Chapter 2
Oncogenes

What Is An Oncogene?

An oncogene is a mutated form of a normal cellular gene – called a *proto-oncogene* – that contributes to the development of a cancer. Proto-oncogenes typically regulate basic processes that direct cell growth and cell differentiation. Most proto-oncogenes are highly conserved in evolutionarily diverse species, underscoring the fact that genes of this class play central roles in fundamental cellular processes. Mutations of proto-oncogenes that cause their conversion to oncogenes cause many of the perturbations in cell growth and differentiation that are commonly seen in cancer cells.

An oncogene is a type of cancer gene. While all cancer genes are created by mutation, oncogenes are unique in that they are caused by mutations that alter, but do not eliminate, the functions of the proteins they encode. Proteins encoded by oncogenes typically show an increased level of biochemical function as compared with the protein products of the corresponding, non-mutated proto-oncogene.

Most proto-oncogenes encode enzymes. The oncogenic forms of these enzymes have a higher level of activity, either because of an altered affinity for substrate or a loss of regulation. To reflect these gains of function, the mutations that convert proto-oncogenes to oncogenic alleles are known as *activating mutations*.

The Discovery of Transmissible Cancer Genes

The first cancer genes to be discovered were oncogenes. Indeed, the oncogene concept was the first redaction of what would eventually become the cancer gene theory.

Oncogenes were initially discovered as intrinsic components of viruses that cause cancer. Present-day molecular oncologists can trace their scientific lineage to the pioneering virologists of the early twentieth century. This group of technologically advanced and elite scientists established many of the laboratory methods and

© Springer Science+Business Media B.V. 2016

F. Bunz, *Principles of Cancer Genetics*, DOI 10.1007/978-94-017-7484-0_2

reagents that are essential to modern cancer research. The early virologists created a scientific infrastructure that would facilitate studies of cells and genes. In a tangible way, the revolution triggered by the germ theory begat a successive revolution in cancer research.

By the early twentieth century, the germ theory was firmly established, as were scientific methods for the systematic study of infectious agents. It was both technically feasible and intellectually compelling to explore whether cancer, like many other common diseases, might have an infectious etiology. Particularly interesting at that time were viruses, which were a new and largely mysterious entity. Viruses were largely uncharacterized, and defined simply as submicroscopic infectious agents present in tissue extracts that would pass through fine filters.

Early experimental observations that laid the foundation for the discovery of oncogenes predated the era of molecular biology. In 1908, Vilhelm Ellerman and Oluf Bang demonstrated that a filtered extract devoid of cells and bacteria could transmit leukemia between chickens. Leukemia was not yet recognized as a form of cancer at that time, so this work initially had little impact. Two years later, Peyton Rous discovered that chicken sarcomas could be serially transmitted from animal to animal by cell-free tumor extracts (Fig. 2.1). The causative agent in the cell filtrates, the Rous sarcoma virus (RSV), was among the first animal viruses to be isolated. The discovery of oncogenic viruses like RSV for the first time led a cancer-causing agent to be studied from a genetic perspective.

The idea that infectious agents cause cancer has a long and tortuous history. The contagious nature of cancer was promulgated in classical times by the widespread

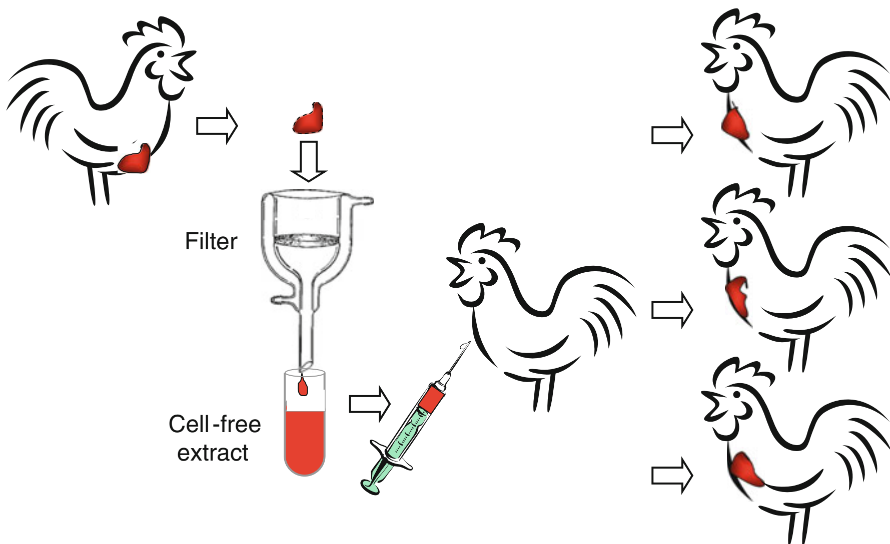


Fig. 2.1 The Rous experiment. Chicken sarcomas can be horizontally transferred between animals via injection of a cell-free filtrate. This experiment demonstrated the infectious nature of this avian cancer

belief that cancer was commonly transmitted between individuals by intimate contact, particularly between spouses, from mothers to children and from patients to caregivers. Such thinking persisted well into the nineteenth century, when they were gradually disproven by rigorous epidemiologic analysis.

A resurgence of interest in infectious agents as common causes of cancer was prompted by the formulation of the germ theory at the end of the nineteenth century. Various bacteria, yeasts, fungi, protozoa, spirochetes and coccidia were, at times, briefly implicated as potential agents that could transmit cancer, but subsequent studies failed to support a positive association. As negative results accumulated, the idea that cancer has an infectious etiology fell out of favor once again.

The initial reports by Rous were therefore met with a considerable amount of skepticism. It was suggested that his cell-free filtrates contained active cell fragments or even submicroscopic cells. The prevailing climate of antipathy towards an infectious cause of cancer substantially delayed full acceptance of Rous' work. The idea that viruses could cause cancer was dogmatically rejected as late as the 1950s, despite intermittent reports showing that other cell-free solutions could induce diverse cancers, including breast cancer, in experimental animals. Eventually, the preponderance of evidence grew too large to discount. Peyton Rous was awarded the Nobel prize in 1966, 55 years after his pioneering work was first published.

Interest in viruses as a cause of human cancer reached a new peak with the discovery of DNA tumor viruses in the 1960s. As the name of this category of viruses suggests, these common viruses – which include the papilloma-, parvo-, and polyoma-viruses – can cause tumors in animals and induce cancer-like characteristics in cultured cells. These findings led to the resurgence of the idea that viruses might be involved in the etiology of human cancer. The contemporary discovery of the DNA tumor virus simian virus 40 (SV40) as a contaminant in polio vaccine stocks that had been previously administered to millions of people was, in this context, troubling. However, as was the case with other infectious agents that had generated interest decades earlier, large follow up studies failed to establish a causal relationship between the DNA tumor viruses and common human cancers. Despite the fact that most of the viruses in this class are not a significant cause of cancer (with the very notable exception of the papillomaviruses), DNA tumor viruses have nonetheless been very useful tools for cancer research. The most widely mutated gene in human cancer, *TP53*, was initially discovered by virtue of its physical association with an SV40 viral protein in cultured cells (see Chap. 3).

As discussed in Chap. 1, most of the viruses that impact the incidence of human cancer stimulate a chronic inflammatory response. Inflammation, in turn, creates a microenvironment that promotes the acquisition, by mutation, of cancer genes and the proliferation of cells that harbor cancer genes. The DNA tumor viruses are different in that they inactivate specific host proteins, and thus create a proliferative advantage for both the cell and the infecting virus.

There is no known virus that causes cancer in humans in the dramatic way that RSV causes cancers in chickens. Nonetheless, the use of RSV to induce chicken tumors provided an invaluable model system that showed how a simple genetic element could cause cells to acquire cancer phenotypes. Prior to the complete

sequencing of the human genome, much of the information contained in the genome was unavailable or inaccessible. Cancer-associated viruses presented researchers with relatively short, well-defined regions of DNA sequence that were known to directly relate to cancer development. Viral genes could be fully sequenced and experimentally manipulated with recombinant DNA technology that was developed in the 1970s and the 1980s. The unraveling of the complex relationship between the genes of cancer-associated viruses and human genes was a pivotal step in the elucidation of the cancer gene theory.

Viral Oncogenes Are Derived from the Host Genome

The sarcoma virus isolated by Rous is one of the most potent carcinogens known. Inoculation of chickens with RSV results in the appearance of tumors within several weeks. This acute onset is in stark contrast to the development of most human tumors, which take decades to develop. Clearly, viruses like RSV have evolved a unique mechanism to trigger the cellular changes that cause cancer.

RSV belongs to a category of viruses now known as the *retroviruses*. Retrovirus particles contain genomes that are in the form of ribonucleic acid (RNA). After infection with RSV, the retroviral RNA genome is copied into DNA by the virus-encoded enzyme *reverse transcriptase*. The viral DNA then integrates into the host genome, and thus becomes a *provirus*. The provirus is replicated along with the host genome by the host DNA replication machinery, and is also transcribed by host RNA polymerase complexes. The proviral RNA transcripts are packaged into new virions, completing the virus life cycle (Fig. 2.2).

Retroviruses can cause cancer in two different ways. Depending upon where they integrate, proviruses can disrupt the functions of host genes, usually by altering their transcriptional regulation. In effect, a proto-oncogene can be changed into an oncogene upon integration of a provirus. Typically, cancers caused by the disruption of a host gene by a provirus have a long latent period and take a long time to develop. The viruses that cause such tumors are accordingly known as *slowly transforming retroviruses*. In contrast, acutely transforming retroviruses such as RSV carry their own cancer genes.

RSV contains a cancer gene known as *SRC* (pronounced "sark"). The protein encoded by *SRC* is an enzyme that localizes near the cell membrane and covalently modifies proteins in response to growth signals (Fig. 2.3). Specifically, *SRC* encodes a protein tyrosine kinase, a class of enzymes that catalyzes the addition of a phosphate group onto the tyrosine residues of multiple protein substrates, thereby altering their function. Each covalent modification catalyzed by the *SRC*-encoded protein is one of a series of enzymatically controlled events that collectively function to mediate signals that promote cell growth and division. In short, the *SRC*-encoded protein signals the cell to grow. The biochemical modes by which the enzymes encoded by cancer genes act as cellular messengers will be discussed in detail in Chap. 5.

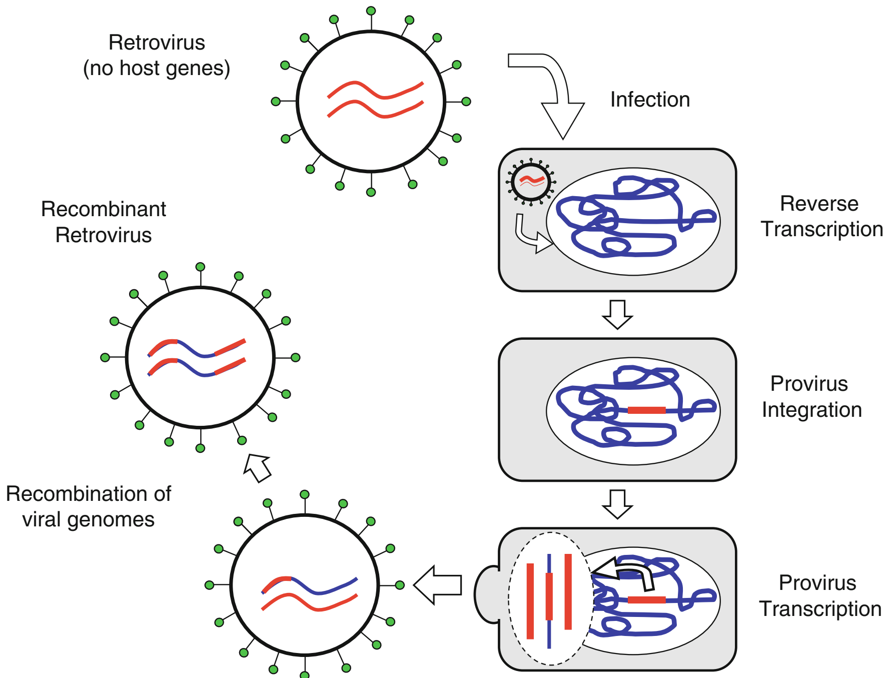
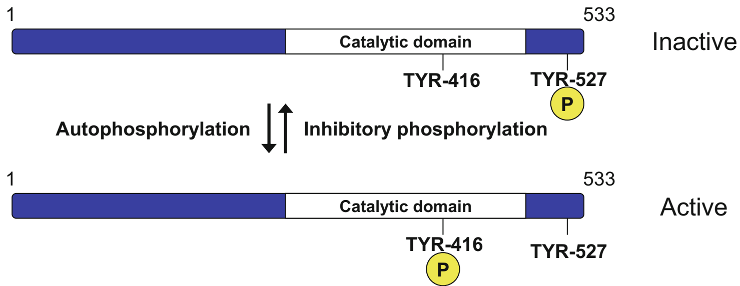


Fig. 2.2 The acquisition of oncogenes by retroviruses. The retrovirus capsule contains 2 copies of the viral RNA genome. After infection, the viral genome is copied into DNA by reverse transcriptase and integrates into the cellular genome as a provirus. If the provirus is integrated in close proximity to exon sequences, proviral transcripts can be spliced with host cell exons. These hybrid transcripts are packaged into a virion, resulting in a heterozygous viral genome. The viral genome undergoes recombination during a second round of infection. The resulting recombinant virus contains coding genetic elements that originated in the host cell

In a landmark study published in 1976, J. Michael Bishop, Harold Varmus and their colleagues demonstrated that the retroviral genes that rapidly trigger the growth of avian cancers are actually variants of genes that are already present in the host genome. There are in effect two related *SRC* genes. The cellular form of the *SRC* gene is a proto-oncogene that encodes a protein containing a tyrosine residue in the carboxy-terminus. This residue is a substrate of an enzyme that regulates growth in concert with the SRC protein (see Chap. 6). When phosphorylated at this tyrosine residue, the *SRC*-encoded protein is rendered functionally inactive and does not transduce growth signals. In contrast, the *SRC* gene carried by RSV, *V-SRC*, encodes a protein that has a truncated carboxy-terminus, and therefore does not contain the tyrosine residue that is the target of the inhibitory signal (Fig. 2.2). The *V-SRC* encoded protein thus is missing a regulatory feature present in the *SRC* encoded protein. The role of SRC and protein phosphorylation in cancer is described in detail in Chap. 6.

How did a host gene come to reside in a retrovirus? The answer lies in the retrovirus life cycle (Fig. 2.2), during which retroviruses shuttle in and out of the host genome. It appears that retroviruses acquire cellular genetic material over the course



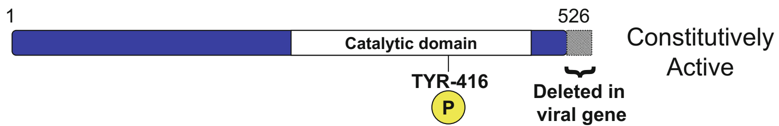


Fig. 2.3 Viral and cellular *SRC* genes. Cellular *SRC* (formerly designated *C-SRC*) is a protein tyrosine kinase that consists of 533 amino acids. Tyrosine autophosphorylation at residue 416 within the kinase domain causes a conformational change and results in the activation of kinase activity. Phosphorylation at tyrosine 527 by upstream inhibitory kinases prevents *SRC* – encoded protein activation. The viral oncogene *V-SRC* does not encode the c-terminal 7 amino acids, and therefore does not contain the negative regulatory element

of these cycles by recombination of the viral DNA with cellular DNA, and incorporate these genes into their own genomes. Evolutionary forces would favor proviruses that can most effectively propagate. Once integrated, the fate of a provirus becomes linked to the fate of the host cell genome. Proviruses that contain genes such as *V-SRC* trigger DNA replication and cell proliferation and thereby promote their own production.

The discovery that cancer-causing retroviruses contain altered forms of host genes fundamentally changed the focus of cancer research. This critical finding showed that the key to understanding cancer lies in the genome of the cancer cell itself. For the first time it was clear that altered cellular genes could cause cancer.

The Search for Activated Oncogenes: The RAS Gene Family

The oncogenes that most often contribute to the development of human cancers are not transmitted by viruses, but rather are acquired by the somatic mutation of proto-oncogenes. The horizontal transfer of cancer by RSV-containing cell extracts does not reflect the means by which human cells acquire oncogenes. Nonetheless, viruses

such as RSV did provide important insight as to what oncogenes look like and to how they might induce cellular changes.

The idea that oncogenes could be transmitted by some viruses fostered creative strategies to isolate additional genes that might have oncogenic potential. Genetic material can be efficiently transferred to cultured cells by chemical techniques that were developed during the 1970s. When introduced into primary cells growing in culture dishes, oncogenes can cause observable changes in growth properties. In a process known as *in vitro* transformation, cells that are experimentally forced to express many types of oncogene undergo changes in morphology, lose contact inhibition and begin to grow in piles known as foci (Fig. 2.4). These quantifiable changes formed the basis of numerous experiments that led to the discovery of several widely mutated oncogenes.

Potent oncogenes were found to be carried by two retroviral strains, the murine Harvey and Kirsten sarcoma viruses. These retrovirus-associated DNA sequences (or RAS genes) were designated *HRAS* and, *KRAS* respectively. The Harvey and Kirsten retroviruses were not naturally occurring pathogens, but had been experimentally derived by repeated passage of murine leukemia viruses through laboratory strains of rats. During the creation of these new, highly carcinogenic viruses, *HRAS* and *KRAS* had been acquired in altered, oncogenic form from the host genome. Using DNA transfer schemes, the laboratories of Robert Weinberg and Geoffrey Cooper, and Mariano Barbacid and Stuart Aaronson independently isolated variants of the RAS gene family directly from human cancer cells.

That retroviral oncogenes are related to the oncogenes created by the somatic mutation of proto-oncogenes was underscored by the discovery of the RAS genes. Activated RAS alleles were the first cancer genes to be found in cells derived from naturally-occurring human cancers. It was shown that the RAS genes isolated from human bladder and lung carcinoma cells were homologous to the RAS genes harbored by the Harvey and Kirsten retroviruses. Soon thereafter, Michael Wigler and colleagues isolated a third RAS gene family member that had no known viral homolog, from a neuroblastoma. The third RAS gene was accordingly designated *NRAS*. These three genes are encoded by distinct loci but are highly related, both structurally and functionally.

The wild type RAS proto-oncogenes do not induce focus formation in the *in vitro* transformation assay. The gain of function that leads to the acquisition of this property is conferred by an activating point mutation. The bladder carcinoma from which the cellular *HRAS* gene was first isolated was found to have a single base substitution that changed codon 12 from GGC (glycine) → GTC (valine). Subsequent DNA sequence analysis of large numbers of human tumors has revealed a high frequency of RAS gene mutations in several common tumor types. The majority of these cancer-associated mutations involve just three codons: 12, 13 and 61. Different tumor types differ greatly in the overall frequency of RAS gene mutations, and also in the RAS family member that is predominantly mutated (Table 2.1).

Interestingly, the first oncogenes discovered were not representative of naturally-occurring activated oncogenes. Although activated *HRAS* was among the first oncogenes to be discovered in a tumor, mutations in this RAS family member are

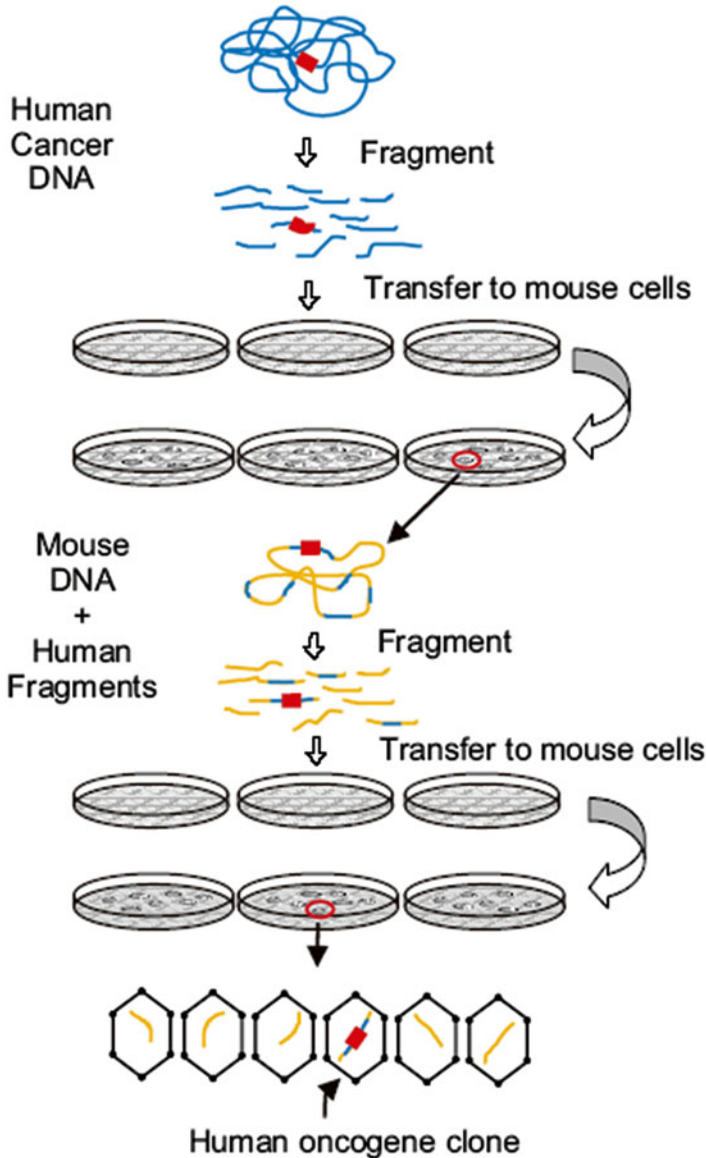


Fig. 2.4 Oncogene discovery by *in vitro* transformation. Genes transferred from human genomic DNA (*blue*) can alter the growth properties of mouse fibroblasts. Genomic DNA is sheared into small fragments, which are introduced into mouse cells grown in monolayer cultures. Appearing after a period of growth, discrete foci represent clones of mouse cells that have altered growth and cell-cell interactions. Genomic DNA from these clones (*yellow*) can contain multiple integrated fragments of human DNA. A second round of transfer allows the isolation of individual human fragments. DNA from the second clone is packaged into a bacteriophage library, which is then screened with a probe corresponding to human genomic DNA-specific repeat elements. Assays of this type were relatively non-specific. Foci can be caused by actual oncogenes that are activated in cancer cells, but also by proto-oncogenes activated by the gene transfer process and growth regulatory genes that are not found to be mutated in cancers

Table 2.1 Mutations in the *RAS* gene family

Cancer type	Mutation frequency (%)	RAS family member
Pancreatic carcinoma	95	*KRAS*
Colorectal carcinoma	50	*KRAS*
Lung carcinoma	40	*KRAS*
Acute Myelogenous Leukemia	25	*NRAS*
Melanoma	15–30	*NRAS*

not widespread in cancers. Similarly, *NRAS* was first isolated from a neuroblastoma, yet subsequent studies have failed to detect *NRAS* mutations in a significant proportion of these tumors. It remains a possibility that the mutated RAS genes identified by *in vitro* transformation arose during the maintenance of tumor-derived cell lines in culture (*in vitro*), rather than by somatic mutation that occurred during tumorigenesis. Nonetheless, the initial identification of the RAS family of oncogenes was an important achievement that paved the way for the systematic analysis of common cancer mutations. Mutations in *RAS* family members are involved in a significant proportion of a number of common malignancies.

RAS genes are ubiquitously expressed and presumably have the same function in all cells. Why then is mutation of *KRAS* a dominant feature of pancreatic tumors and present at much lower frequencies other malignancies? Why are *NRAS* mutations but not other RAS family mutations prevalent in acute myelogenous leukemias? The basis for the tissue specificity of RAS mutations, and indeed of cancer gene mutations in general, remains largely unknown. One might assume that tissue-specific gene alterations arise in cancers at a detectable frequency because they provide a selective advantage under the unique micro-environmental conditions of given cellular compartment.

The cellular role of the RAS-encoded proteins involves the coupling of signals that arise at cell membrane receptors with downstream intracellular signaling molecules. The mutation of conserved codons in the RAS family members affects the regulation of the enzymatic activity of RAS proteins. The nature of RAS protein activity and the cellular functions of the RAS gene family will be discussed in detail in Chap. 6.

Complex Genomic Rearrangements: The MYC Gene Family

The MYC gene family first emerged as a viral gene, *V-MYC*, harbored in the genomes of four independent isolates of avian leukemia virus. Among the tumors caused by these oncogenic retroviruses is myelocytomatosis, a tumor composed mainly of myelocytes, a type of white blood cell. It is from this rare tumor that the name of a commonly activated oncogene family was derived. The cellular homolog

of *V-MYC* is the proto-oncogene *C-MYC*, now simply known as *MYC*. There exist two structurally and functionally related genes that were discovered subsequently, originally designated *N-MYC* and *L-MYC* and now known as *MYCN* and *MYCL*. The latter two genes were isolated as oncogenes from a neuroblastoma and a lung carcinoma, respectively.

In contrast to the genes in the *RAS* family, which are activated by single nucleotide substitutions, *MYC* and related genes are typically activated by larger and more complex genomic rearrangements. The encoded protein product is most often not structurally altered by *MYC* gene activation, but increased in quantity. The consequence of *MYC* activation is an increase in gene expression. Even modest increases in *MYC* expression caused by activating mutations are thought to significantly contribute to tumorigenesis in some tissues.

The *MYC* genes encode transcription factors that directly affect the expression of genes involved in diverse aspects of cell growth and death. The *MYC* genes are sometimes referred to as nuclear proto-oncogenes, reflecting their role in controlling the transcription of genes in the cell nucleus. The function of the *MYC* genes in the alteration of gene expression in cancer cells will be discussed in Chap. 6.

The three *MYC* genes share a common genomic structure that consists of three exons. Including intronic regions, each spans approximately 5 kb. This compact genetic unit has been found to be rearranged in a number of ways that result in the aberrantly high expression of MYC proteins. Studies of *MYC* genes in cancers have revealed several general mechanisms by which proto-oncogenes can be activated.

All of the activating mutations that convert *MYC* genes to their oncogenic forms increase the protein levels. There are several mechanisms by which this occurs. The number of functional *MYC* genes can increase as a result of the amplification of the genomic region containing a *MYC* gene. Alternatively, the level at which a MYC gene is expressed can be altered if that gene is repositioned in proximity to a highly active promoter element, usually as a result of a chromosomal translocation. These genetic changes are types of somatic mutations that are stably propagated by cancer cell clones during their evolution.

Proto-oncogene Activation by Gene Amplification

In normal cells, proto-oncogenes exist as single copy genes. That is, a single genomic locus contains one copy of each exon, intron and regulatory element. Due to the diploid nature of the human genome, a total of two alleles of each gene will be present in each cell.

The copy number of a gene can increase as a result of the amplification of a subchromosomal region of DNA. The increase in gene copy number leads, in turn, to a corresponding increase in the overall expression levels of that gene. The process by which genomic amplification occurs remains incompletely understood, but is thought to involve repeated rounds of DNA replication that occur during a single cell cycle.

The unit of genomic DNA that is amplified is known as the *amplicon*. Amplicons vary in size, but typically range in size between 10^5 and 10^6 base pairs. The number of amplicons found within a region of amplification also varies broadly. An amplicon can contain varying numbers of genes depending on the size and location of the genomic region contained within the amplicon. Overall genomic structure is typically preserved within amplified regions, with amplicons ordered in repetitive arrays in head-to-tail orientation (Fig. 2.5).

If the copy number is high or if an amplicon is particularly large, the amplified region may be directly observable by cytogenetic methods. Amplified regions of the genome can exist in extrachromosomal bodies known as *double minutes*, which are small structures that resemble chromosomes but do not contain centromeres. Double minutes can integrate into a chromosome. The region of integration can often be distinguished cytogenetically as a region that stains homogenously with dyes used to reveal chromosome banding patterns. The integration of double minutes is thought to be reversible. Accordingly, the integrated and extrachromosomal forms of amplified genomic DNA are interchangeable. Double minutes and homogeneous

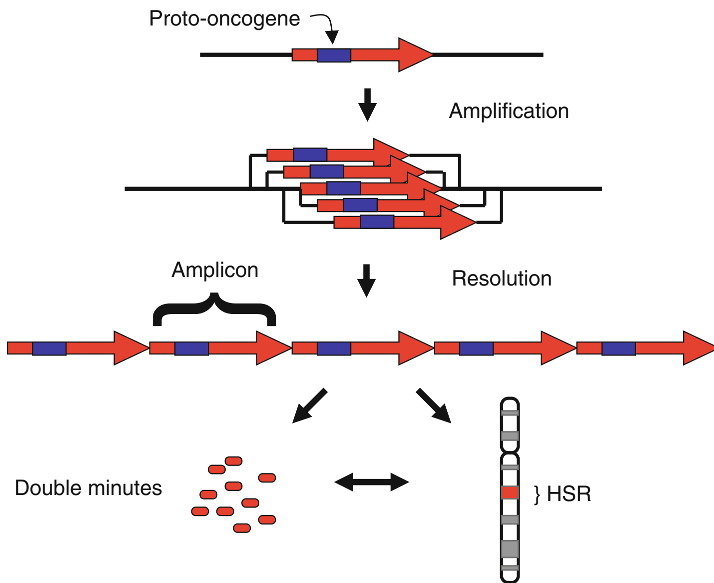


Fig. 2.5 Oncogene activation by gene amplification. A genomic region (*red arrow*) containing a proto-oncogene is amplified as a result of multiple rounds of DNA replication during a single cell cycle. Resolution of the over-replicated region results in a tandem array of amplicons in a head-to-tail orientation. The amplified region can alternatively be maintained as double minutes, or integrated into a chromosome to form a heterogenous staining region (*HSR*). It is believed that these two configurations are interchangeable

staining regions are not seen in cytogenetic analysis of normal cells, but are observed in a significant number of tumor cells.

Upon amplification of a MYC locus, *MYC* is converted from a proto-oncogene to an oncogene. The most notable role for *MYCN* amplification is in the growth of neuroblastomas, tumors that arise from immature nerve cells. These tumors almost exclusively affect young children. Amplification of the genomic region on chromosome 2p24 containing *MYCN* can be detected in about 25 % of neuroblastomas. The degree of amplification of *MYCN* in neuroblastomas can be extensive; as many as 250 copies have been found in some of these cancers. The extent of *MYCN* amplification has been found to correlate with both the stage of the disease, and independently with the rate of disease progression and outcome. These findings provide evidence that *MYCN* amplification directly contributes to neuroblastoma progression.

Amplified *MYC* genes are commonly found in a number of tumors in addition to neuroblastomas. The first example of *MYC* amplification was observed in a myelocytic leukemia, and the gene was named accordingly. Small-cell cancers of the lung have been found to variously contain amplification of one of the three MYC genes, *MYC*, *MYCN* and *MYCL*. *MYC* amplification is found in approximately 20–30 % of breast carcinomas and appears to be correlated with a poor clinical outcome.

Another gene that is commonly amplified in a broad spectrum of cancers is *ERBB2*, previously referred to as *HER2/neu*. *ERBB2* amplification has been found in a significant proportion of breast and ovarian cancers and also in adenocarcinomas arising in the stomach, kidneys and salivary glands.

The *ERBB2* gene was first identified as the cellular homolog of an oncogene, *VERBB2*, carried by the avian erythroblastic leukemia virus, a retrovirus. At around the same time, an oncogene termed *NEU* was isolated from a rat neuroblastoma cell line by *in vitro* transformation, while a gene known as *HER2* was discovered by virtue of its similarity to a previously discovered gene that encodes a cell surface signaling protein called human epidermal growth factor receptor. Efforts to determine the chromosomal locations of these genes suggested – and DNA sequencing subsequently proved – that *HER2/neu* and *ERBB2* are in fact the same gene.

Genetic alterations that activate *ERBB2* are among the most common somatic mutations found in breast cancer, occurring in about 15 % of tumors analyzed. Most of these alterations are gene amplifications that result in increased *ERBB2* expression. The amplicons that include the entire *ERBB2* locus vary between cancers but span a common region of about 280 kb in length. This core amplicon includes several loci in addition to *ERBB2*, but genetic analysis strongly suggests that it is the enhanced expression of *ERBB2* that confers clonal selectivity. Amplified regions typically contain about 20 copies of the *ERBB2* amplicon, but have been found to contain as many as 500 copies. Analysis of the *ERBB2* coding regions has revealed relatively few alterations that affect the open reading frame, confirming that the increase in gene dosage is the primary mode of activation.

ERBB2 encodes a protein that functions as a receptor on the cell surface that transduces growth signals. The activation, by amplification, of this proto-oncogene results in the overexpression of the ERBB2 receptor and a resulting hypersensitivity to growth factors. The *ERRB2*-encoded protein is a prototype of an important class of oncogene-encoded proteins that will be described further in Chap. 5.

Amplification of *ERBB2* in breast cancers is a useful prognostic marker. While amplification of *ERBB2* does not appear to correlate with disease characteristics such as tumor size, there is a significant correlation with the spread of cancer cells to local lymph nodes, which is independently a negative prognostic sign. Breast tumors that harbor *ERBB2* amplification tend grow more aggressively. Statistically, patients with *ERBB2* positive cancers exhibit a significantly shorter time to relapse following standard therapy and reduced long-term survival. The recent development of specific therapy that targets *ERBB2* function makes the identification of patients with *ERBB2* overexpressing tumors a priority. The molecular basis for targeted therapies is discussed in Chap. 8.

Oncogenes activated by gene amplification contribute to many common types of cancer (see Table 2.2).

Table 2.2 Oncogenes frequently amplified in human cancers

Oncogene	Cellular function	Type of cancer	%
MYC	Transcription factor	Breast ca	20
		Ovarian ca.	30–40
		Prostate ca.	15
		Pancreatic ca.	15
CCND1	Cell cycle regulator	Esophageal ca.	35
		Head and Neck ca.	25
		Breast ca	15
		Bladder ca.	10–15
CCNE1	Cell cycle regulator	Uterine serous cell ca.	45
		Ovarian ca.	20
CDK4	Cell cycle regulator	Sarcoma	20
		Glioblastoma	20
EGFR	Growth factor receptor	Glioblastoma	30–50
ERBB2	Growth factor receptor	Breast ca	20–35
		Gastric ca	10
MDM2	Regulation of tumor suppressor protein	Sarcoma	20–25
		Glioblastoma	10–15
MET	Protein tyrosine kinase	Breast ca.	20
PIK3CA	Lipid kinase	Lung squamous cell ca.	40
		Ovarian ca	30
		Esophageal ca	20

Proto-oncogenes Can Be Activated by Chromosomal Translocation

A chromosomal break presents a unique challenge to a growing cell. Cells that contain broken chromosomes cannot continue to grow and divide; proliferation can only continue once a chromosomal break is repaired. The resolution of such breaks is critical to cell survival, but the process of repair frequently results in mutations. One such mutation is the chromosomal translocation.

A translocation is the transfer of a chromosome segment to a new position, often on a nonhomologous chromosome. In some cases the repair process results in the exchange of pieces between nonhomologous chromosomes; such an exchange is termed a reciprocal translocation (see Chap. 1).

Gross structural rearrangements like translocations can juxtapose proto-oncogenes with genetic elements that normally would be unrelated. Proto-oncogenes can be activated by translocations in two ways, depending on the location of the breakpoint. A translocation can put the exons of two separate genes under the control of a single promoter element. The splicing together of previously unrelated exons can then result in the expression of a single hybrid protein that contains elements of each of the two genes involved. Alternatively, a translocation can preserve a complete open reading frame but place it under the control of a more active promoter.

An example of a proto-oncogene that can be activated by chromosomal translocation is *MYC*. The expression of *MYC* is normally tightly regulated. This tight transcriptional control is altered in some lymphomas and leukemia in which the *MYC* gene is repositioned, via translocation, into the vicinity of a highly active promoter. The repositioning of *MYC* into the vicinity of these strong promoters is sufficient to activate *MYC*, and thereby convert it into a functional oncogene.

Chromosomal Translocations in Liquid Tumors

Somatically acquired chromosomal translocations are frequently found in the liquid tumors: the leukemias and lymphomas. Although translocated chromosomes have been found in many solid tumors, translocations are highly prevalent in liquid tumors and are in some cases pathognomonic. Translocations that convert proto-oncogenes to oncogenes have been found in over 50 % of leukemias and in a significant proportion of lymphomas.

Some common genetic alterations are repeatedly observed in cancers of a single type from many different patients. Such alterations are said to be *recurrent*. Many of the recurrent translocations found in liquid tumors are structurally conserved and defined by common breakpoints. These breakpoints often occur in closely spaced clusters. The location of the breakpoints or breakpoint clusters that define translocations is highly disease-specific. In other words, cancers that arise in a particular type of cell will typically harbor similar translocations.

Recurrent translocations, like other genetic alterations, are cell lineage-dependent. The recurrent translocations involving *MYC* indicate why this is the case. The chromosomal translocation resulting in the juxtaposition of *MYC* and highly expressed immunoglobulin genes is a common feature of both B-cell leukemia and Burkitt lymphomas, particularly those arising in children. These cancers arise from a common stem cell, the lymphoid progenitor, in which immunoglobulin gene expression is highly activated. In contrast, *MYC* is activated in T-cell leukemias by translocation and juxtaposition with highly expressed T-cell receptor genes. In these distinct cancers, both the oncogene and the mode by which is it activated are recurrent. Based on these patterns of activation, one can infer that increased expression of *MYC* confers a particularly strong survival advantage in these distinct tissue compartments.

Chronic Myeloid Leukemia and the Philadelphia Chromosome

The activation of a proto-oncogene by a recurrent translocation is best illustrated by the example of chronic myeloid leukemia (CML). In 95 % of CML patients, the cancer cells contain a unique derivative chromosome named after the city in which it was discovered, the Philadelphia chromosome (Fig. 2.6). The Philadelphia chromosome was originally identified in 1960 by Peter Nowell and David Hungerford and upon detailed cytogenetic analysis in 1973 by Janet Rowley, was found to result

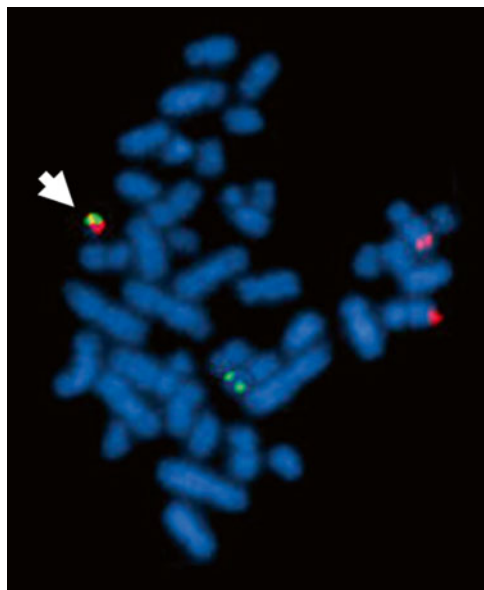


Fig. 2.6 The Philadelphia chromosome. The Philadelphia chromosome (indicated by *arrow*) stained in a mitotic spread. Fluorescence *in situ* hybridization probes are derived from *BCR* (*green*) and *ABL* (*red*). The spots in other chromosomes represent the untranslocated *BCR* and *ABL* genes

from a reciprocal translocation involving chromosomes 9 and 22. The small proportion of CML patients that do not exhibit a typical Philadelphia chromosome have translocations that are structurally more complex, but still ultimately involve the same chromosomal regions. Subsequent to its discovery in CML patients, the Philadelphia chromosome was also found to be present in 3–5 % of children and 30–40 % of adults with acute lympohcytic leukemia (ALL).

CML is a cancer that arises in blood cell progenitors and spreads throughout peripheral blood and bone marrow. CML affects all age groups, but is most common in older adults. The natural history of CML unfolds in clinically defined stages. Within 3–5 years after its detection, CML typically progresses from a relatively benign chronic disease to an acute illness known as blast crisis, which is life-threatening. While the CML cells found during the chronic stage are mature, those found during blast crisis are relatively undifferentiated and resemble those found in patients with acute leukemias.

Interestingly, the only environmental factor known to increase the risk of CML is exposure to high-dose ionizing radiation. It appears that the double strand DNA breaks caused by ionizing radiation can facilitate the translocation that creates the Philadelphia chromosome. In most cases, no predisposing factors are identified and the initiating translocation thus appears to result from a stochastic process. Regardless of the mechanism by which they arise, the rare cells containing the Philadelphia chromosome are then clonally selected and expanded by the process of clonal evolution. The recurrence of a single translocation in CML suggests that this genetic alteration must provide the cancer precursor cells with a unique and essential survival advantage.

At the molecular level, the consequence of the translocation involving chromosomes 9 and 22, denoted t(9;22), is the unique juxtaposition of two genes, *BCR* and *ABL. ABL* is a proto-oncogene homologous to an oncogene originally found in the retroviral genome of the Ableson leukemia virus. In the absence of translocation, the expression of the *ABL* proto-oncogene is tightly regulated. The *BCR* gene, in contrast, was so named because of its location within the breakpoint cluster region on chromosome 22. *BCR* expression is driven by a strong, constitutively active promoter. Strictly speaking, *BCR* is not considered a proto-oncogene, and in fact its normal cellular role is unknown. The *BCR* promoter functions to transcribe *ABL* exons when the two genes are fused by translocation (Fig. 2.7).

The t(9;22) reciprocal translocation results in the creation of two separate fusions between the *BCR* and *ABL* genes. The *BCR-ABL* gene is created on the derivative of chromosome 22, the Philadelphia chromosome, while a corresponding *ABL-BCR* fusion gene is created on the derivative chromosome 9. Numerous experiments have demonstrated that it is the product of the *BCR-ABL* gene that is oncogenic. Like a substantial number of proto-oncogenes, the *ABL* gene encodes a protein tyrosine kinase. The fusion gene encodes the catalytic domain of this enzyme, while the expression of this domain is controlled by the *BCR* promoter. It appears that the BCR peptide mediates oligomerization of the BCR-ABL fusion protein, causing constitutive activation of the protein tyrosine kinase domain in the ABL peptide.

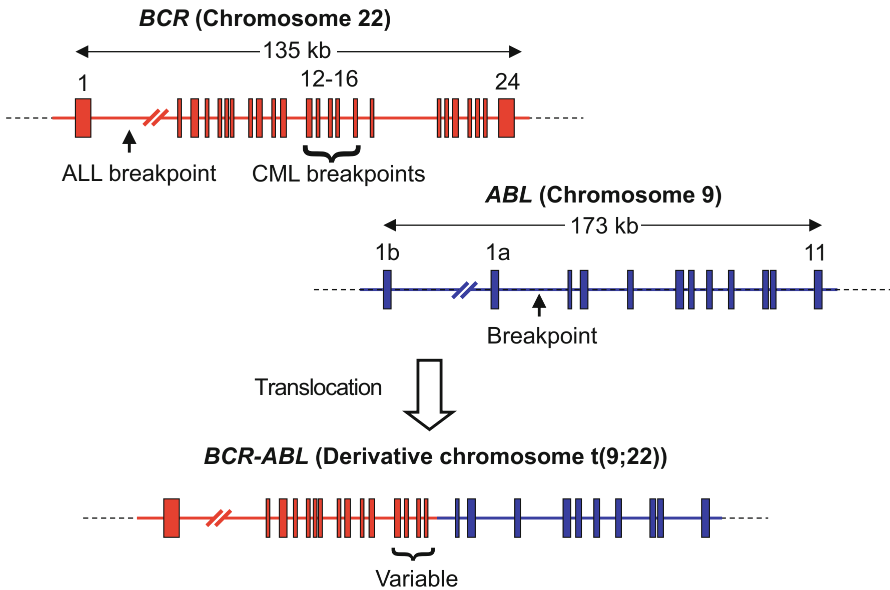


Fig. 2.7 The creation of *BCR-ABL* by translocation. The BCR locus on chromosome 22 spans roughly 135 kilobases (kb) and is composed of 24 exons. Within this gene is a recurring breakpoint, found in acute lymphocytic leukemia-associated translocations, and also a cluster of breakpoints found in chronic myeloid leukemias. The ABL locus on chromosome 9 spans 173 kb and has 11 exons. Two first exons are alternatively utilized. A single recurrent breakpoint occurs upstream of exon 2. In the t(9;22) derivative, the *BCR* and *ABL* genes are fused, and together form a single open reading frame. The different CML-associated breakpoints in *BCR* result in the variable inclusion of *BCR* exons 12–15 in different allelic forms of *BCR-ABL*

The mutational activation of tyrosine kinases and their roles in the cell are discussed in detail in Chap. 6.

The precise junction between chromosome 9 and chromosome 22 sequences varies between different groups of CML patients. While there is a single breakpoint on chromosome 9, the breakpoint on chromosome 22 is actually of cluster of distinct breakpoints variably found in different groups of patients. Accordingly, the portion of *BCR-ABL* that is composed of *ABL* sequence is invariant. However, the existence of multiple breakpoints within the *BCR* locus results in the creation of distinct in-frame fusions. The chimeric proteins encoded by these different gene fusions differ at their N-termini and can be distinguished by their molecular weight (Fig. 2.8). CML is monoclonal in nature, and so only one *BCR-ABL* encoded protein is detectable in each patient. Depending on the site of the break point in the *BCR* gene, the fusion protein can vary in size from 185 to 230 kDa.

The different *BCR-ABL* fusion proteins can be correlated with different clinical outcomes. Most CML patients express the 210 kDa form of the fusion protein. A subgroup of CML patients has been identified that express a 230 kDa BCR-ABL

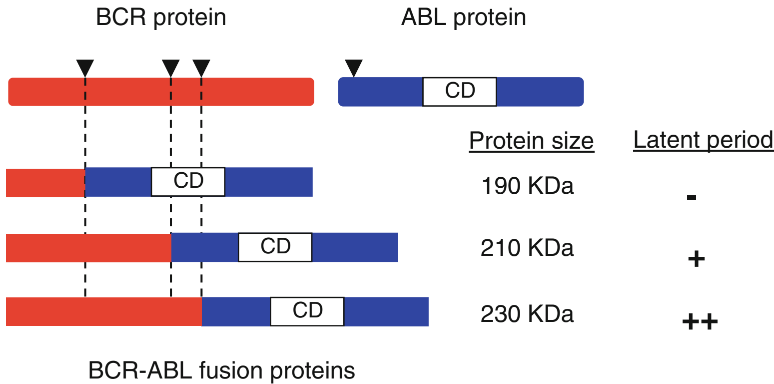


Fig. 2.8 BCR-ABL-encoded proteins. The primary structures of the native BCR and ABL proteins are shown. *Arrowheads* indicate the regions of defined by the recurrent breakpoints. The various breakpoints in *BCR* lead to the appearance of distinct fusion proteins with molecular weights of 190, 210 and 230 kilodaltons (kDa). The 190 kDa protein is restricted to ALL, an acute disease that is not characterized by a latent period. The 210 kDa is the most prevalent CML associated version, while the 230 kDa protein is found in a subset of CML patients that typically exhibit an extended period of disease latency

encoded protein. These patients have a distinct disease course that is typified by decreased numbers of white cells in the peripheral blood and delayed progression to blast crisis. Patients with highly-aggressive ALL express either the 210 kDa form or a unique 190 kDa protein. The 190 kDa protein has been shown to be a more active tyrosine kinase than the 210 kDa protein, suggesting that different levels of activity affect the clinical course of these diseases.

Because the presence of the various gene fusion products correlates with both the type and course of disease, these molecules are useful markers for diagnosis and prognosis. The presence of a chimeric RNA species transcribed from a fusion gene is readily detectable by commonly employed RNA/DNA amplification techniques. Thus, the expression of these unique oncogenes provides a convenient and highly informative marker than can be directly used in the clinic.

The catalytic activity of the *BCR-ABL* encoded tyrosine kinases can be directly inhibited by drugs. Therapy based on this approach has been highly successful at delaying blast crisis and has significantly improved the overall outlook for patients with CML. The fact that specific therapy directed at the *BCR-ABL* gene product is highly effective demonstrates conclusively the central role of the *BCR-ABL* oncogene in CML pathogenesis. The effects of tyrosine kinase activation on cancer cell proliferation will be discussed in Chap. 6; novel therapeutic approaches to specifically target these enzymes will be described in Chap. 8.

Oncogenic Activation of Transcription Factors in Prostate Cancer and Ewing's Sarcoma

Cytogenetic analysis is considerably more difficult in solid tumor samples, and so the role of translocations was first appreciated in the liquid tumors. But recurrent translocations activate oncogenes in solid tumors as well.

Prostate cancers are often caused by driver gene mutations that alter the cells' normal responses to male sex hormones, known as *androgens*. The most frequently detected oncogene in prostate cancers is a fusion gene that is created by a translocation involving two loci on chromosome 21. The translocation creates a novel fusion gene that contains the 5' untranslated region of *TMPRSS2*, an androgen-regulated transmembrane receptor protein and the proto-oncogene *ERG*. ERG encodes a transcriptional regulator that is one of a large family of evolutionarily conserved proteins that mediate several basic cellular functions, including cell division, differentiation and migration, which are normally active during embryonic development. The products of the *ERG* gene had previously been observed to be expressed at high levels in many prostate cancers. The discovery of the *TMPRSS2-ERG* fusion gene in 2005 provided a mechanistic explanation for this androgen-dependent upregulation.

The founding member of the transcription factor family that includes *ERG* is an oncogene that was found to be transduced by the leukemia virus E26, or ETS. A second ETS-family member designated *ETV1* has also been found to be activated by fusion to upstream elements of *TMPRSS2* in a smaller number of prostate cancers. The encoded proteins of the ETS family of genes contain a common protein domain that is important for the protein interactions that facilitate sequence-specific DNA binding.

Ewing's sarcoma is a rare tumor of bone or soft tissue that occurs in children and young adults, most frequently in male teenagers. These highly aggressive tumors can occur in various anatomic sites, but most often are found in the bones of the pelvis, femur, humerus, ribs and clavicle. The cells that compose Ewing's sarcomas are morphologically similar to those found in diverse types of pediatric solid tumors, making accurate diagnosis difficult. This challenge prompted focused investigation into cytogenetic changes that could potentially provide a diagnostically useful marker. A distinguishing characteristic of Ewing's the majority of sarcoma cells was found to be the presence of a reciprocal translocation between chromosomes 11 and 22, abbreviated t(11;22).

Molecular analysis revealed that t(11;22) consistently juxtaposes the *FLI1* gene on chromosome 11 and the *EWS* gene on chromosome 22 (Fig. 2.9). *FLI1* is a proto-oncogene that was originally identified in mice as the integration site common to two retroviruses that cause leukemias and sarcomas, including the Friend leukemia virus for which the locus was named. Like *ERG*, *FLI1* is a gene in the ETS family and shares many of the basic functional attributes of these genes. The *EWS* gene, named for its discovery as the locus at the Ewing sarcoma breakpoint, is the proximal component of the *EWS-FLI* oncogene. The proto-oncogene *EWS* encodes a

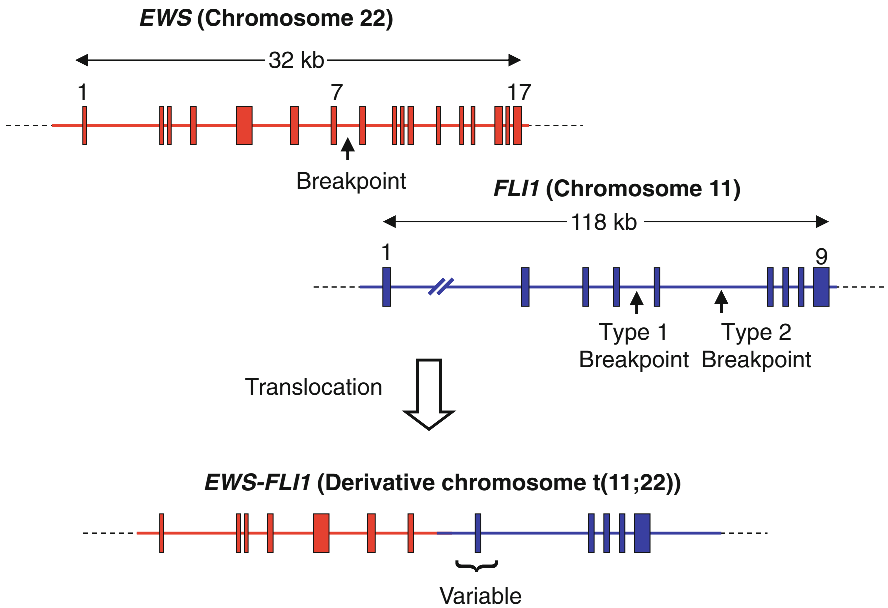


Fig. 2.9 The creation of *EWS-FLI1* by translocation. The *EWS* locus on chromosome 22 spans roughly 32 kb and is composed of 17 exons. In patients with Ewing's sarcoma, a recurring breakpoint is found the 7th intron. The *FLI1* locus on chromosome 11 spans 118 kb and has 9 exons. Within this gene are two recurrent disease-associated breakpoints. In the t(9;22) derivative, the *EWS* and *FLI1* genes are fused, and contain a single open reading frame. The Type 1 and Type 2 fusions result in two distinct *EWS-FLI1* genes that differ in the inclusion of one *FLI1*-derived exon

multifunctional RNA binding protein that is involved in gene expression, cell signaling and RNA processing and transport.

In the Ewing's sarcoma translocation, the chromosomal breakpoints occur within the introns of *FLI1* and *EWS*, and result in the in-frame fusion of the promoter and upstream elements of *EWS* and the downstream elements of *FLI1*. The precise locations of the breakpoints vary from tumor to tumor. The most frequent junction, occurring in 60 % of cases, joins exon 7 of EWS to exon 6 of *FLI1* in what is termed a Type 1 fusion. Approximately 25 % of cases are associated with a so-called Type 2 fusion, which includes exon 5 of *FLI1*. As was seen to be the case in CML, the fusion variants correlate with distinct clinical outcomes. In particular, the Type 1 fusion is associated with a significantly better prognosis than the other fusion types.

In all t(11;22) breakpoints, the RNA-binding domain encoded by *EWS* is replaced with the DNA-binding domain encoded by *FLI1*. The *EWS-FLI1*-encoded fusion protein is thus a chimera. Though the target sequences recognized by the DNA-binding domain of the *EWS/FLI1* gene product are indistinguishable from those recognized by native *FLT1*, the chimeric protein is more active and is found to transactivate 5–10 times more transcription than native *FLI1*. Of direct clinical relevance are functional differences between the alternative forms of *EWS-FLI1*. The protein product of the Type 1 fusion was found to be a less effective transcriptional

transactivator than the other fusion gene products. This difference in activity correlates closely with the more benign clinical course associated with this alteration.

The *EWS-FLI1* fusion is the most common gene product of chromosomal translocation in Ewing's sarcoma, occurring in about 90 % of cases. These alterations are also found in rare tumors that are similar to Ewing's sarcoma. The *EWS* gene has also been found to be fused with several other members of the ETS family of transcription factors in both Ewing's sarcoma and in related disorders. About 10 % of Ewing's sarcomas exhibit fusion of *EWS* with *ERG*, the same gene that is activated in prostate cancers harboring the *TMPRSS2-ERG* fusion.

The discovery of these molecular similarities has led to the reclassification of a group of molecularly and clinically related diseases, which is now referred to as the Ewing's sarcoma-related family of tumors. Cumulatively, these molecular data suggest that the dysregulation of ETS-mediated transcription by *EWS* fusion is a critical step in the clonal evolution of the Ewing's sarcoma family from their stem cell progenitors.

Oncogene Discovery in the Genomic Era: Mutations in *PIK3CA*

The identification of the majority of known oncogenes predated the sequencing of the human genome. The prototypical oncogenes described in previous sections were isolated on the basis of their homology to genes carried by oncogenic retroviruses or on their ability to induce colony formation in an *in vitro* transformation assay. These early oncogenes were not discovered because they were necessarily involved in large numbers of cancers. Rather, they emerged as a consequence of idiosyncratic properties that facilitated their discovery by the tools available at the time. While these groundbreaking discoveries provided a paradigm for understanding how genes cause cancer, the actual genes that emerged were not necessarily those that contributed to the greatest number of cancers. For example, studies of *SRC* provided the first critical link between tumorigenic retroviruses and the activation of host cell genes. While the SRC protein is active in many tumors, the mutational activation *SRC* is not a predominant feature of human cancer.

The complete sequencing of the human genome by 2000 facilitated a shift from the functional approaches to gene discovery that had dominated the field of cancer genetics, to the more direct approach of scanning cancer cell genomes in search of mutations. Cancer gene discovery came to rely less on cellular experimentation and more on informatics, the study and processing of large and complex datasets. In the genomic era, new oncogenes are discovered not on the basis of an idiosyncrasy or serendipity, but on the basis of their frequency of mutation in cancers.

An example of an oncogene identified by high throughput DNA sequencing is *PIK3CA*. *PIK3CA* is a member of a family of genes that encode a class of enzymes known as *phosphatidylinositol 3'-kinases* (PI3Ks). The PI3K enzymes first became

a focus of interest to cancer researchers in the 1980s, when it was found that PI3K activity was linked to the protein products of viral oncogenes, such as SRC. PI3K enzymes function in the signaling pathways involved in tissue homeostasis, including cell proliferation, cell death, and cell motility. The organization of these signaling pathways and the role of PI3Ks in cancer phenotypes will be described in detail in Chap. 6.

The known roles of the lipid kinases in cancer-associated cellular processes and the association of these enzymes with known viral oncogenes formed the rationale for the large scale analysis of all genes in this family. As part of an early attempt to scour the genome for cancer genes, a group at Johns Hopkins University used informatics to identify eight members of the PI3K family, on the basis of similarities in their coding sequences. Each of the PI3K genes identified contained a putative kinase domain at its C-terminus. The research team proceeded to sequence the 117 exons that, in total, encoded the kinase domains of each of the PI3K-family members in a panel of colorectal tumors. Recurrent mutations were found in a single family member, *PIK3CA*. Expanding their analysis to include all *PIK3CA* coding exons in nearly 200 tumor samples, the Johns Hopkins group established that *PIK3CA* is mutated in more than 30 % of colorectal cancers.

The majority of mutations that occur in *PIK3CA* during colorectal tumorigenesis are single nucleotide substitutions that result in missense mutations. These mutations do not occur at random points along the *PIK3CA* open reading frame, but rather occur in clusters known as *hot spots*. Most frequently mutated was a helical domain that largely defines the three dimensional structure of the encoded protein. The C-terminus portion of the lipid kinase domain was also mutated in many cancers. The amino acid residues that are affected by hot spot mutations are highly conserved among evolutionarily-related proteins. Functional studies of *PIK3CA* mutants have shown that hot spot mutations cause an increase in the enzymatic activity of the encoded protein.

Subsequent sequencing studies revealed hot spot mutations of *PIK3CA* in brain tumors, and breast, lung, endometrial, urinary bladder and gastric cancers. *PIK3CA* is amplified in significant numbers of ovarian, esophageal and prostate cancers (Table 2.3 and Fig. 2.10). Overall, the mutated and amplified alleles of *PIK3CA* are among the most prevalent of all cancer genes.

Table 2.3 Activating mutations in *PIK3CA*

Cancer type	Mutation frequency (%)
Breast ca.	30–35
Endometrial (uterine) ca.	35
Colorectal ca.	15–30
Gastric ca.	25
Ovarian ca.	<5
Glioma	10
Lung ca.	5–20

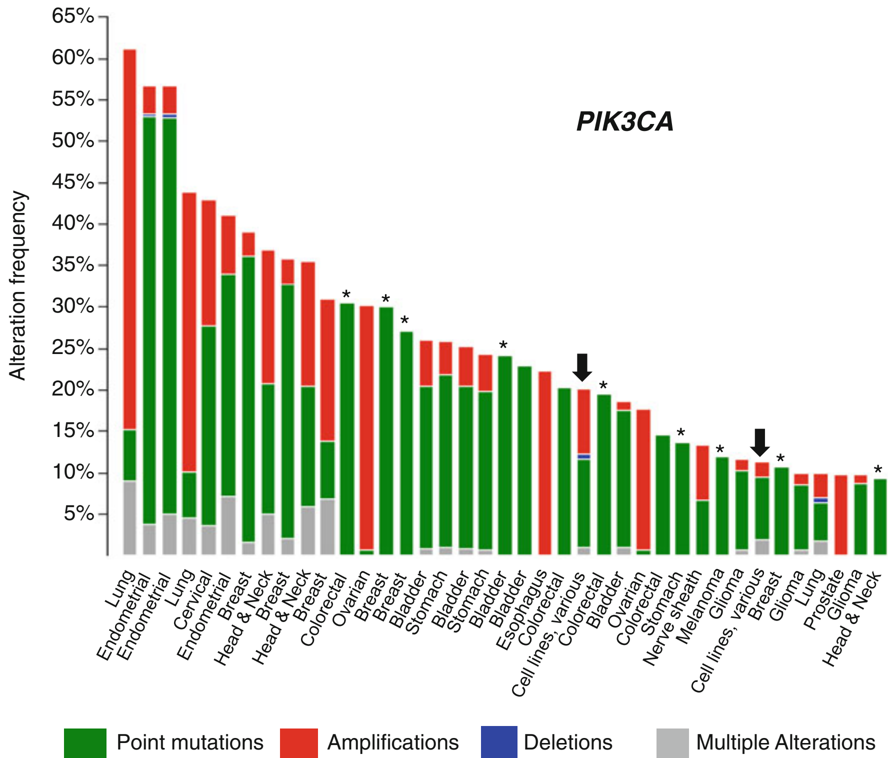


Fig. 2.10 A meta-analysis of *PIK3CA* across tumor types reveals a high frequency of point mutations and amplification. Since its identification as an oncogene in colorectal cancers, *PIK3CA* has been examined in many tumor types. Relatively minor discrepancies between studies can in some cases be attributed to different methodologies and distinct patient populations. However, the frequency of *PIK3CA* mutation in tumors of a given type is largely consistent. Mutations are more prevalent than amplifications. The rare deletions that have been observed are probably incidental or artifactual. * denotes a study in which amplifications were not assessed. ↓ denotes a study of cell line collections that were derived from various tumor types. The results shown here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/

Selection of Tumor-Associated Mutations

Mutations identified via high throughput approaches are not identified on the basis of their function, but on the basis of their sequence. With a high degree of sensitivity and specificity, genomic DNA sequencing can reveal all base changes, passenger mutations and driver mutations alike. The evaluation of the *PIK3CA* mutations found in cancers provides an example of the features that practically distinguish passengers and drivers.

What is the evidence that *PIK3CA* is a cancer gene and not simply a target of passenger mutations? The first and strongest piece of evidence is the high frequency of *PIK3CA* mutations in tumors. Passenger mutations are clonally expanded by chance and thus occur at random, non-recurrent positions. The observed mutations in *PIK3CA* hot spots were found to occur at a rate that was over 100-fold above the background rate of nonfunctional alterations that had previously been observed in colorectal cancer cells. In contrast, analysis of other genes, such as the other members of the PI3K family, has revealed non-recurrent base changes that are consistent with passenger mutations.

A second piece of evidence is the proportion of silent mutations to missense mutations observed. Silent mutations, otherwise known as *synonymous mutations*, should confer no selective advantage because by definition such mutations do not result in changes to the encoded protein. Missense mutations, or *nonsynoymous mutations*, potentially confer a selectable advantage. Among mutations that are propagated by chance alone, nonsynonymous mutations would be expected to occur at a rate that is about two-fold greater than the rate of synonymous mutations. This is simply a function of the numbers of potential bases changes that can occur at random within an open reading frame. The nonsynonymous mutations found in the *PIK3CA* gene occur at a frequency 30 times higher than synonymous mutations in the same gene. This overrepresentation of nonsynonymous mutations suggests that they conferred a selective advantage, and thus contributed to tumorigenesis.

The clustering of *PIK3CA* mutations in evolutionarily conserved hot spots also suggests that these mutations are significant. As described in Chap. 1, evolutionarily-conserved protein elements tend to be fundamental to protein function. Therefore, the frequency of mutation at these key codons provides another convincing piece of evidence that the mutations observed in colorectal tumors are highly likely to confer functional phenotypic changes that, in turn, promote cancer cell growth.

Multiple Modes of Proto-oncogene Activation

There are several ways in which changes to the genome can result in the activation of proto-oncogenes. Whether a mutation results from a small sequence alteration such as a single base substitution, gene amplification, a chromosomal translocation or another more complex gross chromosomal rearrangement, the contribution of an oncogene to tumorigenesis is qualitatively the same. Somatic mutations that activate proto-oncogenes increase the activity of the encoded protein.

Increased protein activity can result from increased levels of gene expression, as we have seen in the examples of the commonly amplified *MYC* and *ERBB2* oncogenes and in the cases in which *MYC* is relocated to a position upstream of a highly active promoter. Alternatively, somatic mutations can result in the expression of a mutant protein. In the case of the *RAS* gene family, activating point mutations cause a loss of regulation and result in constitutive enzymatic activity. In the case of the more complex *BCR-ABL* and *EWS-FLI1* oncogenes, the fusion of unrelated genes

results in both a change in transcriptional activation and in some cases a change in protein structure. Both of these factors can contribute to increased activity of oncogenic proteins.

Another general theme that emerges from a survey of frequently activated oncogenes is that the same oncogene can be activated by different kinds of mutations in different cancers. *MYC* is activated by amplification in a significant proportion of breast and ovarian cancers, but activated by rearrangements in Burkitt lymphoma, and in B-cell and T-cell leukemias. The mechanism of activation in a single cancer type is not always exclusive. While *ERBB2* is most frequently activated by amplification in breast cancers, non-synonymous single base substitutions are found in lower levels in breast, ovarian, gastric, and colorectal cancers. Similarly, *PIK3CA* is activated by single base substitutions in a wide range of carcinomas. While single base substitutions within *PIK3CA* mutations are occasionally found in high grade ovarian carcinomas, a greater proportion of such tumors harbor amplifications of this locus.

In some cases, the causal relationship between a cancer cell lineage and a specific mechanism of proto-oncogene activation is fairly obvious. In the cellular precursors of many leukemias and lymphomas, for example, immune response genes are normally transcriptionally much more active than in any other cell type. It is easy to imagine that a translocation event that results in the juxtaposition of a growth promoting gene such as *MYC* with a transcriptionally active gene would result in a strong selectable advantage, and the outgrowth of that clone.

In most types of cancers, the reason for an apparent bias towards the activation of a proto-oncogene by one mechanism over another is unclear. One important factor, to be discussed in more detail in Chap. 4, is that different types of cancer cells appear to be inherently prone to different kinds of genomic alterations. Some cancers are characterized by gross numerical and/or structural chromosomal abnormalities, while others exhibit a preponderance of changes that occur at the nucleotide level. The acquisition of different forms of genetic instability during tumorigenesis is an important factor in determining the spectrum of somatic mutations present in an advanced cancer.

Oncogenes Are Dominant Cancer Genes

A single nucleotide substitution at a critical position is sufficient to activate a proto-oncogene and convert it to an oncogene. The activating mutation results in a growth advantage, in spite of the continued presence of a normal, unmutated allele in every cell. Because the phenotype conferred by an oncogenic mutation is not influenced by the presence of the remaining wild type allele, oncogenes are, by definition, dominant alleles.

The oncogenic mutations found in a tumor sample are almost never present in the normal cells of that same individual. (The few known exceptions to this pattern are

described in the following section.) Generally, activated oncogenes do not run in the germlines of cancer-prone families. Extensive examination of proto-oncogenes and oncogenes in normal tissues and in cancers has revealed that the mutations that convert proto-oncogenes to oncogenes are almost always acquired by somatic mutation.

Cancer genes can be acquired by somatic mutation or by inheritance. Cancer predisposition is an inherited trait, and therefore the genes that confer this trait must be present in the germline. Oncogenes are not commonly found in the germline and therefore are not a major factor in cancer predisposition. Clearly, this is true for oncogenes that are highly *penetrant*, those that exert strong phenotypic effects regardless of environment or genetic background. Heritable predispositions to cancer development are attributable to different type of cancer gene: the tumor suppressor gene. The nature of these important cancer genes will be described in Chap. 3.

Germline Mutations in *RET* and *MET* Confer Cancer Predisposition

All of the oncogenes described thus far are activated by somatic mutations that occur during tumorigenesis. An interesting exception to this general pattern is the *RET* oncogene, which is somatically mutated in cancers, but is also found in the germline of individuals that are predisposed to inherited cancers of the endocrine system.

Multiple endocrine neoplasia type 2 (MEN2) is a rare, autosomal dominant cancer syndrome. There are several clinically distinct subtypes of this inherited disorder, designated MEN2A, MEN2B and familial medullary thyroid carcinoma (FMTC). Affected individuals most commonly develop an atypical form of thyroid carcinoma which is derived from a population of cells that have an origin in the neural crest. Other endocrine cancers, benign lesions and developmental abnormalities are variably seen in the different MEN2 subtypes.

MEN2-related cancers are caused by germline mutations in the *RET* proto-oncogene. The *RET* proto-oncogene is located on chromosome 10 and contains 21 exons that encode a membrane-bound tyrosine kinase. Like many other oncogenes, *RET* was first discovered during *in vitro* transformation assays using genomic DNA from lymphomas and gastric tumors. The first isolates of this gene were chimeras that had formed during the transfection process, and so the gene was accordingly designated by the acronym for 'rearranged during transfection'. The oncogenic forms of *RET* that have been found in sporadic cancers are similarly rearrangements. These somatic rearrangements vary in different cancers, but commonly put the tyrosine kinase domain in-frame with highly expressed genes, thereby resulting in its constitutive activation. These types of mutations are different from those that cause MEN2.

In contrast to the *RET* mutations found in sporadic cancers, the mutations harbored by individuals affected with MEN2 are usually single nucleotide substitutions. Activating point mutations that convert *RET* into an oncogene typically affect the extracellular domain of the *RET*-encoded protein and lead to ligand-independent activation of the kinase and constitutive activation of downstream mitogenic pathways. (These pathways and the manner in which they related to cancer cell phenotypes will be described in Chap. 5.) Most commonly, mutations in *RET* affect exons 8 and exons 10–16. The precise location of the mutations appears to confer distinct disease phenotypes.

RET is one of a small number of oncogenes that causes an inherited predisposition to cancer. Another oncogene known as *MET* is carried in families affected by hereditary renal cell carcinoma. Like the MEN2 syndomes, hereditary renal cell carcinoma is rare, but highly illustrative of the role that oncogenes can play in some inherited forms of cancer.

The role of the oncogenic forms of *RET* and *MET* in heritable cancers is highly unusual. Activated oncogenes are dominant alleles. As will be extensively described in Chap. 3, the cancer genes that contribute to hereditary forms of cancer are typically recessive alleles that are unmasked during the process of tumorigenesis. Highly penetrant, dominant cancer genes would confer a growth advantage onto every proliferative cell in the body, a phenotype that would presumably compromise the viability of carriers. The role of oncogenes in inherited cancer predisposition is therefore limited.

Proto-oncogene Activation and Tumorigenesis

How do oncogenes fit into the sequence of genetic alterations that underlie tumorigenesis? Activated oncogenes can be found in nearly all colorectal cancers. The oncogenes that commonly contribute to colorectal tumorigenesis are associated with discrete clinico-pathological stages of the disease (Fig. 2.11).

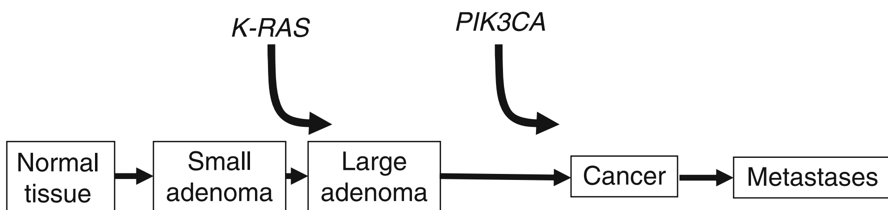


Fig. 2.11 Oncogenes and colorectal cancer progression. Oncogenic mutations in *KRAS* and *PIK3CA* contribute to the later stages of tumor progression. While *KRAS* mutations are occasionally found in very small aberrant crypt foci (ACF), the lesions that harbor these mutations do not appear to progress. In contrast, *KRAS* activation plays an important role in the transition from small to large adenomas, which have significant potential to become malignant. *PIK3CA* mutations are largely found in invasive cancers

The first cancer genes to be firmly associated with colorectal cancers were activated members of the RAS family. Single base substitutions within *KRAS* or less often in *NRAS* are found in approximately 50 % of all colorectal cancers. Among the precancerous lesions, adenomas greater than 1 cm in size exhibit a frequency of *RAS* mutations that is similar to that seen in invasive cancers. In contrast, smaller adenomas (<1 cm) rarely exhibit *RAS* mutations. This finding suggests that *RAS* mutations are acquired during adenoma progression. In support of this hypothesis, dissection of adenomas has revealed small subpopulations in which *RAS* mutations has occurred. Presumably, these subpopulations represent clones that are beginning to progress and that have the potential to give rise to more invasive clones.

Interestingly, *RAS* mutations can be found in some very early lesions arising in the colorectal mucosae. In a distinct histological subset of the earliest lesions, the aberrant crypt foci, *RAS* mutations are found at a high rate. However, such lesions are self-limited and appear to have little, if any, potential for progression. This is a very illuminating finding that underscores a basic principle of tumorigenesis. Clearly, *RAS* mutations can occur in any cell population, but they alone are not sufficient to promote the continued growth of a neoplasm. Rather, the stepwise expansion of tumor cell clones requires a defined sequence of events. While *RAS* mutations appear to be of primary importance in the progression of adenomas to more advanced tumors, this effect is stage-specific and requires prior genetic alterations.

Mutational activation of *PIK3CA* also occurs frequently in colorectal cancers. Like *RAS* mutations, *PIK3CA* mutations are not often found in early stage tumors. Rather, *PIK3CA* mutations most often arise late in the process of tumorigenesis, when a tumor begins to invade surrounding normal tissues. One can infer that the increased activity of the *PIK3CA*-encoded protein provides a survival advantage to cancer cells as they penetrate the barriers that physically separate tissue compartments.

Further Reading

Bishop JM (1981) Enemies within: the genesis of retrovirus oncogenes. Cell 23:5–6

Epstein MA (2001) Historical background. Philos Trans R Soc Lond B Biol Sci 356:413–420

Garraway LA, Sellers WR (2006) Lineage dependency and lineage-survival oncogenes in human cancer. Nat Rev Cancer 6:593–602

Martin GS (2004) The road to Src. Oncogene 23:7910–7917

Mitelman F, Johansson B, Mertens F (2007) The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 7:233–245

Rowley JD (2001) Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer 1:245–250

Sawyers CL (1999) Chronic myeloid leukemia. N Engl J Med 340:1330–1340

Schwab M (1999) Oncogene amplification in solid tumors. Semin Cancer Biol 9:319–325

Thomas RK et al (2007) High-throughput oncogene mutation profiling in human cancer. Nat Genet 39:347–351

Vogt PK (2010) Retroviral oncogenes: a historical primer. Nat Rev Cancer 12:639–664

Weinberg RA (1997) The cat and mouse games that genes, viruses, and cells play. Cell 88:573–575

http://www.springer.com/978-94-017-7482-6

Principles of Cancer Genetics
Bunz, F.
2016, XII, 343 p. 138 illus., 29 illus. in color., Hardcover
ISBN: 978-94-017-7482-6