Review in Advance first posted online on August 8, 2016. (Changes may still occur before final publication online and in print.)

An Evolutionary Genetic Perspective on Cancer Biology

Max Shpak[1,2,3] and Jie Lu[1,*]

[1]NeuroTexas Institute Research Foundation, St. David's Medical Center, Austin, Texas 78705; email: shpak.max@gmail.com

[2]Center for Systems and Synthetic Biology, University of Texas, Austin, Texas 78712

[3]Fresh Pond Research Institute, Cambridge, Massachusetts 02140

Annu. Rev. Ecol. Evol. Syst. 2016. 47:25–49

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

This article's doi:
10.1146/annurev-ecolsys-121415-032109

*Current address: Genetic Sciences Division, Thermo Fisher Scientific, Austin, Texas 78744

Keywords

carcinogenesis, somatic evolution, clonal selection, senescence, pleiotropy, neutral theory

Abstract

Cancer biology can be better understood by drawing upon methods and concepts from evolutionary genetics. Cancer progression proceeds through somatic evolution, being driven by selection on clonal lineages via the differential survival and proliferation of cell lines. This within-patient evolution can be modeled and analyzed using population genetic and phylogenetic tools to identify mutations and genotypes that are under directional selection during tumor growth, spatial differentiation, and metastasis. Evolutionary genetics can also explain the persistence of cancer within populations. A minority of cancers are associated with inherited risk alleles, which are maintained in populations through genetic drift or antagonistic pleiotropy. Finally, cancer biology can be understood from a macroevolutionary perspective as a case study of evolutionary cooperation and conflict between different levels of biological organization.

INTRODUCTION

Like most biological phenomena, cancer can be analyzed from the perspectives of both proximate and ultimate causation (Mayr 1982). The proximate causes for cancer are case specific, contingent upon dysregulation of particular genes, disruption of regulatory and signaling pathways, and particular classes of mutations. Understanding the proximate, mechanistic causes of tumorigenesis and cancer progression involves a systems analysis of the molecular mechanisms that regulate cell growth, apoptosis, as well as cell adhesion and motility.

In contrast, evolutionary biology has informed the search for cancer's ultimate causes and its incidence in populations. For example, cancer has been conceptualized as a conflict between the fitness of a multicellular organism and the fitness of its component cells (Burt & Trivers 2006, Aktipis et al. 2015). The empirical and conceptual bridge between these proximate/functional and ultimate/evolutionary explanations of cancer biology is through population and evolutionary genomics. Mutational changes in individual cells drive the dysregulation of gene expression and signaling pathways characteristic of cancer, and they also provide the raw material for natural selection to act on clonal lineages of cells within an organism. Consequently, both the incidence of cancer and the dynamics of cancer progression are the result of evolutionary processes (Cairns 1975, Greaves 2000, Crespi & Summers 2005, Merlo et al. 2006, Pepper et al. 2009, Thomas et al. 2013). Understanding these evolutionary dynamics will lead to novel approaches to characterizing the genetic and cellular basis of cancers, as well as guiding the development of cancer treatments that are attuned to genetic variation within tumors and their ability to adapt to the host environment and to therapies. Conversely, cancer evolution, whether within affected individuals or populations or over macroevolutionary time, can provide a valuable model system for better understanding fundamental evolutionary principles in populations of clonally reproducing organisms.

We begin with a brief review of the genetic mechanisms and cellular basis of cancer (Hanahan & Weinberg 2000, 2011). We then describe evolutionary perspectives on cancer, starting at small temporal and spatial scales (within-host dynamics) and moving up to larger scales of host population genetics and ultimately macroevolution.

The Genetic and Cellular Hallmarks of Cancer

A unifying feature of cancer is the tendency of tumor cells to proliferate, either because of increased cell division, reduced cell death, or differentiation. These are often the result of mutations that (*a*) make cells insensitive to external signals inhibiting growth (e.g., contact inhibition), (*b*) induce internal signaling pathways that drive autonomous cell division, or (*c*) immortalize cell lines by preventing senescence and inhibiting apoptosis following cell damage. These characteristics confer increased fitness of tumor cell lines (clonal lineages) relative to other cells in the nascent tumor or to normal tissue cells.

The fitness of cancer cells also increases through their ability to modify their microenvironments. Among these abilities is the induction of angiogenesis to provide oxygen and nutrients to cells within large tumors, thus removing surface area/volume constraints on growth (Nishida et al. 2006, Weis & Cheresh 2011). Changes to motility and adhesion allow cancer cells to disperse (local invasion, metastasis) and to invade novel environments (Martin et al. 2000) in response to selective pressures, analogous to the dispersal and subsequent niche construction in free-living organisms. Tumor cells also suppress the host's immune response (Seliger 2005) or alter the configuration of their membrane protein antigens (Gajewski et al. 2013) to evade immune attack. In this respect, cancer cells behave much like microbial pathogens and parasites that suppress or evade the host immune system.

The molecular bases of these traits are beyond the scope of this article, but it is important to highlight several general classes of cancer-related mutations that influence evolutionary dynamics. Cancer-causing mutations can be broadly subdivided into tumor suppressor genes and oncogenes. However, for the purposes of evolutionary models of cancer, it is often of greater importance to classify cancer-associated mutations according to additional criteria, such as their effects on the survival and reproduction of cancer cells, whether the mutations are dominant or recessive, and whether the mutations are somatic or in the germline.

Tumor suppressor genes have a negative, inhibitory role in the cell cycle. The function of most tumor suppressors is to either arrest the cell cycle or induce apoptosis in damaged cells. Others encode proteins that repair mutations or membrane adhesion proteins that inhibit cell growth and invasive dispersal (Sherr 2004). Loss-of-function mutations in tumor suppressor genes drive tumorigenesis. Such mutations are typically recessive, so mutations in both alleles of a tumor suppressor gene are often necessary for tumorigenesis. For example, this two-hit model (Knudson 1971) has been demonstrated for initiation of retinoblastoma. Other examples of tumor suppressor genes whose mutant forms are associated with multiple cancers include *TP53* (tumor protein 53, linked to cancer types including gliomas, cervical cancer, head and neck carcinomas, and breast cancer), *RB1* (retinoblastoma protein 1, associated with its namesake as well as sarcomas, lung, and breast cancers), and *BRCA* (breast cancer genes, also linked to ovarian cancers); see the appendix in Bunz (2008) for a summary.

In contrast, oncogenes initiate carcinogenesis through single gain-of-function mutations. The protein products of proto-oncogenes (unmutated oncogenes) regulate the mitotic cycle. Most proto-oncogenes encode proteins that act as receptors or transducers in signaling pathways (Croce 2008), such as the *RAS* gene family, which regulates signal transduction in the mitogenic *RAS/RAF/MAPK* kinase pathway (Malumbres & Barbacid 2003). Oncogenic mutations amplifying activity in this pathway have been identified in the majority of common cancer types. Other important proto-oncogenes occur in the *WNT* transduction pathway (Polakis 2012), which regulates cell proliferation and migration. Mutations that increase telomerase activity can also promote tumor progression by immortalizing cell lines (Vinagre et al. 2013); however, these mutations are not technically oncogenic because they do not dysregulate the cell cycle (Harley 2002).

CANCER AS A GENETIC DISEASE

Clonal Selection and the Multistage Model

The recognition by Boveri (1914) that tumors are derived from mutated normal cells led to our current understanding of cancer as a genetic disease and of the fact that cancers are primarily the consequence of spontaneous somatic mutations. This realization led to the multistage model of carcinogenesis. Typically, although loss-of-functions in tumor suppressors or increases in oncogene activity may initiate a precancerous tumor, progression to cancer requires a sequence of mutations. This process is accelerated through clonal expansion and positive selection on mutated cell lines with high rates of proliferation and survival (Nowell 1976); **Figure 1** illustrates this process. Early mutations that initiate tumorigenesis are shared by cells throughout the tumor. Later mutations can either sweep to fixation by replacing earlier clonal lineages or give rise to spatially heterogeneous regions of the tumor characterized by distinct subclones defined by somatic mutations at other loci.

Because the multistage model requires multiple sequential mutations, it predicts that cancers should be most prevalent in tissues with extensive cell division and hence more scope for somatic mutation and subsequent clonal selection. This prediction is supported by a recent analysis by Tomasetti & Vogelstein (2015) showing that the incidences of different cancer types are predicted

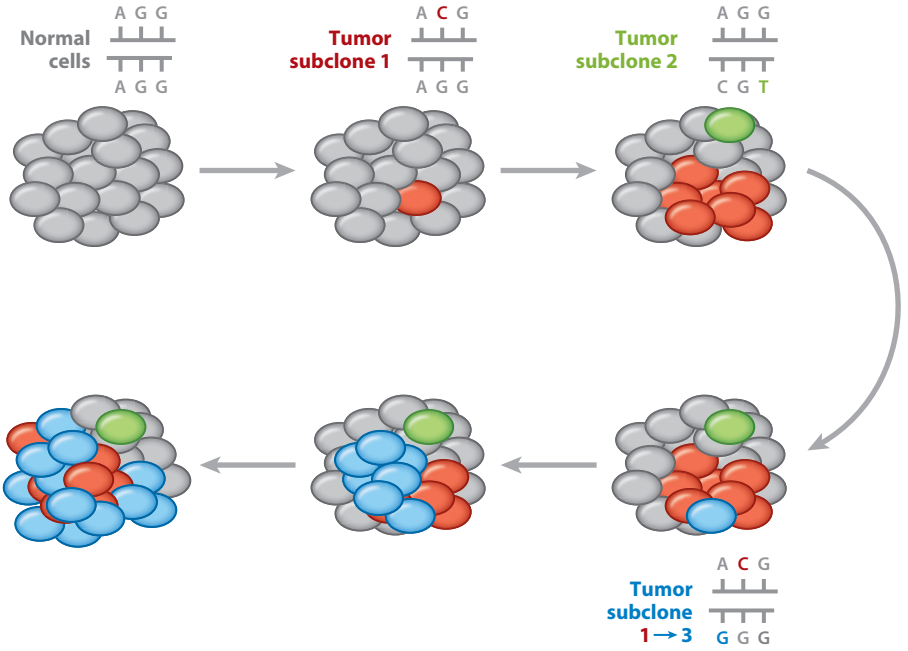

Figure 1

Somatic mutation and clonal selection in cancer. In this hypothetical example, the germline genotype is defined by alleles A,G,G at three sites. Two clonal lineages are initialized by a point mutation to C at the first site (*red*) and to T at the third site (*green*). A third mutation (A to G at first site) occurs in a cell of the red subclone genotype, defining the blue (third) subclonal lineage. The red and blue lineages both undergo higher rates of clonal expansion than green subclone 2.

from the number of stem cell divisions in the source tissue. The lifetime incidence of melanoma and colorectal cancer is ~100-fold higher than for gliomas or cancers of the small intestine, because the rate of cell replacement (and of somatic mutations) is much higher in the former than in the latter tissues, supporting the hypothesis that the potential number of random somatic mutations is a strong predictor of incidence among cancers.

The Multistage Model and Cancer Incidence

Normal cells accumulate somatic mutations during cell divisions (Martincorena et al. 2015). A very small fraction of these mutations are loss-of-function mutations in tumor suppressor genes, and a smaller fraction still are gain-of-function mutations in oncogenes. If initiation of a cancer begins with an oncogenic mutation with instantaneous rate μ, the probability that at least one such mutation is encountered in a population of N dividing cells at time t (scaled with cell division rate) is, following Nowak (2006, chapter 12),

$$P(t) = 1 - e^{-\mu N t} \approx \mu N t. \qquad 1.$$

Assuming somatic selection deterministically drives such a mutation to high frequency, $P(t)$ is the probability of oncogene activation initiating a tumor.

In contrast, assuming a loss-of-function mutation in one allele of a tumor suppressor gene is recessive, such mutations are often neutral and unaffected by somatic selection. The waiting

time until the origin of a clonal lineage with loss-of-function in tumor suppression depends on the rate for the first mutation, μ_1, and the typically higher rate of mutation for the second, μ_2 (i.e., a point mutation or indel is necessary for the first, and the second can occur through loss of heterozygosity). The clonal lineage with the mutation must increase in frequency sufficiently through genetic drift for the second mutation to occur. In the limit of long time intervals and a small population of cells (e.g., colorectal crypt stem cells), Nowak et al. (2004) showed that the approximate probability of having a cell with two loss-of-function mutations in a tumor suppressor gene is approximately

$$P(t) \approx 1 - e^{-\mu_1 t} \approx \mu_1 t. \qquad 2.$$

The fixation probability of the first mutation is $1/N$, so that the waiting time until the first mutation occurs and reaches a sufficiently high frequency is a rate-limiting step for this process. For shorter time intervals and for larger populations of cells, both the first and second mutation rates are rate limiting, and waiting times are exponential functions of both μ_1 and μ_2 (Nowak et al. 2004). The scaling of (1) with N implies shorter waiting times until oncogene-driven cell lineages become common. However, mutations are more likely to incur loss-of-function rather than gain-of-function, so that $\mu_1 \gg \mu$. If this inequality more than compensates for the effect of N in (1), then loss of tumor suppressor gene function is more likely than oncogene activation.

The distribution of waiting times until mutation events can be used to parameterize and predict cancer risk as a function of age or by tissue. Using the simplifying assumption of exponential waiting times before each mutation in the multistage model, Armitage & Doll (1954) derived a power law distribution of cancer incidence as a function of age. If k specific mutations are required for a normal ancestral cell to progress to a malignant tumor cell, and $\lambda \ll 1$ is a somatic mutation rate for the aggregate of precancer cells in the host organism (consistent with either μ or μN above), the probability of a particular mutation occurring by time t is approximately λt. Assuming further that the mutations are independent and equiprobable, the probability that at time t there are $k - 1$ such mutations is approximately $(\lambda t)^{k-1}$. Therefore, the rate at which an individual becomes cancerous, or the incidence of cancer for individuals at age t, is

$$I(t) \approx C(\lambda t)^{k-1}, \qquad 3.$$

where $C = \lambda$ if the order in which mutations occur is not significant and $C = \lambda/(k-1)!$ if, more realistically, mutational order is assumed, so that later mutations are tumorigenic against only preexisting mutational backgrounds. [See Frank (2004, 2007; chapter 6), and Bozic et al. (2010) for generalizations of the multistage model that account for time heterogeneity due to increasing densities of mutated cells due to clonal expansion.] Equation 3 provides a good fit to data on lifetime cancer incidence, with the rate parameter λ being specific to cancer/tissue types (larger values of λ are characteristic of high rates of stem cell division, and smaller values of k are associated with genetic and/or environmental risk factors).

The implication of the multistage model is that all else being equal, larger, long-lived organisms with more cell divisions should have a higher incidence of cancer, just as we expect a higher incidence of cancer in tissue types with intrinsically higher rates of cell division.

Peto's Paradox

Within species, there is some evidence of higher lifetime cancer incidence in individuals with larger bodies, for example, a greater occurrence of cancer in large dog breeds (Greer et al. 2011) and in tall people (Green et al. 2011). However, the predicted relationship between body size and cancer incidence doesn't hold across species, for example, large and long-lived elephants and

whales have lower rates of cancer than smaller mammals (Caulin & Maley 2011). This rarity of cancers in many large animals is known as Peto's paradox (Peto 1977) and is a focus of comparative cancer genomics studies seeking insight into evolved mechanisms of tumor suppression.

The proximate cause of cancer's rarity in elephants, for example, is partly the duplication of the tumor suppressor gene *TP53* (Abegglen et al. 2015). In humans and most other mammals *TP53* exists as a single functional copy, but in elephants there are 20. As a result, the probability of loss-of-function in *TP53* through somatic mutation is negligible. Because elephants remain fertile for the majority of their adult life span, there is strong selective pressure against genetic disease (including cancer susceptibility) in older individuals, favoring the proliferation of tumor suppressor paralogs. In contrast, selection on antioncogenic and tumor suppressive genotypes is comparatively weak in organisms that are short lived or whose reproductive years are short relative to their potential life span.

Another notable anomaly from the standpoint of life history and cancer biology is the naked mole rat (*Heterocephalus glaber*). With a life span of >20 years, they are unusual among rodents of the same size, whose life spans are typically <5 years. Simply by virtue of their long lives, one would expect a very high cancer incidence in naked mole rats, as cancer is common in older mice and rats. In fact, mole rats rarely develop cancer, apparently due to particularly sensitive contact inhibition (the arrest of cell division at high local cell densities) as a consequence of high levels of hyaluronic acid secretion by fibroblasts. Consequently, the *TP53* and *RB* mediated pathways can block *RAS*-activated tumor cells from proliferating (Seluanov et al. 2009). This characteristic may have initially evolved to give these rodents a more flexible integument for their burrowing mode of life. Cancer protection may have been a secondary consequence of this trait, enabling a longer life span, long-term fertility, and perhaps facilitating their unique eusocial organization.

DRIVER MUTATIONS AND GENOMIC INSTABILITY

In cancer genomics, driver mutations are the somatic mutations that induce tumor progression and are under positive selection. Subsequently, the tumor lineage may accumulate selectively neutral passenger mutations that are a consequence rather than a cause of the cancer. Many passenger mutations are the result of genetic hitchhiking (Maynard Smith & Haigh 1974). Clonal reproduction of cancer cells creates complete linkage disequilibrium between a new mutation and its genetic background. Consequently, if a driver somatic mutation undergoes a clonal expansion, all other somatic mutations in its genome will do so as well (in the absence of loss of heterozygosity through mitotic recombination and gene conversion).

Additionally, many somatic mutations in tumors are a secondary consequence of genetic instability in cancer (Cahill et al. 1999). For example, aneuploidy, both partial and involving entire chromosomes, is one of the characteristic features of cancer cells (Sen 2000); aneuploidization plays a prominent role in the tumorigenesis of colorectal adenomas and other cancers (Danielsen et al. 2015). Aneuploidy is self-perpetuating in cell lines: An aneuploid cell has a high probability of gaining or losing additional chromosomes through nondisjunction in subsequent cell divisions. Although gain and loss are equiprobable during cell division, human cancer cells with >46 chromosomes predominate because cells without a complete chromosomal complement are typically inviable.

Genetic instability also occurs at the level of sequence variation; often, much of the molecular machinery that minimizes the occurrence of point mutations is compromised in tumor cells, along with the barriers to the propagation of mutated cells. Loss-of-function mutations in DNA mismatch and excision repair genes increase the incidence of somatic mutations (λ in Equation 3), resulting in passenger mutations. Several hereditary cancers are associated with germline

mutations in DNA repair genes (e.g., melanoma susceptibility through *XPC* mutations), and somatic mutations in repair genes are common in many cancer types.

Both chromosomal instability and increased sequence mutability have the similar consequence of accelerating the rate of somatic evolution. As with any population of replicating organisms, greater mutation rates have conflicting implications for cancer evolution. On the one hand, higher mutation rates result in more genetic variations among cells, in other words, potentially more mutations beneficial to the growing tumor. On the other hand, because beneficial mutations are rarer than deleterious mutations, genomic instability increases the genetic load in the population of tumor cells. Even though many loss-of-function mutations that are deleterious to normal cells are neutral in cancer cells (e.g., loss of endocrine function in a pancreatic cell is deleterious to the host, not the cancer), tumor cells must still maintain a broad range of housekeeping functions so that housekeeping genes remain under purifying selection.

If the rate of deleterious mutation is sufficiently high, populations undergo error catastrophe (Eigen 1971, Eigen & Schuster 1979), a monotonic decrease in mean population fitness, rather than maintaining mutation–selection equilibrium. This decrease in fitness is a consequence of deleterious mutations being introduced into a population at a higher rate than the rate at which they are removed by natural selection. Indeed, many chemotherapies, such as temozolomide treatment of gliomas (Hegi et al. 2005), are mutagenic and function by inducing error catastrophe in tumor cells (Fox & Loeb 2010). In a small population of cells, lineages are also subject to Muller's ratchet (Muller 1964, Felsenstein 1974), in other words, the sequential loss of least mutated and most fit genomes from a population through genetic drift in the absence of recombination. Bignold (2007) has suggested that aneuploidization and mitotic recombination may counter Muller's ratchet in tumor cell populations. This proposal is consistent with the results of Mandegar & Otto (2007), which showed that heterozygosity and gene conversion can accelerate the spread of beneficial mutations and the elimination of deleterious mutations in asexual organisms.

Genomic instability in cancer cells raises the question of whether the mutability is itself an outcome of selection. If higher mutation rates were unconditionally deleterious, we should expect to find selection against aneuploidy and other sources of mutability. However, if mutation rates are sufficiently low to avoid error catastrophe, selection can favor genomic instability (Breivik 2004, Cowperthwaite et al. 2006). In sexually reproducing organisms, a modifier that increases the mutation rate cannot maintain linkage disequilibrium with beneficial mutations due to recombination with other genomes, in other words, a modifier found in the same genome as an induced beneficial mutation in the current generation is most likely going to be associated with deleterious mutations in the next. In contrast, modifiers that increase the mutation rate can be selected in clonally reproducing organisms, because the modifier and any beneficial mutation it induces are maintained in nearly complete linkage disequilibrium (Leigh 1973). Consequently, it is possible that high mutation rates are actively maintained by selection in cancer cells, an example of the evolution of evolvability (Dawkins 1989, Wagner & Altenberg 1996). Whether the genomic instability is just a by-product of cancer biology, or whether the higher mutation rates are what fundamentally drive the cancer progression, remains unknown. A high mutation rate in cancer cells relative to normal tissue is not sufficient evidence for positive selection on mutation rates. However, the elevated mutation rate is evidence that selection against frequent somatic mutations is weak or absent in tumors.

SOMATIC SELECTION AND TUMOR HETEROGENEITY

Tumor cells experience numerous selective pressures, from selection on cell lines that establish the tumor to selection on mutations that help cancer cells evade the host immune system and adapt

to chemotherapies. Because the genes under strong selection are often those most physiologically relevant to cancer progression, characterizing the genes and mutations under somatic selection is a focus of cancer biology.

A conventional heuristic approach to identifying genes under selection has been to find highly mutated genes, in other words, genes in which the relative frequency of nonsynonymous mutations is significantly higher than in the genomic background. This is the basis for analyses using tools such as MutSigCV (Lawrence et al. 2013), which has been the standard for identifying so-called cancer genes in The Cancer Genome Atlas (TCGA 2012a–c) consortium analyses of cancer mutational profiles.

There are several potential weaknesses to this method. A highly mutated gene may have many mutations that are passengers rather than drivers (see previous section). Furthermore, identification of highly mutated genes is more effective at discovering loss-of-function mutations (e.g., tumor suppressor genes) than gain-of-function mutations (e.g., oncogenes). This is because there are many ways in which nonsynonymous mutations can cause loss-of-function in tumor suppressors (e.g., frame-shift or premature stop codon mutations), whereas gain-of-function is often site-specific. Consequently, many oncogenes won't be highly mutated and will escape detection when hypermutation is used as a criterion. Finally, tallying hypermutated genes cannot identify functionally significant genes that are under purifying selection. These considerations suggest that methods grounded in evolutionary models will provide more information-rich inferences of the functional significance of genes and mutations.

Phylogenetic Approaches

Somatic evolution is a cumulative process: The earliest driver mutations that initiate tumorigenesis are generally inherited by all tumor cells, and subsequent driver mutations increase the fitness of tumor cells by enhancing proliferation, inducing angiogenesis, etc. Some of these mutations contribute to genetic heterogeneity in the tumor (Burrell et al. 2013), whereas others increase in frequency through selective sweeps and displace other clonal lineages. Later mutations facilitate invasion of local tissues and metastasis, as metastatic cells have to adapt to new environments, following the seed and soil models for metastatic invasion of new organs (Paget 1889, Langley & Fidler 2011). Consequently, levels of genetic heterogeneity in precancerous tumors have proven to be strong predictors of which nonmalignant tumors progress to cancer (Maley et al. 2006).

Genetic heterogeneity in tumors also provides the heritable variation that allows cancer cells to evade chemotherapies, including evolved resistance to multidrug therapies that impose simultaneous selection pressures on cancer cells by targeting different biochemical pathways (Persidis 1999, Burrell & Swanton 2014). Consequently, recurrent tumors are often characterized by evolved resistance to initial chemotherapies (Harris 1985, Luqmani 2005). Furthermore, human cancers experience extreme population bottlenecks following treatment (Tsao et al. 2000); together with selective pressures, these population bottlenecks result in very different genetic profiles in recurrent versus initial tumors.

Phylogenetic analysis allows us to reconstruct the genealogical history of tumor genotypes and provides an approach to characterizing genetic heterogeneity within and among tumors based on this history. One can identify genetic features of specific subclonal lineages and their point of origin during the history of the primary tumor, as well as those of descendant metastases and recurrent tumors. In solid tumors, reconstructing the phylogeny of subclones requires sampling and sequencing a large number of small sectors from spatially distinct regions to capture the range of clonal diversity. Including samples from metastatic and recurrent tumors in a phylogeny makes it possible to identify mutations and genotypes in the primary tumor that seeded the metastatic cell

lines and the lineages that survived drug and other standard therapies. For example, Johnson et al. (2014) found that many of the characteristic mutations of the initial glioma tumor (e.g., mutations in *TP53*, etc.) were absent in the recurrent tumor following temozolomide treatment. Similarly, Gerlinger et al. (2012) reconstructed phylogenies of primary and metastatic clones of renal cell carcinomas and found spatial heterogeneity within tumors with respect to mutations in *mTOR*, which in turn predict which regions of the tumor will respond to chemotherapeutics that inhibit *mTOR* pathway signaling. Such examples illustrate how phylogenetic analyses of clonal heterogeneity within tumors can be used to identify the molecular basis of response to therapies, both as a prognostic tool and (potentially) for developing personalized therapies based on the profile of somatic mutations in a patient's tumor.

Owing to limited recombination, linked early neutral passenger mutations will also be among the synapomorphies at basal nodes. However, because these passengers are random with respect to tumor fitness, they are expected to be unique to individual tumors. In contrast, shared mutations in unrelated individuals suggest similar selective pressures in common driver genes. Furthermore, mutations that are synapomorphies of metastatic cell lines will potentially include genetic alterations that facilitate invasion, metastasis, and local adaptation. Several phylogenetic analyses of tumor subclones have leveraged this approach to identify genes and mutations that are characteristic of metastasis (Yachida et al. 2010, Gerlinger et al. 2012, Gundem et al. 2015, M.H.K. Hong et al. 2015, Schwarz et al. 2015).

Reconstructing the phylogeny of clonal lineages and mapping the distribution and order of mutations pose additional technical challenges. In the absence of an exhaustive reconstruction of clonal phylogeny within the primary tumor, it is often difficult to determine whether the metastatic subclones represent a monophyletic or a polyphyletic lineage. Such information is necessary to determine whether the metastatic subclones are actually unique to the metastases, as opposed to having been inherited from unsampled subclones in the primary tumor (**Figure 2**). If the metastatic subclones have an evolutionary prehistory within the primary tumor, one cannot unambiguously infer the functional role of these mutations in metastasis (Fidler & Kripke 1977, W.S. Hong et al. 2015). Additionally, heterogeneity within primary tumors and linked passenger mutations may confound efforts to use the order of mutations on a phylogeny to determine genomic markers of metastatic tumors. Using data from different cancer types, Zhao et al. (2016) found that in some cases the metastatic tumor subclones are more phylogenetically basal than dominant subclones in the primary tumor. Their results challenge the linear model of tumor progression in which metastatic genotypes are assumed to be highly derived relative to primary tumor cells. In fact, in at least some cases, phylogenetically basal primary tumor cells have the potential for invasion and metastasis, consistent with the parallel model for metastatic progression (Turajlic & Swanton 2016).

Nevertheless, tree-based approaches to cancer gene identification have advantages over methods that simply identify the most highly mutated genes. These approaches provide a time dimension that is absent from mutation counts and frequencies and allow the application of comparative methods (Felsenstein 1980) that distinguish covarying characteristics of tumor cells that are due to relatedness among subclones from shared characteristics that are the consequence of common selective pressures (Zhang et al. 2010). Furthermore, mapping the distribution of mutations onto a phylogeny allows estimation of the order, direction, and number of mutational changes that have occurred. Nielsen & Yang (2003) and Tamuri et al. (2012) used a Markov model of the substitution rate of nucleotide i to j at site k under positive selection, in other words:

$$q_{ij,k} = \mu_{ij} \frac{S_{ij,k}}{1 - Exp[-S_{ij,k}]}, \quad 4.$$

where μ_{ij} is the mutation rate multiplied by the fixation probability of a beneficial allele ($S_{ij,k}$ is the selection coefficient). The substitution rate $q_{ij,k}$ is estimated from the substitution frequencies in

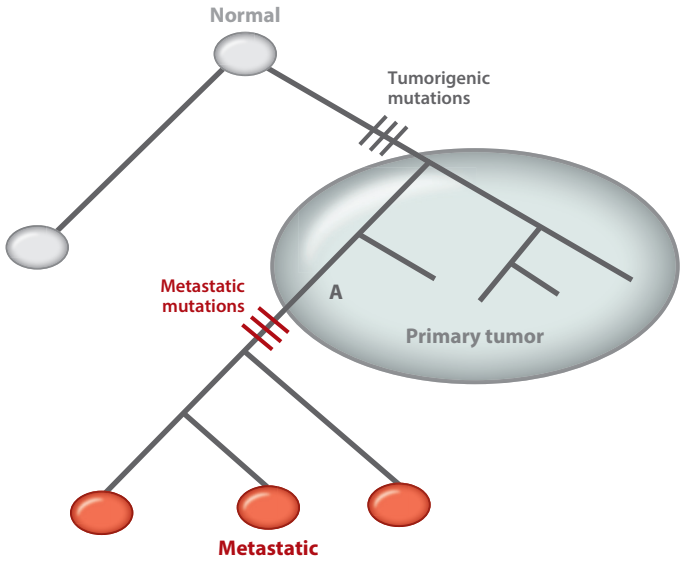

Figure 2

A phylogeny of primary and metastatic tumor subclones. Tumorigenic mutations are among the synapomorphies defining the most basal node of the tumor phylogeny with respect to an outgroup of normal cells, whereas other mutations are unique (autapomorphic) to individual subclones within the primary tumor. If we assume that specific mutations are necessary to allow metastasis, such mutations are synapomorphies of the metastatic lineage subtree. Note that in spite of this scenario being otherwise consistent with a linear model of tumor progression, subclone A in the primary tumor is more closely related to the metastatic lineage than it is to the remaining subclones within the primary tumor.

the coding regions of cancer cell exomes, whereas the background mutation rate μ_{ij} is approximated from background substitution frequencies in noncoding regions. Their models allow a maximum likelihood estimator for S to be computed for each site. A similar approach, albeit with mutation rates estimated without a phylogeny, was used by Foo et al. (2015) to distinguish driver mutations from linked neutral passengers in colorectal cancer genomes.

Mutation Counts and Parallel Mutation

Another approach to identifying cancer drivers is to compare the frequencies of silent versus nonsynonymous somatic mutations in a gene (Nei & Gojobori 1986) by estimating the parameter $\omega = dN/dS$, where dN is the fraction of nonsynonymous mutations relative to the number of nonsynonymous sites in a gene and dS is the relative fraction of synonymous (silent) mutations. The Nei-Gojobori test for selection has been applied in recent publications (Ostrow et al. 2014, Shpak et al. 2015) to identify genes under directional selection in tumors. A similar method, comparing intron/exon junction sites (Chen et al. 2015) has also been used to identify selection on splice sites in cancer genomes. When human sequences are compared with mammalian outgroups, most genes were found to be under strong purifying selection ($\omega \ll 1$) in the germline, whereas the same genes in cancer cells have ω close to 1. A small number of the genes have statistically significant $\omega > 1$, indicating positive selection. For the majority, however, the ω values indicate that purifying selection is relaxed in cancer genomes for all but a small fraction of mutated genes, in other words, many genes that are under negative selection in the germline are evolving neutrally in cancer. The Nei-Gojobori method also identifies a small subset of genes that are under significant

purifying selection in tumors, potentially indicating tumor-specific function when the purifying selection is stronger in cancer genomes than in the germline.

Comparisons of ω between tumors also have the potential to identify specific selection pressures. For example, observing $\omega \gg 1$ for certain genes in recurrent (post-chemotherapy) tumors that are close to or less than 1 in the initial tumor may be indicative of genes under directional selection in response to treatment. Similarly, finding $\omega \gg 1$ in metastatic tumors but not in a primary tumor may provide a method of identifying genes related to invasion, metastasis, and adaptation to new tissue environments.

A related method for identifying driver mutations is to tally recurrently mutated genes or sites across samples (Mwenifumbo & Marra 2013). If a mutation is a secondary consequence of linkage to a driver or a high mutation rate, it is likely to be unique to a single tumor, because fixation of a variant at the same site in multiple tumors is unlikely in the absence of similar selection pressures. Furthermore, identification of mutational recurrence can be applied both at the level of mutated genes/coding sequences and to specific mutated sites. The latter provides the potential to identify specific mutations in driver genes, especially specific gain-of-function mutations in oncogenes that are not otherwise highly mutated. For example, Melton et al. (2015) and Smith et al. (2015) identified recurrent mutations in the regulatory regions of several cancers, including the promoter regions of the telomerase *TERT* and in various oncogenes. A similar approach was used in analyses of somatic mutations in the exomes of several cancer types, including breast cancer (Ellis et al. 2012, Shah et al. 2012), melanoma (Berger et al. 2011), and glioma (Shpak et al. 2015). The joint occurrence of multiple site-specific mutations allows the classification of tumors within a cancer type into subclasses with characteristic mutational profiles, analogous to the classification of subtypes based on expression profiles.

The frequency at which a recurrent mutation occurs across samples can be used as an estimator of fixation probability and of the scaled selection coefficient s (from a modified version of Equation 9 below). Likelihood-based methods for estimating selection coefficients from recurrent mutations in a single population were discussed in Ezawa et al. (2013), and similar methods may be applied to recurrent mutations in different populations, such as recurrent mutations in unrelated tumors. Quantifying the direction and strength of selection on specific sites has the potential to identify functionally significant regions of cancer genes and their proteins, providing a novel approach for the discovery of therapeutic targets.

Genetic Heterogeneity and Coalescent Theory

The utility of ω as an estimate of selection is undermined by polymorphism (Kryazhimskiy & Plotkin 2008) and codon usage bias (Spielman & Wilke 2015). Additionally, tests based on comparisons of silent versus nonsynonymous mutations fail to leverage allele or haplotype frequencies and reference explicit models of neutral evolution as null hypotheses.

Genetic variation within tumors is partitioned among mutational subclones, so that segregating variation (polymorphism) at a site/locus within a tumor means that some cells in the tumor have germline genotypes at the variable sites, whereas others have a copy of the somatic mutation. Somatic mutations are fixed in the tumor when all cells inherit the somatic mutation (regardless of homo- or heterozygosity). Sampling and sequencing spatially distinct sectors of a solid tumor allow one to estimate the number of segregating sites, average genetic distances between polymorphic genes/regions within a tumor, and compute mutation clone frequency spectra, in other words, the number of mutations that are represented by $k-$tuples ($k = 1 \ldots n$) in a sample of n tumor sectors. These observations can then be compared with expected observations under neutral evolutionary models to determine the extent to which genetic heterogeneity within tumors is driven by selection versus random genetic drift.

Under a neutral coalescent model, we expect the clonal genealogy to have a topology consistent with an exponential distribution of waiting times between coalescent events—in other words, under a Wright-Fisher model of genetic drift, the expected coalescent time is equal to the effective number of cells N_e in a population (even though the cells are diploid, there is no segregation of chromosomes, so there are effectively N rather than $2N$ genotypes). Consequently, the expected pairwise distance between genotypes is $\theta = 2N_e\mu$ for mutation rate μ (defined with respect to a single gene, a chromosomal region, or the entire genome). The expected number of segregating sites will also be proportional to θ. A number of tests for selection, such as Tajima's D (Tajima 1989, Korneliussen et al. 2013), compare estimates of θ based on the number of variant sites to the value estimated from pairwise differences. A related approach to evaluating neutrality is to consider the spectrum of allele or haplotype frequencies in a population. Under a neutral model, the number of mutations η_i represented by i in a sample of n cells is $E[\eta_i] = \theta/i$. In other words, we expect there to be θ variants represented by a single individual, $\theta/2$ represented by 2 in the sample, etc. (e.g., Durrett 2008). Significant deviations from these expected frequency spectra reflect the effects of clonal selection.

There are several potential challenges with direct application of these neutral models to cancer genomic data. The assumption of constant population size is not a good model for the demography of tumor cells, because tumors grow. As a first approximation, it is reasonable to assume exponential growth in the number of cells, even though the number of dividing cells in a tumor probably increases at a slower than exponential rate due to necrosis and other factors. Population growth leads to longer branch lengths near the tips and shorter branch lengths near the root, so that the expected number of mutations η_i represented by i in a sample of n cells given an exponential growth rate r is:

$$E[\eta_i] \approx \frac{\mu}{r}\frac{n}{i(i-1)} \text{for } 2 \leq \mathrm{i} < n \qquad 5.$$

(Durrett 2013), and the mutation clone frequency spectra can be estimated numerically for other models of cell growth. In a recent analysis, Ling et al. (2015) compared mutation clone spectra from hepatocellular carcinoma using Equation 5 and under slower (power-law) models of population growth and found that the observed numbers of mutation clones out of 289 samples were consistent (based on goodness of fit) with a neutral model. Specifically, allelic diversity in the hepatocarcinoma was higher than would be expected under purifying selection or recent selective sweeps.

Ling et al.'s (2015) work is one of the few applications of coalescent theory to test for selection in tumors, and its use of nonequilibrium population models represents a major advance over naïve applications of models with constant population size. However, these methods ignore another important complication of neutral theory in tumor biology: namely, the spatial partitioning of clonal diversity. Neutral theory assumes a fully mixed population with randomly distributed genomes, so that any two individuals can coalesce to a common ancestor in the previous generation. In most solid tumors, cell migration is limited, so to a first approximation, only adjacent cells are likely to have shared a common ancestor in the previous generation. A recent simulation-based study by Waclaw et al. (2015) showed that spatial constraints on the coalescent lead to lower predicted genetic heterogeneity than in an unconstrained model.

To see how the constraints of spatial proximity affect coalescent times, consider a population of N cells on a line (**Figure 3**), with the constraint that only neighboring cells can coalesce in the previous generation (to avoid endpoint effects, assume the first and last cell can coalesce, as on a ring). If we impose a Moran-like model in which one cell is selected to die and a distance = 1 neighbor is selected to reproduce per generation, the probability of a pair of cells coalescing one

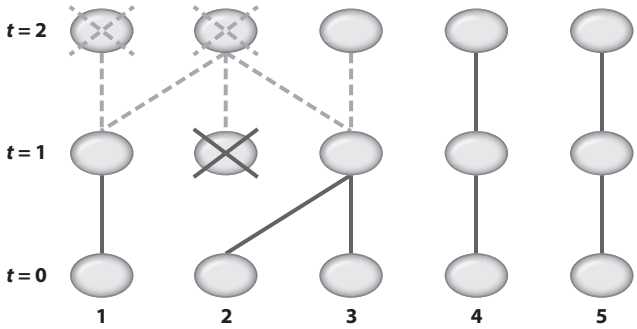

Figure 3

A simple illustration of why spatial structure constrains coalescent events. Assume that in every generation, one cell is selected to die and one of its $d = 1$ neighbors is selected to reproduce (allowing cells 1 and 5 to be neighbors to avoid edge effects). If a cell in position 2 dies at time $t = 1$ and is replaced by one of the two descendants of cell 3 while other cells persist (*solid lines*), it is impossible for cell 1 to share a common ancestor with cells 2 or 3 at $t = 2$, regardless of whether a cell at position 1 or 2 dies $t = 2$ time steps ago (*dashed crosses* represent hypothetical cell deaths and *dashed lines* represent hypothetical descendant cells under the two scenarios). In contrast, this type of interference is absent without spatial constraints. In other words, in a mixed population, any pair of cells can have a common ancestor at time $t = 2$ regardless of cell death or replacement at $t = 1$.

generation ago is the same as for an unstructured population. In other words:

$$P_c(t = 1) = \frac{2}{N(N-1)}. \quad 6.$$

However, the probability that two randomly selected cells share a common ancestor exactly two generations ago is

$$P_c(t = 2) = \frac{N-2}{N^2}\frac{2}{N-1} + \frac{2}{N^2}F(N), \quad 7.$$

for $F(N) = 2/(N-1)$ for $N > 5$, $2/N$ for $N = 5$, and $2(N-2)/(N-1)$ for $N = 4$ (M. Shpak, unpublished material). In contrast, using an unstructured Moran model (e.g., Durrett 2008), the probability of a randomly selected pair coalescing in exactly two time steps is

$$P_c(t = 2) = \frac{2}{N(N-1)}\left(1 - \frac{2}{N(N-1)}\right). \quad 8.$$

This is the result of an interference effect. For example, if two cells coalesce with their neighbor to the left t generations ago, that precludes one of them coalescing with a rightside neighbor $t + 1$ generations ago. This result leads to longer expected branch lengths and greater pairwise genetic distances in a structured model. Similar constraints apply to more realistic 2- and 3-dimensional spatial arrangements of cells. Coalescent models of neutral genetic variation in solid tumors will have to consider these spatial effects to identify deviations from neutrality caused by selection.

GERMLINE MUTATIONS: CANCER RISK ALLELES

In addition to clonal somatic evolution described above, evolution within populations affects the incidence of certain cancers. The discovery of inherited cancer risk through pedigree analyses led epidemiologists and oncologists to propose that some cancers are Mendelian genetic diseases (e.g., Carter et al. 1992). Hereditary cancers are associated with one or a small number of inherited germline mutations that predispose individuals to cancer, and these cancers account for a minority

of cases in comparison with sporadic cancers. For example, <10% of breast and ovarian cancers are due to hereditary incidence (Andersen 1992).

Cancer risk alleles typically have incomplete penetrance because multistage progression requires several mutational changes. A single risk allele is almost always insufficient for cancer to occur; however, inherited mutations substantially increase the lifetime risk of cancer. Using the Armitage-Doll model (Equation 3) with a constant transition rate λ from $i-1$ to i mutations, the expected ratio of relative incidence in those with a single inherited risk mutation at age t compared with those without the inherited risk is $\sim 1/\lambda$, which for small λ leads to lifetime incidence orders of magnitude higher than in the background population and a younger expected age of onset.

Among the best-documented examples of inherited cancer risk are hereditary retinoblastoma, breast cancer, and ovarian cancer. Individuals with a single mutant copy of the *RB1* gene often develop pediatric retinoblastoma, because they need only a single loss-of-function somatic mutation, as opposed to two mutations for noncarriers. Similarly, women who inherit a mutation in the *BRCA1* or *BRCA2* genes have a 5-fold higher lifetime incidence of breast cancer and 10- to 20-fold higher incidence of ovarian cancers than *BRCA*− individuals (Petrucelli et al. 1998); inherited loss-of-function mutations are associated with a 40–85% lifetime risk of breast cancer and a 16–64% lifetime risk of ovarian cancer, versus approximately 12% and 1.2% lifetime risks for the background population, respectively. As with *RB1*, the *BRCA* genes are tumor suppressors, so that the high incidence of breast and ovarian cancer among *BRCA*+ carriers is also due to the necessity of only one mutation to inactivate the gene. As predicted, carriers of mutations in *BRCA* and *RB1* also develop cancers at much younger ages. For example, *RB1*+ carriers develop pediatric retinoblastoma, versus adult-onset spontaneous incidence, and the median age of onset for *BRCA*+ sporadic breast cancers is 42, versus 66 for *BRCA*− sporadic breast cancers (Litton et al. 2012).

Inherited oncogenic mutations are much less common and tend to occur at very low frequencies. Examples include *EGFR* mutations in hereditary lung cancers (Gazdar et al. 2014), gliomas (Wang et al. 2015), and gastric cancers (Torres-Jasso et al. 2015); *RET* mutations in endocrine neoplasia (Margraf et al. 2009); and *P16/CDK4* mutations in hereditary melanomas (Jenkins et al. 2013).

The Population Genetics of Cancer Risk Alleles

To account for the incidence and persistence of cancer risk alleles described above, we can apply population genetic models describing the dynamics of deleterious alleles in finite populations. Consider a population with diploid effective population size N_e, where the initial frequency of the cancer risk allele a is p and the fitness of genotypes are $W_{aa} = 1-s$, $W_{Aa} = 1-hs$, $W_{AA} = 1$. The probability that the risk allele becomes fixed in the population, $U(p)$, is

$$U(p) = 1 - e^{4N_e hsp} \Big/ 1 - e^{4N_e hs}, \qquad 9.$$

whereas the probability of eventual loss of a (ignoring recurrent mutations) from the population is $1-U(p)$. Fixation probabilities for completely recessive ($h=0$) mutations can be similarly estimated from diffusion models (Kimura 1964). The fixation probability of a deleterious (cancer-causing) allele increases with small population size and/or $s \ll 1$, so that in the limit of a selectively neutral allele ($s=0$), we have $U(p)=p$. This is also the approximate fixation probability when $s < 1/2N_e$. The value of h depends on the gene and mutation. For example, oncogene activations are typically dominant ($h \approx 1$), so they are less likely to reach high frequencies. The value of s depends on the penetrance of the cancer risk, in other words, the probability that an individual inheriting the allele actually develops cancer, and on the expected age of onset. The selection coefficient approaches $s=1$ with early onset and high penetrance and 0 with late onset and/or low penetrance.

The three population genetic scenarios we consider for the incidence and persistence of cancer risk alleles in a population are: (*a*) unconditionally deleterious alleles, (*b*) effectively neutral alleles, and (*c*) antagonistic pleiotropy.

Unconditionally Deleterious Risk Alleles

Some cancer risk alleles have high penetrance because the total number of additional mutations necessary to transition from normal cells to malignant tumors is comparatively low in certain cancers, (e.g., many pediatric cancers, which often have fewer somatic mutations than adult cancers; Radtke et al. 2009). Even with low penetrance, alleles that have no fitness benefit to individuals while conferring the cancer risk (to reproductive/prereproductive individuals) are unconditionally deleterious, and it follows from Equation 9 that the probability of eventual loss is close to 1 in a large population, although for weak selection, the time until loss can be quite long, for example, of the order $t \sim 2ln(2N) - ln(2Ns)$ generations for a recessive allele (Kimura & Ohta 1969). Furthermore, many cancer risk alleles are associated with genetic diseases other than cancer, leading to still stronger negative selection. For example, mutations in *NF1* that disrupt the *Ras* signaling pathway are linked to neurofibromatosis as well as to various cancers of the nervous system, and Li-Fraumeni syndrome is linked to germline mutations in *TP53* (Bunz 2008).

Deleterious cancer risk alleles are expected to have population dynamics resembling those of recessive lethal genetic disease traits such as hemophilia, so most hereditary cancer risk alleles occur at frequencies $\ll 1$ in the general population. Even with survival times prolonged by modern medicine, we should not expect unconditionally deleterious risk alleles to reach high frequencies in large populations. Therefore, the long-term persistence of deleterious mutations is a consequence of mutation-selection equilibrium rather than neutral drift. If wild-type alleles mutate to risk alleles at rate μ per generation, the equilibrium frequency of a recessive deleterious allele with selection coefficient s will be $\hat{p} = \sqrt{\mu/s}$, versus the smaller μ/s for dominant risk alleles, which explains why most hereditary cancers associated with typically recessive tumor suppressor genes are more common than those with typically dominant oncogene mutations.

There are exceptions to the expected low frequency of risk alleles, such as the high incidence of *BRCA1+* alleles among Ashkenazi Jews (8–10% versus ~4% frequency in Caucasians generally). This may be a consequence of founder effects and genetic bottlenecks where N_e is very small (Levy-Lahad et al. 1997) and in which even deleterious alleles can reach high frequency. Alternatively, the high incidence of these risk alleles, which are expected to be under purifying selection due to breast cancer onset among some reproductive individuals, may in fact be maintained by antagonistic pleiotropy (see section on Antagonistic Pleiotrophy below).

Senescence and Neutral Evolution

Purifying selection on risk alleles is nearly absent if the cancer onset is in a postreproductive individual, regardless of high penetrance. The alleles may increase the risk of cancer incidence $I(t)$ by a factor $\sim 1/\lambda$, but if the values of $I(t)$ remain close to 0 for young individuals carrying the risk allele, the mutation is effectively neutral. This follows from our understanding of the evolution of senescence through the accumulation of mutations that increase the probability of death or morbidity in older individuals while having no deleterious effect on the young. Senescence-associated mutations are not negatively selected, because pathological phenotypes in postreproductive individuals do not contribute to the fitness of an organism.

This original heuristic argument (Medawar 1952) was formalized (Hamilton 1966) by defining the Malthusian parameter r (instantaneous growth rate) as a function of the age-specific survival rates: b_i (probability of surviving from ith to $i + 1$ age class) and the expected number of offspring at age i, f_i. If a mutation decreases b_i or f_i in reproductive and prereproductive ages, we have

$\partial r/\partial b_i$ and $\partial r/\partial f_i < 0$ and negative selection coefficients on the risk allele. In postreproductive individuals, f_i is near 0, therefore $\partial r/\partial b_i \approx 0$, so that cancer risk alleles are effectively neutral ($s \approx 0$), even if the mutation causes b_i to be 0 for postreproductive i. With effectively neutral selection coefficients $s < 1/2N_e$, even risk alleles that occasionally cause cancer at reproductive ages will not be under purifying selection with sufficiently low occurrence among the young.

Even though such inherited cancer risk is deleterious from the standpoint of human health, it is neutral with respect to natural selection, so that the dynamics of risk alleles will be driven by random genetic drift rather than selection. Unlike deleterious risk alleles associated with early onset, we expect to observe such risk alleles at appreciable frequencies in populations, and the probability of such a risk allele's fixation is equal to its initial frequency p. The low incidence of cancer in large, long-lived organisms that remain fertile through most of their lives, such as elephants (see above discussion of Peto's paradox), illustrates an opposite scenario of strong selection against late-onset cancers.

Weak selection may also explain the high frequencies of alleles predictive of cancer risk in genome-wide association (GWAS) studies. For example, the GWAS analyses of gliomas (Wrensch et al. 2009, Melin et al. 2013) identified risk alleles linked to *RTEL* (a telomerase) and to *CDKN2B* that had background frequencies in the general population ranging from 0.10 to 0.40. These high frequencies can be explained by the fact that their associated risk odds ratios are typically of the order 1.2–1.5 for the risk alleles individually, in spite of a high familial incidence of glioma when multiple risk alleles co-occur. The appropriate population genetic model for the dynamics of multiple low-risk alleles is selection on a polygenic phenotypic trait. The polygenic model accounts for many familial cancers that don't show Mendelian patterns of inheritance but nevertheless correlate with an individual's genetic background.

In polygenic traits in which the contributions of individual loci are small, selection at each locus will be weak. For a quantitative trait with phenotypic variance V under directional selection with coefficient S, the strength of selection on an allele at locus i contributing g_i to the phenotype is $s_i = Sg_i/V$, assuming that the contribution of each locus is additive and statistically independent (Kimura & Crow 1978, Lynch 1984). Consequently, if all g_i of all alleles contributing to cancer risk are small, the strength of selection on each allele will be negligible. The strength of selection will be weaker still if the contribution of the alleles at each site is conditional on a particular genetic background, in other words, if the alleles predispose to cancer only in the presence of other risk alleles. Therefore, the risk alleles individually will often be effectively neutral with respect to purifying selection.

The existence of weakly predictive risk alleles suggests that a fixed multi-hit model for each cancer type is an oversimplification. For most cancers, there are multiple mutational paths that can lead to malignant phenotypes. Indeed, surveys of the TCGA (2012a–c) data for somatic mutations in breast, colorectal, and squamous cell lung cancers have revealed that individual tumors within subtypes of each cancer (defined by common histology, clinical properties, and gene expression profiles) can nevertheless have very different somatic mutation profiles. The implication is that there are many mutations, both somatic and germline, that are neither necessary nor sufficient for progression to cancer but that potentially enhance the probability of carcinogenesis given a background set of mutations, consistent with risk alleles with individually weak effects.

Antagonistic Pleiotropy

Although neutral evolution may account for the high frequency of cancer risk alleles with low penetrance or late age of onset, a more interesting scenario occurs if an allele that increases the risk of cancer also has pleiotropic fitness benefit to the organism. Such instances of antagonistic pleiotropy have been invoked as an alternative to the neutral model of aging, for example, a

mutation that increases survival and fecundity coefficients b_i or f_i while decreasing b_j or f_j, (typically for ages $j > i$), such that r increases despite the trade-off. In contrast to the selectively neutral r scenarios, j need not be in a postreproductive age class for the antagonistic effect to increase fitness (Williams 1957).

Antagonistic pleiotropy has been proposed as the mechanism maintaining a high frequency of cancer risk mutations such as *BRCA1* in some populations. To account for the persistence of *BRCA1/2+* alleles at comparatively high frequencies in some populations, French et al. (2006) proposed that *BRCA1* mutations lead to longer telomeres and argued that this enhances oocyte production and female fertility. If true, the high incidence of breast and ovarian cancer in *BRCA+* women is potentially a pleiotropic by-product of selection for higher fertility. Some natural fertility studies of *BRCA+* families (Kwiatkowski et al. 2015) have supported this hypothesis, although other analyses have challenged it (Smith et al. 2013) and argued that *BRCA+* mutations may in fact diminish fertility due to a failure of DNA repair mechanisms.

Mutations enhancing the expression and action of growth factors such as *IGF-1* also result in increased cancer incidence. The fitness advantages of greater *IGF* activity are manifold, from a more effective ability to extract maternal resources in the placenta (Constancia 2000) to the fitness benefits of larger body size in many mammals. However, mutations upregulating *IGF* are known to be associated with sarcomas and other cancers in humans (Arnaldez & Helman 2012, O'Neill et al. 2013) and other animals. One of the reasons for shorter life spans in large dog breeds is a greater predisposition to cancer (Greer et al. 2011).

Selection on telomerases may also lead to increased cancer risk. The shortening of telomeres through multiple generations of meiosis is thought to be one of the underlying causes of cell death and tissue aging (Hornsby 2007). Natural selection favors telomere elongation to counter the effects of aging; in other words, telomerase upregulation can extend the life span of cell lines by preventing their senescence. One of the characteristics of cancer cells is their stem cell–like immortality due to telomere elongation. Consequently, positive selection for germline mutations that upregulate telomerase may increase cancer risk, accounting for the high frequency of telomerase mutations identified in GWAS studies of various cancer types (Keefe & Liu 2009).

Unconditionally deleterious and neutral cancer risk alleles are easily distinguished, because only the latter can reach high frequencies in a population. In contrast, both genetic drift of selectively neutral alleles and antagonistic pleiotropy can maintain risk alleles at high frequencies. The effects of selection can be disentangled from those of genetic drift using statistical tests that compare observed haplotype and allele frequencies with those predicted by neutral models (e.g., Fay & Wu 2001, Nielsen 2001). For example, Cheng et al. (2014) analyzed evidence of germline selection in glioma-associated genes by comparing synonymous versus nonsynonymous sites and found that genes linked to cancer experience strong purifying selection in the germline relative to background exome regions.

CANCER AND MULTILEVEL SELECTION

The evolution of life on Earth has been characterized by increases in biological complexity through the aggregation of replicating entities at one level of organization into higher-level replicating entities. The outcomes of such cooperative associations include the origin of eukaryotic cells through endosymbiosis, the origin of multicellular organisms from colonies of unicellular eukaryotes, and the origin of complex social organizations from groups of multicellular organisms (Buss 1988, Maynard Smith & Szathmary 1995). Conflicts occur between higher-level replicators and the nested lower-level replicators whenever both have the ability to mutate and reproduce. For example, clonal lineages of cells evolve through somatic selection because cancer cells, like

individual organisms, are self-replicating and because somatic mutations provide a source for heritable variation in cell phenotype and fitness. The proliferation of so-called selfish cancer cells typically reduces the fitness of the host multicellular organism.

Consequently, cancer has been characterized by many evolutionary biologists as a case study in conflict between two classes of replicating entities—cells and organisms (Burt & Trivers 2006, chapter 11)—and the various mechanisms that reduce genetic variation among cells limit the potential for somatic selection. The strongest constraint on somatic evolution is the result of somatic tissue in most multicellular organisms being derived from a single fertilized zygote, so that the cells are genetically identical apart from somatic mutations. Furthermore, following Weissman (1893), animal germline cells are separated from somatic tissue during early ontogenesis, so that somatic mutations are not transmitted to offspring. The limited genetic variation among cells creates a greater genetic incentive via kin selection for cells to cooperate and fewer genetic incentives for defection. Multicellular organisms also have numerous cellular fail-safe mechanisms that reduce the chances of cells with somatic mutations progressing to tumors, and many selective pressures maintain fidelity of DNA replication. DNA repair ensures a comparatively low frequency of somatic mutations, even in cancer cells. Surveys of cancer genomes by the TCGA (2012a–c) consortium have typically found ∼100 unique somatic point mutations in the exomes of breast cancer, lung cancers, and gliomas. Contact inhibition of cell proliferation via tumor suppressor genes further restricts the potential for clonal selection. Finally, the number of actively dividing (stem) cells in any tissue that can potentially progress to cancer is much lower than the census population size of cells, which limits the potential for somatic mutations to occur in most tissues (Cairns 2002).

Frequently in the cancer literature (e.g., Aktipis et al. 2015), analogies are made between the mechanisms that multicellular organisms use to repress cancers and the means that social groups have of preventing or punishing noncooperative defectors (such as inhibition of worker reproduction by queen wasps). However, there is an important difference between the evolutionary conflict of cells versus organism caused by cancer and actual instances of multilevel selection in other biological systems. Reducing the fitness of one class of replicators through selection on replicators at another level of organization is not sufficient for multilevel selection per se. In true multilevel selection, the lower-level replicator has to have the potential to contribute to the genotype or phenotype of the next generation (with respect to the higher-level replicator's generation time). This includes the potential to drive genotype frequencies in opposite directions across generations. In fact, somatic mutations are an evolutionary dead end across generations, so that multilevel selection (sensu stricto) on cancer cells is rarely possible (Gardner 2015).

To see that this is formally the case, consider Price's equation (Price 1970, Frank 1995), a mathematical representation of multilevel selection on a trait value x (which can be an allele frequency, a quantitative trait, etc.):

$$\Delta\bar{x} = \frac{Cov(W, x) + E[W\Delta x]}{\overline{W}}, \qquad 10.$$

where $\Delta\overline{x}$ is the change in mean genotype or phenotype value across generations of a biological replicator, such as a multicellular organism; W is the fitness effect of the trait for the organism ($\overline{W}$ is the mean population fitness); and Δx represents the deviation in trait value due to selection at some other level of organization, such as meiotic drive, transposon replication, etc. For example, if x is the number of transposable elements in the germline genome, natural selection acting on organisms may favor smaller genomes [i.e., $Cov(W, x) < 0$], whereas replication of transposons within the genome tends to bloat genome size [i.e., $E(w\Delta x) > 0$]. Replicated transposons within

the genome are transmitted to the next generation when the organism reproduces, while selection on organisms preferentially removes genomes with a high transposon load.

In contrast, consider the case where x represent the presence or absence of a mutation associated with a cancer cell line and $\overline{x}$ is its frequency among organisms in the population. Because somatic mutations aren't typically transmitted to offspring, positive somatic selection still results in $E(w\Delta x) = 0$ for x between generations. Therefore, describing cancer as an instance of multilevel selection is misleading, except in the rare cases of transmissible sarcomas in certain animals or of heritable tumorigenic mutations in plants that lack germline segregation.

Among the known transmissible cancers are canine transmissible venereal sarcoma in dogs, transmissible facial sarcoma in Tasmanian devils (*Sarcophilus harrisii*) (Murchison 2009, Welsh 2011), and a transmissible leukemia in *Mya arenaria* clams (Metzger et al. 2015). The contagiousness of these cancers was originally believed to be from the transmission of an oncogenic virus (such as cervical cancer induced by human papilloma virus). In fact, analyses of the cancers of recently infected animals indicated that the genomes of the tumor cells were related to those of prior hosts rather than to the genomes of the current hosts—clear evidence that the tumor cells are able to infect new host animals directly. These analyses showed that the canine sarcoma cancer cells are transmitted sexually, the Tasmanian devil facial sarcomas are spread from one animal to another by biting, and the cancer cells in the clams disperse into seawater to infect host siphons. From the standpoint of epidemiology, such cancer cells all behave as infectious unicellular parasites.

DISCUSSION AND FUTURE DIRECTIONS

Evolutionary theory's potential to inform basic and applied cancer research has only recently begun to be tapped. In particular, the use of phylogenies and tests for selection based on neutral models offer a potentially great advance over simple tallying of mutations as a means of identifying cancer drivers. Similarly, advances in evolutionary genomics will continue to provide insights into the genetic basis of cancer progression and other aspects of oncobiology, including clinically relevant characteristics such as evasion of host immune response and adaptive resistance to chemotherapies. Cancer therapies will increasingly leverage data on tumor genetic heterogeneity and adaptive evolution in cancer cells to create somatic genotype-specific targeted treatments (e.g., Jamal-Hanjani et al. 2014) and multidrug therapies that circumvent the ability of tumor cell populations to evolve resistance to single drugs (Burrell & Swanton 2014).

To effectively identify targets of selection in cancer genomes, it will be necessary to develop population genetic models that take into consideration the particular characteristics of tumors, including the constraints imposed by spatial structure, genome-wide linkage disequilibria due to clonal selection, and deviations from equilibrium models due to cell population dynamics. Otherwise, direct application of existing population genetic theory to identify loci under selection or estimates of the strength and direction of selection will provide crude approximations at best and misleading results at worst. This is particularly true of any methods that construct null hypotheses under the assumptions of neutral coalescent models.

Studies of how neoplasms evolve in response to their host environments and therapies also offer researchers an in situ model system of evolution on human timescales (like microbial model systems and unlike most animal and plant models). Just as evolutionary theory can inform our understanding of cancer biology, advances in cancer biology may offer insights into evolution theory. For example, although this review focused specifically on mutation-driven cancer evolution, recent advances in evolutionary epigenetics have also informed our understanding of adaptation in cancer cells. Organisms adapt to their environment not only through genomic mutation followed by Darwinian selection but also through epigenetic changes (i.e., phenotypic plasticity). Much

of the dysregulation in gene expression found in cancer cells is the consequence of epigenetics rather than of mutation, in other words, differential gene expression arising from methylation patterns and disruptions of the various signaling pathways by which cells respond to environmental stress. Brock et al. (2009) proposed a genetic assimilation model in which tumor cells show cancer phenotypes through epigenetic effects and phenotype plasticity that are later canalized through mutation. This example illustrates how cancer biology may prove to be a useful model system for integrating systems biology and epigenetics with evolutionary theory.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank Daniel Bolnick, Matthew Cowperthwaite, and Jeffrey Townsend for helpful discussion and the St. David's Foundation for financial support.

LITERATURE CITED

Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, et al. 2015. Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA* 314(17):1850–60

Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC. 2015. Cancer across the tree of life: cooperation and cheating in multicellularity. *Philos. Trans. R. Soc. B* 370:20140219

Anderson DE. 1992. Familial versus sporadic breast cancer. *Cancer* 70(6 Suppl):1740–46

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

Arnaldez FI, Helman LJ. 2012. Targeting the insulin growth factor receptor 1. *Hematol. Oncol. Clin. North Am.* 26(3):527–42, vii–viii

Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, et al. 2011. The genomic complexity of primary human prostate cancer. *Nature* 470(7333):214–20

Bignold L. 2007. Aneuploidy as a counter to Muller's ratchet in cancer cells. *Cancer Genet. Cytogenet.* 178(2):173–74

Boveri T. 1914. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J. Cell Sci.* 121(Suppl):1–84

Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, et al. 2010. Accumulation of driver and passenger mutations during tumor progression. *PNAS* 107(43):18545–50

Breivik J. 2004. The evolutionary origin of genetic instability in cancer development. *Semin. Cancer Biol.* 15:51–60

Brock A, Chang H, Huang S. 2009. Non-genetic heterogeneity–a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* 10(5):336–42

Bunz F. 2008. *Principles of Cancer Genetics*. New York: Springer

Burrell RA, McGranahan N, Bartek J, Swanton C. 2013. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501(7467):338–45

Burrell RA, Swanton C. 2014. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol. Oncol.* 8(6):1095–111

Burt A, Trivers R. 2006. *Genes in Conflict*. Cambridge, MA: Harvard Univ. Press

Buss L. 1988. *The Evolution of Individuality*. Princeton, NJ: Princeton Univ. Press

Cahill DP, Kinzler KW, Vogelstein B, Lengauer C. 1999. Genetic instability and darwinian selection in tumours. *Cell* 15(12):M57–60

Cairns J. 1975. Mutation, selection, and the natural history of cancer. *Nature* 255:197–200

Cairns J. 2002. Somatic stem cells and the kinetics of mutagenesis and carcinogeneis. *PNAS* 99:10567–70
Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC. 1992. Mendelian inheritance of familial prostate cancer. *PNAS* 89(8):3367–71
Caulin AF, Maley CC. 2011. Peto's paradox: evolution's prescription for cancer prevention. *Trends Ecol. Evol.* 26(4):175–82
Chen H, Xing K, He X. 2015. The dJ/dS ratio test reveals hundreds of novel putative cancer drivers. *Mol. Biol. Evol.* 32(8):2181–85
Cheng F, Jia P, Wang Q, Lin C-C, Li W-H, Zhao Z. 2014. Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. *Mol. Biol. Evol.* 31(8):2156–69
Constancia M. 2000. Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nat. Genet.* 26(2):230–36
Cowperthwaite MC, Bull JJ, Ancel Meyers L. 2006. From bad to good: fitness reversals and the ascent of deleterious mutations. *PLOS Comput. Bio.* 2:e141
Crespi B, Summers K. 2005. Evolutionary biology of cancer. *Trends Ecol. Evol.* 20(10):545–52
Croce CM. 2008. Oncogenes and cancer. *N. Engl. J. Med.* 358(5):502–11
Danielsen HE, Pradhan M, Novelli M. 2015. Revisiting tumour aneuploidy—the place of ploidy assessment in the molecular era. *Nat. Rev. Clin. Oncol.* 13:291–304
Dawkins R. 1989. The evolution of evolvability. In *Artificial Life*, ed. Langton C, pp. 201–20. New York: Addison Wesley
Durrett R. 2008. *Probability Models for DNA Sequence Evolution*. New York: Springer
Durrett R. 2013. Population genetics of neutral mutations in exponentially growing cancer cell populations. *Ann. Appl. Probab.* 23(1):230–50
Eigen M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58(10):465–523
Eigen M, Schuster P. 1979. *The Hypercycle*. Berlin: Springer-Verlag
Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, et al. 2012. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486(7403):353–60
Ezawa K, Landan G, Graur D. 2013. Detecting negative selection on recurrent mutations using gene genealogy. *BMC Genet.* 14(1):37
Fay JC, Wu CI. 2001. The neutral theory in the genomic era. *Curr. Opin. Genet. Dev.* 11(6):642–46
Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* 78(2):737–56
Felsenstein J. 1980. Phylogenies and the comparative method. *Am. Nat.* 125:1–15
Fidler IJ, Kripke ML. 1977. Metastasis results from preexisting variant cells within a malignant tumor. *Science* 197(4306):893–95
Foo J, Liu LL, Leder K, Riester M, Iwasa Y, Lengauer C, Michor F. 2015. An evolutionary approach for identifying driver mutations in colorectal cancer. *PLOS Comput. Biol.* 11(9):e1004350
Fox EJ, Loeb LA. 2010. Lethal mutagenesis: targeting the mutator phenotype in cancer. *Sem. Cancer Biol.* 20:353–59
Frank SA. 1995. George Price's contributions to evolutionary genetics. *J. Theor. Biol.* 175(3):373–88
Frank SA. 2004. A multistage theory of age-specific acceleration in human mortality. *BMC Biol.* 2:16
Frank SA. 2007. *Dynamics of Cancer*. Princeton, NJ: Princeton Univ. Press
French JD, Dunn J, Smart CE, Manning N, Brown MA. 2006. Disruption of *BRCA1* function results in telomere lengthening and increased anaphase bridge formation in immortalized cell lines. *Genes Chromosomes Cancer* 45(3):277–89
Gajewski TF, Schreiber H, Fu Y-X. 2013. Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* 14(10):1014–22
Gardner A. 2015. The genetical theory of multilevel selection. *J. Evol. Biol.* 28(2):305–19
Gazdar A, Robinson L, Oliver D, Xing C, Travis WD, et al. 2014. Hereditary lung cancer syndrome targets never smokers with germline *EGFR* gene T790M mutations. *J. Thorac. Oncol.* 9(4):456–63
Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, et al. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366:833–92
Greaves M. 2000. *Cancer: The Evolutionary Legacy*. Cambridge, UK: Cambridge Univ. Press

Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. 2011. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol.* 12(8):785–94

Greer KA, Hughes LM, Masternak MM. 2011. Connecting serum IGF-1, body size, and age in the domestic dog. *Age* 33(3):475–83

Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, et al. 2015. The evolutionary history of lethal metastatic prostate cancer. *Nature* 520(7547):353–57

Hamilton WD. 1966. The moulding of senescence by natural selection. *J. Theor. Biol.* 12(1):12–45

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* 100(1):57–70

Hanahan D, Weinberg RA. 2011. The hallmarks of cancer: The next generation. *Cell* 144:646–74

Harley CB. 2002. Telomerase is not an oncogene. *Oncogene* 21(4):494–502

Harris AL. 1985. DNA repair and resistance to chemotherapy. *Cancer Surv.* 4(3):601–24

Hegi ME, Diserens A-C, Gorlia T, Hamou M-F, de Tribolet N, et al. 2005. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* 352(10):997–1003

Hong MKH, Macintyre G, Wedge DC, Van Loo P, Patel K, et al. 2015. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nat. Commun.* 6:6605

Hong WS, Shpak M, Townsend JP. 2015. Inferring the origin of metastases from cancer phylogenies. *Cancer Res.* 75(19):4021–25

Hornsby PJ. 2007. Telomerase and the aging process. *Exp. Gerontol.* 42(7):575–81

Jamal-Hanjani M, Hackshaw A, Yenting N, Shaw J, Dive C, Quezada S, et al. 2014. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLOS Biol.* 12:e1001906

Jenkins NC, Jung J, Liu T, Wilde M, Holmen SL, Grossman D. 2013. Familial melanoma-associated mutations in p16 uncouple its tumor-suppressor functions. *J. Invest. Dermatol.* 133(4):1043–51

Johnson BE, Mazor T, Hong C, Barnes M, Aihara K, McLean CY, et al. 2014. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* 343(6167):189–93

Keefe DL, Liu L. 2009. Telomeres and reproductive aging. *Reprod. Fertil. Dev.* 21(1):10–14

Kimura M. 1964. Diffusion models in population genetics. *J. Appl. Probab.* 1(2):177–232

Kimura M, Crow JF. 1978. Effect of overall phenotypic selection on genetic change at individual loci. *PNAS* 75(12):6168–71

Kimura M, Ohta T. 1969. The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* 63:701–9

Knudson AGJ. 1971. Mutation and cancer: statistical study of retinoblastoma. *PNAS* 68(4):820–23

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinform.* 14(1):289

Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. *PLOS Genet.* 4(12):e1000304

Kwiatkowski F, Arbre M, Bidet Y, Laquet C, Uhrhammer N, Bignon Y-J. 2015. BRCA mutations increase fertility in families at hereditary breast/ovarian cancer risk. *PLOS ONE* 10(6):e0127363

Langley RR, Fidler IJ. 2011. The seed and soil hypothesis revisited—the role of tumor-stroma interactions in metastasis to different organs. *Int. J. Cancer* 128(11):2527–35

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–18

Leigh EGJ. 1973. The evolution of mutation rates. *Genetics* 73(Suppl 73):1–18

Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, Hornreich G, et al. 1997. Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. *Am. J. Hum. Genet.* 60(5):1059–67

Ling S, Hu Z, Yang Z, Yang F, Li Y, et al. 2015. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *PNAS* 112(47):E6496–505

Litton JK, Ready K, Chen H, Gutierrez-Barrera A, Etzel CJ, et al. 2012. Earlier age of onset of *BRCA* mutation-related cancers in subsequent generations. *Cancer* 118(2):321–25

Luqmani YA. 2005. Mechanisms of drug resistance in cancer chemotherapy. *Med. Princ. Pract.* 14(Suppl 1):35–48

Lynch M. 1984. The selective value of alleles underlying polygenic traits. *Genetics* 108(4):1021–33

Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, et al. 2006. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet.* 38:468–73

Malumbres M, Barbacid M. 2003. RAS oncogenes: the first 30 years. *Nat. Rev. Cancer* 3(6):459–65

Mandegar MA, Otto SP. 2007. Mitotic recombination counteracts the benefits of genetic segregation. *Proc. Biol. Sci.* 274(1615):1301–7

Margraf RL, Crockett DK, Krautscheid PMF, Seamons R, Calderon FRO, et al. 2009. Multiple endocrine neoplasia type 2 *RET* protooncogene database: repository of MEN2-associated *RET* sequence variation and reference for genotype/phenotype correlations. *Hum. Mutat.* 30(4):548–56

Martin TA, Ye L, Sanders A, et al. 2000. *Cancer Invasion and Metastasis: Molecular and Cellular Perspective*. Austin, Tex.: Landes Bioscience

Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, et al. 2015. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348(6237):880–86

Maynard Smith J, Haigh J. 1974. The hitchhiking effect of a favourable gene. *Genet. Res.* 23:23–35

Maynard Smith J, Szathmary E. 1995. *The Major Transitions in Evolution*. Cambridge, UK: Cambridge Univ. Press

Mayr E. 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, MA: Belknap Press of Harvard Univ. Press

Medawar P. 1952. *An Unsolved Problem of Biology*. London: H.K. Lewis

Melin B, Dahlin AM, Andersson U, Wang Z, Henriksson R, et al. 2013. Known glioma risk loci are associated with glioma with a family history of brain tumours—a case-control gene association study. *Int. J. Cancer* 132(10):2464–68

Melton C, Reuter JA, Spacek DV, Snyder M. 2015. Recurrent somatic mutation in regulatory regions of human cancer genomes. *Nat. Genet.* 47(6):710–16

Merlo LMF, Pepper JW, Reid BJ, Maley CC. 2006. Cancer as an evolutionary and ecological process. *Nat. Rev. Cancer* 6(12):924–35

Metzger MJ, Reinisch C, Sherry J, Goff SP. 2015. Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell* 161:255–63

Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat. Res.* 106:2–9

Murchison EP. 2009. Clonally transmissible cancers in dogs and Tasmanian devils. *Oncogene* 27(S2):S19–30

Mwenifumbo JC, Marra MA. 2013. Cancer genome-sequencing study design. *Nat. Rev. Genet.* 14(5):321–32

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3(5):418–26

Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* 86 (Pt 6):641–47

Nielsen R, Yang Z. 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* 20(8):1231–39

Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. 2006. Angiogenesis in cancer. *Vasc. Health Risk Manag.* 2(3):213–19

Nowak MA. 2006. *Evolutionary Dynamics*. Cambridge, MA: Harvard Univ. Press

Nowak MA, Michor F, Komarova NL, Iwasa Y. 2004. Evolutionary dynamics of tumor suppressor gene inactivation. *PNAS* 101(29):10635–38

Nowell PC. 1976. The clonal evolution of tumor cell populations. *Science* 194:23–28

O'Neill A, Shah N, Zitomersky N, Ladanyi M, Shukla N, et al. 2013. Insulin-like growth factor 1 receptor as a therapeutic target in Ewing sarcoma: lack of consistent upregulation or recurrent mutation and a review of the clinical trial literature. *Sarcoma* 2013:450478

Ostrow SL, Barshir R, DeGregori J, Yeger-Lotem E, Hershberg R. 2014. Cancer evolution is associated with pervasive positive selection on globally expressed genes. *PLOS Genet.* 10(3):e1004239

Paget S. 1889. The distribution of secondary growths in cancer of the breast. *Lancet* 133(3421):571–73

Pepper JW, Scott Findlay C, Kassen R, Spencer SL, Maley CC. 2009. Cancer research meets evolutionary biology. *Evol. Appl.* 2(1):62–70

Persidis A. 1999. Cancer multidrug resistance. *Nat. Biotechnol.* 17:94–95

Peto R. 1977. Epidemiology, multistage models, and short-term mutagenicity tests. In *Cold Springs Harbor Conf. Cell Prolif.* Cold Spring Harb. Lab.

Petrucelli N, Daly MB, Feldman G. 1998. *BRCA1* and *BRCA2* hereditary breast and ovarian cancer. In *GeneReviews*, ed. RA Pagon, MP Adam, HH Ardinger, SE Wallace, A Amemiya, et al. Seattle: Univ. Washington

Polakis P. 2012. Wnt signaling in cancer. *Cold Spring Harb. Perspect. Biol.* 1:a008052

Price GR. 1970. Selection and covariance. *Nature* 227(5257):520–21

Radtke I, Mullighan CG, Ishii M, Su X, Cheng J, et al. 2009. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. *PNAS* 106(31):12944–49

Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, et al. 2015. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLOS Med.* 12(2):e1001789

Seliger B. 2005. Strategies of tumor immune evasion. *BioDrugs* 19(6):347–54

Seluanov A, Hine C, Azpurua J, Feigenson M, Bozzella M, et al. 2009. Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *PNAS* 106(46):19352–57

Sen S. 2000. Aneuploidy and cancer. *Curr. Opin. Oncol.* 12(1):82–88

Shah SP, Roth A, Goya R, Oloumi A, Ha G, et al. 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486(7403):395–99

Sherr CJ. 2004. Principles of tumor suppression. *Cell* 116(2):235–46

Shpak M, Goldberg MM, Cowperthwaite MC. 2015. Rapid and convergent evolution in the glioblastoma multiforme genome. *Genomics* 105(3):159–67

Smith KR, Hanson HA, Hollingshaus MS. 2013. BRCA1 and BRCA2 mutations and female fertility. *Curr. Opin. Obstet. Gynecol.* 25(3):207–13

Smith KS, Yadav VK, Peders BS, Shaknovich R, Geraci MW, et al. 2015. Signatures of accelerated somatic evolution in gene promoters in multiple cancer types. *Nucleic Acids Res.* 43(10):5307–17

Spielman SJ, Wilke CO. 2015. The relationship between *dN/dS* and scaled selection coefficients. *Mol. Biol. Evol.* 32(4):1097–1108

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–95

Tamuri AU, dos Reis M, Goldstein RA. 2012. Estimating the distribution of selection coefficients from phylogenetic data using mutation-selection models. *Genetics* 190:1101–15

The Cancer Genome Atlas (TCGA). 2012a. Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61–70

The Cancer Genome Atlas (TCGA). 2012b. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489(7417):519–25

The Cancer Genome Atlas (TCGA). 2012c. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407):330–37

Thomas F, Fisher D, Fort P, Marie J-P, Daoust S, et al. 2013. Applying ecological and evolutionary theory to cancer: a long and winding road. *Evol. Appl.* 6(1):1–10

Tomasetti C, Vogelstein B. 2015. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347(6217):78–81

Torres-Jasso JH, Marin ME, Santiago-Luna E, Leoner JC, Torres J, et al. 2015. *EGFR* gene polymorphisms -216G>T and -191C>A are risk markers for gastric cancer in Mexican population. *Genet. Mol. Res.* 14(1):1802–7

Tsao J-L, Yatabe Y, Salovaara R, Jarvinen HJ, Mecklin J-P, Alltonen LA, et al. 2000. Genetic reconstruction of individual colorectal tumor histories. *PNAS* 97(3):1236–41

Turajlic S, Swanton C. 2016. Metastasis as an evolutionary process. *Science* 352:169–75

Vinagre J, Almeida A, Populo H, Batista R, Lyra J, et al. 2013. Frequency of TERT promoter mutations in human cancers. *Nat. Commun.* 4:2185

Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. 2015. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* 525(7568):261–64

Wagner GP, Altenberg L. 1996. Perspective: complex adaptations and the evolution of evolvability. *Evolution* 50(3):967–76

Wang X, Zhang H, Wang D, Li X. 2015. Association of genetic polymorphisms of *EGFR* with glioma in a Chinese population. *Genet. Test. Mol. Biomark.* 19(1):59–62

Weis SM, Cheresh DA. 2011. Tumor angiogenesis: molecular pathways and therapeutic targets. *Nat. Med.* 17(11):1359–70

Weissman A. 1893. *The Germ-Plasm: A Theory of Heredity*. New York: Scribner

Welsh JS. 2011. Contagious cancer. *Oncologist* 16(1):1–4

Williams G. 1957. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 11:398–411

Wrensch M, Jenkins RB, Chang JS, Yeh R-F, Xiao Y, et al. 2009. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat. Genet.* 41(8):905–8

Yachida S, Jones S, Bozic I, Antal T, Leary R, et al. 2010. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467(7319):1114–17

Zhang Y, Italia MJ, Auger KR, Halsey WS, Van Horn SF, et al. 2010. Molecular evolutionary analysis of cancer cell lines. *Mol. Cancer Ther.* 9 (2):279–91

Zhao ZM, Zhao B, Bai Y, Iamarino A, Gaffney SG, et al. 2016. Early and multiple origins of a metastatic lineage with primary tumors. *PNAS* 113(8):2140–45