Exploring the Genomes of Cancer Cells: Progress and Promise

Michael R. Stratton*

The description and interpretation of genomic abnormalities in cancer cells have been at the heart of cancer research for more than a century. With exhaustive sequencing of cancer genomes across a wide range of human tumors well under way, we are now entering the end game of this mission. In the forthcoming decade, essentially complete catalogs of somatic mutations will be generated for tens of thousands of human cancers. Here, I provide an overview of what these efforts have revealed to date about the origin and behavioral features of cancer cells and how this genomic information is being exploited to improve diagnosis and therapy of the disease.

Much of our current understanding of cancer is based on the central tenet that it is a genetic disease, arising as a clone of cells that expands in an unregulated fashion because of somatically acquired mutations (1). These somatic mutations include base substitutions, insertions and deletions (indels) of bases, rearrangements caused by breakage and abnormal rejoining of DNA, and changes in the copy number of DNA segments. They also often include epigenetic changes that are stably inherited during mitotic DNA replication, for example, alterations in methylation of cytosine residues (2).

Whether a mature cancer clone emerges in an individual person is influenced by environmental and life-style factors, as well as by the set of genomic sequence variants present in the fertilized egg from which the individual develops and that are therefore found in all somatic cells. These so-called constitutional or "germline" mutations can influence cancer susceptibility in a number of ways, including directly altering growth of the cancer clone, altering the mutation rate in somatic cells, or modulating the metabolism of carcinogens.

Somatic mutations are thought to occur in the genomes of all normal cells as they proceed through the rounds of cell division that take place during development in utero and during replenishment of body tissues in postnatal life. Additional somatic mutations continue to accumulate in cancer cells as they divide. The rate of acquisition and the types of somatic mutation that accrue can be increased by exogenous and endogenous exposures that cause DNA damage and are mitigated by DNA repair processes. Indeed, in the event that DNA repair fails, the somatic mutation rate may also increase.

Somatic mutations are more or less randomly distributed throughout the genome. However, in the cell that undergoes clonal expansion to become a cancer, a subset—termed "driver mutations"—have by chance fallen in a set of key genes, called "cancer genes," and have thus subverted normal control of cell proliferation, differentiation, death, and other homeostatic interactions with the tissue microenvironment (3). Driver mutations confer growth advantage upon the neoplastic clone, allowing it to expand more than normal cells from the same tissue, invade into surrounding tissue, and, in many cases, metastasize. The number of driver mutations in a cancer cell reflects the number of mutated cancer genes and thus the deregulation of cell biological processes required to convert a normal cell into a symptomatic cancer clone. The remaining—and often the large majority of—mutations are "passengers," which, by definition, do not confer growth advantage. The number of passenger mutations in a cancer genome primarily reflects the number of mitotic cell divisions between the fertilized egg and the cancer cell and the mutation rate at each of these cell divisions. Thus, the catalog of somatic mutations in the genome of a cancer cell represents genomic changes that usually accumulate over several decades. It includes the mutations responsible for conferring the various aspects of the neoplastic phenotype and bears the imprints of the mutational processes that caused the disease in the first place.

Cataloging Mutations in Human Cancer Genomes

Over the past half-century a series of technologies have been deployed to characterize systematically, at ever-increasing levels of resolution, the state of cancer genomes across the range of cancer types (Fig. 1). The earliest, and still one of the most influential in its impact on cancer science, was cytogenetic studies of chromosomes from cancer cells. These revealed abnormalities of chromosome copy number and the presence of somatically acquired rearrangements (chromosomal translocations). They showed that some cancer types had very disordered genomes whereas others displayed few genomic abnormalities. They also yielded evidence that certain positions in the genome were recurrently rearranged in particular cancer types, from which it was inferred that a cancer gene resided at the rearrangement breakpoints. After the widespread adoption of recombinant DNA technology in the 1980s, it became possible to isolate and sequence the genome in the vicinity of these recurrently rearranged regions, leading to the identification of many rearranged cancer genes, particularly in leukemias, lymphomas, and sarcomas (4).

The next major suite of technologies primarily provided evidence of copy number change in cancer genomes, but at higher resolution than was generally possible by cytogenetics. These approaches confirmed the variation in extent of copy number change between individual cancer genomes and highlighted regions showing recurrent increases or reductions in copy number. Subsequent studies focusing on these recurrently abnormal regions provided a further harvest of new cancer genes (5, 6).

These technologies had their limitations. Most obviously, they could not directly detect base substitutions or small indels. The emergence of the draft human genome sequence in 2000 empowered the study of cancer genomes in many ways. In particular, it provided a template for the design of polymerase chain reaction (PCR) primer pairs to amplify and sequence (by conventional sequencing technology) the coding exons of large numbers of protein-coding genes. This facilitated more extensive sequencing of cancer genomes, including whole gene families and subsequently most coding exons (7–28). These studies systematically sampled cancer genomes for somatic base substitutions and small indels, providing, for the first time, insights into their prevalence. However, exploration of noncoding areas of the genome and larger numbers of cases was still restricted by high cost and limited sequencing capacity.

The recent arrival of second-generation DNA sequencing technologies (29) has further transformed investigation of cancer genomes. These technologies are being applied in a number of ways. Because most of the currently known driver mutations change the coding sequences of protein-coding genes and because protein-coding exons account for only ~1% of the human genome, sequencing is often being thriftily targeted at these positions. After the widespread adoption of recombinant DNA technology in the 1980s, it became possible to isolate and sequence the genome in the vicinity of these recurrently rearranged regions, leading to the identification of many rearranged cancer genes, particularly in leukemias, lymphomas, and sarcomas (4).

The next major suite of technologies primarily provided evidence of copy number change in cancer genomes, but at higher resolution than was generally possible by cytogenetics. These approaches confirmed the variation in extent of copy number change between individual cancer genomes and highlighted regions showing recurrent increases or reductions in copy number. Subsequent studies focusing on these recurrently abnormal regions provided a further harvest of new cancer genes (5, 6).

These technologies had their limitations. Most obviously, they could not directly detect base substitutions or small indels. The emergence of the draft human genome sequence in 2000 empowered the study of cancer genomes in many ways. In particular, it provided a template for the design of polymerase chain reaction (PCR) primer pairs to amplify and sequence (by conventional sequencing technology) the coding exons of large numbers of protein-coding genes. This facilitated more extensive sequencing of cancer genomes, including whole gene families and subsequently most coding exons (7–28). These studies systematically sampled cancer genomes for somatic base substitutions and small indels, providing, for the first time, insights into their prevalence. However, exploration of noncoding areas of the genome and larger numbers of cases was still restricted by high cost and limited sequencing capacity.

The recent arrival of second-generation DNA sequencing technologies (29) has further transformed investigation of cancer genomes. These technologies are being applied in a number of ways. Because most of the currently known driver mutations change the coding sequences of protein-coding genes and because protein-coding exons account for only ~1% of the human genome, sequencing is often being thriftily targeted at these positions. After the widespread adoption of recombinant DNA technology in the 1980s, it became possible to isolate and sequence the genome in the vicinity of these recurrently rearranged regions, leading to the identification of many rearranged cancer genes, particularly in leukemias, lymphomas, and sarcomas (4).
Cancer Crusade at 40

Hundreds of cancers have been sequenced (34–36). This approach will report substitutions in genes that have sufficiently high levels of mRNA and can report rearrangements that are transcribed. Again, however, abnormalities of noncoding regions will generally be missed, and protein-truncating mutations may be difficult to find if they activate nonsense-mediated RNA decay.

In the longer term, however, the major impact of these remarkable technology shifts will be to permit the sequencing of whole cancer genomes (37–43). This strategy, in which genomic DNA from a cancer (and, in parallel, DNA isolated from normal tissue of the same person) is randomly fragmented and hundreds of millions of fragments are sequenced, can reveal all classes of somatic change (base substitutions, indels, rearrangements, copy number changes, and even potentially epigenetic alterations) in all sectors of the genome (exons, introns, and intergenic regions). It has paved the way to the generation of almost complete catalogs of somatic mutation for individual cancers (39, 40), allowing us to set aside our preconceptions of where the important mutations that cause the disease might lie and, by acquisition of large numbers of mutations from individual cases, empowering deeper study of the mutational processes that have been operative. A few hundred whole cancer genomes have already been generated by sequencing machines worldwide and are in the process of being analyzed.

The Number of Mutations in Cancer Genomes

As noted above, cytogenetic and copy number studies revealed that the number of genomic rearrangements and copy number changes can differ markedly between individual cancers. Until the recent advent of systematic sequencing studies, however, we had little insight into the numbers of somatic base substitutions and indels and the extent of their variation.

We now know that there are usually between 1000 and 10,000 somatic substitutions in the genomes of most adult cancers, including breast, ovary, colorectal, pancreas, and glioma (10, 21). There are cancer types that generally carry relatively few mutations—for example, medulloblastomas, testicular germ cell tumors, acute leukemias, and carcinoids (10, 16)—whereas others, such as lung cancers and melanomas, have many more mutations (occasionally more than 100,000) (9, 10, 27, 39, 40). Even within a particular cancer type, individual tumors often display wide variation in the prevalence of base substitutions.

Two major factors account for these differences in mutation prevalence: differences between individual cancers in mutation rate at the cell divisions that have taken place between the fertilized egg and the cancer cell and differences in the number of mitoses in this lineage. The basis for the high prevalence of somatic substitutions observed in some cancers is likely to be overwhelming mutagenic exposure such as ultraviolet (UV) light (in melanoma) or tobacco carcinogens (in lung cancer); the presence of defective DNA repair mechanisms (e.g., in colorectal, stomach, and other cancers with defective DNA mismatch repair); and therapy with DNA-damaging agents (e.g., in gliomas treated with the alkylating agent temozolomide) (10, 11). However, there are individual cancer cases in which the large number of base substitution mutations remains unexplained (18).

The reason that some cancer types have relatively few mutations is not completely clear. Some are tumors of children or young adults, and therefore it is conceivable that the neoplastic cell has been through relatively few DNA replications. Alternatively, it may be that most cancers, including those with the typical mutation prevalence, have experienced an elevated mutation rate compared to normal cells and that cancers with low mutation prevalence are the exceptions that have evolved without it. Because we currently know little about the prevalence of somatic base

---

**Fig. 1.** Time line showing key events in the investigation of the cancer genome.
substitutions in normal cells, the importance of an elevated base substitution mutation rate in cancer development remains controversial (44, 45). However, for indels in cancer cases with DNA mismatch-repair deficiency and for rearrangements and copy number changes in cancers that have large numbers of these changes, an increased mutation rate is likely to have been operative.

Systematic sequencing studies have also provided our first comprehensive insights into the proportions of driver and passenger mutations. Thus far, the large majority of base substitutions in most cancer genomes appear to be passenger (10, 21). However, these studies also suggest that there may be many more drivers than can be unambiguously identified by current approaches. If the latter interpretation is correct, a substantial number of cancer genes remain to be discovered, albeit many contributing infrequently to cancer development (10, 21).

The Repertoire of Human Cancer Genes

Studies of driver mutations in cancer genes have yielded many insights into the molecular and cellular events that convert a healthy cell into a cancer cell. In recent years, the proteins altered by driver mutations have become targets for successful anticancer drug development (46). Identification of new mutated cancer genes is, therefore, one of the most important deliverables that emanates from exploration of cancer genomes.

The primary analytic approach to the identification of driver mutations and cancer genes has assumed that passenger mutations are randomly distributed throughout the genome, whereas drivers (by definition) are clustered in the subset of genes that are cancer genes. The strategy is thus to search in a large number of samples of a specific cancer type, for genes that have a higher prevalence of somatic mutations than would be expected by chance alone, followed by verification of biological activity in experimental systems. This basic strategy has been highly effective over decades and remains the mainstay of cancer gene discovery today. However, it has become clear that passenger mutations are not always randomly distributed in the genome; their clustering can mimic that of driver changes, and thus additional filtering strategies may sometimes be required to avoid errors in cancer gene identification (47).

The search for cancer genes through systematic exploration of cancer genomes by cytogenetics, copy number analyses, and sequencing has been supplemented by targeted somatic mutational analyses of genes previously identified as cancer susceptibility genes, by mutational analyses of biologically plausible candidates, and by biological assays of transforming activity, notably DNA transfection through NIH3T3 cells. Collectively, these varied approaches have identified ~400 somatically mutated cancer genes that contribute to neoplastic change in one or more types of cancer. This number corresponds to roughly 2% of the protein-coding genes in the human genome (5, 6) (www.sanger.ac.uk/genetics/CGP/Census/).

Cancer genes are often classified according to whether they function in a dominant or recessive manner at the level of the cancer cell. Dominant cancer genes require only one of the two parental alleles present in a normal cell to be mutated, and the encoded protein is usually constitutively activated by the mutations. Recessive cancer genes (also known as tumor suppressor genes) require mutation of both parental alleles, and these usually result in inactivation of the encoded protein. More than 80% of the currently known cancer genes are dominantly acting; mostly these are genes that are rearranged in the myriad recurrent chromosomal translocations particularly found in leukemias, lymphomas, and sarcomas (www.sanger.ac.uk/genetics/CGP/Census/). The current predominance of dominantly acting cancer genes is in part due to ascertainment bias, and the real balance remains to be determined.

Most of the known cancer genes were found through primary cytogenetic analyses, with the wave of ever higher resolution copy number studies bringing a further substantial yield. Recent systematic sequencing of cancer genomes has provided a new harvest of cancer genes identified directly through an elevated prevalence of base substitutions and small indels. These include several dominant cancer genes, such as BRAF, EGFR, ERBB2, PIK3CA, IDH1, IDH2, EZH2, FOXL2, PPP2R1A, and JAK2 (8, 12–15, 17, 34, 36, 48, 49) (some of which were also found by alternative approaches). Some are on biological pathways previously implicated in cancer development. Others—for example, IDH1, which encodes isocitrate dehydrogenase 1, a component of the Krebs cycle; or FOXL2, which encodes a tissue-specific transcription factor—would not have featured on many candidate gene lists.

Several recessive cancer genes (and others for which the dominant or recessive status is unclear) have also emerged through systematic sequencing, including SETD2, KDM6A, KDM5C, PBRM1, BAP1, ARID1A, DNMT3A, GATA3, DAXX, ATRX, and MLL2 (7, 12, 16, 19, 20, 30–32, 35, 50). Many of the proteins encoded by this set of genes (and, in addition, EZH2 and IDH1 among the dominant cancer genes mentioned above) are involved in chromatin modification and remodeling. For example, SETD2, EZH2, and MLL2 are histone H3 methylationases, whereas KDM6A and KDM5C are histone H3 demethylases. ARID1A, PBRM1, BAP1, ATRX, and DAXX are components of protein complexes that restructure chromatin, and DNMT3A is involved in maintenance of cytosine methylation in DNA. Although this sector of cell biology was previously known to be disrupted through mutation in some cancers, these discoveries have placed new emphasis on its role in a range of adult and childhood solid tumor types and highlighted a potentially important link between somatic mutation and epigenetic changes that are present in many cancers. This area of biology promises to be a major focus of activity in the development of new cancer therapeutics.

Some newly discovered cancer genes—for example, BRAF, JAK2, ARID1A, EZH2, BAP1, PBRM1, and DNMT3A—are mutated in a substantial proportion of cases of a particular cancer type. Others, such as SETD2 and KDM5C, are mutated in only a small fraction of cancers of any class. This appears to be an emerging feature of the landscape of somatically mutated cancer genes, of which a relatively limited set are commonly mutated and a substantial number mutated infrequently. From the standpoint of novel drug discovery, the latter presents obvious challenges.

An important perspective on the evolution of the cancer clone is provided by the number of mutated cancer genes required to generate an individual human cancer. It is often speculated that five mutated cancer genes are necessary (51). However, higher estimates have been suggested, and for some hematopoietic neoplasms fewer may be required. The presence of two to four driver mutations has been demonstrated in many cases of various cancer types. In a few years, we will be able to estimate this core metric of cancer biology, and the extent to which it varies, more accurately. Once large numbers of cancer genomes have been completely sequenced with all classes of somatic mutation harvested and once most cancer genes have been identified, robust direct assessments of the number of mutated cancer genes in individual cancers will become achievable.

The Cancer Genome and Drug Discovery

The central role of mutated cancer genes in the genesis and maintenance of cancer clones renders them potential “Achilles’ heels” to be exploited for drug discovery. There are now several celebrated examples of anticancer drugs that act by inhibiting the aberrantly activated proteins encoded by mutated cancer genes (46). A paradigm of such strategies is the development of imatinib and subsequent generations of small-molecule inhibitors of the constitutively activated ABL kinase engendered by the chromosome 9:22 translocation in chronic myeloid leukemia (CML) (52). This advance has transformed the treatment of CML and, on the way, has helped to revolutionize cancer therapeutics. Small-molecule drugs against mutated versions of EGFR, ERBB2, KIT, PDGFRα, PML-RARA, MET, and ALK are either already in clinical use or being evaluated in clinical trials (46, 53). Similarly, a therapeutic antibody (trastuzumab) directed against HER2, the protein encoded by a gene amplified in about 2% of breast cancers, has had a major impact on treatment of these cancers (54).

An illustrative example of the combined power of modern genomics, biology, and drug discovery is that of BRAF. Somatic mutations of BRAF were discovered in an early systematic sequenc-
Cancer Crusade at 40

ing screen in 2002 (8). BRAF encodes a serine-threonine kinase and is mutated in 50 to 70% of malignant melanomas, 10 to 15% of colorectal cancers, 50% of papillary thyroid cancers, and at a lower frequency in other cancer types. A single mutation, V600E (substitution of valine 600 with glutamic acid), accounts for more than 90% of mutations and results in constitutive activation of the BRAF kinase. Being a kinase (with a deep ATP-binding pocket in which inhibitors can sit) and being activated by its mutations made mutated BRAF an attractive target for drug development. Inhibitors of V600E mutant BRAF have been tested in phase 1 trials and have produced encouraging responses in 80% of patients with metastatic malignant melanomas carrying the V600E mutation (55).

Unfortunately, minor subclones resistant to BRAF inhibitors appear to be present in many V600E-positive malignant melanomas, and these grow out as recurrences. Nevertheless, investigation into the genomes of recurrences has already identified some of the mutations that confer resistance, proffering new avenues for therapeutic intervention (56, 57). Thus, in the decade since the discovery of BRAF as a mutated cancer gene, the field has seen small-molecule inhibitors identified and developed into orally available drugs, the drugs put through clinical trials and shown to have anticancer activity, and mechanisms of resistance to the drugs elucidated. Although we collectively aspire to even more rapid progress in the future, this is a remarkable achievement.

In some cancers, direct targeting and inhibition of constitutively activated proteins encoded by mutated cancer genes may not be possible. For example, in clear cell renal cancer, mutated and activated kinases have not been found. Indeed, all the operative cancer genes appear to be recessive (7, 20, 31). Because the proteins encoded by these genes are already inactivated by their mutations, other strategies—for example, the development of drugs that exhibit synthetic lethality with particular mutated cancer genes (58)—will have to be adopted.

Genomic Evidence of Mutagenic and Repair Processes

The patterns of somatic mutation found in a cancer genome reflect the DNA damage and mutagenic processes that have been operative and the repair mechanisms that have mitigated their impact. Thus, the cancer genome can be likened to an archaeological record bearing the imprint of these processes. The mutational patterns (often called mutational spectra) incorporate many types of information, including the numbers of each class of mutation, the DNA sequences around each mutated base, and, in transcribed regions, whether the transcribed or the untranscribed strand is preferentially mutated.

In the past, mutational spectra have been assembled with the use of mutations found in frequently mutated cancer genes, notably the tumor suppressor gene TP53. In such studies, each informative cancer case usually contributes a single mutation, and the spectrum is established by grouping together mutations from multiple cases of the same cancer type. This approach demonstrated that lung cancers exhibit many C>G>A:T transversion mutations, a pattern similar to that induced in experimental systems by tobacco carcinogens; that hepatocellular cancers also show C>G>A:T mutations that are likely to be induced by aflatoxins, known etiological agents in liver cancer development; and that skin cancers predominantly show C>G>A:T mutations of the pattern known to be caused by UV light (59). However, a major limitation of these studies is that mutational spectra generated in this way are composites of all the mutational spectra present in a tumor class. Thus, although well powered to report a strong exposure that dominates a particular cancer type, they cannot untangle the diverse mutational processes and patterns that may be present in some cancer types.

By contrast, partial or complete catalogs of mutations from individual cancer genomes, which usually number several thousand somatic mutations per case, can report with extraordinary resolution the mutational spectra of individual cancers, thus revealing the diverse mutational and repair processes operative within a class of cancer and even within individual cases. This type of analysis is in its infancy, but some examples illustrate its potential. Early systematic sequencing studies revealed the presence in some breast cancers of a mutational process characterized by C>T and C>G mutations that occur almost exclusively at cytosines that follow a thymine [i.e., at TpC dimucleotides (78)]. The nature of the mutagenic process underlying this pattern of mutations remains mysterious, but future epidemiological studies correlating its presence with exogenous exposures and studies replicating the pattern by examining the effects of chemicals or DNA repair defects in experimental systems may elucidate its origin. In the case of melanoma and lung cancers, the use of essentially complete catalogs of thousands of mutations from the genomes revealed the predominant mutational classes expected of the known exposures underlying these cancers (39, 40). Although UV exposure accounted for most mutations in melanoma, there was evidence for at least one additional mutational process, the spectrum of which suggested that it may have been due to reactive oxygen species (39). Similarly, by examining the DNA sequences around somatically mutated bases in a case of lung cancer, it was possible to tease out multiple distinct mutational processes that may reflect the complexity of the carcinogen mixture present in cigarette smoke (40).

Traces of DNA repair processes are also embedded in mutational spectra. For example, in individual melanoma and lung cancer cases, evidence has been found for past activity of transcription-coupled repair, a subclass of nucleotide excision repair that is directed at the transcribed strand of each gene (38–40). These completely sequenced cancer genomes also revealed that nucleotide excision repair had been preferentially deployed to the untranscribed strand of genes, that repair correlated with the expression level of the target gene, and that 5′ ends of genes had been more effectively repaired than 3′ ends.

Sequencing of cancer genomes has revealed unexpected features of mutational processes beyond those that cause base substitutions. For example, some cancers display many more genomic rearrangements than would have been predicted on the basis of cytogenetic studies (60). Indeed, different types of rearrangement architecture predominate in different cancer types. In some breast cancers, for example, there are frequent tandem duplications of DNA (60), whereas in pancreatic cancers this pattern is rare (61). The genetic defects, or possibly environmental exposures, that underlie these distinctive patterns of genomic rearrangement are unknown.

Insights have also emerged with respect to the timing of mutations. In principle, some mutational processes may cause steady accumulation of mutations over decades, whereas others may be characterized by a sharp burst over a short period. A small proportion of cancer genomes exhibit a distinctive pattern characterized by extraordinary numbers of rearrangements localized to a small segment of the genome. In such regions, the genome appears to have been shattered and subsequently reassembled by the cell, albeit in a disordered manner. The structure of these dense aggregates suggests that the rear-

"The cancer genome can be likened to an archaeological record bearing the imprint of [mutagenic and DNA repair] processes."
rangements occurred more or less synchronously, possibly within a single catastrophic cell cycle, rather than in a serial manner over many cell divisions (62).

Revealing the Tree of Clonal Evolution in Cancer

Although each cancer derives from a single normal cell, the population of neoplastic cells constituting the final cancer often has a complex evolutionary history, which can be visualized in the following way. Multiple waves of clonal expansion are thought to be required, each brought about by an additional driver mutation, to generate the dominant subclone that manifests as the symptomatic cancer. Along the way, additional branching subclones with further drivers may have been spun off that failed to outcompete the dominant subclone. The dominant subclone itself may have spawned a further minor subclone with an additional driver mutation that in time would dominate. Some of these minor subclones may have been completely extinguished, but others may persist. Thus, the final population of cancer cells is composed of the dominant subclone accompanied by relics of its evolutionary past, outstripped rivals, and portents of its future.

Somatic mutations acquired by cancer cells as they divide can serve as markers of clonal origin and thus allow retrospective reconstruction of the evolutionary tree of individual cancers. These analyses have revealed the complex subclonal structure of certain cancers and have demonstrated that minor subclones at initial presentation of the cancer are often the source of the major clone that recurs after treatment (63–66).

Metastases have been analyzed by similar methods, and these studies indicate that they are usually subclones of the primary cancer. Comparison of somatic changes in metastases to those of the primary tumors from which they originated has revealed that many likely passed through a clonal bottleneck, continued to acquire somatic changes, and diverged from the primary cancer (37, 41, 61, 67). Comparing genomic changes in different metastases from the same patient has yielded provocative insights. For example, in some cases multiple distinct metastases apparently originated from the same minor subclone of the primary tumor, suggesting that this subclone possessed enhanced metastatic potential compared to the bulk of the primary tumor (37, 61). Unexpected relationships between metastases have also surfaced. For example, in some pancreatic cancers metastatic to the lung and to the abdomen, the lung metastases shared a set of somatic changes with each other, and the abdominal metastases shared a different set (in addition to the mutations both groups of metastases shared with the primary cancer). These results suggest that, rather than each metastasis being a direct offshoot of the primary cancer, a single seed of pancreatic cancer rapidly expanding area of cancer diagnostics that seems certain to be integrated into future clinical practice.

There are additional ways in which knowledge of the cancer genome can potentially be used to improve patient care. Many cancers leak DNA into the circulation as cells die. Detection of somatic changes present in the cancer genome can, in principle, distinguish circulating DNA originating from the cancer from circulating DNA derived from normal cells. Such tests would potentially allow monitoring of tumor burden from measurements on blood samples and might have utility in a variety of circumstances, including evaluation of response to treatment and early detection of recurrence. Similar approaches have been used for many years to monitor disease burden in several types of leukemia. In these diseases, they have been applicable because of the presence of common recurrent driver rearrangements (translocations) that allow design of PCR assays across the rearrangement junction; these assays serve as sensitive and specific tests that are straightforward to implement clinically. This mode of diagnostic has not generally been employed in solid tumors, in part because there are relatively few examples of common recurrent rearrangements. However, most solid tumors do carry multiple passenger rearrangements that are specific to each individual cancer. The possibil-

![Cancer genome analysis](https://www.sciencemag.org/content/specialsections/331/6121/1557/F2.large.jpg)

**Fig. 2.** Cancer genome analysis is expected to have a far-reaching impact on our understanding of cancer biology and will likely prompt new approaches to the detection, diagnosis, treatment, and possibly prevention of the disease.
approaches, its utility in clinical practice will ultimately depend on its sensitivity, false-positive rate, and impact on mortality.

The Future
The march toward exhaustive sequencing of cancer genomes across the range of tumor classes is now under way. This global enterprise is being conducted under the auspices of the International Cancer Genome Consortium (70) and currently includes large-scale sequencing initiatives in Australia, Canada, China, France, Germany, India, Italy, Japan, Mexico, Spain, the United Kingdom, and the United States. Sequencing of several hundred cases of each major cancer subtype is envisaged to provide sufficient statistical power to detect cancer genes that may be operative in only 5% of cases. Over the next 5 to 7 years, it is realistically anticipated that tens of thousands of cancer genomes will be sequenced. Initially, there will continue to be a diversity of approaches, with some studies sequencing the DNA of exons of protein-coding genes while others continue to analyze transcriptomes. It is likely, however, that these large-scale initiatives will ultimately converge on whole-genome sequencing, coupled to the exploration of the transcriptome and epigenome from the same cases. This convergence will be encouraged by the falling cost of whole-genome sequencing; by the convenience of harvesting all known exposures or epidemiological risk factors, from early preinvasive lesions, from metastases, from disease pathogenesis, progression, and mechanisms of drug resistance.

These studies will also establish a new, comprehensive, and biologically rational classification of human cancer based on genomic abnormalities. As with any new classification, it will be necessary to evaluate, in a further wave of research, the ability of this genomic classification scheme to predict the key features of tumor behavior of most concern to us, notably progression and response to therapeutics. We already know that the presence of certain mutated genes determines response to some therapies, and mutational testing of specific genes has already been introduced into some clinical trials, particularly when the mutated gene is the target of the new therapy being evaluated. However, with an essentially complete set of cancer genes to be revealed in a few years, and the plausibility that some are likely to influence the clinical behavior of cancer, the ultimate goal should be to examine the prognostic and predictive effectiveness of all mutated cancer genes present in each cancer type, much as similar waves of research in the past have correlated cancer behavior with clinical parameters, pathology, or specific biomarkers. In principle, this assessment should be implemented systematically for both existing and new patient treatment protocols.

What sort of test design could accomplish this? Each cancer type is driven by different, although often overlapping, sets of mutated cancer genes. Thus, customized tests for each cancer class might be an option. However, a single test that could be applied to all types of cancer and access all the relevant information in each type is an especially attractive prospect. The complete catalog of somatic mutations provided by the sequence of the cancer genome fits that description. Although currently expensive for routine implementation, it is unlikely to remain so for long, and the costs of performing a cancer genome sequence in 10 years will be insignificant compared to other aspects of conducting clinical trials. Thus, a full cancer genome sequence may well turn out to be a pragmatic test design for this next phase of research. One should not, however, underestimate the technical, scientific, and analytic challenges intrinsic to this proposal. Moreover, such a test is unlikely to replace all other intrinsic predictors of cancer behavior. Nevertheless, given the rich seam of information that we know is buried in each cancer genome, the extraordinary pace of technological advance in sequencing, and the practical advantage of using a single test in diverse clinical contexts (including many outside oncology), it seems reasonable to look forward to a time in the not-so-distant future when we will consider a cancer genome sequence as a routine adjunct in clinical trials and a test we will perform on many newly diagnosed cancers.

References and Notes