Mutational Heterogeneity in Human Cancers: Origin and Consequences

Jesse J. Salk, Edward J. Fox, and Lawrence A. Loeb

Joseph Gottstein Memorial Cancer Research Laboratory, Department of Pathology, University of Washington, Seattle, Washington 98195; email: laloeb@u.washington.edu

Annu. Rev. Pathol. Mech. Dis. 2010. 5:51–75

First published online as a Review in Advance on September 10, 2009

The *Annual Review of Pathology: Mechanisms of Disease* is online at pathmechdis.annualreviews.org

This article's doi:
10.1146/annurev-pathol-121808-102113

1553-4006/10/0228-0051$20.00

Key Words

evolution, cancer genome, DNA sequencing

Abstract

Cancer recapitulates Darwinian evolution. Mutations acquired during life that provide cells with a growth or survival advantage will preferentially multiply to form a tumor. As a result of The Cancer Genome Atlas Project, we have gathered detailed information on the nucleotide sequence changes in a number of human cancers. The sources of mutations in cancer are diverse, and the complexity of those found to be clonally present in tumors has increasingly made it difficult to identify key rate-limiting genes for tumor growth that could serve as potential targets for directed therapies. The impact of DNA sequencing on future cancer research and personalized therapy is likely to be profound and merits critical evaluation.

INTRODUCTION

The early historical reliance on surgical excision for treatment of cancers suggests that tumorigenesis was believed to be a fundamentally irreversible process. The multiplicity of chromosomal aberrations associated with abnormal cellular morphology in many human tumors was noted by pathologists more than a century ago and has continued to serve as a means of identifying malignant cells and stratifying the aggressiveness of certain cancers (**Figure 1**). As early as 1902, Boveri (1) suggested that the intercellular cooperation required between cell types during embryogenesis was disrupted in tumors as a result of chromosomal aberrations. Currently, we no longer consider chromosomes, or even

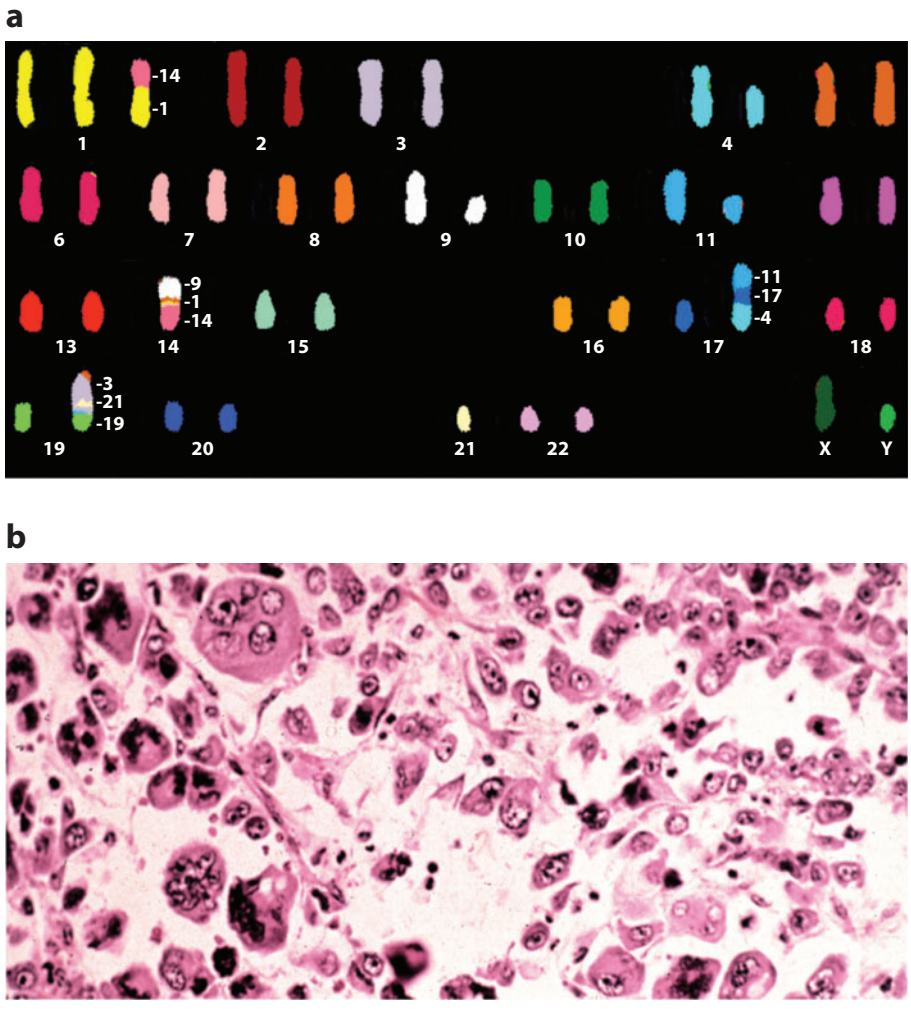

Figure 1

Heterogeneity in cancer. (*a*) Chromosomal heterogeneity. Spectral karyotype from an acute myelogenous leukemia cell demonstrating aneuploidy and multiple chromosomal rearrangements. Reproduced courtesy of Dr. Karen Swisshelm, Department of Pathology, University of Colorado, Denver. (*b*) Morphologic heterogeneity. Hematoxylin and eosin section from a large-cell, undifferentiated lung cancer demonstrating a highly pleiomorphic cellular population. Reproduced courtesy of Dr. Ray Monnat, Department of Pathology, University of Washington, Seattle.

individual genes, as the irreducible genetic units of cancer. Instead, we can identify single-base changes buried among billions of faithfully replicated nucleotides in the human genome. Modern biochemical approaches allow us not only to identify mutated cancer genes, but also to infer how specific mutations affect gene function. As a result, we can catalog unique nucleotide changes that contribute to malignant phenotypes or that increase the susceptibility of individuals to develop specific tumors.

The primacy of DNA as the critical macromolecule involved in the etiology of cancer is strongly supported by the inherited human diseases that are associated with, and increase the incidence of, specific cancers (2). In addition, multiple genetic changes in cancer cells have frequently been documented by microscopic examination of chromosomes or by hybridization with specific probes. These changes include deletions, insertions, amplifications, and translocations, and they frequently involve millions of nucleotides. Chromosomal studies on adenocarcinoma of the breast (3) and ovary (4), as well as in leiomyosarcoma (5), have documented tumors harboring more than 20 different chromosomal alterations. Measurements of loss of heterozygosity in tumors using polymerase chain reaction (PCR)-amplified gene fragments have revealed an even greater number of changes (6). Klein et al. (7) used both of these techniques and demonstrated the presence of multiple alterations within single tumor cells (7). These chromosomal changes have often been considered to result from chromosomal instability and in some tumors may occur sequentially (8).

The technology for dissecting chromosomes into their finest nucleotide elements has exponentially improved in recent years, both in terms of throughput and cost-effectiveness. It is now possible to analyze the processes that generate mutations in normal and malignant cells and begin the ambitious task of cataloging cancer-associated nucleotide changes by DNA sequencing. For some, the surprise has been the unexpectedly large number and diversity of mutations present in human tumors. In light of this emerging mutational complexity, it seems timely to review mechanisms that guarantee the high fidelity of DNA replication in normal human cells, to consider how mechanisms for preventing mutations may be altered in tumors, and to interpret the recently reported results on mutations identified in different human cancers.

THE ACCURACY OF DNA REPLICATION IN NORMAL HUMAN CELLS

DNA is a living molecule; it continually breathes and is exposed to modifications. Yet, in normal cells it is faithfully copied during each division cycle. Each human cell contains more than 6 billion nucleotides that are replicated with exceptionally high accuracy. Approximately one point mutation is introduced into DNA during each division cycle. Most mutation rate measurements have been carried out at the *hgprt* locus because it is present as a single copy on the X chromosome. Spontaneous mutations in this gene render a cell dominantly resistant to the toxic effects of the nucleoside analog 6-thioguanine and form countable colonies under appropriate culture conditions. A tabulation of data derived from *hgprt* studies indicates that the overall mutation frequency in mammalian cells varies from 10^{-5} to 10^{-7} mutations per gene, or approximately 10^{-9} to 10^{-10} substitutions per DNA nucleotide (9–12). It should be noted that cancers likely arise in stem cells, and detailed studies of mouse embryonic stem cells indicate that mutation rates in these pluripotent progenitors are as much as 100-fold lower than those observed in cultured fibroblasts derived from adult tissues (13). Based on the conservative assumption that the accuracy of DNA replication in stem cells is similar to that in other cells, it can be estimated that each stem cell would amass, on average, only one to two mutant genes during 100 cell divisions in a normal human life span (14).

This remarkably high accuracy results from sequential processes, each contributing a 100- to 1000-fold increase in the fidelity

of DNA replication. First, based on simple thermodynamics, the difference in free energy of hydrogen bonding between complementary and noncomplementary base pairings during DNA synthesis can provide an accuracy of base selection down to approximately one error per 10^2 nucleotides incorporated (15–16). Second, DNA polymerases are believed to undergo allosteric transformations with each nucleotide-addition step that tightens the bonding of complementary nucleotides at the active site on the polymerase (17). Third, replicative DNA polymerases have an associated "proofreading" $3' \rightarrow 5'$ exonuclease activity that preferentially excises noncomplementary nucleotides prior to incorporation of the next nucleotide (18). Fourth, remaining noncomplementary nucleotides are removed by mismatch repair after the replication fork has passed (19). Together, these processes have the potential to synthesize DNA in vitro with an accuracy that approximates the fidelity of DNA replication observed in vivo. Notably, however, experimental values are based on reactions carried out under simplified, optimal conditions that may not exist in cells; other components of the replicative machinery are likely to play a role in replication fidelity. Also notable is that the error rates of DNA polymerases are proportional to the concentrations of noncomplementary to complementary nucleotides in the reaction (20). This finding suggests that size of cellular nucleotide pools has a significant impact on the accuracy of DNA replication. Alterations in the accuracy of DNA polymerases by mutation, damage, or imbalances in nucleotide pools could therefore have profound effects on the overall fidelity of DNA replication (17).

DNA DAMAGE BY ENDOGENOUS AND ENVIRONMENTAL AGENTS

DNA is subject to attack by both endogenous and exogenous reactive molecules. A major source of DNA modification in human cells is reactive oxygen species (ROS) arising as a by-product of energy metabolism in mitochondria. (22). On the basis of measurements of 8-oxo-dG and other oxidative modifications, it has been estimated that more than 10,000 nucleotide residues in DNA are altered by ROS per cell per day (23). Other modifications include methylation, alkylation, inter-/intrastrand cross-links, and apurinic sites; there may be as many as 50,000 alterations per cell per day that result solely from normal cellular metabolism (24). Many human cancers arise in the setting of chronic inflammation (25–28), where extracellularly derived ROS are likely to contribute to the burden of DNA damage in affected tissues (29). Mutagens are ubiquitous in our environment, and it is important to recognize their contribution to spontaneous human cancer. Exposure to high concentrations of mutagens has frequently been associated with an increased incidence of cancer (30–32), and the recognition of tobacco smoke as a human carcinogen (33–34) may have led to the most significant and successful effort at reducing cancer incidence in human history (35).

THE REPAIR OF DNA DAMAGE IN HUMAN CELLS

Working against this onslaught of DNA damage is an armamentarium of repair systems with overlapping specificities (**Figure 2**). These systems continuously monitor the genome and correct many forms of DNA damage. So far, more than 100 repair genes have been identified (36). Pathways of DNA repair include base-excision repair, nucleotide-excision repair, transcription-coupled repair, mismatch repair, double-strand-break repair, and even direct reversal of adduct-mediated lesions (37). DNA damage by environmental agents is predominantly a stochastic process. Recognition of damage is generally dependent on the nature of the lesion and is less a function of sequence context. Small adducts on bases are excised by both short-patch and long-patch pathways, and resynthesis of the excised segment is carried out by DNA polymerase β and presumably DNA polymerases δ and ε, respectively (17). Bulky adducts such as thymine dimers, resulting from ultraviolet irradiation, or benzo[a]pyrene, resulting from tobacco products, are subject to

nucleotide-excision repair. In the presence of DNA damage, many sensing systems signal to cell-cycle control genes, such as *TP53*, to arrest the cycle and allow additional time for DNA repair (38). The high efficiency of these mechanisms for DNA repair guarantees that only a few of the tens of thousands of DNA lesions generated persist prior to DNA replication and have the potential to cause mutations.

INHERITED CANCERS FREQUENTLY INVOLVE MUTATIONS IN DNA-REPAIR GENES

A number of rare inherited diseases caused by germline mutations in genes involved in DNA repair are associated with elevated risks of specific cancers. Investigation of many of these diseases has been instrumental in deciphering the different cellular mechanisms for DNA repair. The seminal findings on the defects of ultraviolet-induced DNA-damage repair in patients with xeroderma pigmentosum highlighted the association of DNA repair with suppression of carcinogenesis and provided a powerful tool for identifying genes involved in nucleotide-excision repair (2). Subsequently, inherited defects in members of several other DNA-repair pathways have been shown to underlie a variety of cancer syndromes, including mismatch repair [hereditary nonpolyposis colorectal cancer (39)], base-excision repair [*MYH*-associated polyposis (40)], homologous recombination [early-onset breast cancer (41)], nonhomologous end-joining [LIG4 syndrome (42)] and translesion synthesis [xeroderma pigmentosum variant (43)].

Hereditary mutations in other DNA maintenance enzymes are also associated with cancer. Defects in genes encoding members of the RecQ helicases are found in Bloom and Werner syndromes—rare inherited diseases characterized by developmental abnormalities and a high incidence of specific cancers (44–45). Mutations in *TP53* are found in Li–Fraumeni syndrome (46), a highly cancer-prone condition most frequently associated with sarcomas and breast adenocarcinomas. Additionally, polymorphisms in a wide variety of DNA-repair genes, including *OGG1* and *XRCC1*, are increasingly being considered as risk factors for cancer (47). Why defects in particular DNA-repair genes result in specific types of cancer remains largely unresolved.

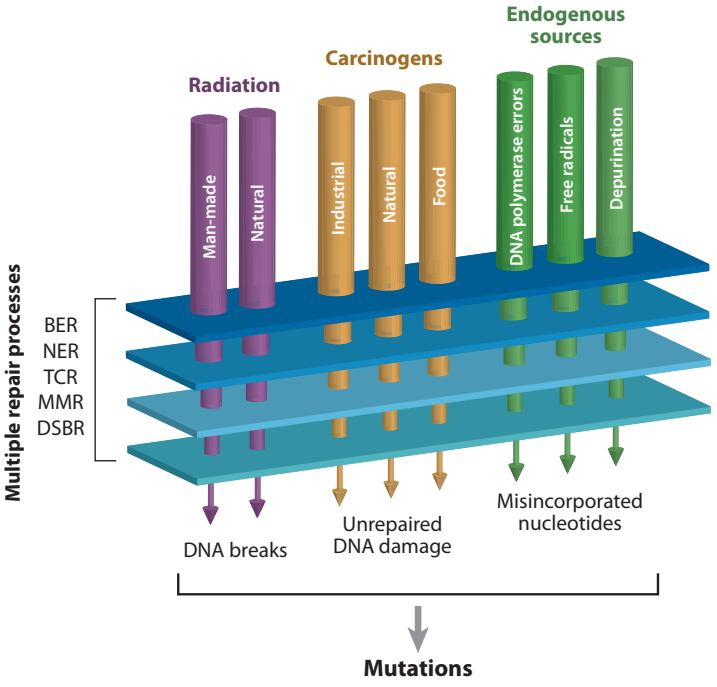

Figure 2

Mutational homeostasis. In each human cell, DNA is damaged thousands of times per day by both exogenous and endogenous sources. Most alterations are corrected by cellular mechanisms, including base-excision repair (BER), nucleotide-excision repair (NER), transcription-coupled repair (TCR), mismatch repair (MMR), and double-strand-break repair (DSBR). Lesions that escape repair have the potential to cause mutations during DNA replication.

PREVENTION OF MUTATIONS BY GROWTH LIMITATION, CLONAL HIERARCHY, AND PROGRAMMED CELL DEATH

Despite these extensive DNA-repair mechanisms, every time a cell divides there remains an opportunity for fixation of new mutations through miscopying across unrepaired damage, missegregation of replicated chromosomes, and/or failure to recognize improperly

repaired sequences. Thus, a key mechanism for preventing the accumulation of DNA mutations is limiting the number of cell divisions that occur. The importance of proliferation in oncogenesis has been demonstrated by experiments showing that liver regeneration is associated with an increased incidence of cancer (48). The initiation of skin cancer in mice by mutagens was markedly accelerated by the subsequent application of phorbol esters that promote cell proliferation (49). Many of the molecular mechanisms that control cellular growth control were first identified in the context of their disruption in cancer. The identification of replication-enhancing avian viral oncogenes (50), and the discovery that these represent constitutively active versions of endogenous genes irreversibly activated through mutation in some cancers (51), led to the characterization of one of the earliest known growth-control pathways in human cells. Extensive work during the past three decades has revealed the cellular network of defenses preventing unregulated proliferation to be staggeringly complex, with many redundant protections (52).

The necessity of replacing worn out or damaged tissues must be carefully balanced against the risk of proliferation-induced mutations. To allow cellular repopulation with minimal risk of mutation, tissues in the body are frequently organized hierarchically, whereby the ability to continuously proliferate is relegated to a specialized subset of cells (53, 53a). Stem cells are believed to have an inherently lower rate of mutation than the majority of their daughter cells, which have only limited replication potential (13). Among the most studied examples of this hierarchical organization is colonic epithelium. Here, a small number of long-lived stem cells reside at the base of each crypt and produce progeny that migrate luminally to populate the upper levels of the crypt—first as transiently amplifying cells, then as terminally differentiated colonocytes destined to slough off after several days (54). Because the majority of mutations that arise during division occur in short-lived daughter cells, most mutant cells are rapidly purged from the population. It has been hypothesized that the same so-called immortal strand of DNA is maintained in the parental stem cell and that the (potentially imperfect) newly replicated strand is always transferred to the daughter (55), although at least one recent study (21) suggests that this is not the case. Whether tumors derive from abnormally replicating stem cells, or from dedifferentiated progeny, remains an open question (56).

CANCER AS A SOMATIC EVOLUTIONARY PROCESS

November, 2009 marks the 150th anniversary of the publication of Darwin's seminal work, *On the Origin of Species*. Therein, Darwin postulated that heritable phenotypic variation underlies natural selection and is responsible for adaptation as well as for the emergence of new species. Although it was initially used to describe how organisms evolve across generational time, the idea that evolutionary forces also drive intraorganismal neoplastic development has frequently been noted (57–59). In this model, individual cancer cells, rather than complete metazoans, are considered the reproductive units within a population (**Figure 3**). New mutations are acquired somatically, and genetic alterations that bestow a growth advantage upon the cancer cell enable them to clonally expand. Additional mutations that arise in the expanding population generate further selectable phenotypes, such as the ability to invade adjacent tissues, recruit a blood and lymphatic supply, overcome nutritional deficiencies, and resist immune attack. After bypassing all antineoplastic defense mechanisms, tumor growth may continue indefinitely until the death of the host.

THE NUMBER OF MUTATIONS TO CANCER

Given that carcinogenesis may be viewed as an evolutionary process that sequentially increases neoplastic cell fitness through a series of (epi)genome-modifying events, an important question arises: How many mutations are needed to produce a tumor? The increased

incidence of most human cancers as the fifth or sixth power of age (60) has been taken to indicate that there are five or six events (presumably mutations) that drive the carcinogenic process, each event increasing the probability of the next. Exceptions include (*a*) certain pediatric tumors, such as retinoblastoma, in which significantly fewer mutations appear to be necessary (61), and (*b*) some late-onset adult tumors, such as those of the prostate, that may require as many as 10 to 12 events (62). Weinberg and colleagues (63) have demonstrated that at least three or four altered genes are required for the expression of the malignant phenotype in cultured cells. Passage in tissue culture or as implanted xenografts may, in itself, select for additional mutations as highlighted by the recent work of Mahale et al. (64). Thus, if cancer requires as many as 12 different rate-limiting mutations to arise, if the normal per-division mutation rate of human stem cells is as low as calculated, and if the number of long-lived stem cell divisions is limited, how can a cancer possibly occur within the human lifetime? It is hypothesized that early in the neoplastic process at least one, and likely several, of the mechanisms for preventing mutations must be reduced.

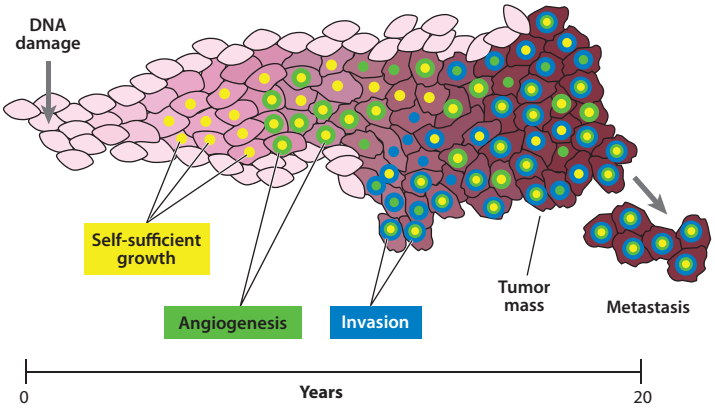

Figure 3

Cancer recapitulates evolution. Within a developing tumor, mutations accumulate over time as a result of unrepaired DNA damage. Most of these mutations are either neutral or detrimental; only a small number bestow growth and survival benefits upon a cell. Cells with these beneficial variants preferentially multiply and additional mutations occur that may undergo further selection and expansion. Advantageous phenotypes for tumor growth include, among others, the ability to divide independently of extracellular signals (*yellow*), the ability to recruit a blood supply (*green*), and the ability to invade adjacent and distant tissues (*blue*).

ANTIMUTATIONAL DEFENSES: PRIMARY VERSUS SECONDARY MECHANISMS

From a simplified perspective, one may divide antimutational processes that suppress the emergence of new genetic diversity into two classes. The first class consists of primary mechanisms, which act at the level of DNA to prevent genetic and epigenetic mutations from occurring or persisting until cell division, such as proofreading by DNA polymerases, DNA-repair processes, ability to quench ROS, and other means of limiting the per–cell division mutation rate. The second class includes secondary mechanisms, which in contrast prevent the accumulation of mutations in the population at large by limiting the total number of divisions in long-lived cells. These mechanisms include means of controlling cell growth, confinement of reproductive abilities to stem cells in a well-protected niche, and culling of irreversibly damaged cells through programmed cell death. Secondary mechanisms do not affect the per-division rate at which individual cells accumulate mutations; instead, they limit the overall number of different mutations in the population as a whole.

For evolution to occur in a somatic setting, (*a*) there must be heritable genetic or epigenetic variation within a population of cells, and (*b*) these variants must be able to undergo selection (i.e., through more efficient division and survival) in response to advantageous environments. If either of these features is rate-limiting to the process, and if the factor responsible for the limitation is genetically encoded within the evolving cells, disruption of the responsible genes or regulatory sites through mutation accelerates the adaptive process. The question of which parameter is most limiting in different stages of tumorigenesis is complex and, rightly, has been the subject of extensive debate (65–66). Given that both therapeutic interventions and preventative measures might be better

Mutator phenotype: an increased (per–cell division) mutation rate resulting from heritable cellular defects, generally in DNA maintenance machinery

Clonal expansion: multiplication of a single cell to produce a population of genetically related progeny

directed if this was known, the question is not merely of academic interest.

Within the view that somatic evolution drives carcinogenesis, at least two schools of thought have emerged regarding the relative importance of overcoming primary and secondary antimutational defenses. One hypothesis asserts that the mutation frequency existing within nonmalignant cells provides sufficient genetic diversity to fuel tumorigenesis (65, 67–68). Some of these variants in a normal population can confer a phenotype of increased proliferative abilities that enables subsequent clonal outgrowth and consequent generation of more mutations, upon which further selection may act. Not only does escape from growth limitations facilitate a cell acting upon preexisting genetic diversity through selection, it also enhances the production of additional variants by increasing the number of fallible DNA-replication cycles. This school of thought conceptually favors the importance of defects in secondary mechanisms but also relies upon the fact that primary mechanisms are imperfect in nonneoplastic cells. Eventually, after enough successive rounds of mutation, selection, and expansion, the threshold number of events required to drive carcinogenesis is reached.

A second school of thought emphasizes the importance of defects in the primary class of antimutational mechanisms for fueling neoplastic evolution (69–70). Deficits in these cellular components render DNA replication more error prone and increase the number of potentially advantageous variants produced per generation. It is argued that, even with additional mutations generated from increased cell division, the number of variants in the population will still be a limiting parameter for evolutionary adaptation. Proponents of this view suggest that mutation-prone variants would overcome this limitation and therefore be likely to emerge during clonal selection. A genetically encoded mutator phenotype need not, in itself, be a driver that directly increases reproductive fitness. Instead, by virtue of increasing the probability of advantageous new mutations, this phenotype would be passively carried along on the resulting clonal expansion as a passenger. Because of this unique position, generation of a mutator phenotype is unlikely to be the very first "hit" on the path to cancer. As with inherited deficiencies of DNA repair, it is only when the phenotype is expressed in a plurality of cells that it can meaningfully increase the total number of genetic variants in the population. Thus, defects in secondary antimutational defenses leading to increased proliferation remain a critical component of this model of tumor development.

Investigators on both sides of the debate have made arguments for (69) and against (71–72) the necessity of a primary mutator phenotype in carcinogenesis on the basis of calculations that assume the number of mutations needed for cancer, the mutation rate in normal cells, and the estimated number of divisions that occur between conception and a late-stage tumor. Given that these values are, themselves, not easily quantified, the disparate results are not surprising. An alternate approach taken by our group is to consider, instead, the relative efficiency of mutator and nonmutator pathways to cancer (73). Modeling of fitness landscapes suggests that in spite of the cost of an extra step to produce the phenotype, a primary mutator pathway is generally a more efficient trajectory to cancer, as long as the total number of mutations required exceeds three to five (73a). Moreover, although mutator lineages are more likely to suffer deleterious mutations that terminate their lineage through negative clonal selection (74), this negative effect does not predominate until mutation rates become very high (73a). An additional consideration that has been frequently overlooked is that newly arising mutations, including those with a fitness advantage, have a high probability of becoming extinct from random drift (58). Depending on the population size and precise fitness advantage, a given mutation may have to arise on multiple independent occasions before it can expand to a clinically meaningful size. Hence, calculating the mutation rate required for a defined number of genetic events to occur once per tumor leads to underestimation of the

rate required for each mutation to arise and expand.

Making a sharp distinction between primary and secondary mechanisms of mutation suppression is conceptually interesting but ultimately artificial, given the mechanisms' intimate link within the cell. It is likely that both mechanisms are operative; the relative contribution of each may depend on the tumor. Cell-cycle progression, DNA repair, and programmed cell death are coordinately regulated. *TP53*, for example, the most commonly mutated gene identified to date in human cancer, encodes a multifunctional protein that acts as a network hub to integrate information about the genome state from more than a dozen sources (75). Many types of DNA damage can trigger activation of p53, which may lead to cell-cycle arrest, upregulation of DNA-repair processes or activation of programmed cell death if damage is severe. *TP53* therefore exemplifies a gene that is directly involved in both primary and secondary antimutational pathways. It has been argued that for such dual-function tumor-suppressor genes the pressure for loss in neoplastic cells stems from the immediate proliferative advantage gained, rather than from an increased mutation frequency (65). Although this is likely to be the case given that a primary mutator phenotype is not itself selectably advantageous, disruption of such genes nevertheless increases the genetic diversity of a developing cancer through both primary and secondary mutational pathways and facilitates continued evolution.

DETERMINISTIC VERSUS PLASTIC TUMORIGENESIS

DNA damage by chemical agents and physical processes is predominantly stochastic. For the most part, damage occurs randomly throughout the genome, and mechanisms governing the correction of affected nucleotides are specified primarily by the nature of the chemical alterations and not by the surrounding nucleotide sequence. Exceptions to this random hit model are nucleotide sequences that can form alternative DNA structures (76) that are resistant to DNA repair and highly repeated sequences of the genome (77)—both of which constitute mutational hot spots. The specific role of histones and other DNA-associated proteins in preventing DNA damage has not been fully delineated. Similarly, transcriptional status (78) as well as local replication timing (79) have general, yet incompletely predictive, influences on local mutation rate. Only a small fraction of all mutations that occur confer a selectable fitness advantage with the ability to initiate or promote tumor growth. Given that different individuals acquire a unique, yet random, selection of mutations prior to and during tumorigenesis, the order and specific nature of advantageous variants are determined stochastically. Just how similar are two tumors with respect to specific genes mutated and the order in which the selected mutations arise? Do tumors evolve along a pathway characterized by sequential rounds of mutation and selection of a defined set of targets, or is the process more variable, selecting instead for phenotypes that may be encoded by many different loci (**Figure 4**)?

There is considerable evidence for sequential mutations during tumor progression. In melanomas (80), colon cancer (8), and esophageal (25) cancer, a series of genetic alterations frequently characterize different phases of neoplastic progression. These changes have been most extensively documented in adenocarcinomas of the large intestine, where the order of DNA alterations has been correlated with tumor grade and stage (8). From a clinical perspective, a chronologically ordered series of mutations driving malignancy is particularly attractive, as it implies that the evolutionary process must bottleneck through a defined set of genes that could be therapeutically targeted. Unfortunately, even in the most-studied model, early investigations indicated that fewer than 10% of advanced colon cancers simultaneously bear mutations in the three most frequently mutated genes (81).

It has long been clear from traditional genetic and molecular methods that despite some commonalities, the profile of clonal somatic

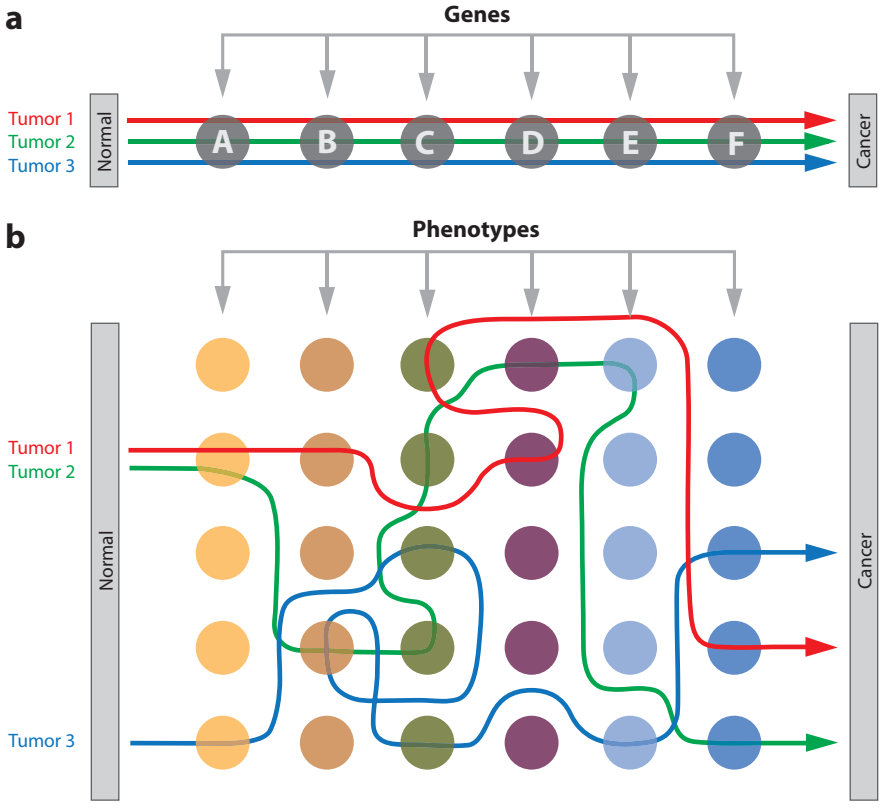

Figure 4

Pathways to cancer. (*a*) The deterministic pathway. In this model, different tumors of the same cancer type occur reproducibly through sequential mutation of each gene within a defined series. Although mutation occurs randomly, the order of selection is fixed. (*b*) The plastic pathway. In this model, different tumors evolve along highly variable pathways, selecting for specific cancer phenotypes that may be achieved through mutation or epimutation of many possible sites in the genome. Although some mutated loci may be shared by different tumors, most are not, and the order of selection is predominantly stochastic.

mutations occurring in different tumors is highly variable overall. The resolution of these early findings was inherently limited to a moderate number of genes by the biochemical tools available. The complete human genome, however, comprises more than 6 Gb of sequence information and, until very recently, methods for high-throughput analysis were inadequate for the task of whole-genome exploration. Many long-standing questions remain to be answered: How frequently do different tumors overcome hardwired barriers to neoplasia in the same way, in terms of altering both specific genes and specific pathways? Does this process vary from one cancer type to another? What, if any, mutational changes differ between metastatic lesions and their primary tumor? What is the relative importance of different types of mutations, such as single-base changes versus deletions, rearrangements, and epigenetic phenomena? Can the mutational spectrum of a tumor inform us of likely environmental contributors or specific dysfunctional antimutational pathways? The emergence of high-throughput capillary sequencing robotics and more recent next-generation sequencing methods provide an exciting opportunity to delve more deeply into these questions.

THE HUMAN CANCER GENOME ATLAS

Within the past two years, numerous DNA sequences from human cancers have been published as part of The Cancer Genome Atlas. Included among these is the first complete genome of a human cancer and its paired normal (82). With the passing of this milestone, it is important to consider the likely implications of this data and how they might frame both basic and clinical research in the near future. Prevailing models of tumorigenesis stress that tumor progression is the result of sequential mutations in a few key cancer genes, each mutation driving a new round of clonal proliferation. The effort to systematically tabulate mutations found in different human cancers encompasses the expectation that a cancer's most significant mutated genes will be potent targets for chemotherapy. This supposition has been reinforced by the success of targeted treatments in some hematological cancers (83) and by the hope that identification of analogous key mutational events in solid tumors might allow specific therapeutic targets to be similarly identified and exploited. However, an increasingly complex picture has emerged from nearly 20 studies detailing the genome of many solid tumors: The findings suggest that the extent of prevalent, new targets may be far more limited than anticipated.

INITIAL STUDIES ON NUCLEOTIDE VARIATION WITHIN HUMAN CANCERS

The first large-scale efforts to systematically screen individual tumors for somatic mutations identified remarkably few previously unknown genes that were mutated in a significant proportion of specific cancers (84–86). The relatively limited sequence coverage of these initial studies prompted more comprehensive screens of larger portions of the genome. The first complete sequencing of all predicted coding exons, conducted in breast and colon cancer, concluded that these cancers respectively contain a median of 84 and 76 clonal mutations that are likely to alter protein function (87–88). Although nearly one-tenth of the 18,197 genes sequenced were detectably mutated in at least one specimen, each tumor displayed a unique and diverse profile of mutated genes. Other than those previously known, such as *TP53*, *APC*, and *KRAS*, no new prevalently mutated genes were identified. The authors of these studies proposed that the cancer genome may be considered as a landscape composed of a handful of commonly mutated gene "mountains" and a larger number of less frequently mutated gene "hills"—a view consistent with multiple pathways to cancer. The authors additionally affirmed that tumor-to-tumor heterogeneity of clonally present mutations may underlie the wide variation in tumor behavior and responsiveness to therapy.

The initial studies of Sjöblom et al. (87) and Wood et al. (88) served to highlight several important technical challenges faced by The Cancer Genome Atlas Project. First, with large amounts of tumor sequence data comes significant experimental noise that complicates detection of true clonal mutations. Such noise derives from PCR-introduced mutations, automated base-calling errors, mutations arising in the germline rather than somatically, and previously unknown single nucleotide polymorphisms (SNPs). The most rigorous cancer-genotyping approach would entail sequencing of a matched nonneoplastic sample for every tumor to rule out germline variation as well as automatic resequencing of every tumor-specific mutation identified. Unfortunately, even the highest-throughput capillary sequencing systems are cost- and time limited when used on the scale of multiple human exomes. Thus, in these initial studies, as well as in several that followed, compromises to preferred protocols have been made; for example, noncoding changes and known SNPs were eliminated prior to confirming the small fraction of mutations remaining against a tumor's corresponding control sample.

A second impediment to high-throughput capillary sequencing is the substantial amount

The Cancer Genome Atlas: the complete catalog of genetic and epigenetic alterations found in cancers of all types

Clonal mutation: mutation present in the majority of cells in a tumor; detectable by conventional sequencing

Driver mutation: mutation that provides a selectable fitness advantage to the cell and facilitates its clonal expansion in the population

Passenger mutation: mutation that has no effect on a cell's fitness and that clonally expands in the population as a result of a different driver mutation

Exome: encompasses all coding sequences within the genome

of DNA required for the hundreds of thousands of PCRs. One way to overcome this limitation has been to expand tumor cells in culture or as mouse xenografts. Such ex vivo passaging poses the possibility of artifactually introducing new mutations as a result of artificial growth conditions (64). Direct biochemical methods of whole-genome amplification have also been used to extend DNA samples. It has been suggested that the differing mutational spectrums reported within the discovery and validation phases of experiments of Sjöblom et al. (87) are likely reflective of such differences in ex vivo treatments of samples (A.F. Rubin & P. Green, manuscript submitted).

A third, and arguably most significant, challenge to cancer genome sequencing lies in the complex problem of determining which mutations in a tumor are causative and which are merely present by chance. Mutagenesis is largely a random process; only a small subset of mutations confer a proliferative advantage to their host cell. Differentiating so-called driver mutations from neutral passenger mutations (or hitchhiker mutations) that happen to co-occur and be swept along with the same expanding genome poses a formidable challenge to deciphering megabases of sequencing data. The practical approach taken by these and later studies has been to sequence multiple tumors of the same variety and look for genes that are commonly mutated. The appropriate statistical methods to be used for determining which genes are found clonally mutated more frequently than would be expected by chance alone has been heavily debated (90–93). More importantly, implicit in this approach is the de facto assumption that a limited subset of genes frequently drive tumorigenesis. The alternate hypothesis—that a large number of loci may combinatorially serve as weak drivers and that any one may arise only infrequently—cannot easily be addressed by such methods, given that rare drivers are filtered out as probable passengers. The number of tumors needing to be sequenced to resolve minority drivers increases substantially as the prevalence of the driver declines. Researchers have also attempted to identify likely driver mutations by bioinformatically predicting the probable functional impact of specific mutations (98). Although such approaches are useful in a limited number of instances, the current technology for accurately modeling the resulting changes to protein activity remains limited. Despite these complexities, the demonstrated ability of exome sequencing to reidentify the majority of previously characterized genes known to play a role in colorectal cancer was a noteworthy technical achievement that set the stage for the cancer genome–sequencing studies that followed.

MULTIPLE MUTATIONS IN DIVERSE TUMORS

Whereas the initial studies of Sjöblom et al. (87) and Woods et al. (88) focused on the exhaustive comparison of mutations in two tumor types, a second milestone project analyzed an extensive gene family in a wide variety of different cancers. Greenman et al. (94) sequenced 518 protein kinase genes in 210 tumors of diverse origin, including breast, colorectal, lung, brain, and blood tumors. They observed 1007 likely driver mutations, of which 921 were single-base substitutions. As in previous studies (95–97), there was substantial variation in the number of genes mutated per tumor regardless of type; again, few commonly mutated genes were found in any of the cancer types examined. Although these studies generated an extensive catalog of somatic point mutations, only a small number of prevalently mutated genes were identified. The data reinforced the notion that mutational patterns of solid tumors evolve stochastically and are highly diverse, in contrast to the relatively predictable stepwise patterns of cytogenetic abnormalities in some hematological cancers.

Subsequent studies (**Table 1**) have increasingly relied on associating sequence data with other complementary genomic information. This trend has been paralleled by a shift to a more integrated interpretation of the significance of individual mutations: from one of specific genes into one of pathways and processes.

Table 1 Cancer genome–sequencing studies

Study type	Number of genes screened	Total number of mutations	Number of genes mutated	Average number of mutations per tumor	Estimated number of driver mutations	Reference(s)
Exomic						
Breast ($n = 11$)	18,191	1243	1137	84	140	87, 88
Colorectal ($n = 11$)	18,191	942	848	76	140	87, 88
Diverse ($n = 210$)	518	798	581	–	119	94
Pancreatic ($n = 24$)	20,661	1163	1007	48	160	98
Glioblastoma ($n = 21$)	20,661	748	685	47	155	102
Glioblastoma ($n = 91$)	601	453	223	–	8	103
Lung ($n = 188$)	623	1013	348	–	26	108
Genomic						
Acute myeloid leukemia ($n = 1$)	–	500–1000	10	Not applicable	10	82

One follow-up study focused on mutations arising during the progression of adenoma to carcinoma to metastasis (97a). No metastasis-specific mutations were detected in the vast majority of specimens, and as expected, the number of mutations was markedly increased in the carcinoma compared with its matched precursor adenoma. Building on the observation that individual tumors express unique immune profiles (99), Segal et al. (100) demonstrated that the diverse mutational pattern of breast and colorectal tumors likely underlies their immunological heterogeneity. Leary et al. (101) examined homozygous deletions and focal amplifications in the breast and colorectal cancer genomes. Each of these studies further confirmed the heterogeneity and intertumor diversity among breast and colorectal cancer genomes.

PARADIGM SHIFT

This change of focus, away from the search for key, sequentially mutated genes that govern cancer progression toward a more systems-oriented description, is evident in recent studies on pancreatic cancer. Using a two-part discovery and prevalence-determination strategy, along with copy-number and transcriptomic analyses, Jones et al. (98) concluded that pancreatic cancers contain an average of 63 clonal genetic alterations, of which the majority are point mutations. As with breast and colorectal cancer, there was considerable variation both in the number of mutations and in the specific genes mutated among different cancer specimens; again, no new prevalently mutated genes were identified. Because nearly all of the predicted protein-coding genes in the human genome were evaluated, these data provided the opportunity to investigate groups of genes operating through specific signaling pathways and processes in a relatively unbiased manner. The authors concluded that pancreatic cancer results from genetic dysregulation of 12 core pathways and processes, including apoptosis, DNA-damage control, and regulation of invasion. Although these 12 processes are genetically altered in the majority of pancreatic cancers, the specific components mutated in individual tumors were largely different. It was proposed that agents be designed to target the physiological effects of the altered pathways and processes rather than individual genes. Although this therapeutic logic is reasonable, and the analyses do demonstrate enrichment for specific cellular processes, the granularity of the results does not extend much beyond reinforcing the general hallmarks of cancer (52).

Glioblastoma multiforme (GBM) was the first cancer type to be screened systematically by two independent groups (102–103). Both

Indel: a mutation entailing the insertion or deletion of one or more bases

studies integrated gene sequencing, identification of focal amplifications and deletions through comparative hybridization arrays, and expression analysis to comprehensively interrogate the GBM genome. One study, focusing on a group of 601 selected genes in 92 predominantly primary GBMs, found no novel commonly mutated genes among different tumors (103). Interestingly, the number of gene alterations in GBMs was smaller than that previously reported for colorectal and breast cancers. The second study focused on exhaustively sequencing all likely coding exons in a discovery screen, then determined the prevalence of any identified variants in a secondary screen (102). The discovery of one novel recurrent mutation [isocitrate dehydrogenase 1 (*IDH1*)], mutated in 12% of all GBMs and strongly associated with secondary GBMs in particular, was cited as a validation of the utility of genome-wide genetic analysis of tumors. Indeed, as of this writing, this finding is among the most significant to be unearthed by cancer genome sequencing. Two follow-up reports have indicated that active site mutations in *IDH1*, and occasionally its homolog *IDH2*, are found in more than 70% of certain central nervous system (CNS) tumors including grade II and III astrocytomas and oligodendrogliomas and secondary glioblastomas, although they are rarely found in primary glioblastomas and have not been found in any of the tested tumors from outside the CNS (104–105). Reinforced by functional studies of these mutations in cultured cells indicating lowered enzymatic activity (105–106), this work has unequivocally identified an important new pathway for a specific subset of CNS tumors. Nevertheless, both initial sequencing studies independently concluded that for the bulk of GBMs dysregulation of three core pathways, based around the already well-studied genes *TP53*, *CDKN2A*, and *EGFR*, is central to tumor progression.

Analysis of the lung cancer genome identified 1013 mutations in 188 cases of lung cancer (107–108). Twenty-five cases harbored no mutations in the 623 genes analyzed, and only four genes had point mutations in more than 10% of tumors. By examining the distribution of genes across cellular pathways, the authors identified five key pathways in which components were frequently mutated. By far the most commonly affected of these was the MAPK pathway, in which 70% of tumors sequenced had at least one mutation altering known MAPK proteins. This pathway, however, encompasses 56 genes, most of which, individually, are mutated in fewer than 1% of lung cancers. Again, the most significant findings of this study are the mutational heterogeneity among tumors and the absence of prevalently mutated genes.

THE FIRST COMPLETE CANCER GENOME

The characterization of the first hematopoietic cancer genome represented an important methodological milestone in cancer genomics—truly whole-genome sequencing of a tumor specimen (82). Prior efforts at resequencing tyrosine kinase genes in acute myeloid leukemia (AML) had yielded few mutations (109–110). By exhaustively screening the entire genome of a paired set of cancer and normal samples from a single AML patient through use of massively parallel sequencing, the authors identified 500–1000 nonsynonymous somatic changes uniquely present in the cancer. Of these, only ten mutations occurred in protein-coding genes, including two previously described indels (within *FLT3* and *NPM1*) known to occur at high frequency in AML. Importantly, none of the eight newly identified genes was found to be mutated in 187 additional cases of AML.

This study marked a number of significant advances in large-scale cancer genome sequencing. First, unbiased whole-genome sequencing is inherently a more complete means of cataloging all the clonal alterations in a cancer genome. With the increasing recognition that so-called intergenic "junk" DNA and intronic sequences contain functional elements such as regulatory regions and noncoding RNAs, exon-centered genotyping may be missing important drivers occurring in regions of the genome

not previously explored. Even if most of these noncoding mutations turn out to be noncausative, the spectrum and total number of clonal mutations borne by a tumor may provide important information about sources of mutations as well as about a tumor's life history.

Second, unlike previous studies in which matched normal DNA was only sequenced to specifically validate candidate mutations, this study was the first to simultaneously apply an unbiased analysis to paired normal DNA, albeit with less sequence depth. One criticism of the earlier strategies was the exclusion of any sequence variant coincident with a previously described SNP, without considering it a possible de novo event. New mutations in known SNP sites, in fact, may represent some of the most likely selectable drivers given the strong familial component to many cancers for which no specific genes have yet been implicated. Only by sequencing of paired normal DNA can clonal mutations at polymorphic loci be scored.

Third, next-generation sequencing platforms have several advantages that will likely make them the preferred technologies for future studies. The most obvious is the significantly lower cost per base pair sequenced. In addition, the minimal-input DNA requirement obviates the need for expanding tumor cells in culture or as xenografts. Three such devices are presently in commercial production, and many more are on the horizon. Genotyping is accomplished by randomly shearing the genome into many pieces, clonally amplifying individual fragments on a solid matrix, and then sequencing these immobilized clones using one of a variety of chemistries. Although enriching for certain portions of the genome is possible, typically the fragments to be sequenced are generated randomly rather than by user-selected sequencing primers. An important benefit to these random fragments is the greater ease with which breakpoints resulting from large indels or other rearrangements can be identified. Conventional targeted resequencing by standard capillary methods alone is likely to miss many such events because PCR amplification of a given region cannot occur if one or more primer sites is lost or distantly relocated.

Fourth, and finally, the markedly lower cost of next-generation methods also means that automatic confirmation of mutant bases can be reasonably built into the sequencing protocol. With such platforms, this process is simply a matter of sequencing enough random fragments to have a high probability of genotyping every base several times. The lower throughput of capillary methods has often necessitated the initial filtering of mutant calls such that only those deemed to have a high likelihood of being drivers are retested. For example, the first major study by Sjöblom et al. (87) triaged 260,000 noncoding changes without further confirmation. Although this may have been unavoidable from the logistical standpoint of traditional sequencing, synonymous changes within tumors may be of importance, given that they can influence both transcription and translation (89).

Next-generation sequencing: refers to new methods of high-throughput DNA sequencing carried out on amplified single DNA molecules affixed to a solid matrix

IMPLICATIONS OF THE CANCER GENOME ATLAS: INTERTUMOR HETEROGENEITY

The primary goal of cancer genome–sequencing studies has been the identification of the genes and pathways that play a causal role in the neoplastic process (98). It was the expectation that sufficiently detailed genetic analysis would lead to the identification of a small set of commonly mutated genes that drive tumor progression and thus present new therapeutic targets. Collectively, the initial studies described above constitute the most systematic characterization of a disease genome ever undertaken and demonstrate the feasibility of producing a compendium of clonally altered somatic sequences. Additional analyses that use complementary approaches, including those assessing rearrangements (111), deletions (112), and epimutations (113), as well as the impact of mutations in noncoding sequences, are within the scope of our current technology and will soon provide an even more complete

description of changes within the cancer genome.

Subclonal mutation: mutation present in a single cell or a minority of cells in a tumor; not detectible by conventional sequencing

The overarching conclusion to be drawn from completed cancer genome–sequencing studies is that most cancer types display substantial intertumor mutational heterogeneity. Individual solid organ tumors harbor, on average, more than 50 nonsilent clonal mutations in the coding regions of different genes, yet only a small fraction of these genes are mutated in a high proportion of tumors. Although certain clonally disrupted genes are more prevalently represented within specific types of cancer, there remains a great deal of overlap. The large number and breadth of diversity in genes mutated among individual tumor specimens emphasize the fundamentally stochastic nature of cancer evolution.

The therapeutic implications of these findings are considerable. Preliminary studies focusing on kinases, the most druggable portion of the genome, explored the possibility of identifying commonly mutated genes that might be exploited with targeted approaches (86). However, after the number of samples profiled increased, it has become clear that any one of these genes is only mutated in a small fraction of tumors. To synthesize and test enough small-molecule inhibitors to combat even half of the kinase class of suspected tumor drivers would be a daunting undertaking on a scale that is arguably beyond our current drug-developing and -regulatory capacities. The alternative option of targeting general pathways rather than specific mutant proteins may be more feasible.

NONCLONAL MUTATIONS AND INTRATUMOR HETEROGENEITY

A significant limitation of these studies lies not in the complexity of the clonal mutations they are attempting to annotate, but in the fact that, by design, they are unable to address deeper heterogeneity within individual tumors. Most investigations to date have been concerned only with identifying mutations in the dominant clone. A tumor is itself genomically heterogeneous, as each cell has a different mutational signature reflective of its unique lineage history within the evolving neoplasm. Analysis of disseminated single cells in minimal residual disease has demonstrated a high level of genomic heterogeneity within individual lesions as well as between primary tumors and metastatic cells (114). Irrespective of the predominant forces generating mutations during tumorigenesis, it is of critical importance to recognize that mutations occur randomly and that only a tiny subset of these are likely to be selectably advantageous at a given stage of development. This leaves a much larger number of unexpanded mutations to act as a dormant repository of genetic diversity. A tumor is a dynamic entity that never ceases to evolve. The specific fitness of a cancer cell depends on the context of its tumor environment. As the environment changes over time, new cellular stresses such as hypoxia, nutrient depletion, and immune recognition arise, and cancer cells with the requisite phenotypes are selected. Thus, subclonal heterogeneity is of paramount importance to tumor progression.

The clinical importance of subclonal mutations arises from the fact that genetic variants encoding resistance to all single-target drugs are likely to preexist in a tumor cell population (70). Imatinib, the prototypical targeted therapy for chronic myelogenous leukemia (CML) bearing activating mutations in the *ABL* gene, frequently loses clinical efficacy due to the emergence of resistant clones (83). The basis for this resistance is frequently attributable to new point mutations in the *ABL* kinase domain that decrease its affinity for the drug (115–116). It has been specifically demonstrated that resistance mutations can be found in CML prior to the initiation of therapy (117). A method for overcoming the problem of resistance, which has long been standard in the clinic with existing chemotherapy regimens, is to use multidrug cocktails. A group of agents directed at multiple, unrelated tumor-relevant pathways is more likely to prevent or delay resistance, given that the combinatorial probability of a single cell simultaneously bearing resistance mutations to each pathway is very low. Although this

concept is not new, the degree of intertumor heterogeneity revealed by cancer genome studies hints that an even larger cocktail of drugs than is presently used may be advantageous to combat resistance arising from intratumor genetic variants (118).

The ability to quantify subclonal genetic diversity may provide important clinical information about the likelihood of a tumor becoming resistant to specific therapies. Exactly how many unexpanded random mutations are there in a tumor? This has historically been a difficult question to answer because of the technical challenges facing low-frequency measurements (**Figure 5**). Standard capillary sequencing technology measures average population genotype and only detects minority clones down to approximately 25% with routine, automated use. Next-generation sequencing methods are much more sensitive, given that they genotype the amplified product of individual molecules. Sequencing of many fragments from a given region (i.e., deep sequencing) produces a digital histogram representing the frequency of different genotypes in a population of molecules. However, because of imperfections in detection hardware and chemistry, as well as the need for amplification steps by fallible polymerases, sensitivity is currently limited to approximately 1 in 5000 under the most ideal circumstances (119) and probably nearer to 1 in 100 with routine use. Although there do exist exquisitely sensitive methods of mutation detection, including cell culture–based fluctuation assays (10) and systems involving transgenic animals bearing reporter genes, neither method is amenable to the direct examination of human tumors.

Our group has recently developed a method for the detection of random mutations that offers unprecedented sensitivity; one mutation can be identified among 10^8 wild-type nucleotides in nuclear DNA (120). The system is based on the concept that spontaneous mutations occurring in a noncoding, TaqI restriction endonuclease recognition site render it noncleavable by this enzyme. After multiple rounds of enzymatic digestion, only the mutant sequences from a larger population remain intact and are amplifiable by PCR primers flanking the restriction site. We (121) and others (122) have used this approach to demonstrate a markedly elevated frequency of random, unexpanded mutations in several types of cancer. Although highly sensitive, this approach can interrogate only four bases out of the entire

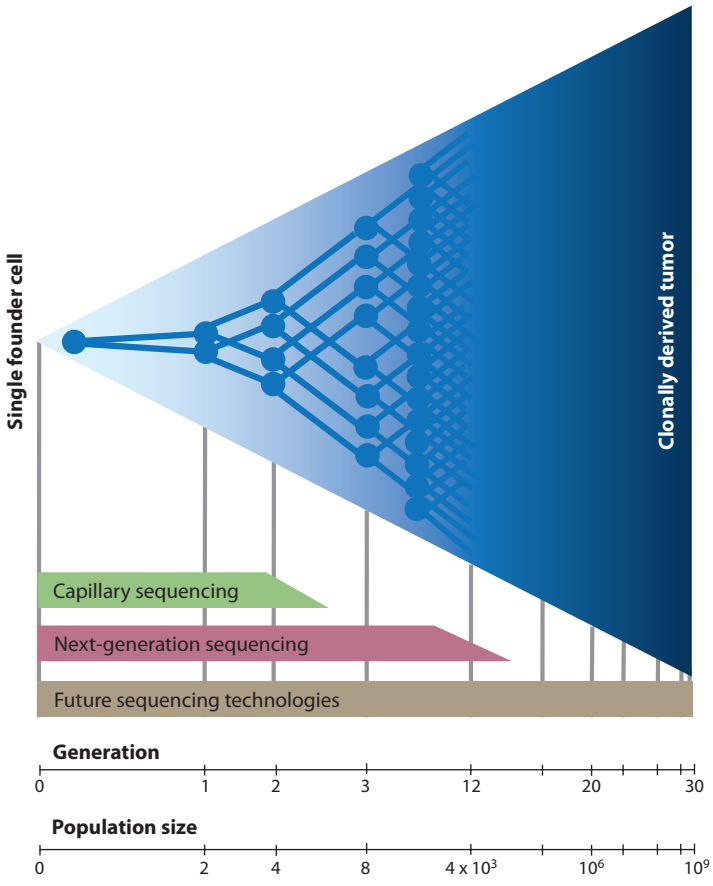

Figure 5

Limit of subclonal detection. Depicted is the clonal expansion of a single cell into a population of 1 billion cells. In a hypothetical scenario in which no cell death occurs, this expansion requires approximately 30 generations of division. Current capillary methods of DNA sequencing only detect mutations that have clonally expanded to represent 25% or more of a tumor. Only mutations that are present in the founding cell or that arise within the first two generations of division can be identified. Deep sequencing on current next-generation sequencing platforms has been reported to detect subclonal mutations down to a frequency of 1 in 5000 (119) (i.e., those mutations that arise within the first 12 generations after founding). Sensitivity is limited by the error rate of polymerase chain reaction amplification steps and that of the sequencing chemistry itself. Future technologies may eventually enable ultra-accurate, high-throughput detection of mutations that arise during any stage of clonal expansion.

Deep sequencing: sequencing of many individual DNA fragments from an identical portion of the genome to identify subclonal mutations

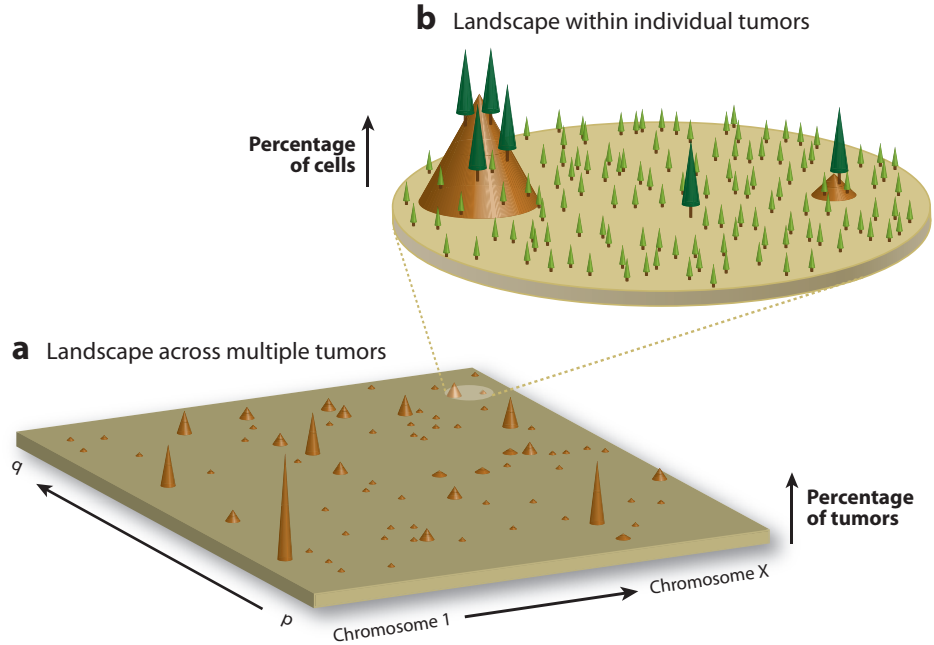

Figure 6

The mutational landscape of the cancer genome. (*a*) The cancer genome landscape proposed by Wood et al. (88) graphically represents the mutational heterogeneity among different tumors of a single cancer type. The height of each brown peak indicates the percentage of tumors found to carry a clonal mutation in a particular gene. The landscape comprises a small number of "mountains"—genes that are clonally mutated in a large fraction of individual cancers—and a significantly greater number of "hills"—genes clonally mutated in only one or a few tumors. Although 50 or more genes may be clonally mutated within the genome of an individual tumor, most genes are rarely mutated in more than a few tumors. (*b*) An additional level of the mutational landscape exists within individual tumors due to differences among the genomes of single tumor cells. Although a small number of mutations—"trees"—are clonally present in the majority of cells in an individual tumor, an exponentially larger number—"seedlings"—exist subclonally in only one or a few cells. Among this vast reservoir of nonclonal mutations are many therapy-resistant variants.

genome at once. Hopefully the future will bring even more sophisticated methods that combine the throughput of present next-generation sequencing platforms with the ultrahigh sensitivity needed to accurately identify single mutants.

The cancer genomic landscape has been previously described from a multiple-tumor perspective, in which the abovementioned "mountains" and "hills" represent, respectively, frequently and infrequently mutated genes (**Figure 6*a***) (87). We posit that a complete description of the cancer genome must necessarily include a provision for the intratumor heterogeneity of individual neoplasms. To this landscape we add a small number of "trees" to represent clonally mutated genes present in a large number of cells and, surrounding these, a much larger number of "seedlings" representing mutations present in only one or a few cells (**Figure 6*b***). We argue that it is this forest, undetected by the cancer genome–sequencing studies described above, that provides much of the basis for the wide variations in tumor behavior and responsiveness to therapy and that represents one of the most clinically important features of the cancer genome: When an old tree falls or is logged, many seedlings are poised to grow and take its place.

FINAL THOUGHTS AND FUTURE DIRECTIONS

Currently we recognize the unidirectional fate of neoplastically transformed cells to be the result of unrepaired mutational events that become permanently fixed in the genome and epigenome of subsequent generations of progeny. The path that has brought us to our current understanding of the genetic basis for cancer has been long and remarkable, punctuated by a breadth of discoveries. The ability of most cells in the human body to prevent the build-up of cancer-causing mutations is an impressive tribute to the billions of years of evolution leading up to the emergence of multicellular organisms. Mechanisms for mutation prevention and suppression are, nevertheless, imperfect; progressive accumulation of new genetic and epigenetic variants provides the fuel for evolution on a cellular level and forms the basis of tumorigenesis.

New technology has allowed us to begin to tabulate the mutations of cancer. First-generation cancer genome–sequencing studies were driven by the expectation that clonal mutations in a limited set of key genes would be commonly found in different tumors and might provide new drug-responsive targets. However, the results so far indicate a more complex picture than initially hoped. Very few genes that have not been previously identified by other means are prevalently mutated in specific cancers. Many genes likely to be involved in driving tumorigenesis are altered in only a small fraction of tumors. The presence of many thousands of clonally expanded passengers, although playing no causal role in the cancer, serve as a reminder of the invisible legions formed by the exponentially larger number of unexpanded variants, many of which are drug resistant and are awaiting the opportunity to selectively proliferate upon induction of new treatments. In light of this emerging complexity, it is becoming increasingly difficult to envision how it will be possible to develop a realistic number of targeted chemotherapies to be directed against a discrete panel of commonly mutated cancer genes. The findings described herein substantiate the concept that simultaneous use of multiple agents against different general pathways may be the most efficacious approach.

Although sobering, the cancer genome studies performed so far have established an important baseline of information from which to advance. As technology improves and comes down in cost, large-scale genome-analysis methods will become tractable to smaller research groups, who will be able to explore innovative and higher-risk approaches. As we move forward in these endeavors, we must not lose sight of the need to confirm the functional status of mutations identified. Although genome sequencing remains a powerful tool, it cannot address all the questions of cancer research, so we cannot neglect to spread our resources among many complementary means of identifying novel features of the disease and ways to prevent and target them. Most importantly, we need to recognize the many levels of heterogeneity inherent to cancer and ensure that this reality be integrated into future studies.

SUMMARY POINTS

1. Cancer is a disease of somatic evolution that occurs on a cellular level. Random mutations occur throughout an organism's life, and only a small subset of mutations that bestow growth and survival benefits upon a cell clonally expand to form a tumor.
2. Mutations result from errors in DNA synthesis and failure of DNA repair. Mutations accumulate more rapidly during tumorigenesis from increased cell proliferation as well as from defects in DNA maintenance pathways. The relative importance of these two mechanisms is likely to vary from one tumor to another and remains the subject of debate.

3. Identifying causative mutations within a tumor's genome by DNA sequencing is a complex problem from both a technical and an analytical standpoint. Many mutations are clonally present by chance alone, and differentiating these neutral passengers from causal drivers presents a significant challenge.
4. Recent large-scale cancer genome–sequencing studies have indicated a great deal of variation in the clonal mutations found among different tumors. Very few genes are commonly mutated in any type of cancer, and this finding suggests that it will be difficult to design a limited number of widely usable targeted therapies that focus on specific genes.
5. Current methods of DNA sequencing cannot accurately portray the many mutations in a developing cancer that are present in only a minority of tumor cells. These subclonal mutations constitute a tumor's potential for overcoming therapeutic interventions, and their presence argues for multiple and simultaneous chemotherapeutic approaches for tumor ablation. Characterization of this intratumor heterogeneity will be of clinical importance.

DISCLOSURE STATEMENT

L.A.L. is a member of the advisory board for Stratos, Inc. The other authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Grant support: National Cancer Institute, CA102029 and CA115802 (L.A.L.); U.S. Department of Defense, W81XWH-07-0029 (L.A.L.); National Institute of Aging, AG033485 (J.J.S.); National Cancer Institute/Health Research Board of Ireland, Joint Research Fellowship (E.J.F.).

LITERATURE CITED

1. **Boveri T. 1902. Über mehrpolige Mitosen als Mittel zur Analyze des Zellkerns.** ***Verh. Dtsch. Zool. Ges. Würzburg*** **35:67–90**
2. **Cleaver JE. 1968. Defective repair replication of DNA in xeroderma pigementosum.** ***Nature*** **218:652–56**
3. Kallioniemi A, Kallioniemi O-P, Piper J, Tanner M, Stokke T, et al. 1994. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl. Acad. Sci. USA* 91:2156–60
4. Riopel MA, Spellerberg A, Griffin CA, Perlman EJ. 1998. Genetic analysis of ovarian germ cell tumors by comparative genomic hybridization. *Cancer Res.* 58:3105–10
5. El-Rifai W, Sarlomo-Rikala M, Knuutila S, Miettinen M. 1998. DNA copy number changes in development and progression in leiomyosarcomas of soft tissues. *Am. J. Pathol.* 153:985–90
6. Kerangueven F, Noguchi T, Coulier F, Allione F, Wargniez V, et al. 1997. Genome-wide search for loss of heterozygosity shows extensive genetic diversity of human breast carnomas. *Cancer Res.* 57:5469–74
7. Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. 1999. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl. Acad. Sci. USA* 96:4494–99
8. Vogelstein B, Fearon ER, Kern SE, Hamilton SR, Preisinger AC, et al. 1988. Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* 319:525–32

1. Classic paper; first to speculate about cancer as a chromosomal disease.

2. First paper to establish the association between defects in DNA repair and human cancer.

9. Albertini RJ, Nicklas JA, O'Neill JP, Robison SH. 1990. In vivo somatic mutations in humans: measurement and analysis. *Annu. Rev. Genet.* 24:305–26
10. Bielas JH, Heddle JA. 2003. Elevated mutagenesis and decreased DNA repair at a transgene are associated with proliferation but not apoptosis in p53-deficient cells. *Proc. Natl. Acad. Sci. USA* 100:12853–58
11. Chu EHY, Boehnke M, Hanash SM, Kuick RD, Lamb BJ, et al. 1988. Estimation of mutation rates based on the analysis of polypeptide constituents of cultured human lymphoblastoid cells. *Genetics* 119:693–703
12. DeMars R, Held KR. 1972. The spontaneous azaguanine-resistant mutants of diploid human fibroblasts. *Humangenetik* 16:87–110
13. Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ. 2002. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc. Natl. Acad. Sci. USA* 99:3586–90
14. Jackson AL, Loeb LA. 1998. The mutation rate and cancer. *Genetics* 148:1483–90
15. Ferrin LJ, Beckman RA, Loeb LA, Mildvan AS. 1986. Kinetic and magnetic resonance studies of the interaction of Mn^{2+}, substrates and templates with DNA polymerases. In *Manganese in Metabolism and Enzyme Function*, ed. VL Schramm, FC Wedler, pp. 259–73. New York: Academic
16. Petruska J, Sowers LC, Goodman MF. 1986. Comparison of nucleotide interactions in water, proteins and vacuum: model for DNA polymerase fidelity. *Proc. Natl. Acad. Sci. USA* 83:1559–62
17. Loeb LA, Monnat R. 2008. DNA polymerases and human disease. *Nat. Rev. Genet.* 9:594–604
18. Uyemura D, Lehman IR. 1976. Biochemical characterization of mutant forms of DNA polymerase I from *Escherichia coli*. I. The polA12 mutation. *J. Biol. Chem.* 251:4078–84
19. Modrich P, Lahue R. 1996. Mismatch repair in replication fidelity, genetic recombination and cancer biology. *Annu. Rev. Biochem.* 65:101–33
20. Kunkel TA, Schaaper RM, Loeb LA. 1983. Depurination-induced infidelity of DNA synthesis with purified DNA replication proteins in vitro. *Biochemistry* 22:2378–84
21. Kiel MJ, He S, Ashkenazi R, Gentry SN, Teta M, et al. 2007. Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. *Nature* 449:238–42
22. Nishimura S. 2006. 8-hydroxyguanine: from its discovery in 1983 to the present status. *Proc. Jpn. Acad.* 82:127–41
23. Floyd RA. 1995. Measurement of oxidative stress in vivo. In *The Oxygen Paradox*, ed. KJA Davies, F Ursini, pp. 89–103. Padova, Italy: CLEUP Univ. Press
24. Lindahl T, Wood RD. 1999. Quality control by DNA repair. *Science* 286:1897–905
25. Barrett MT, Sanchez CA, Prevo LJ, Wong DJ, Galipeau PC, et al. 1999. Evolution of neoplastic cell lineages in Barrett esophagus. *Nat. Genet.* 22:106–9
26. Bronner MP, O'Sullivan JN, Rabinovitch PS, Crispin DA, Chen L, et al. 2008. Genomic biomarkers to improve ulcerative colitis neoplasia surveillance. *Am. J. Pathol.* 173:1853–60
27. Baik SC, Youn HS, Chung MH, Lee WK, Cho MJ, et al. 1996. Increased oxidative DNA damage in *Heliocobacter pylori*–infected human gastric mucosa. *Cancer Res.* 56:1279–82
28. Fausto N, Webber EM. 1993. Mechanisms of growth regulation in liver regeneration and hepatic carcinogenesis. In *Progress in Liver Diseases*, ed. JL Boyer, RK Okner, 11:115–37. Philadelphia: W.B. Saunders
29. Feig DI, Loeb LA. 1994. Oxygen radical induced mutagenesis is DNA polymerase specific. *J. Mol. Biol.* 235:33–41
30. Ames BN, Durston WE, Yamasaki E, Lee FD. 1973. Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc. Natl. Acad. Sci. USA* 70:2281–85
31. Auerbach C, Robson JM. 1946. Chemical production of mutations. *Nature* 154:302
32. Loeb L, Harris CC. 2008. Advances in chemical carcinogenesis: a historical review and prospective. *Cancer Res.* 68:6863–72
33. Hecht SS. 1999. Tobacco smoke carcinogens and lung cancer. *J. Natl. Cancer Inst.* 91:1194–210
34. Wynder EL, Graham EA. 1950. Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. *JAMA* 143:329–36
35. Loeb LA, Ernster VL, Warner KE, Abbotts J, Laszlo J. 1984. Smoking and lung cancer: an overview. *Cancer Res.* 44:5940–58
36. Friedberg EC, Aguilera A, Gellert M, Hanawalt PC, Hays JB, et al. 2006. DNA repair: from molecular mechanism to human disease. *DNA Repair* 5:986–96

30. The test described within is the standard of reference for detection of mutagenic chemicals.

35. Position paper of the American Association of Cancer Research on the association between smoking and cancer.

37. Hoeijmakers JH. 2001. Genome maintenance mechanisms for preventing cancer. *Nature* 411:366–74
38. Harris CC. 1993. p53: at the crossroads of molecular carcinogenesis and cancer risk assessment. *Science* 262:1980–81
39. Kolodner RD, Marsischky GT. 1999. Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.* 9:89–96
40. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, et al. 2002. Inherited variants of MYH associated with somatic G:C–>T:A mutations in colorectal tumors. *Nat. Genet.* 30:227–32
41. King MC, Marks JH, Mandell JB. 2003. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* 302:643–46
42. Riballo E, Critchlow SE, Teo SH, Doherty AJ, Priestley A, et al. 1999. Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr. Biol.* 9:699–702
43. Masutani C, Araki M, Yamada A, Kusumoto R, Nogimori T, et al. 1999. Xeroderma pigmentosum variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity. *EMBO J.* 18:3491–501
44. Ellis NA, Groden J, Ye T-Z, Straughen J, Lennon DJ, et al. 1995. The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* 83:655–66
45. Kamath-Loeb AS, Lan L, Nakajima S, Yasui A, Loeb LA. 2007. The Werner syndrome protein interacts functionally with translesion DNA polymerases. *Proc. Natl. Acad. Sci. USA* 104:10394–99
46. Levine AJ. 1997. p53, the cellular gatekeeper for growth and division. *Cell* 88:323–31
47. Goode EL, Ulrich CM, Potter JD. 2002. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol. Biomark. Prev.* 11:1513–30
48. Webber EM, Wu JC, Wang L, Merlino G, Fausto N. 1994. Overexpression of transforming growth factor-α causes liver enlargement and increased hepatocyte proliferation in transgenic mice. *Am. J. Pathol.* 145:398–408
49. Hecker E. 1967. Phorbol esters from croton oil. Chemical nature and biological activities. *Naturwissenschaften* 54:282–84
50. Rous P. 1911. Transmission of a malignant new growth by means of a cell-free filtrate. *JAMA* 56:198
51. Varmus HE, Shank PR, Hughes SE, Kung HJ, Heasley S, et al. 1979. Synthesis, structure, and integration of the DNA of RNA tumor viruses. *Cold Spring Harb. Symp. Quant. Biol.* 43(Part 2):851–64

52. Presents a detailed analysis of the required phenotypes for tumorigenesis.

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

53. Reya T, Morrison SJ, Clarke MF, Weissman IL. 2001. Stem cells, cancer, and cancer stem cells. *Nature* 414:105–11

53a. Pepper JW, Sprouffske K, Maley CC. 2007. Animal cell differentiation patterns suppress somatic evolution. *PLoS Comput. Biol.* 3(12):e50

54. Humphries A, Wright NA. 2008. Colonic crypt organization and tumorigenesis. *Nat. Rev. Cancer* 8:415–24
55. Cairns J. 2006. Cancer and the immortal strand hypothesis. *Genetics* 174:1069–72
56. Prindull GA, Fibach E. 2007. Are postnatal hemangioblasts generated by dedifferentiation from committed hematopoietic stem cells? *Exp. Hematol.* 35:691–701

57. Seminal paper describing cancer an as evolutionary process.

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

58. Merlo LM, Pepper JW, Reid BJ, Maley CC. 2006. Cancer as an evolutionary and ecological process. *Nat. Rev. Cancer* 6:924–35
59. Cairns J. 1975. Mutation selection and the natural history of cancer. *Nature* 255:197–200
60. Armitage P, Doll R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* 8:1–12
61. Knudson AG. 1996. Hereditary cancer: two hits revisited. *J. Cancer Res. Clin. Oncol.* 122:135–40
62. Renan MJ. 1993. How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol. Carcinog.* 7:139–46
63. Rangarajan A, Hong SJ, Gifford A, Weinberg RA. 2004. Species- and cell type–specific requirements for cellular transformation. *Cancer Cell* 6:171–83
64. Mahale AM, Khan ZA, Igarashi M, Nanjangud GJ, Qiao RF, et al. 2008. Clonal selection in malignant transformation of human fibroblasts transduced with defined cellular oncogenes. *Cancer Res.* 68:1417–26
65. Bodmer W. 2008. Genetic instability is not a requirement for tumor development. *Cancer Res.* 68:3558–61

66. Loeb LA, Bielas JH, Beckman RA. 2008. Cancers exhibit a mutator phenotype: clinical implications. *Cancer Res.* 68:3551–57
67. Lengauer C, Kinzler KW, Vogelstein B. 1998. Genetic instabilities in human cancers. *Nature* 396:643–49
68. Moolgavkar SH, Luebeck EG. 2003. Multistage carcinogenesis and the incidence of human cancer. *Genes Chromosomes Cancer* 38:302–6
69. Loeb LA, Springgate CF, Battula N. 1974. Errors in DNA replication as a basis of malignant change. *Cancer Res.* 34:2311–21
70. Loeb LA, Loeb KR, Anderson JP. 2003. Multiple mutations and cancer. *Proc. Natl. Acad. Sci. USA* 100:776–81
71. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, et al. 2007. Genetic progression and the waiting time to cancer. *PLoS Comput. Biol.* 3:e225
72. Tomlinson IP, Novelli MR, Bodmer WF. 1996. The mutation rate and cancer. *Proc. Natl. Acad. Sci. USA* 93:14800–3
73. Beckman RA, Loeb LA. 2006. Efficiency of carcinogenesis with and without a mutator mutation. *Proc. Natl. Acad. Sci. USA* 103:14140–15
73a. Beckman RA. 2009. Mutator mutations enhance tumorigenic efficiency across fitness landscapes. *PLoS One* 4:e5860
74. Beckman RA, Loeb LA. 2005. Negative clonal selection in tumor evolution. *Genetics* 171:2123–31
75. Robins H, Alexe G, Harris SL, Levine AJ. 2005. The first twenty-five years of p53 research. In *Twenty-Five Years of p53 Research*, ed. P Hainaut, KG Wiman, pp. 1–26. Dordrecht: Springer
76. Wells RD, Sinden RR. 1993. Defined ordered sequence DNA, DNA structure, and DNA-directed mutation. In *Genome Analysis, Genome Rearrangement and Stability*, 7:107–38. Plainview, NY: Cold Spring Harb. Lab. Press
77. Farber RA, Petes TD, Dominska M, Hudgens SS, Liskay RM. 1994. Instability of simple sequence repeats in a mammalian cell line. *Hum. Mol. Genet.* 3:253–56
78. Hanawalt PC, Spivak G. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* 9:958–70
79. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. 2009. Human mutation rate associated with DNA replication timing. *Nat. Genet.* 41:393–95
80. Balaban GB, Herlyn M, Clark WH Jr, Nowell PC. 1986. Karyotypic evolution in human malignant melanoma. *Cancer Genet. Cytogenet.* 19:113–22
81. Smith G, Carey FA, Beattie J, Wilkie MJV, Lightfoot TJ, et al. 2002. Mutations in APC, Kirsten-ras, and p53—alternative genetic pathways to colorectal cancer. *Proc. Natl. Acad. Sci. USA* 99:9433–38
82. Ley TJ, Mardis E, Ding L, Fulton B, McLellan M, et al. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456:66–72
83. Druker BJ. 2004. Imatinib: paradigm or anomaly? *Cell Cycle* 3:833–35
84. Wang TL, Rago C, Silliman N, Ptak J, Markowitz S, et al. 2002. Prevalence of somatic alterations in the colorectal cancer cell genome. *Proc. Natl. Acad. Sci. USA* 99:3076–80
85. Wang Z, Shen D, Parsons DW, Bardelli A, Sager J, et al. 2004. Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science* 304:1164–66
86. Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, et al. 2003. Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* 300:949
87. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, et al. 2006. The concensus coding sequences of human breast and colorectal cancers. *Science* 314:268–74
88. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science* 318:1108–13
89. Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324:255–58
90. Getz G, Hofling H, Mesirov JP, Golub TR, Meyerson M, et al. 2007. Comment on "The Consensus Coding Sequences of Human Breast and Colorectal Cancers." *Science* 317:1500
91. Rubin AF, Green P. 2007. Comment on "The Consensus Coding Sequences of Human Breast and Colorectal Cancers." *Science* 317:1500

69. First paper to propose that cancers exhibit a mutator phenotype.

82. Provided the first complete whole-genome sequence of a tumor and matched normal and identified nearly 1000 nonsynonymous somatic mutations.

87–88. First cancer genome studies to sequence mutations in all known coding exons of any cancer.

92. Forrest WF, Cavet G. 2007. Comment on "The Consensus Coding Sequences of Human Breast and Colorectal Cancers." *Science* 317:1500
93. Parmigiani G, Lin J, Boca S, Sjöblom T, Jones S, et al. 2007. Response to comments on "The Consensus Coding Sequences of Human Breast and Colorectal Cancers." *Science* 317:1500
94. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, et al. 2007. Patterns of somatic mutation in human cancer genomes. *Nature* 446:153–58
95. Stephens P, Edkins S, Davies H, Greenman C, Cox C, et al. 2005. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat. Genet.* 37:590–92
96. Davies H, Hunter C, Smith R, Stephens P, Greenman C, et al. 2005. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res.* 65:7591–95
97. Bignell G, Smith R, Hunter C, Stephens P, Davies H, et al. 2006. Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults. *Genes Chromosomes Cancer* 45:42–46
97a. Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, et al. 2008. Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. USA* 105:4283–88
98. Jones S, Zhang X, Parsons D, Lin J, Leary R, et al. 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321:1801–6
99. Prehn RT, Main JM. 1957. Immunity to methylcholanthrene-induced sarcomas. *J. Natl. Cancer Inst.* 18:769–78
100. Segal N, Parsons D, Peggs K, Velculescu V, Kinzler K, et al. 2008. Epitope landscape in breast and colorectal cancer. *Cancer Res.* 68:889–92
101. Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, et al. 2008. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl. Acad. Sci. USA* 105:16224–29
102. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science* 321:1807–12
103. Cancer Genome Atlas Network. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455:1061–68
104. Balss J, Meyer J, Mueller W, Korshunov A, Hartmann C, von Deimling A. 2008. Analysis of the IDH1 codon 132 mutation in brain tumors. *Acta Neuropathol.* 116:597–602
105. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, et al. 2009. IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* 360:765–73
106. Zhao S, Lin Y, Xu W, Jiang W, Zha Z, et al. 2009. Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1α. *Science* 324:261–61
107. Weir BA, Woo MS, Getz G, Perner S, Ding L, et al. 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450:893–98
108. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, et al. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455:1069–75
109. Tomasson MH, Xiang Z, Walgren R, Zhao Y, Kasai Y, et al. 2008. Somatic mutations and germline sequence variants in the expressed tyrosine kinase genes of patients with de novo acute myeloid leukemia. *Blood* 111:4797–808
110. Loriaux MM, Levine RL, Tyner JW, Frohling S, Scholl C, et al. 2008. High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. *Blood* 111:4788–96
111. Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, et al. 2008. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* 40:722–29
112. Lai LA, Paulson TG, Li X, Sanchez CA, Maley C, et al. 2007. Increasing genomic instability during premalignant neoplastic progression revealed through high resolution array-CGH. *Genes Chromosomes Cancer* 46:532–42
113. Shames DS, Girard L, Gao B, Sato M, Lewis CM, et al. 2006. A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PLoS Med.* 3:e486

94. The most extensive characterization to date of mutations in a single gene family across different cancer types.

114. Stoecklein NH, Hosch SB, Bezler M, Stern F, Hartmann CH, et al. 2008. Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. *Cancer Cell* 13:441–53
115. Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, et al. 2005. Dynamics of chronic myeloid leukaemia. *Nature* 435:1267–70
116. Radich JP, Dai H, Mao M, Oehler V, Schelter J, et al. 2006. Gene expression changes associated with progression and response in chronic myeloid leukemia. *Proc. Natl. Acad. Sci. USA* 103:2794–99
117. Talpaz M, Shah NP, Kantarjian H, Donato N, Nicoll J, et al. 2006. Dasatinib in imatinib-resistant Philadelphia chromosome–positive leukemias. *N. Engl. J. Med.* 354:2531–41
118. Golub TR. 2003. Mining the genome for combination therapies. *Nat. Med.* 9:510–11
119. Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, et al. 2008. Subclonal phylogenetic structures in cancer revealed by ultradeep sequencing. *Proc. Natl. Acad. Sci. USA* 105:13081–86
120. Bielas J, Loeb LA. 2005. Quantification of random genomic mutations. *Nat. Methods* 2:285–90
121. Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA. 2006. Human cancers express a mutator phenotype. *Proc. Natl. Acad. Sci. USA* 103:18238–42
122. Zheng L, Dai H, Zhou M, Li M, Singh P, et al. 2007. Fen1 mutations result in autoimmunity, chronic inflammation and cancers. *Nat. Med.* 13:812–19

119. Demonstrates the extent of subclonal mutational heterogeneity identifiable by current sequencing technologies.

Annual Review of Pathology: Mechanisms of Disease

Volume 5, 2010

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Pathology, Mechanisms of Disease* articles may be found at http://pathol.annualreviews.org