

DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology: Principles & Practice of Oncology

ELEVENTH EDITION

Publishing November 2018



Activate your eBook

Updated Quarterly

DeVita, Hellman, and Rosenberg's **Cancer** Principles & Practice of Oncology **11th edition**

Vincent T. DeVita, Jr.
Theodore S. Lawrence
Steven A. Rosenberg



Wolters Kluwer

PREVIEW: FIRST 10 PAGES OF CHAPTER 1

Visit www.lww.co.uk
for more information



Wolters Kluwer

When you have to be right

DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology: Principles & Practice of Oncology

ELEVENTH EDITION

By Vincent DeVita, MD and Theodore Lawrence, MD, PhD and Steven Rosenberg, MD, PhD

ISBN 978-1-4963-9463-7

Pages 2432

Publishing November 2018 [Preview: First 10 Pages of Chapter 1](#)

Features include:

This unique, continually updated reference remains the gold standard reference in the field

Describes the interrelationship of other imaging technologies with echocardiography and the role of echo as part of a multimodality approach to patients.

Contains more than 1,600 high-quality illustrations – 600 in full colour.

Covers the use of 3D echocardiography and perfusion imaging; current AHA/ACC guidelines, appropriate use guidelines, and the mechanics and utility of strain and strain rate imaging.

Keeps you up to date in this fast-changing field. Every quarter, your eBook will be updated with late-breaking developments in oncology, including new drugs, clinical trials, and more.



Table of contents

Part I BIOLOGY OF CANCER

Part II ETIOLOGY AND EPIDEMIOLOGY OF CANCER

Section 1 Etiology of Cancer

Part III ETIOLOGY AND EPIDEMIOLOGY OF CANCER

Part IV. CANCER PREVENTION AND SCREENING

Part V. PRACTICE OF ONCOLOGY

Section 1 Cancer of the Head and Neck

Section 2 Cancer of the Thoracic Cavity

Section 3 Cancers of the Gastrointestinal Tract

Section 4 Cancers of the Genitourinary System

Section 5 Gynecologic Cancers

Section 6 Cancer of the Breast

Section 7 Cancer of the Endocrine System

Section 8 Sarcomas of Soft Tissue and Bone

Section 9 Cancers of the Skin

Section 10 Neoplasms of the Central Nervous System

Section 11 Cancers in Adolescents and Young Adults

Section 12 Lymphomas in Adults

Section 13 Leukemias and Plasma Cell Tumors

Section 14 Other Cancers

Section 15 Oncologic Emergencies

Section 16 Treatment of Metastatic Cancer

Section 17 Stem Cell Transplantation

Section 18 Management of Adverse Effects of Treatment

PART VI. PALLIATIVE AND ALTERNATIVE CARE

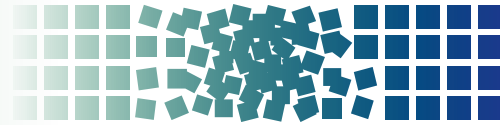
Section 1 Supportive Care and Quality of Life

Section 2 Complementary, Alternative, and Integrative Therapies

PART I

Molecular Biology of Cancer

1 The Cancer Genome



Yardena Samuels, Alberto Bardelli, Yochai Wolf, and Carlos López-Otin

INTRODUCTION

There is a broad consensus that cancer is, in essence, a genetic disease and that accumulation of molecular alterations in the genome of somatic cells is the basis of cancer progression (Fig. 1.1).¹ In the past 10 years, the availability of the human genome sequence and progress in DNA sequencing technologies has dramatically improved knowledge of this disease. These new insights are transforming the field of oncology at multiple levels:

1. The genomic maps are redesigning the tumor taxonomy by moving it from a histologic- to a genetic-based level.
2. The success of cancer drugs designed to target the molecular alterations underlying tumorigenesis has proven that somatic genetic alterations are legitimate targets for therapy.
3. Tumor genotyping is helping clinicians individualize treatments by matching patients with the best treatment for their tumors.
4. Tumor-specific DNA alterations represent highly sensitive biomarkers for disease detection and monitoring.
5. Finally, the ongoing analyses of multiple cancer genomes will identify additional targets, whose pharmacologic exploitation will undoubtedly result in new therapeutic approaches.

This chapter reviews the progress that has been made in understanding the genetic basis of sporadic cancers. An emphasis is placed on an introduction to novel integrated genomic approaches that allow a comprehensive and systematic evaluation of genetic alterations that occur during the progression of cancer. Using these powerful tools, cancer research, diagnosis, and treatment are poised for a transformation in the next years.

CANCER GENES AND THEIR MUTATIONS

Cancer genes are broadly grouped into oncogenes and tumor suppressor genes. Using a classical analogy, oncogenes can be compared to a car accelerator so that a mutation in an oncogene would be the equivalent of having the accelerator continuously pressed.² Tumor suppressor genes, in contrast, act as brakes² so that when they are not mutated, they function to inhibit tumorigenesis. Oncogene and tumor suppressor genes may be classified by the nature of their somatic mutations in tumors. Mutations in oncogenes typically occur at specific hotspots, often affecting the same codon or clustered at neighboring codons in different tumors.¹ Furthermore, mutations in oncogenes are almost always missense, and the mutations usually affect only one allele, making them heterozygous. In contrast, tumor suppressor genes are usually mutated throughout the gene; a large number of the mutations may truncate the encoded protein and generally affect both alleles, causing loss of heterozygosity (LOH). Major types of somatic mutations present in malignant tumors include nucleotide substitutions, small insertions and deletions (*indels*), chromosomal rearrangements, and copy number alterations.

IDENTIFICATION OF CANCER GENES

The completion of the Human Genome Project marked a new era in biomedical sciences.³ Knowledge of the sequence and organization of the human genome now allows for the systematic analysis of the genetic alterations underlying the origin and evolution of tumors. Before elucidation of the human genome, several cancer genes, such as *KRAS*, *TP53*, and *APC*, were successfully discovered using approaches based on an oncovirus analysis, linkage studies, LOH, and cytogenetics.^{4,5} The first curated version of the Human Genome Project was released in 2004³ and provided a sequence-based map of the normal human genome. This information, together with the construction of the HapMap, which contains single nucleotide polymorphisms, and the underlying genomic structure of natural human genomic variation,^{6,7} allowed an extraordinary throughput in cataloging somatic mutations in cancer. These projects now offer an unprecedented opportunity: the identification of all the genetic changes associated with a human cancer. For the first time, this ambitious goal is within reach of the scientific community. Already, a number of studies have demonstrated the usefulness of strategies aimed at the systematic identification of somatic mutations associated with cancer progression. Notably, the Human Genome Project, the HapMap project, as well as the candidate and family gene approaches (described in the following paragraphs), utilized capillary-based DNA sequencing (first-generation sequencing, also known as Sanger sequencing).⁸ Figure 1.2 clearly illustrates the developments in the search of cancer genes, its increased pace, as well as the most relevant findings in this field.

Cancer Gene Discovery by Sequencing Candidate Gene Families

The availability of the human genome sequence provides new opportunities to comprehensively search for somatic mutations in cancer on a larger scale than previously possible. Progress in the field has been closely linked to improvements in the throughput of DNA analysis and in the continuous reduction in sequencing costs. What follows are some of the achievements in this research area as well as how they affected knowledge of the cancer genome.

A seminal work in the field was the systematic mutational profiling of the genes involved in the RAS-RAF pathway in multiple tumors. This candidate gene approach led to the discovery that *BRAF* is frequently mutated in melanomas and is mutated at a lower frequency in other tumor types.⁹ Follow-up studies quickly revealed that mutations in *BRAF* are mutually exclusive with alterations in *KRAS*,^{9,10} genetically emphasizing that these genes function in the same pathway, a concept that had been previously demonstrated in lower organisms such as *Caenorhabditis elegans* and *Drosophila melanogaster*.^{11,12}

In 2003, the identification of cancer genes shifted from a candidate gene approach to the mutational analyses of gene families. The first gene families to be completely sequenced were those that

A metastatic cancer genome requires decades to develop

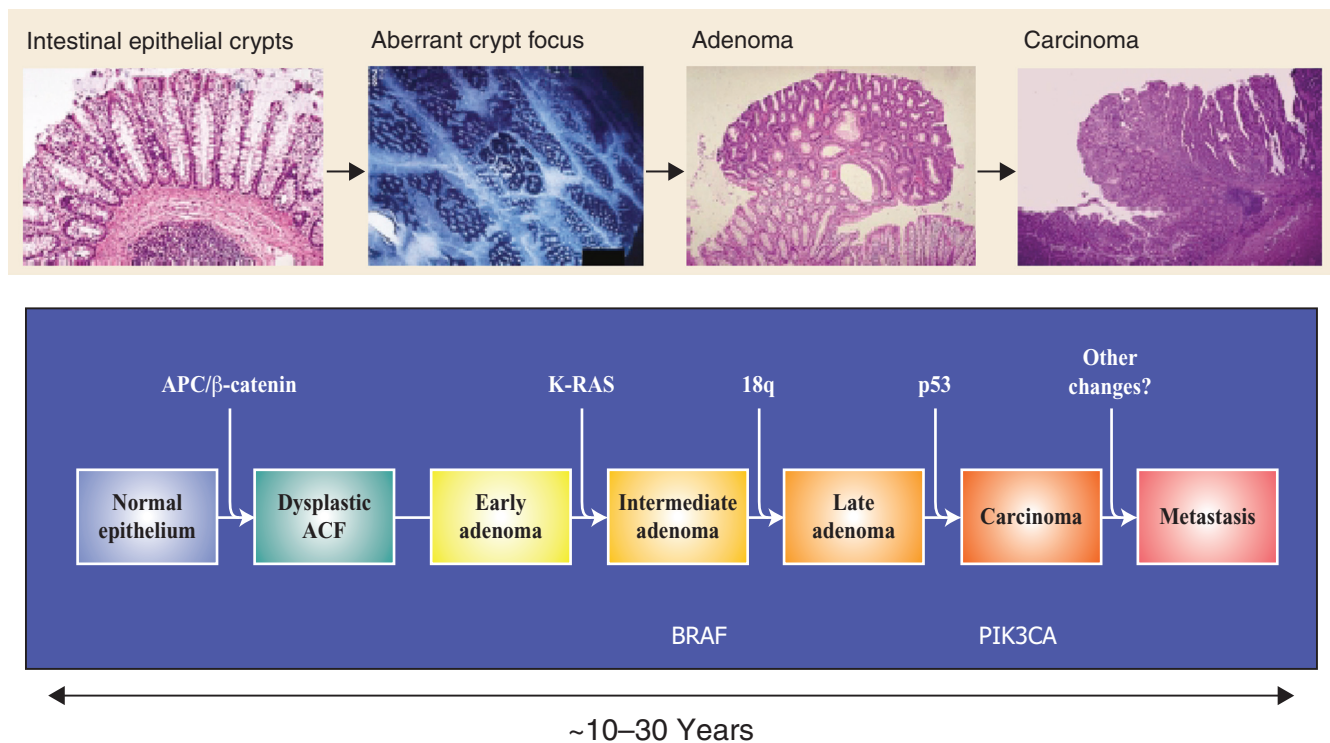


Figure 1.1 Schematic representation of the genomic and histopathologic steps associated with tumor progression: from the occurrence of the initiating mutation in the founder cell to metastasis formation. It has been convincingly shown that the genomic landscape of solid tumors such as that of pancreatic and colorectal tumors requires the accumulation of many genetic events, a process that requires decades to complete. This timeline offers an incredible window of opportunity for the early detection, which is often associated with an excellent prognosis, of this disease. ACF, aberrant crypt focus.

involved protein^{13,14} and lipid phosphorylation.¹⁵ The rationale for initially focusing on these gene families was threefold:

- The corresponding proteins were already known at that time to play a pivotal role in the signaling and proliferation of normal and cancerous cells.
- Multiple members of the protein kinases family had already been linked to tumorigenesis.
- Kinases are clearly amenable to pharmacologic inhibition, making them attractive drug targets.

The mutational analysis of all the tyrosine-kinase domains in colorectal cancers (CRCs) revealed that 30% of cases had a mutation in at least one tyrosine-kinase gene, and overall mutations were identified in eight different kinases, most of which had not previously been linked to cancer.¹³ An additional mutational analysis of the coding exons of 518 protein kinase genes in 210 diverse human cancers, including breast, lung, gastric, ovarian, renal, and acute lymphoblastic leukemia, identified approximately 120 mutated genes that probably contribute to oncogenesis.¹⁴ Because kinase activity is attenuated by enzymes that remove phosphate groups called phosphatases, the rational next step in these studies was to perform a mutation analysis of the protein tyrosine phosphatases. A combined analysis of the protein tyrosine kinases and the protein tyrosine phosphatases showed that 50% of CRCs had mutations in a tyrosine-kinase gene, a protein tyrosine phosphatase gene, or both, further emphasizing the pivotal role of protein phosphorylation in neoplastic progression. Many of the identified genes had previously been linked to human cancer, thus validating the unbiased comprehensive mutation profiling. These landmark studies led to additional gene family surveys.

The phosphatidylinositol 3-kinase (PI3K) gene family, which also plays a role in proliferation, adhesion, survival, and motility, was also comprehensively investigated.¹⁶ Sequencing of the exons

encoding the kinase domain of all 16 members belonging to this family pinpointed *PIK3CA* as the only gene to harbor somatic mutations. When the entire coding region was analyzed, *PIK3CA* was found to be somatically mutated in 32% of CRCs. At that time, the *PIK3CA* gene was certainly not a newcomer in the cancer arena because it had previously been shown to be involved in cell transformation and metastasis.¹⁶ Strikingly, its staggeringly high mutation frequency was discovered only through systematic sequencing of the corresponding gene family.¹⁵ Subsequent analysis of *PIK3CA* in other tumor types identified somatic mutations in this gene in additional cancer types, including 36% of hepatocellular carcinomas, 36% of endometrial carcinomas, 25% of breast carcinomas, 15% of anaplastic oligodendrogliomas, 5% of medulloblastomas and anaplastic astrocytomas, and 27% of glioblastomas.^{17–21} It is known that *PIK3CA* is one of the two (the other being *KRAS*) most commonly mutated oncogenes in human cancers. Further investigation of the PI3K pathway in CRC showed that 40% of tumors had genetic alterations in one of the PI3K pathway genes, emphasizing the central role of this pathway in CRC pathogenesis.²²

Although most cancer genome studies of large gene families have focused on the kinome, recent analyses have revealed that members of other families highly represented in the human genome are also a target of mutational events in cancer. This is the case of proteases, a complex group of enzymes consisting of at least 569 components that constitute the so-called human degradome.²³ Functional studies have also revealed that beyond the initial recognition of proteases as prometastatic enzymes, they play dual roles in cancer, as assessed by the identification of a growing number of tumor-suppressive proteases.^{24–27}

These findings emphasized the possibility that mutational activation or inactivation of protease genes occurs in cancer. A systematic analysis of genetic alterations in breast and CRCs revealed that proteases from different catalytic classes were somatically mutated

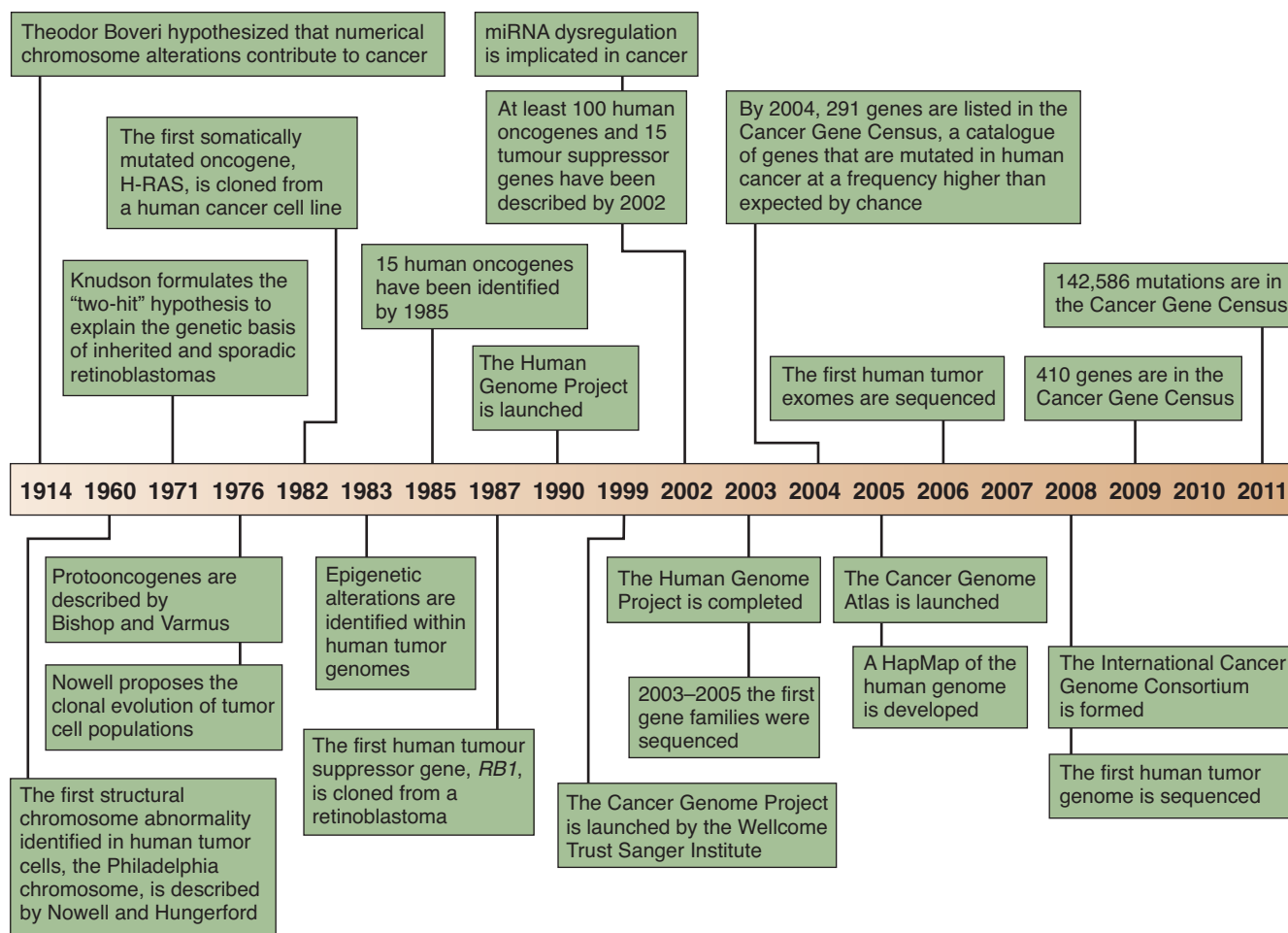


Figure 1.2 Timeline of seminal hypotheses, research discoveries, and research initiatives that have led to an improved understanding of the genetic etiology of human tumorigenesis within the past century. The consensus cancer gene data were obtained from the Wellcome Trust Sanger Institute Cancer Genome Project Web site (<http://www.sanger.ac.uk/genetics/CGP/>). miRNA, microRNA. (Redrawn from Bell DW. Our changing view of the genomic landscape of cancer. *J Pathol* 2010;220:231–243.)

in cancer.²⁸ These results prompted the mutational analysis of entire protease families such as matrix metalloproteinases (MMP), a disintegrin and metalloproteinase (ADAM), and ADAMs with thrombospondin domains (ADAMTS) in different tumors. These studies led to the identification of protease genes frequently mutated in cancer, such as *MMP8*, which is mutated and functionally inactivated in 6.3% of human melanomas.^{29,30}

Mutational Analysis of Exomes Using Sanger Sequencing

Although the gene family approach for the identification of cancer genes has proven extremely valuable, it still is a candidate approach and thus biased in its nature. The next step forward in the mutational profiling of cancer has been the sequencing of exomes, which is the entire coding portion of the human genome (18,000 protein-encoding genes). The exomes of many different tumors—including breast, colorectal, pancreatic, and ovarian clear cell carcinomas; glioblastoma multiforme; and medulloblastoma—have been analyzed using Sanger sequencing. For the first time, these large-scale analyses allowed researchers to describe and understand the genetic complexity of human cancers.^{29,31–34} The declared goals of these exome studies were to provide methods for exomewide mutational analyses in human tumors, to characterize their spectrum and quantity of somatic mutations, and, finally, to discover new genes involved in tumorigenesis as well as novel

pathways that have a role in these tumors. In these studies, sequencing data were complemented with gene expression and copy number analyses, thus providing a comprehensive view of the genetic complexity of human tumors.^{32–35} A number of conclusions can be drawn from these analyses, including the following:

- Cancer genomes have an average of 30 to 100 somatic alterations per tumor in coding regions, which was a higher number than previously thought. Although the alterations included point mutations, small insertions, deletions, or amplifications, the great majority of the mutations observed were single-base substitutions.^{32,35}
- Even within a single cancer type, there is a significant inter-tumor heterogeneity. This means that multiple mutational patterns (encompassing different mutant genes) are present in tumors that cannot be distinguished based on histologic analysis. The concept that individual tumors have a unique genetic milieu is highly relevant for personalized medicine, a concept that will be further discussed.
- The spectrum and nucleotide contexts of mutations differ between different tumor types. For example, over 50% of mutations in CRC were C:G to T:A transitions, and 10% were C:G to G:C transversions. In contrast, in breast cancers, only 35% of the mutations were C:G to T:A transitions, and 29% were C:G to G:C transversions. Knowledge of mutation spectra is vital because it allows insight into the mechanisms underlying mutagenesis and repair in the various cancers investigated.

- A considerably larger number of genes that had not been previously reported to be involved in cancer were found to play a role in the disease.
- Solid tumors arising in children, such as medulloblastomas, harbor on average 5 to 10 times less gene alterations compared to a typical adult solid tumor. These pediatric tumors also harbor fewer amplifications and homozygous deletions within coding genes compared to adult solid tumors.

Importantly, to deal with the large amount of data generated in these genomic projects, it was necessary to develop new statistical and bioinformatic tools. Furthermore, an examination of the overall distribution of the identified mutations allowed for the development of a novel view of cancer genome landscapes and a novel definition of cancer genes. These new concepts in the understanding of cancer genetics are further discussed in the following paragraphs. The compiled conclusions derived from these analyses have led to a paradigm shift in the understanding of cancer genetics.

A clear indication of the power of the unbiased nature of the whole-exome surveys was revealed by the discovery of recurrent mutations in the active site of *IDH1*, a gene with no known link to gliomas, in 12% of tumors analyzed.³⁵ Because malignant gliomas are the most common and lethal tumors of the central nervous system, and because glioblastoma multiforme (GBM; World Health Organization grade IV astrocytoma) is the most biologically aggressive subtype, the unveiling of *IDH1* as a novel GBM gene is extremely significant. Importantly, mutations of *IDH1* predominantly occurred in younger patients and were associated with a better prognosis.³⁶ Follow-up studies showed that mutations of *IDH1* occur early in glioma progression; the R132 somatic mutation is harbored by the majority (greater than 70%) of grades II and III astrocytomas and oligodendrogliomas as well as in secondary GBMs that develop from these lower grade lesions.^{36–42} In contrast, less than 10% of primary GBMs harbor these alterations. Furthermore, analysis of the associated *IDH2* revealed recurrent somatic mutations in the R172 residue, which is the exact analog of the frequently mutated R132 residue of *IDH1*. These mutations occur mostly in a mutually exclusive manner with *IDH1* mutations,^{36,38} suggesting that they have equivalent phenotypic effects. Subsequently, *IDH1* mutations have been reported in additional cancer types, including hematologic neoplasias.^{43–45}

Next-Generation Sequencing and Cancer Genome Analysis

In 1977, the introduction of the Sanger method for DNA sequencing with chain-terminating inhibitors transformed biomedical research.⁸ Over the past 30 years, this first-generation technology has been universally used for elucidating the nucleotide sequence of DNA molecules. However, the launching of new large-scale projects, including those implicating whole-genome sequencing of cancer samples, has made necessary the development of new methods that are widely known as next-generation sequencing technologies.^{46–48} These approaches have significantly lowered the cost and the time required to determine the sequence of the 3×10^9 nucleotides present in the human genome. Moreover, they have a series of advantages over Sanger sequencing, which are of special interest for the analysis of cancer genomes.⁴⁹ First, next-generation sequencing approaches are more sensitive than Sanger methods and can detect somatic mutations even when they are present in only a subset of tumor cells.⁵⁰ Moreover, these new sequencing strategies are quantitative and can be used to simultaneously determine both nucleotide sequence and copy number variations.⁵¹ They can also be coupled to other procedures such as those involving paired-end reads, allowing for the identification of multiple structural alterations, such as insertions, deletions, and rearrangements, that commonly occur

in cancer genomes.⁵⁰ Nonetheless, next-generation sequencing still presents some limitations that are mainly derived from the relatively high error rate in the short reads generated during the sequencing process. In addition, these short reads make the task of de novo assembly of the generated sequences and the mapping of the reads to a reference genome extremely complex. To overcome some of these current limitations, deep coverage of each analyzed genome is required and a careful validation of the identified variants must be performed, typically using Sanger sequencing. As a consequence, there is a substantial increase in both the cost of the process and in the time of analysis. Therefore, it can be concluded that whole-genome sequencing of cancer samples is already a feasible task but not yet a routine process. Further technical improvements will be required before the task of decoding the entire genome of any malignant tumor of any cancer patient can be applied to clinical practice.

The number of next-generation sequencing platforms has substantially grown over the past few years and currently includes technologies from Roche/454, Illumina/Solexa, Life/APG's SOLiD3, Helicos BioSciences/HeliScope, and Pacific Biosciences/PacBio RS.⁴⁸ Noteworthy also are the recent introduction of the Polonator G.007 instrument, an open source platform with freely available software and protocols; the Ion Torrent's semiconductor sequencer; as well as those involving self-assembling DNA nanoballs or nanopore technologies.^{52–54} These new machines are driving the field toward the era of third-generation sequencing, which brings enormous clinical interest because it can substantially increase the speed and accuracy of analyses at reduced costs and can facilitate the possibility of single-molecule sequencing of human genomes. A comparison of next-generation sequencing platforms is shown in Table 1.1. These various platforms differ in the method utilized for template preparation and in the nucleotide sequencing and imaging strategy, which finally result in their different performance. Ultimately, the most suitable approach depends on the specific genome sequencing projects.⁴⁸

Current methods of template preparation first involve randomly shearing genomic DNA into smaller fragments, from which a library of either fragment templates or mate-pair templates are generated. Then, clonally amplified templates from single DNA molecules are prepared by either emulsion polymerase chain reaction (PCR) or solid-phase amplification.^{55,56} Alternatively, it is possible to prepare single-molecule templates through methods that require less starting material and that do not involve PCR amplification reactions, which can be the source of artifactual mutations.⁵⁷ Once prepared, templates are attached to a solid surface in spatially separated sites, allowing thousands to billions of nucleotide sequencing reactions to be performed simultaneously.

The sequencing methods currently used by the different next-generation sequencing platforms are diverse and have been classified into four groups: cyclic reversible termination, single-nucleotide addition, real-time sequencing, and sequencing by ligation (Fig. 1.3).^{48,58} These sequencing strategies are coupled with different imaging methods, including those based on measuring bioluminescent signals or involving four-color imaging of single molecular events. Finally, the extraordinary amount of data released from these nucleotide sequencing platforms is stored, assembled, and analyzed using powerful bioinformatic tools that have been developed in parallel with next-generation sequencing technologies.⁵⁹

Next-generation sequencing approaches represent the newest entry into the cancer genome decoding arena and have already been applied to cancer analyses. The first research group to apply these methodologies to whole cancer genomes was that of Ley et al.,⁶⁰ who reported in 2008 the sequencing of the entire genome of a patient with acute myeloid leukemia (AML) and its comparison with the normal tissue from the same patient, using the Illumina/Solexa platform. As further described, this work allowed for the identification of point mutations and structural alterations

TABLE 1.1
Comparative Analysis of Next-Generation Sequencing Platforms

Platform	Library/ Template Preparation	Sequencing Method	Average Read-Length (Bases)	Run Time (Days)	Gb Per Run	Instrument Cost (U.S. \$)	Comments
Roche 454 GS FLX	Fragment, mate-pair Emulsion PCR	Pyrosequencing	400	0.35	0.45	500,000	Fast run times High reagent cost
Illumina HiSeq 2000	Fragment, mate-pair Solid phase	Reversible terminator	100–125	8 (mate- pair run)	150–200	540,000	Most widely used platform Low multiplexing capability
Life/APG's SOLiD 5500xl	Fragment, mate-pair Emulsion PCR	Cleavable probe, sequencing by ligation	35–75	7 (mate- pair run)	180–300	595,000	Inherent error correction Long run times
Helicos BioSciences HeliScope	Fragment, mate-pair Single molecule	Reversible terminator	32	8 (fragment run)	37	999,000	Nonbias template representation Expensive, high error rates
Pacific Biosciences PacBio RS	Fragment Single molecule	Real-time sequencing	1,000	1	0.075	NA	Greatest potential for long reads Highest error rates
Polonator G.007	Mate pair Emulsion PCR	Noncleavable probe, sequencing by ligation	26	5 (mate- pair run)	12	170,000	Least expensive platform Shortest read lengths

PCR, polymerase chain reaction; NA, not available.
Data represent an update of information provided in Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010;11:31–46.

of putative oncogenic relevance in AML and represented proof of principle of the relevance of next-generation sequencing for cancer research.

Whole-Genome Analysis Utilizing
Second-Generation Sequencing

The sequence of the first whole cancer genome was reported in 2008, where AML and normal skin from the same patient were described.⁶⁰ Numerous additional whole genomes, together with the corresponding normal genomes of patients with a variety of malignant tumors, have been reported since then.^{43,50,61–72}

The first available whole genome of a cytogenetically normal AML subtype M1 (AML-M1) revealed eight genes with novel mutations along with another 500 to 1,000 additional mutations found in non-coding regions of the genome. Most of the identified genes had not been previously associated with cancer. However, validation of the detected mutations did not identify novel recurring mutations in AML.⁶⁰ Concomitantly, with the expansion in the use of next-generation sequencers, many other whole genomes from a number of cancer types started to be evaluated in a similar manner (Fig. 1.4).⁷³

In contrast to the first AML whole genome, the second did observe a recurrent mutation in *IDH1*, encoding isocitrate dehydrogenase.⁴³ Follow-up studies extended this finding and reported that mutations in *IDH1* and the related gene *IDH2* occur at a 20% to 30% frequency in AML patients and are associated with a poor prognosis in some subgroups of patients.^{66,67,74} A good example illustrating the high pace at which second-generation technologies and their accompanying analytical tools are found is demonstrated by the following finding derived from a reanalysis of the first AML whole genome. Thus, when improvements in sequencing techniques were available, the first AML whole genome (described previously), which identified no recurring mutations and had a 91.2% diploid coverage, was reevaluated by deeper sequence coverage, yielding 99.6% diploid coverage of the genome. This improvement, together with more advanced mutation calling algorithms, allowed for the discovery of several nonsynonymous mutations that had not been identified in the initial sequencing. This included a frameshift mutation in the DNA

methyltransferase gene *DNMT3A*. Validation of *DNMT3A* in 280 additional de novo AML patients to define recurring mutations led to the significant discovery that a total of 22.1% of AML cases had mutations in *DNMT3A* that were predicted to affect translation. The median overall survival among patients with *DNMT3A* mutations was significantly shorter than that among patients without such mutations (12.3 months versus 41.1 months; *P* < .001).

Shortly after this study, complete sequences of a series of cancer genomes, together with matched normal genomes of the same patients, were reported.^{43,65,70,75} These works opened the way to more ambitious initiatives, including those involving large international consortia, aimed at decoding the genome of malignant tumors from thousands of cancer patients. Thus, over the last few years, many whole genomes of different human malignancies have been made available.^{61–63}

In addition to direct applications of next-generation sequencing technologies for the mutational analysis of cancer genomes, these methods have an additional range of applications in cancer research. Thus, genome sequencing efforts have begun to elucidate the genomic changes that accompany metastasis evolution through a comparative analysis of primary and metastatic lesions from breast and pancreatic cancer patients.^{64,68,69,71} Likewise, massively parallel sequencing has been used to analyze the evolution of a tongue adenocarcinoma in response to selection by targeted kinase inhibitors.⁷⁶ Detailed information of several of these whole-genome projects is found in the following paragraph.

The first solid cancer to undergo whole-genome sequencing was a malignant melanoma that was compared to a lymphoblastoid cell line from the same individual.⁷⁵ Impressively, a total of 33,345 somatic base substitutions were identified, with 187 nonsynonymous substitutions in protein-coding sequences, at least one order of magnitude higher than any other cancer type. Most somatic base substitutions were C:G > T:A transitions, and of the 510 dinucleotide substitutions, 360 were CC.TT/GG.AA changes, which is consistent with ultraviolet light exposure mutation signatures previously reported in melanoma.¹⁴ (Such results from the most comprehensive catalog of somatic mutations not only provide insight into the DNA damage signature in this cancer type but also is useful in determining the relative order of some acquired mutations.) Indeed, this study shows that a significant correlation exists between

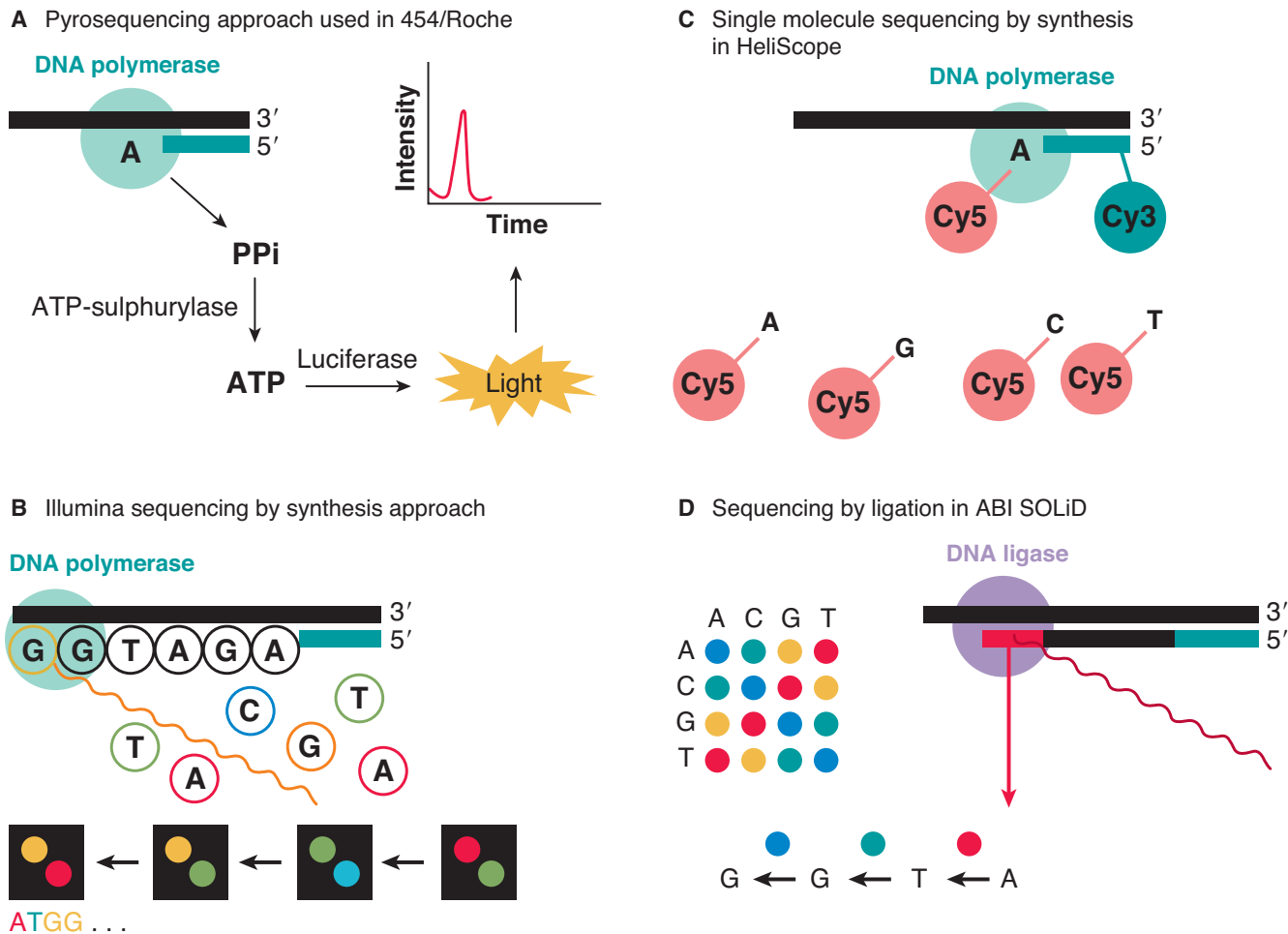


Figure 1.3 Advances in sequencing chemistry implemented in next-generation sequencers. **A:** The pyrosequencing approach implemented in 454/Roche sequencing technology detects incorporated nucleotides by chemiluminescence resulting from PPi release. ATP, adenosine triphosphate. **B:** The Illumina method utilizes sequencing by synthesis in the presence of fluorescently labeled nucleotide analogs that serve as reversible reaction terminators. **C:** The single-molecule sequencing by synthesis approach detects template extension using Cy3 and Cy5 labels attached to the sequencing primer and the incoming nucleotides, respectively. **D:** The sequencing by oligonucleotide ligation and detection (SOLiD) method sequences templates by sequential ligation of labeled degenerate probes. Two-base encoding implemented in the SOLiD instrument allows for probing each nucleotide position twice. ABI, Applied Biosystems. (From Morozova O, Hirst M, Marra MA. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* 2009;10:135–151.)

the presence of a higher proportion of C/A/G/T transitions in early (82%) compared to late mutations (53%). Another important aspect that the comprehensive nature of this melanoma study provided was that cancer mutations are spread out unevenly throughout the genome, with a lower prevalence in regions of transcribed genes, suggesting that DNA repair occurs mainly in these areas.

An interesting and pioneering example of the power of whole-genome sequencing in deciphering the mutation evolution in carcinogenesis was seen in a study in which a basal-like breast cancer tumor, a brain metastasis, a tumor xenograft derived from the primary tumor, and the peripheral blood from the same patient were compared (Fig. 1.5).⁷¹ This analysis showed a clear overlap in mutation incidence between the metastatic and xenograft cases, suggesting that xenografts undergo similar selection as metastatic lesions and, therefore, are a reliable source for genomic analyses. The main conclusion of this whole-genome study was that although metastatic tumors harbor an increased number of genetic alterations, the majority of the alterations found in the primary tumor are preserved. Further studies have confirmed and extended these findings to metastatic tumors from different types, including renal and pancreatic carcinomas.⁷⁷

The importance of performing whole-genome sequencing has also been emphasized by the recent identification of somatic

mutations in regulatory regions, which can also elicit tumorigenesis, such as alterations in the *telomerase reverse transcriptase* (*TERT*) promoter region in melanoma. (In combination, these *TERT* mutations are seen in a greater frequency than *BRAF*- and *NRAS*-activating mutations. They occur in a mutually exclusive manner and in regions that do not show a large background mutation rate, all suggesting that these mutations are important driver events contributing to oncogenesis.)

As the *TERT* promoter mutation discovery shows, regions of the genome that do not code for proteins are just as vital in our understanding of the biology behind tumor development and progression. Another class of non-protein-coding regions in the genome are the noncoding RNAs. One class of noncoding RNAs are microRNAs (miRNAs). Discovered 20 years ago, miRNAs are known to be expressed in a tissue- or developmentally specific manner, and their expression can influence cellular growth and differentiation along with cancer-related pathways such as apoptosis or stress response. miRNAs do this through either overexpression, leading to the targeting and downregulation of tumor suppressor genes, or inversely through their own downregulation, leading to increased expression of their target oncogene. miRNAs have been extensively studied in cancer, and their functional effects have been noted in a wide variety of cancers like glioma⁷⁸ and breast cancer,⁷⁹ to name just two.

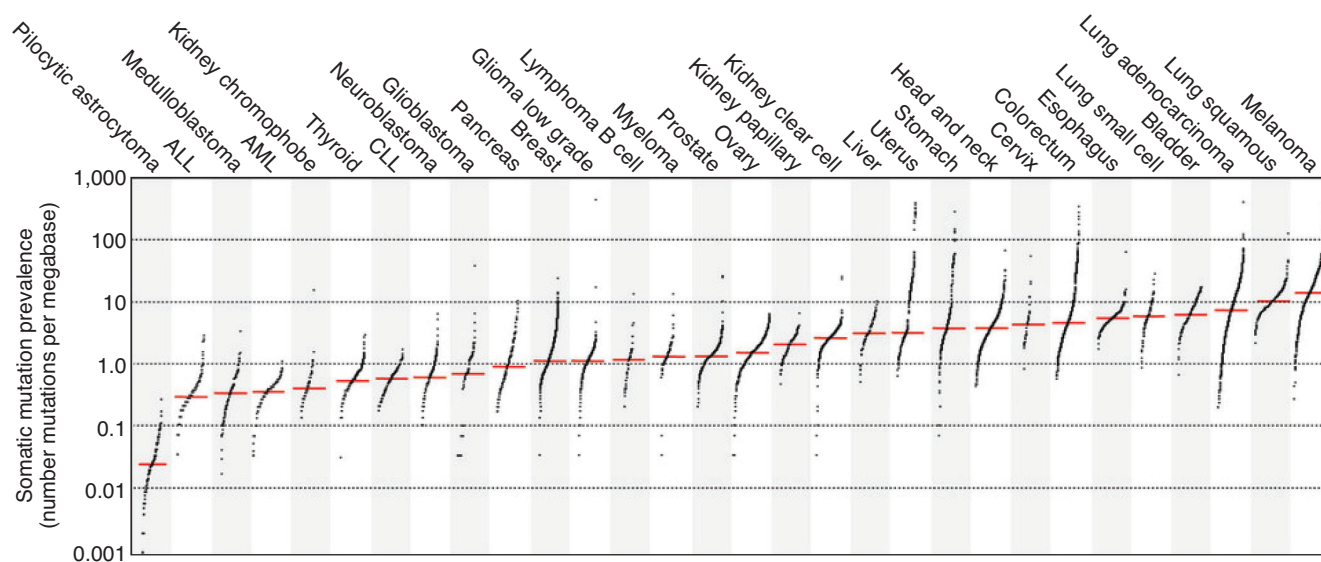


Figure 1.4 The prevalence of somatic mutations across human cancer types. Every *dot* represents a sample, whereas the *red horizontal lines* are the median numbers of mutations in the respective cancer types. The vertical axis (log scaled) shows the number of mutations per megabase, whereas the different cancer types are ordered on the horizontal axis based on their median numbers of somatic mutations. ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia. (Used with permission from Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–421.)

Another class of noncoding RNAs (ncRNA) are the long noncoding RNAs (lncRNA). These RNAs are typically greater than 200 bp and can range up to 100 kb in size. They are transcribed by RNA polymerase II and can undergo splicing and polyadenylation. Although much less extensively studied when compared to miRNAs for their role in cancer, lncRNAs are beginning to come under much more scrutiny. A recent study of the steroid receptor RNA activator revealed two transcripts, an lncRNA (steroid receptor RNA activator) and a translated transcript (steroid receptor RNA activator protein), that coexist within breast cancer cells. However,

their expression varies within breast cancer cell lines with different phenotypes. It was shown that in a more invasive breast cancer line, higher relative levels of the noncoding transcript were seen.⁸⁰ Because this ncRNA acts as part of a ribonucleoprotein complex that is recruited to the promoter region of regulatory genes, it has been hypothesized that this shift in balance between both noncoding and coding transcripts may be associated with growth advantages. When this balance was shifted in vitro, it led to a large increase in transcripts associated with invasion and migration. The results of this study highlight the importance of the investigation

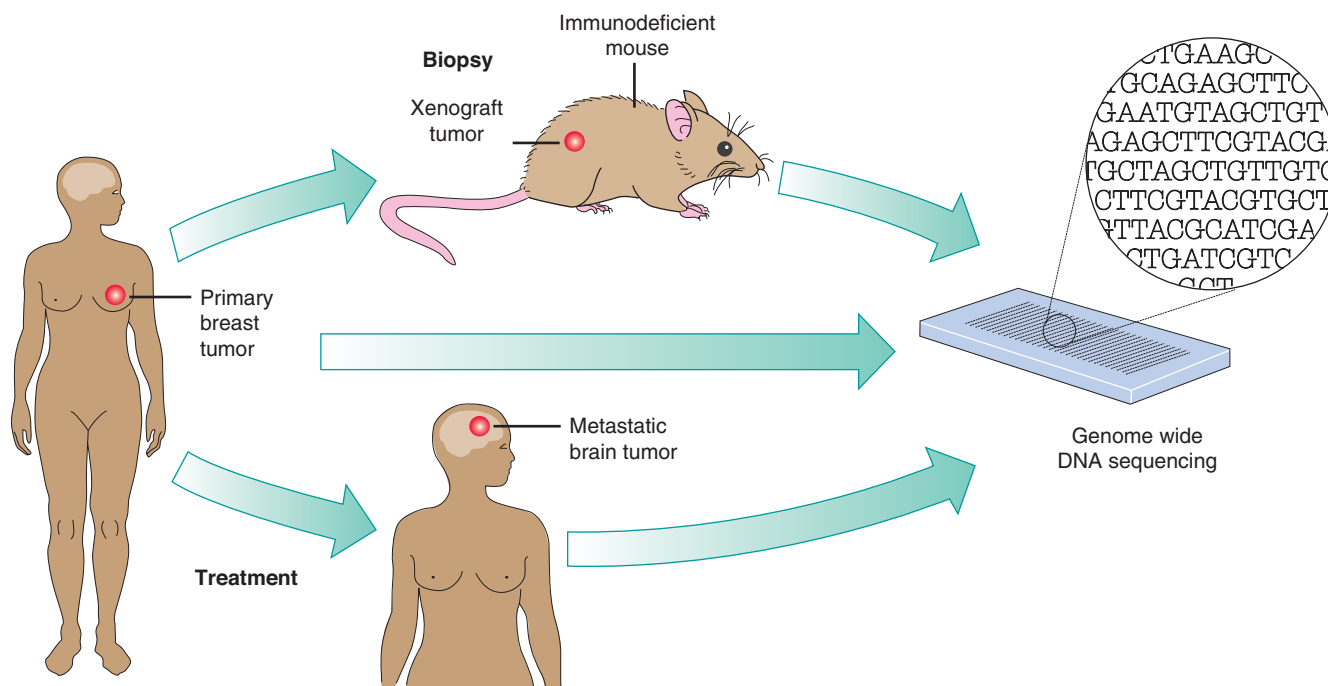


Figure 1.5 Covering all the bases in metastatic assessment. Ding et al.⁷¹ performed a genomewide analysis on three tumor samples: a patient's primary breast tumor; her metastatic brain tumor, which formed despite therapy; and a xenograft tumor in a mouse, originating from the patient's breast tumor. They find that the primary tumor differs from the metastatic and xenograft tumors mainly in the prevalence of genomic mutations. (With permission from Gray J. Cancer: genomics of metastasis. *Nature* 2010;464:989–990.)

into the roles of ncRNA in tumor development or progression and confirm again that the study of coding variants is not sufficient in determining the full genomic spectrum of cancer.

It must be also noted that the recent analysis of whole genomes of many different human tumors has provided additional insights into cancer evolution. Thus, it has been demonstrated that multiple mutational processes are operative during cancer development and progression, each of which has the capacity to leave its particular mutational signature on the genome. A remarkable and innovative study in this regard was aimed at the generation of the entire catalog of somatic mutations in 21 breast carcinomas and the identification of the mutational signatures of the underlying processes. This analysis revealed the occurrence of multiple, distinct single- and double-nucleotide substitution signatures. Moreover, it was reported that breast carcinomas harboring *BRCA1* or *BRCA2* mutations showed a characteristic combination of substitution mutation signatures and a particular profile of genomic deletions. An additional contribution of this analysis was the identification of a distinctive phenomenon of localized hypermutation, which has been termed *kataegis*, and which has also subsequently been observed in other malignancies distinct from breast carcinomas.⁷³

Whole-genome sequencing of human carcinomas has also allowed for the ability to characterize other massive genomic alterations, termed *chromothripsis* and *chromoplexy*, occurring across different cancer subtypes.⁸¹ Chromothripsis implies a massive genomic rearrangement acquired in a one-step catastrophic event during cancer development and has been detected in about 2% to 3% of all tumors but is present at high frequency in some particular cases, such as bone cancers.⁸² Chromoplexy has been originally described in prostate cancer and involves many DNA translocations and deletions that arise in a highly interdependent manner and result in the coordinate disruption of multiple cancer genes.⁸³ These newly described phenomena represent powerful strategies of rapid genome evolution, which may play essential roles during carcinogenesis.

Whole-Exome Analysis Utilizing Second-Generation Sequencing

Another application of second-generation sequencing involves utilizing nucleic acid “baits” to capture regions of interest in the total pool of nucleic acids. These could either be DNA, as described previously,^{84,85} or RNA.⁸⁶ Indeed, most areas of interest in the genome can be targeted, including exons and ncRNAs. Over the last few years, thousands of cancer samples have been subjected to whole-exome sequencing. These studies, combined with data from whole-genome sequencing, have provided an unprecedented level of information about the mutational landscape of the most frequent human malignancies.^{61–63} In addition, whole-exome sequencing has been used to identify the somatic mutations characteristic of both rare tumors and those that are prevalent in certain geographical regions.⁶³

Overall, these studies have provided very valuable information about mutation rates and spectra across cancer types and subtypes.^{73,87,88} Remarkably, the variation in mutational frequency between different tumors is extraordinary, with hematologic and pediatric cancers showing the lowest mutation rates (0.001 per Mb of DNA) and melanoma and lung cancers presenting the highest mutational burden (more than 400 per Mb). Whole-exome sequencing has also contributed to the identification of novel cancer genes that had not been previously described to be causally implicated in the carcinogenesis process. These genes belong to different functional categories, including signal transduction, RNA maturation, metabolic regulation, epigenetics, chromatin remodeling, and protein homeostasis.⁶¹ Finally, a combination of data from whole-exome and whole-genome sequencing has allowed for the identification of the signatures of mutational processes operating in different cancer types.⁷³ Thus, an analysis of a dataset of about 5 million mutations from over 7,000 cancers from

30 different types has allowed for the extraction of more than 20 distinct mutational signatures. Some of them, such as those derived from the activity of APOBEC cytidine deaminases, are present in most cancer types, whereas others are characteristic of specific tumors. Known signatures associated with age, smoking, ultraviolet light exposure, and DNA repair defects have been also identified in this work, but many of the detected mutational signatures are of cryptic origin. These findings demonstrate the impressive diversity of mutational processes underlying cancer development and may have enormous implications for the future understanding of cancer biology, prevention, and treatment.

SOMATIC ALTERATION CLASSES DETECTED BY CANCER GENOME ANALYSIS

Whole-genome sequencing of cancer genomes has an enormous potential to detect all major types of somatic mutations present in malignant tumors. This large repertoire of genomic abnormalities includes single nucleotide changes, small insertions and deletions, large chromosomal reorganizations, and copy number variations (Fig. 1.6).

Nucleotide substitutions are the most frequent somatic mutations detected in malignant tumors, although there is a substantial variability in the mutational frequency among different cancers.⁴⁷ On average, human malignancies have one nucleotide change per million bases, but melanomas reach mutational rates 10-fold higher, and tumors with mutator phenotype caused by DNA mismatch repair deficiencies may accumulate tens of mutations per million nucleotides. By contrast, tumors of hematopoietic origin have less than one base substitution per million. Several bioinformatic tools and pipelines have been developed to efficiently detect somatic nucleotide substitutions through comparison of the genomic information obtained from paired normal and tumor samples from the same patient. Likewise, there are a number of publicly available computational methods to predict the functional relevance of the identified mutations in cancer specimens.⁴⁷ Most of these bioinformatic tools exclusively deal with nucleotide changes in protein coding regions and evaluate the putative structural or functional effect of an amino acid substitution in a determined protein, thus obviating changes in other genomic regions, which can also be of crucial interest in cancer. In any case, current computational methods used in this regard are far from being optimal, and experimental validation is finally required to assess the functional relevance of nucleotide substitutions found in cancer genomes.

For years, the main focus of cancer genome analyses has been on identifying coding mutations that cause a change in the amino acid sequence of a gene. The rationale behind this is quite sound because any mutation that creates a novel protein or truncates an essential protein has the potential to drastically change the cellular environment. Examples of this have been shown earlier in the chapter with *BRAF* and *KRAS* along with many others. With the advancements in next-generation sequencing, larger studies are able to be conducted. These studies give the power to detect mutations occurring in the cancer genome at a lower frequency. Interesting to note is that these studies are leading to the discovery that recurrent synonymous mutations occur in cancer. Previously believed to be merely neutral mutations that maintain no functional role in tumorigenesis, these mutations were largely ignored, but a recent study shows⁸⁹ that simply dismissing these mutations as silent may be premature.

In a review of only 29 melanoma exomes and genomes, 16 recurring synonymous mutations were discovered. When these mutations were screened in additional samples, a synonymous mutation in the gene *BCL2L12* was discovered in 12 out of 285 total samples. The observed frequency of this recurrent mutation is greater than expected by chance, suggesting that it has undergone some type of selective pressure during tumor development.⁸⁹ Noting that *BCL2L12* had previously been linked to tumorigenesis,

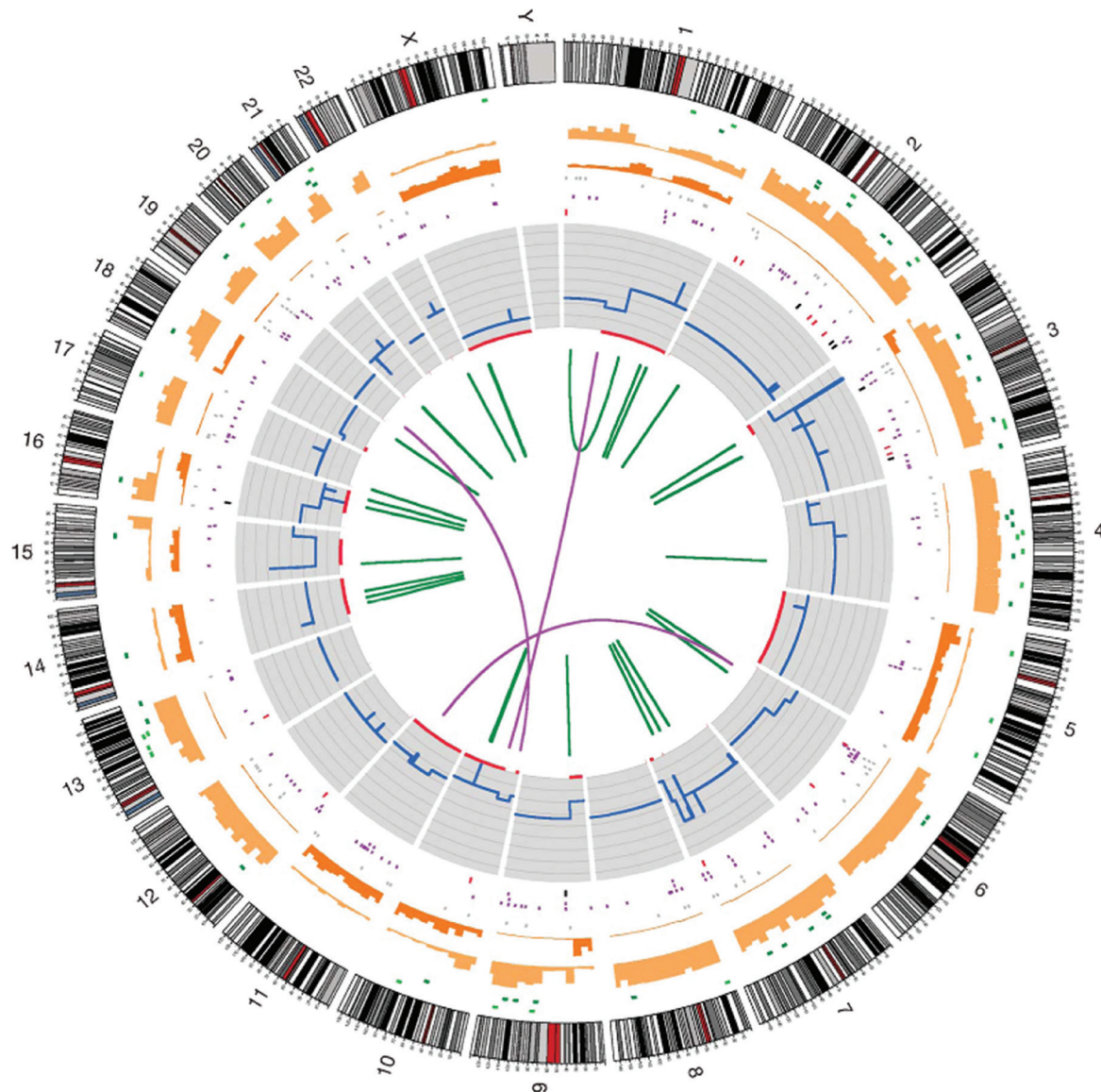


Figure 1.6 The catalog of somatic mutations in COLO-829. Chromosome ideograms are shown around the *outer ring* and are oriented pter–qter in a clockwise direction with centromeres indicated in *red*. Other tracks contain somatic alterations (from *outside to inside*): validated insertions (*light green rectangles*), validated deletions (*dark green rectangles*), heterozygous (*light orange bars*) and homozygous (*dark orange bars*) substitutions shown by density per 10 megabases, coding substitutions (*colored squares*: silent in gray, missense in purple, nonsense in red, and splice site in black), copy number (*blue lines*), regions of loss of heterozygosity (*red lines*), validated intrachromosomal rearrangements (*green lines*), and validated interchromosomal rearrangements (*purple lines*). (From Pleasance ED, Cheetham RK, Stephens PJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010;463:191–196.)

the mutation was further evaluated for its functional effect, with the finding that it led to an abrogation of the effect of an miRNA, leading to the deregulated expression of *BCL2L12*. *BCL2L12* is a negative regulator of the gene p53, which functions by binding and inhibiting apoptosis in glioma.⁹⁰ Accordingly, the dysregulation observed in *BCL2L12* led to a reduction in p53 target gene expression.

Small insertions and deletions (*indels*) represent a second category of somatic mutations that can be discovered by whole-genome sequencing of cancer specimens. These mutations may not only be about 10-fold less frequent than nucleotide substitutions but may also have an obvious impact in cancer progression. Accordingly, specific bioinformatic tools have been created to detect these *indels* in the context of the large amount of information generated by whole-genome sequencing projects.⁹¹

The systematic identification of large chromosomal rearrangements in cancer genomes represents one of the most successful

applications of next-generation sequencing methodologies. Previous strategies in this regard had mainly been based on the utilization of cytogenetic methods for the identification of recurrent translocations in hematopoietic tumors. More recently, a combination of bioinformatics and functional methods has allowed for the finding of recurrent translocations in solid epithelial tumors such as *TMPRSS2-ERG* in prostate cancer and *EML4-ALK* in non-small-cell lung cancer (NSCLC).^{92,93} Now, by using a next-generation sequencing analysis of genomes and transcriptomes, it is possible to systematically search for both intrachromosomal and interchromosomal rearrangements occurring in cancer specimens. These studies have already proven their usefulness for cancer research through the discovery of recurrent translocations involving genes of the *RAF* kinase pathway in prostate and gastric cancers and in melanomas.⁹⁴ Likewise, massively parallel paired-end genome and transcriptome sequencing has already been used

to detect new gene fusions in cancer and to catalog all major structural rearrangements present in some tumors and cancer cell lines.^{50,95–97} The ongoing cancer genome projects involving thousands of tumor samples will likely lead to the detection of many other chromosomal rearrangements of relevance in specific subsets of cancers. It is also remarkable that whole-genome sequencing may also facilitate the identification of other types of genomic alterations, including rearrangements of repetitive elements, such as active retrotransposons, or insertions of foreign gene sequences, such as viral genomes, which can contribute to cancer development. Indeed, a next-generation sequencing analysis of the transcriptome of Merkel cell carcinoma samples has revealed the clonal integration within the tumor genome of a previously unknown polyomavirus likely implicated in the pathogenesis of this rare but aggressive skin cancer.⁹⁸

Finally, next-generation sequencing approaches have also demonstrated their feasibility to analyze the pattern of copy number alterations in cancer because they allow researchers to count the number of reads in both tumor and normal samples at any given genomic region and then to evaluate the tumor-to-normal copy number ratio at this particular region. These new methods offer some advantages when compared with those based on microarrays, including much better resolution, precise definition of the involved breakpoints, and absence of saturation, which facilitates the accurate estimation of high copy number levels occurring in some genomic loci of malignant tumors.⁴⁷

PATHWAY-ORIENTED MODELS OF CANCER GENOME ANALYSIS

Genomewide mutational analyses suggest that the mutational landscape of cancer is made up of a handful of genes that are mutated in a high fraction of tumors, otherwise known as *mountains*, and most mutated genes are altered at relatively low frequencies, otherwise known as *hills* (Fig. 1.7).²⁸ The mountains probably give a high selective advantage to the mutated cell, and the hills might provide a lower advantage, making it hard to distinguish them from passenger mutations. Because the hills differ between cancer types, it seems that the cancer genome is more complex and heterogeneous than anticipated. Although highly heterogeneous, bioinformatic studies suggest that the mountains and hills can be grouped into sets of pathways and biologic processes. Some of these pathways are affected by mutations in a few pathway members

and others by numerous members. For example, pathway analyses have allowed for the stratification of mutated genes in pancreatic adenocarcinomas to 12 core pathways that have at least 1 member mutated in 67% to 100% of the tumors analyzed (Fig. 1.8).³² These core pathways deviated to some that harbored one single highly mutated gene, such as in *KRAS* in the G1/S cell cycle transition pathway and pathways where a few mutated genes were found, such as the transforming growth factor signaling pathway. Finally, there were pathways in which many different genes were mutated, such as invasion regulation molecules, cell adhesion molecules, and integrin signaling. Importantly, independent of how many genes in the same pathway are affected, if they are found to occur in a mutually exclusive fashion in a single tumor, they most likely give the same selective pressure for clonal expansion.

The idea of genetically analyzing pathways rather than individual genes has been applied previously, revealing the concept of mutual exclusivity. Mutual exclusivity has been shown elegantly in the case of *KRAS* and *BRAF*, where a *KRAS*-mutated cancer generally does not also harbor a *BRAF* mutation because *KRAS* is upstream of *BRAF* in the same pathway.⁹ A similar concept was applied for *PIK3CA* and *PTEN*, where both mutations do not usually occur in the same tumor.²²

Passenger and Driver Mutations

By the time a cancer is diagnosed, it is composed of billions of cells carrying DNA abnormalities, some of which have a functional role in malignant proliferation; however, many genetic lesions acquired along the way have no functional role in tumorigenesis.¹⁴ The emerging landscapes of cancer genomes include thousands of genes that were not previously linked to tumorigenesis but are found to be somatically mutated. Many of these changes are likely to be *passengers*, or neutral, in that they have no functional effects on the growth of the tumor.¹⁴ Only a small fraction of the genetic alterations are expected to drive cancer evolution by giving cells a selective advantage over their neighbors. Passenger mutations occur incidentally in a cell that later on in parallel develops a *driver* mutation but are not ultimately pathogenic.⁹⁹ Although neutral, cataloging passenger mutations is important because they incorporate the signatures of the previous exposures the cancer cell underwent as well as DNA repair defects the cancer cell has. In many cases, the passenger and driver mutations occur at similar frequencies, and the identification of drivers versus the passenger is of utmost relevance and remains a pressing challenge

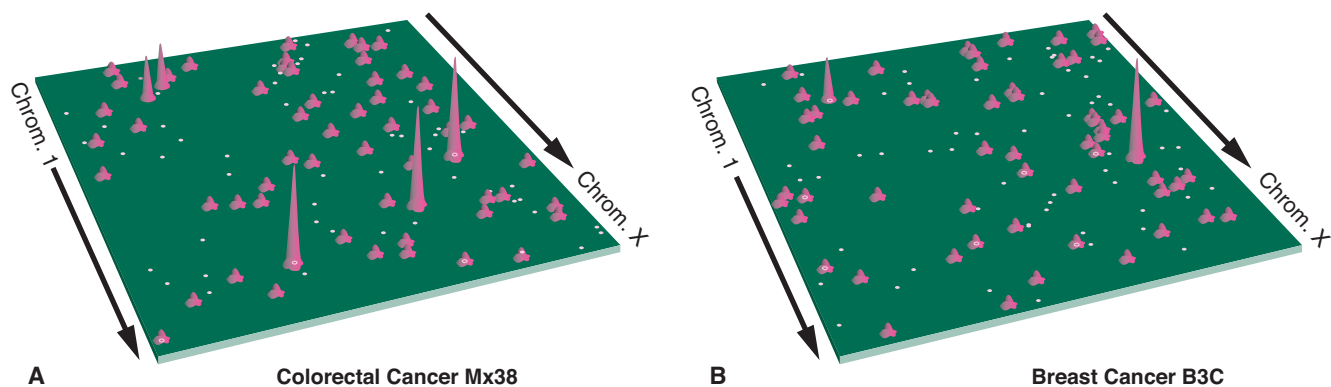


Figure 1.7 Cancer genome landscapes. Nonsilent somatic mutations are plotted in a two-dimensional space representing chromosomal positions of RefSeq genes. The telomere of the short arm of chromosome 1 is represented in the rear left corner of the *green plane*, and ascending chromosomal positions continue in the direction of the *arrow*. Chromosomal positions that follow the front edge of the plane are continued at the back edge of the plane of the adjacent row, and chromosomes are appended end to end. Peaks indicate the 60 highest ranking CAN genes for each tumor type, with peak heights reflecting cAMP scores. The *dots* represent genes that were somatically mutated in the individual colorectal (Mx38) (A) or breast tumor (B3C) (B). The *dots* corresponding to mutated genes that coincided with hills or mountains are black with white rims; the remaining *dots* are white with red rims. The mountain on the right of both landscapes represents *TP53* (chromosome 17), and the other mountain shared by both breast and colorectal cancers is *PIK3CA* (upper left, chromosome 3). (Redrawn from Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;318:1108–1113. Reprinted with permission from the American Association for the Advancement of Science.)

Notes:

25 Canada Square,
Citi Building,
41st Floor,
Canary Wharf,
London, E14 5LQ

www.wolterskluwer.com



Wolters Kluwer

When you have to be right