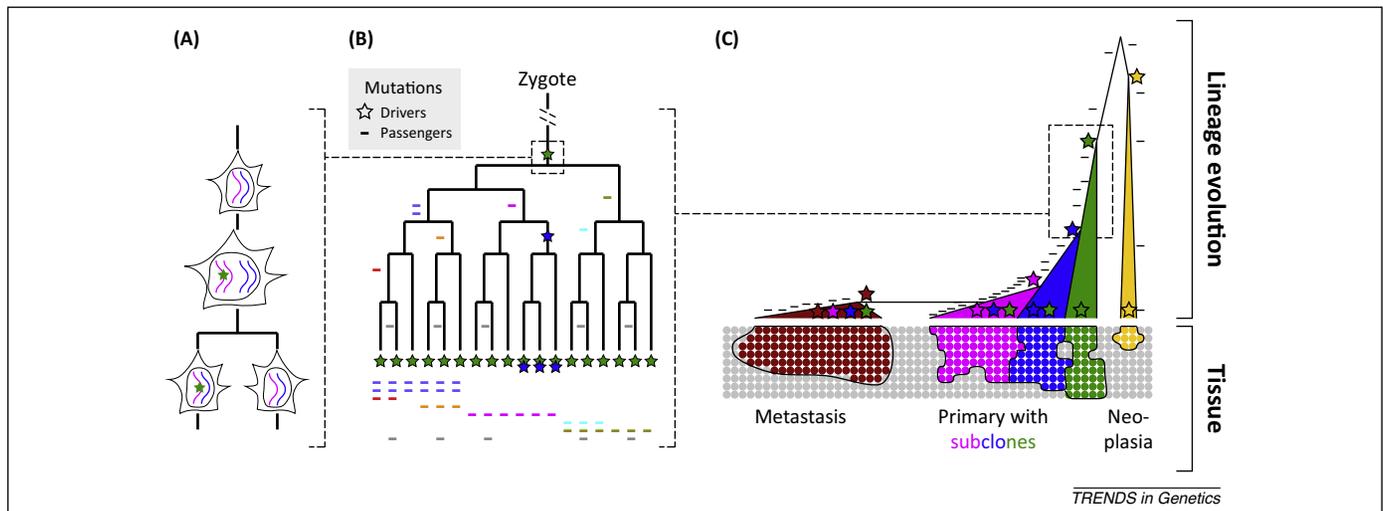


Figure 1. Passenger mutations serve as lineage markers, and driver mutations lead to subclones whose size is a function of the strength and timing of the driver. (A) Mutations (star) that arise during DNA replication will be passed on in a heterozygous state to one of the daughter cells. Mutations that arise during interphase (not shown), such as incorrectly repaired DNA damage, will likely be passed on to both daughter cells. (B) Mutations serve as lineage markers because they are passed on to only their progeny and are not horizontally transferred. (C) Drivers initiate proliferation and then increase the growth rate, and this leads to subclones (or metastases) that are marked by ever-increasing numbers of mutations. When the ancestor is old, as in slowly-growing neoplasias such as columnar cell lesions in breast, the number of mutations that are shared by all cells of the neoplasia is comparatively lower than the number of mutations in a fast-growing tumor whose ancestral cell is more recent. Note that, in the primary tumor, the subclones (blue and magenta) must grow more quickly to be detectable compared to the already existing and still-expanding original clone (green).

availability and the accumulation of toxic byproducts, is likely due to the microenvironment and not due to common ancestry. Cells of a clone behave independently from one another due to their cell-autonomy. The growth rates of the cells in any given clone may be similar because of common ancestry, and they may be greater than the growth rates of cells of another clone, but there is no reason to think that the cells of earlier clones stop growing and dividing merely because new clones arise.

Mutations do not fix; selection neither sweeps nor purifies

Clonal mitotic growth, cell-autonomy, and lack of recombination/mating combine to produce population genetic and evolutionary effects in cancer that are distinct from, and simpler than, those in germline evolution. Most importantly, the somatic variants present in the cancer cell ‘population’ are locked into each individual cell such that allele frequencies are dependent on ‘population size’ (the number of cells in a tumor). In organismal populations, allele frequency is generally independent of population size, but, in cancer, variants can only rise in frequency as a result of clonal (‘population’) growth, and ‘positive selection’ cannot take an allele to high frequency without increasing the number of cells that carry it. Similarly, genetic drift (the stochastic process that changes the frequency of neutral variants in the population) is largely irrelevant because neutral variants (passenger mutations) deterministically change frequency as a function of the cell growth caused by drivers.

Purifying selection, a useful term in population genetics describing the loss of an allele because of lower fitness, has little relevance to cancer growth because cells with a decreased growth rate trivially end up contributing negligible mass to a tumor. Perhaps cancer therapies can be thought of as agents of purifying selection, but the regimen is so different from the usual action of organismal purifying selection that there is little to be learned by the analogy. Organismal purifying selection acts to gradually reduce allele frequencies in a population, whereas the goal in cancer therapy is to achieve a 100% kill rate and eliminate the entire population of tumor cells.

The forced application of terms and concepts from organismal population genetics can distract from the fundamental simplicity of cancer evolution. For example, calling a driver an ‘advantageous allele’, or invoking ‘natural selection’ instead of clonal proliferation, can lead to unnecessary conceptual complications, such as in the case of discussions about ‘fixation of advantageous alleles’. Fixation refers to the rise of a variant in frequency in the population, to 100%. In cancer, fixation would mean that all cells of a tumor carry a particular somatic variant, such as a strong driver mutation. Nevertheless, because mutations in cancer generally work cell-autonomously, sibling cells lacking a new driver mutation continue replicating, albeit at the original slower rate. Understanding these dynamics without confounding concepts from organismal population genetics is key to understanding the reasons for tumor heterogeneity.

Tumor heterogeneity is clonal (and virtually infinite)

Cancer genome data unequivocally demonstrate that the bulk of the tumor (excepting stroma and other ‘normal’

cells) derives from a single ancestral cell, and that there are not parallel lineages originating from several initially normal cells that proliferate at a similar rate to give a heterogeneous mass. Instead, tumor heterogeneity is generated by occasional, individual tumor cells that acquire increased proliferative potential and whose greater growth rate leads to a mass within a mass (see above; [Figure 1C](#)). Among the cells of each clone there is additional genomic heterogeneity, as evidenced by single cell sequencing [20–22]. It is due, especially at later stages of tumor growth when mutator drivers have taken hold, to mutagenic potential in every cell division.

These two types of intratumor heterogeneity (among clones vs among cells within a clone) are best kept conceptually separate. The former embodies the proven proliferative ability as a result of drivers that caused the clones to grow, and, although drivers are vastly outnumbered by passengers, they are guaranteed to be present. In the latter, the private mutations present in individual cells of a clone are not enriched for drivers. Among the tremendous number of rare (i.e., not ancestrally shared) variants present in the billion or more cells that may make up a growth are mutations that do confer a proliferative advantage. However, overall the variation in an individual cell that is not shared with most other cells of a clone is highly enriched for irrelevant genomic changes that do not confer a proliferative advantage. One practical implication of this is that sequencing samples of larger numbers of cells may be preferable in many circumstances because it conveniently identifies those mutations that arose in their common ancestral lineage – which are comparatively enriched for the drivers that made those cells proliferate.

Heterogeneity is key to understanding the response of a tumor to drug treatment. Even in the absence of accelerated mutation rates, the genomic diversity present in the billions of cells of a moderate-sized tumor can harbor mutations conferring drug resistance, resulting in the high rates of relapse following chemotherapy typical of many types of cancer [23].

Models of tumor growth

Many mathematical models of tumor evolution have been proposed, of varying levels of complexity [24–28]. Our intent here is to explore the simplest possible model that gives interpretable results that are consistent with known dynamics of tumor evolution. Our uncontroversial premise is that, in all tissues of the adult, homeostatic (non cell-autonomous) mechanisms keep bulk cell growth in check such that the number of new cells roughly equals the number of cells that die. Cancer begins when the nonautonomous mechanisms suppressing cell growth are overruled by cell-autonomous changes (driver mutations) in the genome of the founding cancer cell.

Given that cells grow by binary division there are two opposite, extreme scenarios that bound plausible models of how tumors grow: we refer to these as the strict stem cell (SSC) model, and the perfectly proliferating cell (PPC) model. In the SSC model, division of a true stem cell regenerates a stem cell and one nondividing, differentiated daughter cell. In this model, one division produces one additional cell and growth is linear with the number of cell

divisions. By contrast, the PPC model is perfectly exponential where every daughter cell divides and the number of cells doubles in each mitotic generation. Both models have biological underpinnings. Stem cells exist in many natural systems, and the PPC occurs in cell culture.

Under the SSC, the number of cell divisions required to produce n cells is n . It would take a million consecutive cell divisions of one stem cell lineage to generate a 1 mm^3 neoplasm. This is impossible within a human lifespan. More generally, it is linear growth that makes the SSC or relaxed versions of it implausible. Allowing some limited proliferative potential of the non-stem daughter cells does not change the linearity of stem cell models. For example, if each of the non-stem daughter cells produces 100 progeny, the number of stem cell divisions would still need to be $10^6/10^2 = 10^4$ for a 1 mm^3 neoplasm, or 10^7 for a 1 cm^3 tumor. Thus, relaxing the SSC does not yield a plausible variant of the model, fundamentally because it is linear.

At the other extreme, the PPC model implies rapid tumor growth: a million cells would be generated by a mere 20 cell divisions. This in itself does not make the model implausible because the time between divisions could be very long, but cell death (or terminal differentiation) decreases the growth rate from a perfect doubling per generation, making the PPC in its extreme form implausible. However, its essential feature is exponential growth, a much more plausible dynamic than linear growth.

A simple and interpretable model of proliferation

Cell-autonomy and mitotic division suggest a biologically straightforward and parsimonious view of tumor cell divisions: tumor cells divide symmetrically to give rise to daughter cells that are very much like their mother cell, not in a deterministic but in a probabilistic sense. Daughter cells inherit their probability of division from their mother. This probability of division is cell-autonomously set by the genome and epigenome, and new drivers increase this probability. When homeostatic mechanisms are in place, an average daughter cell has a 0.5 probability of dividing again or dying. If instead of dying, the cell lives, growth is stem-cell-like and linear because only half of the daughters of a generation will divide again, exactly as in asymmetric stem cell divisions. At the other end of the extreme, a daughter that has a 1.0 probability of dividing again is a PPC because the cell number doubles in every generation. Let's call this probability f , for the fraction of cells in a generation that will go on to divide again. A similar, slightly more complex model has been shown to accurately model the progression of known tumor types [24].

First stochasticity, then determinism

Homeostatic mechanisms initially keep f at 0.5. The first driver must set $f > 0.5$, meaning that the next cell division has a somewhat higher probability to give two proliferating daughters rather than only one. However, the first driver occurs in a single cell, and there is a chance that this cell does not divide (equal to $1-f$), or, if it divides, that neither daughter divides [equal to $(1-f)^2$], in which case the driver is lost. Therefore, the chance that the driver is still around in the third generation is only $f[1-(1-f)^2]$, or, for example,

0.44 if $f = 0.55$. The consequence is that, if f is close to 0.5, it takes many independent occurrences for a first driver to succeed, as some simple simulations show. We explored the parameter space of f above 0.50 to find the number of tries it takes to eventually generate a neoplasm of 10^6 cells, or roughly 1 mm^3 in volume. These simulations show that it takes 25 tries at $f = 0.51$ and still almost three tries at $f = 0.6$ (Figure 2A). We found that a simple transformation of the probability of a driver's survival in a previous model [24] gives exactly the same results.

The universe of candidate driver mutations is composed of drivers with a range of effect sizes [29]. Some may tweak f ever so slightly, some much more. Further, the same driver mutation may well behave differently in different individuals, depending on genetic background. Thus, depending on genetic background and the effect of candidate driver mutations, cancers with a strong inherited component and some

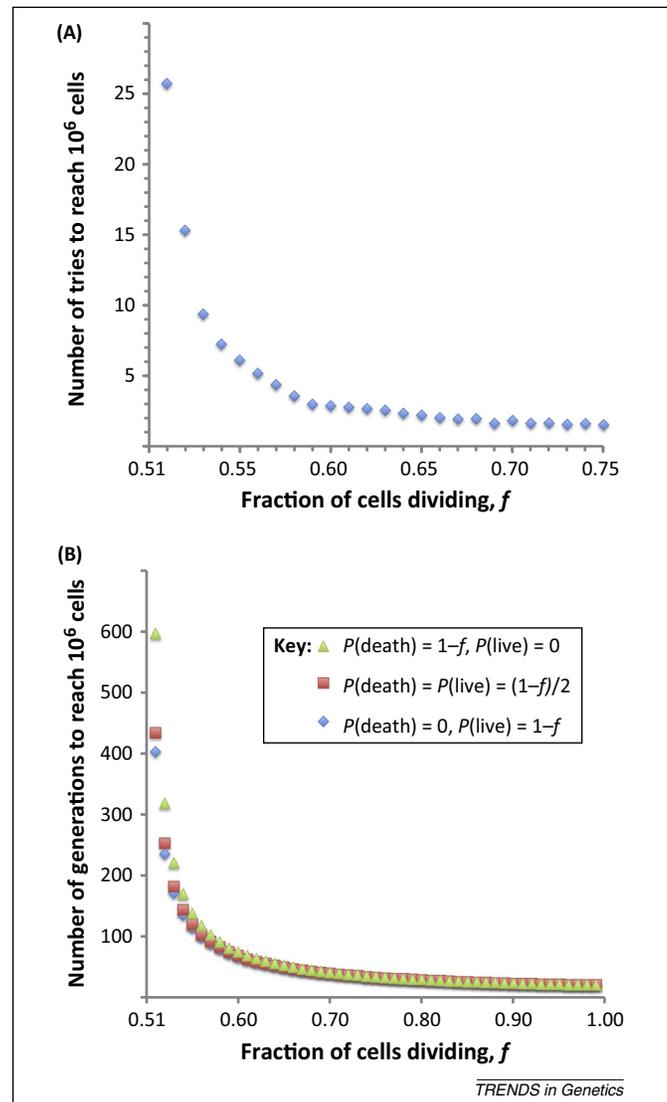


Figure 2. Simulations of tumor growth based on a simple model of proliferation. (A) Stochasticity in the initial phase of tumor growth, illustrated by the number of tries it takes (y axis) for a clone to grow to 10^6 cells (as opposed to becoming extinct), as a function of the strength of the initial driver mutation (x axis). (B) Results from a simulation exploring the effect of the balance between cell death and 'terminal differentiation' of nondividing cells. Only when drivers are weak (x axis) is there an appreciable difference in the number of generations it takes to reach 10^6 cells.

sporadic cancers with strong-effect drivers develop early, whereas some individuals without predisposition or better somatic luck will never develop a tumor.

Once a large enough number of cells is produced, the stochastic process of tumor growth gradually approximates a deterministic one because there are now sufficient cells to ensure that the population practically never becomes extinct. Furthermore, with an average $f \gg 0.5$ and a large number of cells, a tumor is eventually guaranteed.

To be or not to be, it appears to matter not

Now let us consider the path that has generated an incipient tumor mass of 1 mm^3 . Intuition suggests that f is not the only important parameter that determines how many cell generations it takes to produce the mass, but that it might also matter whether the nondividing cells live or die. Thus, the simple model of dividing versus nondividing cells can be extended to dividing versus dying versus nondividing cells, such that $f + d + l = 1$, where d is the probability that a cell dies and l is the probability that a cell lives but does not divide. We performed simulations where f again ranges from 0.5 to 1, under three different values of d : $d = 0$ (no death, all remain alive), $d = l = (1-f)/2$ (equal probability of dying and living), and $d = 1-f$ (certain death, if destined to not divide).

The results confound the intuition that the balance between cell death and terminal differentiation must be an important parameter (Figure 2B). Only when the probability of division is close to 0.5 does it matter whether the nondividing cells live or die, and even then it does not matter much. At $f = 0.51$, it takes 400 cell generations to produce 10^6 cells if all nondividing cells live and 600 generations if all of them die. At $f = 0.55$, it takes 138 and 114 generations, respectively. At $f = 0.60$, the difference practically disappears. Why is this? Consecutive cell division is an exponential process, and the growth of every generation progressively and dramatically diminishes the contribution of the previous generation to total cell number. The dividing cells are those that matter; the nondividing cells statistically disappear even if they live – thus, whether they die makes no difference. The greater the probability of division, the greater the fraction of dividing cells in the next generation, and the greater the irrelevance of the nondividing cells.

Time

So far we have only considered the number of cell generations required to produce a particular number of cells, not actual time. That is because real time and generation time are related linearly, and, given the above discussions, it is less confusing to only model cell generations than to bring in the time per generation. Hence – how long in real time would it take to produce a clone of a particular size, and are these estimates consistent with what happens in cancer? We let t , the average time between divisions, range from two per week to one per month, and calculated the number of weeks it takes to form a clone of 10^3 , 10^6 , or 10^9 cells in the same range of f as before. For simplicity we set $d = l = (1-f)/2$; that is, we let the same number of nondividing cells live and die, and started with a single cell as before.

The resulting times are entirely within the range of plausible values [30]. At the slow end, a clone with $f = 0.55$ and $t = 4$ weeks/division takes 3.2 years to produce 10^3 cells, or 8.9 years to produce 10^6 cells. In the middle, a clone with $f = 0.65$ and $t = 2$ weeks/division takes 22 months to produce 10^6 cells, or 34 months to produce 10^9 cells. An aggressive clone with $f = 0.75$ and $t = 2$ divisions/week takes 16 weeks to produce 10^6 cells, or 25 weeks to produce 10^9 cells.

In principle, this simple model could also accommodate the increase in mutation rates many tumors experience (via genome instability or other mechanisms) at later stages of progression, or mutator phenotypes. Increasing mutation rates decrease the wait-time to a driver mutation appearing somewhere in the clone, whether they are caused by genetic predisposition (Lynch syndrome), the environment (carcinogens in lung cancer), or somatic events in the course of tumor development (mutagenic drivers such as APOBEC3 – apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3 [31]). In the interest of simplicity, we chose here not to model mutation rates because their increase only compresses clock time between drivers and not the fundamental cell biology of drivers themselves.

Progression (evolution)

Constraints on the cellular machinery and the robustness of biological networks suggest that the first successful driver is likely to set f to be only a little above 0.5, as opposed to immediately increasing the probability of division to a high value [24]. Subsequent drivers may continue to increase f , but a decrease in t would also cause a higher proliferative rate, and for each subsequent driver to succeed it must either increase f or decrease t . The cells carrying these new drivers will only contribute substantially to the tumor mass if the magnitude of the change is great enough, because the earlier genotype has a head-start in the number of cells that carry it (Figure 3). Thus, unless drivers occur in very rapid succession, functional heterogeneity due to differential presence of drivers (not only lineage heterogeneity, which involves accumulation of passengers and happens regardless of drivers; Figure 1B) is ensured within the tumor (Figure 1C).

That tumor evolution (progression) should follow this simple scenario, which explains intratumor heterogeneity, is entirely consistent with the nature of driver mutations. As comprehensively explored in a recent review on cancer [30], decades of functional studies as well as recent genome-sequencing efforts have shown that there is a limited number, perhaps between two and eight, of large-effect driver genes in most tumors. There is still ascertainment bias towards point mutations and very large chromosomal changes, such as recurrent chromosome 1q gain in estrogen receptor (ER)-positive breast cancers, and the number may therefore be higher once all genomic changes are reliably identified. There may also be a large number of small-effect drivers that are sometimes lucky enough to help generate a clone [32]. However, the fact that there are unlikely to be many more large-effect driver changes per cancer is consistent with the idea that, in any given cell type, there is a limited number of

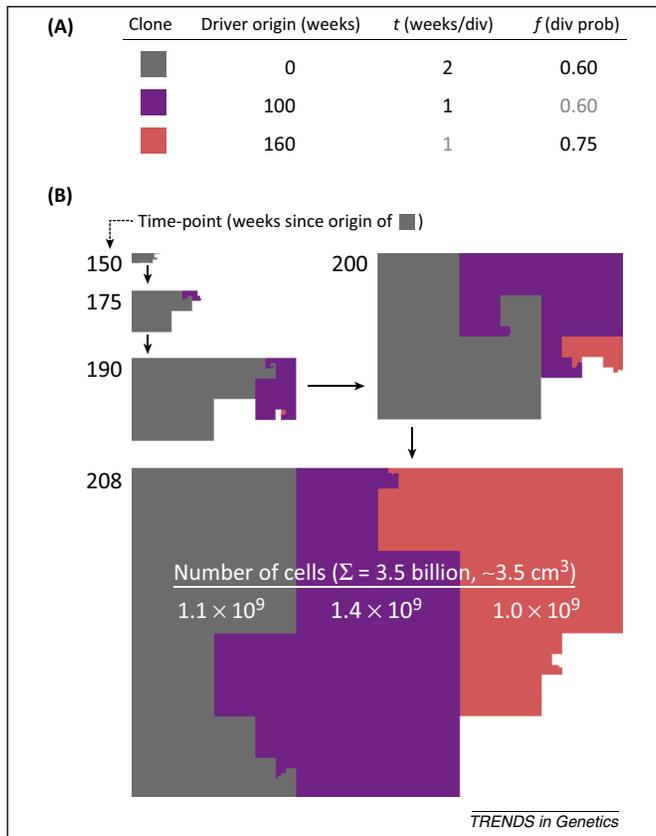


Figure 3. Visual representation of evolution and resulting clonal heterogeneity in a tumor that grows to 3.5 billion cells over the course of three driver mutations that occur within 4 years. (A) Parameters of the simulation. The grey clone originates with the first driver mutation that sets $f = 0.60$, and $t = 2$ weeks. 100 weeks later the purple clone originates due to a driver that decreases t to 1 week (f does not change). A further 60 weeks later, the orange clone arises by a driver that sets f to 0.75. Abbreviations: div, division; prob, probability. (B) Five time-points in the evolution of this tumor. Area sizes are proportional to the number of cells in each clone. After 4 years (208 weeks), all clones are approximately the same size, 10^9 cells each, constituting a solid tumor of about 3.5 cm³.

pathways that when mutated confer a big growth advantage. Small-effect drivers may also accumulate, but by definition they will not contribute the majority of the tumorigenic phenotype.

Simplicity of basics, complexity of instantiations

Cancer is a dumb disease, a disease of cell number alone and their uncontrolled growth, a regression to a more primitive state that overcomes the intricate homeostatic mechanisms that have evolved in multicellular organisms. Perhaps it is not surprising that a very simple model with only two parameters, f and t , can provide an explanatory basis for its progression. Cancer is also capricious, however, because mutation is random, and even the process of clonal growth is governed by stochasticity because each driver mutation is not guaranteed to be successful in causing a sufficient number of cell divisions to make a persisting clone. Synthetic interactions among drivers, or between drivers and genetic background, add genetic variance. Microenvironment, its interaction with genetics, local nutrient supply, and other organismal factors add further stochasticity. As a consequence, the variation in cancer incidence, speed of progression, likelihood of recurrence or metastasis, and eventual outcome is huge.

However, while every tumor is an individual experiment in evolution's morbid playground, basic principles apply – and the finite number of underlying mechanisms gives hope that an exact 100% kill rate, not rounded up, will be achieved reliably in the future.

Acknowledgments

Our cancer evolution work has been supported by a Stanford Pathology Department seed grant and by the National Cancer Institute. N.S. is supported by a National Research Council fellowship from the National Institute of Standards and Technology. We thank an anonymous reviewer for pointing out the equivalence of our 'number of tries for a driver to succeed' with a transformation of the Bozic *et al.* model [24].

References

- Darwin, C.R. (1859) *On the Origin of Species by Means of Natural Selection, Or the Preservation of Favoured Races in the Struggle for Life*, John Murray
- Zuckerklund, E. and Pauling, L.B. (1962) Molecular disease, evolution, and genetic heterogeneity. In *Horizons in Biochemistry* (Kasha, M. and Pullman, B., eds), pp. 189–225, Academic Press
- Dobzhansky, T. (1973) Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teacher* 35, 125–129
- Nowell, P.C. (1976) The clonal evolution of tumor cell populations. *Science* 194, 23–28
- The Cancer Genome Atlas Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550
- Nik-Zainal, S. *et al.* (2012) The life history of 21 breast cancers. *Cell* 149, 994–1007
- Yashida, S. *et al.* (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467, 1114–1117
- Gerlinger, M. *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366, 883–892
- Newburger, D.E. *et al.* (2013) Genome evolution during progression to breast cancer. *Genome Res.* 23, 1097–1108
- Sottoriva, A. *et al.* (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *PNAS* 110, 4009–4014
- Gerlinger, M. *et al.* (2014) Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* 46, 225–233
- Baca, S.C. *et al.* (2013) Punctuated evolution of prostate cancer genomes. *Cell* 153, 666–677
- Wang, X. *et al.* (2014) Widespread genetic epistasis among cancer genes. *Nat. Commun.* 5, 4828
- Grasso, C.S. *et al.* (2012) The mutational landscape of lethal castration-resistant prostate cancer. *Nature* 487, 239–243
- Diaz, L.A. (2012) The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486, 537–540
- Marusyk, A. *et al.* (2014) Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* 7520, 54–58
- Stephens, P.J. *et al.* (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144, 27–40
- Rausch, T. (2012) Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* 148, 59–71
- Greaves, M. and Maley, C.C. (2012) Clonal evolution in cancer. *Nature* 481, 306–313
- Navin, N. *et al.* (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94
- Potter, N.E. *et al.* (2013) Single-cell mutational profiling and clonal phylogeny in cancer. *Genome Res.* 23, 2115–2125
- Wang, Y. *et al.* (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512, 155–160
- Johnson, B.E. *et al.* (2014) Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* 343, 189–193
- Bozic, I. *et al.* (2010) Accumulation of driver and passenger mutations during tumor progression. *PNAS* 107, 18545–18550
- Gupta, P.B. *et al.* (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146, 633–644

- 26 Molina-Pena, R. (2012) A simple mathematical model based on the cancer stem cell hypothesis suggests kinetic commonalities in solid tumor growth. *PLoS ONE* 7, e26233
- 27 Gentry, S.N. and Jackson, T.L. (2013) A mathematical model of cancer stem cell driven tumor initiation: implications of niche size and loss of homeostatic regulatory mechanisms. *PLoS ONE* 8, e71128
- 28 Tomasetti, C. (2013) Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *PNAS* 110, 1999–2004
- 29 Davoli, T. *et al.* (2013) Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* 155, 948–962
- 30 Vogelstein, B. *et al.* (2013) Cancer genome landscapes. *Science* 339, 1546–1558
- 31 Burns, M.B. *et al.* (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494, 366–370
- 32 Lawrence, M.S. *et al.* (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501
- 33 Yang, L. *et al.* (2013) Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 153, 919–929
- 34 Shah, S. *et al.* (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486, 395–399
- 35 Govindan, R. *et al.* (2012) Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* 150, 1121–1134
- 36 The Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70
- 37 Lawrence, M.S. *et al.* (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218
- 38 Ojesina, A.I. *et al.* (2014) Landscape of genomic alterations in cervical carcinomas. *Nature* 506, 371–375
- 39 Broad Institute Genome Data Analysis Center (GDAC) (2014) *Firehose Computational Framework for The Cancer Genome Atlas (TCGA) Database*, April 2014 release. <http://gdac.broadinstitute.org>
- 40 Hoadley, K.A. *et al.* (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158, 929–944