Abstract Unsatisfactory progress in cancer medicine and prevention calls for new research approaches. Research can broaden its view of cancer to include not only specific molecular elements, but also the process that explains their origin and dynamics. This process is Darwinian evolution of somatic cells. Applicable modeling techniques are available from process-oriented systems biology. We review relevant concepts and techniques, and their application to four key open questions in cancer prevention research. Helpful concepts are transferable from classical evolutionary biology and ecology, while useful techniques include computational agent-based modeling. The research questions we review include (1) why do benign neoplasms often progress to malignancy? (2) what is the chronological sequence of molecular events in cancer progression? (3) how can we find reliable molecular biomarkers for cancer? and (4) will evolved drug resistance stymie efforts at a long-term cancer chemoprevention? We conclude that molecular analysis can be usefully augmented with process-oriented systems biology to guide empirical research into the most productive directions.

Keywords systems biology; cellular evolution; clonal evolution; cancer prevention

1. Introduction

Despite the US “war on cancer” declared in 1971, cancer incidence and mortality have fallen less than initially hoped and expected [7]. Even within the most recent five years, cancer incidence in the US declined less than 1%, and cancer mortality declined less than 2% [64]. This has led some researchers to argue that new approaches to cancer medicine are needed, including a conceptual framework that can explicitly address the computational complexity that arise from the cellular evolution and adaptation that occurs in cancer as it interacts with its own microecology [18,26].

Compared to the areas in which medical research had its most dramatic successes, cancer presents fundamentally different challenges, because it arises through a process that is endogenous to human tissues. The research strategy of using molecular reductionism to define a fixed disease-causing entity has a brilliant track record in the fields of infectious disease and inherited genetic disease, but has been less successful in cancer biology. Molecular analysis quickly revealed the consistent genetic nature of nonhuman pathogen species, such as bacteria, and of inherited genetic defects. In contrast, cancer reflects a complex process of ongoing genetic change in human somatic cells, and it usually has no fixed molecular basis. Thus, in important ways, a cancer is not simply a thing to be removed or destroyed, but also a process to be prevented, controlled or managed.

Although cancer risk can be greatly increased by inherited genetic factors or infectious pathogens, neither of those is necessary or sufficient on its own to be the cause of most cancer. Because cancer arises from a process that is endogenous to human tissues, it is often independent of inherited genes or external insults. Cancer biology can most naturally be organized around an understanding of this endogenous process of somatic cellular evolution, which consists of mutations causing differential reproduction, with inheritance, among somatic cells. Somatic evolution is a special case of the Darwinian process of evolution by natural selection. Process-oriented approaches from systems biology can help to encompass the complexity of somatic cellular evolution, making cancer more comprehensible and tractable.

2. The central process of cancer biology: somatic cellular evolution

As early as the 1970s, Cairns [12] and Nowell [50] proposed that somatic cell mutation, selection, and resulting Darwinian evolution are the fundamental processes by which neoplasms arise, acquire malignancy, and evade therapy. This hypothesis lay dormant for decades before being further developed and confirmed with empirical results from modern cancer biology [14,21,29,56,57,67]. Currently though, “cancer development at the cellular level is widely regarded as a Darwinian evolutionary process involving “natural selection” of genetically variant cells in the context of a complex microenvironmental ecology” [6]. Our only amendment to the foregoing quote would be to...
explicitly recognize that relevant mechanisms of cellular heredity include both genetic and epigenetic variation [57].

In common terminology, “cellular” selection and evolution are equivalently referred to as “clonal” selection and evolution, because each dividing somatic cell produces a clone of progeny with a shared genome and a shared fate.

As molecular techniques have provided an increasing wealth of data, the hypothesis of cellular evolution in cancer has been supported in multiple cancer types. For example, cellular evolution as a basis for neoplastic progression and changes in cancer has been documented in molecular detail in breast cancer [47, 62, 61], pancreatic cancer [13], renal cancer [28], ovarian cancer [48], leukemia [6, 70], colorectal cancer [32], and head and neck cancer [42].

The hypotheses and observations of cellular evolution have led to an emerging scientific theory of cancer dynamics, which elucidates origination, progression, metastasis, and response to treatment. This development is greatly assisted by a mature literature of concepts and techniques from the classical study of Darwinian evolution in populations of organisms. One optimistic assessment is that, despite the extreme complexity of cancer, “most, if not all, of this complexity can be explained by classical evolutionary principles” [29].

3. Computational tools for studying evolutionary processes

Armed with an understanding of the basic process behind the complex dynamics of cancer, researchers are increasingly adopting systems biology approaches that have proven effective for understanding complex dynamic processes such as cellular evolution. These approaches allow researchers to reveal not only the molecular components of cancer cells, but also the reasons why molecular components vary among cells, and why they change over time during cancer progression. Cancer is arguably the ultimate complex biological system [31]. For coping with such complexity, systems biology approaches have proven effective in much of biomedicine, including cancer biology [8, 44, 60], and these approaches fall into two distinct categories. One consists of techniques for analyzing large “omics” data sets [1, 33]. The other category consists of mathematical and computational models for representing underlying processes, as opposed to a specific set of resulting data [4, 5, 31]. Here, we refer to this second category as “process-oriented systems biology”. Without using that label for it, other authors have suggested that process-oriented systems biology has the potential to play a central role in cancer prevention [27].

One specific technique for conducting process-oriented systems biology is the use of “agent-based” computer models [39]. In these models, each cell can be explicitly represented by a computational “agent” (a piece of software) that represents a cell as it interacts with its microenvironment, mutates, divides, passes on acquired somatic mutations, and thereby founds a new clone. In this technique (also called “bottom-up” simulation, or “microsimulation”), the traits and behaviors of cells are specified in advance, but larger scale patterns and processes are not. Only as the model is run to simulate, often with a large and heterogeneous population of virtual cells, do larger scale patterns and processes emerge, often with collective outcomes that were not intuitively obvious in advance. This can generate a process of “digital evolution” inside a computer, which can provide a detailed, transparent, and predictive model of cellular evolution and progression under a specified set of conditions [2, 52]. The results of such simulations are not empirical data, but they can generate incisive new hypotheses that are empirically testable, and that can guide empirical research into productive new directions. Far from substituting for empirical research, computer models are most effective when they are closely integrated with experimentation, and are used to refine hypotheses and better focus experiments [17, 22].

Other techniques are also useful for modeling in process-oriented systems biology. For example, the authors in [9], among others, have represented evolutionary dynamics using a game theoretic framework for mathematical analysis. Their results elucidated the role of tissue interactions in directing the progression of prostate cancer into distinct courses, ranging from indolent to highly life-threatening.

4. Process-oriented systems biology in cancer prevention research

The process of somatic cellular evolution can rarely be observed directly, and most empirical data can only represent a “snapshot” in time from this ongoing process. Explicitly defining the process thought to underlie such observable data extends their explanatory power. Below, we consider four current questions in cancer prevention research, and discuss a process-oriented approach to each of them.

4.1. Why do benign neoplasms often progress to malignancy?

The defining (and deadly) feature of cancer, as opposed to benign growths, is local tissue invasion, ultimately leading to distant metastasis. Because this process emerges from interactions among multiple cells and their microenvironments, process-oriented system models are particularly well suited to its study. For this purpose, tools have been borrowed from both classical ecology and evolutionary biology.

Ecologists have used computational studies of spatiotemporal dynamics to compare tissue invasion by cancer cells with the invasion of natural ecosystems by artificially introduced species [40]. These authors posit that their results
4.2. What is the chronological sequence of molecular events for reduced cell proliferation and tumor growth [3].

Independent biomedical researchers have also found that computer simulations of mathematical models closely resemble the actual morphologies and spatial patterns of tumor growth and invasion, and that predictions of the model are met by observations of real tumors in vivo [23]. This group agreed with the ecologists cited above that results from process-oriented models can make tumor invasion more recognizable, and they further predicted that different invasion morphologies would correspond to different stages of tumor progression [10]. More recently, agent-based models have been calibrated to match the precancerous breast tumors of specific patients, to better track and predict their progression, as observed by mammography [38].

Beyond mere prediction, a central goal of cancer prevention is to block progression from benign premalignancies to cancer. At a cellular level, this progression entails the acquisition of motility by formerly sedentary neoplastic cells. In seeking ways to block this change, it will help to understand what causes it. Building on classical evolutionary ecology, one research team [3] hypothesized that the intense local competition for resources caused by abnormally rapid cell proliferation in neoplasms selects for motile cells that can thrive by leaving the resource-depleted zones where other neoplastic cells languish and die, and that this selection drives the evolution of somatic cell motility. Agent-based computational models provided support for this hypothesis. Their results also suggest that therapeutic agents designed to block cell motility, instead of killing cancer cells, will be more robust against acquired drug resistance, and while blocking malignancy, will also select for reduced cell proliferation and tumor growth [3].

4.3. How can we find reliable molecular biomarkers for cancer?

Another major research direction in cancer prevention is the development of cancer biomarkers for early detection and for prognosis. To date, most biomarker development has not been informed by evolutionary dynamics, but instead has focused on specific molecules. This strategy has identified many candidate somatic genetic and epigenetic “biomarkers” of cancer or cancer risk. However, very few specific marker molecules have proven to be reproducibly effective for identifying cancer [59]. Similarly, no specific molecule has yet been validated as a robust predictor of progression to cancer, or been useful to reduce cancer mortality [35].

One likely reason for this lack of success is that the process of cellular evolution is fueled by stochastic mutation, which is unpredictable and varies among individuals. Because of shared selective pressures for survival and proliferation, different cancer cases tend to converge on the same “hallmark” cell traits [30]. However, because somatic mutation is stochastic, the molecular basis for these traits can differ, and at a molecular level, each cancer can be unique in many ways [29], making it difficult to find molecular biomarkers that will be consistent for different patients.

Instead of seeking consistency across individuals in cancer-specific molecules, a different approach is to recognize that the cellular genetic diversity of neoplasms can help explain the patterns of genetic diversity observed both within and among tumors in a patient. They also describe a “distinct geometrical signature” of invasion dynamics in both ecological landscapes and tumors, which they suggest may someday be used along with advanced imaging technologies to accurately discriminate malignancies from benign growths [41]. Independently, biomedical researchers have also found that computer simulations of mathematical models closely resemble the actual morphologies and spatial patterns of tumor growth and invasion, and that predictions of the model are met by observations of real tumors in vivo [23]. This group agreed with the ecologists cited above that results from process-oriented models can make tumor invasion more recognizable, and they further predicted that different invasion morphologies would correspond to different stages of tumor progression [10]. More recently, agent-based models have been calibrated to match the precancerous breast tumors of specific patients, to better track and predict their progression, as observed by mammography [38].

Beyond mere prediction, a central goal of cancer prevention is to block progression from benign premalignancies to cancer. At a cellular level, this progression entails the acquisition of motility by formerly sedentary neoplastic cells. In seeking ways to block this change, it will help to understand what causes it. Building on classical evolutionary ecology, one research team [3] hypothesized that the intense local competition for resources caused by abnormally rapid cell proliferation in neoplasms selects for motile cells that can thrive by leaving the resource-depleted zones where other neoplastic cells languish and die, and that this selection drives the evolution of somatic cell motility. Agent-based computational models provided support for this hypothesis. Their results also suggest that therapeutic agents designed to block cell motility, instead of killing cancer cells, will be more robust against acquired drug resistance, and while blocking malignancy, will also select for reduced cell proliferation and tumor growth [3].

4.3. How can we find reliable molecular biomarkers for cancer?

Another major research direction in cancer prevention is the development of cancer biomarkers for early detection and for prognosis. To date, most biomarker development has not been informed by evolutionary dynamics, but instead has focused on specific molecules. This strategy has identified many candidate somatic genetic and epigenetic “biomarkers” of cancer or cancer risk. However, very few specific marker molecules have proven to be reproducibly effective for identifying cancer [59]. Similarly, no specific molecule has yet been validated as a robust predictor of progression to cancer, or been useful to reduce cancer mortality [35].

One likely reason for this lack of success is that the process of cellular evolution is fueled by stochastic mutation, which is unpredictable and varies among individuals. Because of shared selective pressures for survival and proliferation, different cancer cases tend to converge on the same “hallmark” cell traits [30]. However, because somatic mutation is stochastic, the molecular basis for these traits can differ, and at a molecular level, each cancer can be unique in many ways [29], making it difficult to find molecular biomarkers that will be consistent for different patients.

Instead of seeking consistency across individuals in cancer-specific molecules, a different approach is to recognize that the cellular genetic diversity of neoplasms
fuels their evolution, and to measure this diversity itself as a marker, instead of any single molecule. Classical evolutionary theory states that the rate of evolutionary change is proportional both to trait variance and to fitness variance in the members of the evolving population [58]. Thus greater genetic variation among cells in a neoplasm is expected to drive faster progression. There is substantial empirical evidence supporting this hypothesis. In one study [71], five different cell lines representing various types of cancers were analyzed for karyotypic diversity among cells. In each of these lines, the highest level of chromosomal diversity was coupled with the strongest tumorigenicity (Figure 2). This pattern is not limited to karyotype, or to in vitro cell lines. High cellular diversity accurately predicted risk in patients monitored for progression from Barrett’s esophagus to esophageal cancer, and simultaneously monitored by whole-genome sequencing of repeated biopsies. Cell diversity was measured in SNPs, copy number, loss of heterozygosity, and aneuploidy [36]. Every tested measure of genetic diversity among cells produced highly significant ($P < .001$) predictors of progression to cancer [43]. Similar patterns have been reported in other human tumors. In a study of human breast carcinomas, an index of cellular genetic diversity based on copy number ratios was higher in invasive than in situ components from the same section of the same tumor [51]. Cellular genetic diversity was also positively correlated with tumor grade [51, supplemental Table 7]. (Note that these authors considered this latter trend to be only “suggestive,” needing confirmation in a larger sample set.)

Taken together, these experimental results and observational clinical results suggest that the genetic diversity that fuels somatic cellular evolution may be a robust general biomarker of cancer risk and prognosis. Such a diversity marker, if proven in translational research, could simultaneously help to address two major challenges:
Comparison of the spectral karyotypes of two cell sublines, each derived in vitro from the same cell line (MCF10A3, from a human breast carcinoma in situ). The high-diversity subline CSC-3 (panel (b)) produced tumors in immunodeficient mice, while the low-diversity subline CSC-1 (panel (a)) did not. Only subline CSC-3 with xenograft tumorigenicity (panel (b)) contained many low-frequency chromosome aberrations not found in both sublines (indicated by the yellow shaded boxes), in addition to the five high-frequency chromosome aberrations found in both (indicated by blue-shaded boxes).

achieving early detection of truly dangerous neoplasms, and avoiding overdiagnosis of nonthreatening growths.

4.4. Will evolved drug resistance stymie efforts at long-term cancer chemoprevention?

Developing drugs for chemoprevention is a major line of research in cancer prevention that may face pitfalls from cancer evolution. These potential pitfalls have already been demonstrated in cancer therapy. Many agents have been developed and successfully used for selectively killing cancer cells. A very general limitation on this approach, however, is the acquired drug resistance that results from cellular evolution [54]. This makes resistant relapse a routine sequel to initial tumor shrinkage during therapy [15, 54], and most patients who die of cancer are killed by a cancer that has evolved drug resistance. The problem of acquired resistance may prove to be even more limiting for a long-term chemoprevention than it is for therapy. While successful therapy is completed within a shorter time frame, cancer prevention might ideally continue throughout life. Thus, any drugs that can be administered effectively only for a limited time may not be entirely satisfactory.

There is clinical evidence that the time limits on effective drug administration seen during cancer therapy can also arise in chemoprevention. This is illustrated by the selective estrogen receptor modulators (SERMs) tamoxifen and raloxifene. Both of these have recently been tested and approved as chemoprevention agents for breast cancer in high-risk women [24,68,69]. However, the problem of acquired drug resistance in chemoprevention has not yet been adequately addressed. In cancer therapy, many initially responsive tumors later develop acquired resistance to tamoxifen [45,46]. A similar acquired resistance to raloxifene, as well as cross-resistance between these two SERMs, has been demonstrated in vitro with breast carcinoma cells [25,37,66]. This established pattern of acquired resistance to SERMs appears to arise in chemoprevention as well as in chemotherapy. Controlled chemoprevention trials showed a reduction in breast cancer incidence in healthy high-risk women after using tamoxifen for five years [20]. However, in the adjuvant treatment setting, longer administration does not confer further benefit in preventing new primary tumors [20]. This result suggests that long-term intervention to prevent a new cancer may suffer from the same limitations as long-term chemotherapy to prevent recurrence, and for the same reason of cellular evolution of acquired drug resistance. Achieving life-long protection through chemoprevention may require longer periods of effective drug administration, and thus may not be possible with any drug that selects for drug-resistant variants. Whenever somatic cells are dividing and subject to mutation, resistant variants are likely to arise. Any drug that kills or suppresses individual drug-sensitive cells thereby “selects” more resistant cells to survive with reduced competition after the removal of drug-sensitive neighbors [53]. This leads to a rapid out-growth, or “clonal expansion,” of a lineage of resistant cells which quickly takes the place of the drug-sensitive cells removed by treatment [54].

Providing the best possible protection through chemoprevention may require ways to delay cellular evolution of acquired resistance. Application of evolutionary theory from other contexts suggests specific strategies [53]. Therapeutics that change the tumor microenvironment to make it less hospitable to tumor growth and invasion can impede all cancer cells equally, without selecting among them and thereby driving the evolution of resistance. For example, nonsteroidal anti-inflammatory drugs meet this criterion, and there is some evidence that they may remain effective indefinitely for long-term chemoprevention [16].

Other potential drug classes could also exert less selection on existing variation, and thus slow the evolution of acquired drug resistance. Potential new drugs could exploit the fact that cancer cells construct their own microenvironment by producing, and depending on, many shared substances that increase the capacity of their shared
initially normal human cells. This process is best understood as somatic cellular evolution. Process-oriented systems biology provides tools for modeling this process to generate the most crucial hypotheses for empirical testing.

Acknowledgments For helpful input, the authors thank Carlo C. Maley and Brian J. Reid.

References

Figure 3: Simulation results reprinted from a previous work [55]. Frequency of drug-resistant cells after 1,000 cell generations of mutation and evolution. High transfer coefficients correspond to highly diffusible drug targets produced by one cancer cell that provide benefits to other nearby cancer cells. A transfer coefficient of zero corresponds to a cell-intrinsic target molecule that is not shared among neighboring cells. Each marker represents the mean for a different drug dosage: filled circle = 100% of full dose, open circle = 75%, triangle = 50%. Error bars show standard error across 10 simulation runs using different seed values for the pseudorandom number generator.


