Topic Introduction

Oncogenomics Methods and Resources

Simon J. Furney, Gunes Gundem, and Nuria Lopez-Bigas

Today, cancer is viewed as a genetic disease and many genetic mechanisms of oncogenesis are known. The progression from normal tissue to invasive cancer is thought to occur over a timescale of 5–20 years. This transformation is driven by both inherited genetic factors and somatic genetic alterations and mutations, and it results in uncontrolled cell growth and, in many cases, death. In this article, we review the main types of genomic and genetic alterations involved in cancer, namely copy-number changes, genomic rearrangements, somatic mutations, polymorphisms, and epigenomic alterations in cancer. We then discuss the transcriptomic consequences of these alterations in tumor cells. The use of “next-generation” sequencing methods in cancer research is described in the relevant sections. Finally, we discuss different approaches for candidate prioritization and integration and analysis of these complex data.

INTRODUCTION

The role of genetic alterations in tumor cells was first introduced in the early 20th century (Ponder 2001). By the early 1970s, it had been showed that viruses could promote cellular transformation in vitro, and Knudson (1971) had described his hypothesis of two genetic events in the rare cancer retinoblastoma, eventually shown to be caused by loss of both alleles of the tumor suppressor gene (Cavenee et al. 1983). Further experimental studies showed point mutations to be the mechanism of activation in oncogenes (Reddy et al. 1982; Tabin et al. 1982). However, at this stage, environmental influences were still viewed as the cause of common cancers (Doll and Peto 1981; Peto 2001). Now, cancer is thought of as a genetic disease and many genetic mechanisms of oncogenesis have been described (Vogelstein and Kinzler 2004).

THE GENETIC BASIS OF CANCER

The transformation of a normal cell into a cancer cell is a multistep process, with each intermediate stage conferring a selective advantage on the cell (Vogelstein and Kinzler 1993). These changes result primarily from irreversible aberrations in the DNA sequence or structure (e.g., translocations, mutations, and copy-number alterations) (Fig. 1). However, cancer alterations also include potentially reversible changes, known as epigenetic modifications, to the DNA and/or histone proteins, which are closely associated to the DNA in chromatin (Esteller 2008). Normal cellular homeostasis and division are tightly controlled processes that incorporate signals from many pathways to regulate the expression of the appropriate genes. Mutations or alterations to genes involved in these processes can contribute to cellular transformation by unbalancing the natural physiological equilibrium of a cell.

Adapted from *Genetics of Complex Human Diseases* (ed. Al-Chalabi and Almasry). CSHL Press, Cold Spring Harbor, NY, USA, 2009.

© 2012 Cold Spring Harbor Laboratory Press

Cite this article as *Cold Spring Harb Protoc*; 2012; doi:10.1101/pdb.top069229

FIGURE 1. Main genomic and epigenomic alterations identified in tumor samples. These alterations have consequences at the levels of gene expression and alteration of protein functions. These aberrations and their consequences allow cancer cells to acquire key capabilities for their status (Hanahan and Weinberg 2000).

Indeed, cancer progression is the accumulation of a series of genetic alterations in a somatic cell (Vogelstein and Kinzler 2004).

Although there are many different types of cancer (and even subtypes within the same tissue) that result from the action of different sets of genes (Dyrskjot et al. 2003), it has been suggested that the combinations of genes required for oncogenesis can be reduced to six essential changes in cellular physiology (Hanahan and Weinberg 2000): self-sufficiency in growth signals, insensitivity to growth inhibitory signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. The requirement for uncoupling of the normal processes that result in these six alterations shows the complex genetic nature of cancer. It has further been proposed that the chronological order of these alterations is not fixed and can vary among different cancer types (Hanahan and Weinberg 2000).

The genetic alterations that lead to cancer occur only in certain genes. Cancer-causing genes have been traditionally classified as either proto-oncogenes (e.g., the genes for MYC, ERBB2 [Her-2/neu], and EGFR) or tumor suppressor genes such as the genes that encode TP53, CDKN2A, and RB. Proto-oncogenes normally function as proliferative agents, and when mutated or misregulated in cancer, they promote uncontrolled cell growth. Usually they are phenotypically dominant, requiring a gain-of-function mutation or chromosomal gain to become oncogenic. Conversely, tumor suppressor genes are endowed with antiproliferative properties and generally require inactivation of both alleles to induce cancer. This can occur, for example, by point mutation, deletion, or epigenetic silencing. In addition to proto-oncogenes and tumor suppressor genes, stability genes (e.g., base excision repair and mismatch repair genes), which keep genetic alteration to a minimum, have been proposed more recently as an additional type of cancer gene (Vogelstein and Kinzler 2004).

In the last decade, the study of the genetic basis of cancer has undergone a profound transformation. Until recently, most cancer genes had been identified by positional cloning (Futreal et al. 2004), and scientists were focused on studying particular candidate genes involved in oncogenesis. Today, high-throughput techniques allow scientists to simultaneously analyze a large number of genes and their alterations. Cytogenetic methods such as comparative genome hybridization (CGH) have been used to analyze structural changes and genome-wide gains and losses. The use of cDNA microarrays to simultaneously analyze the expression of thousands of genes in tumor samples has become prevalent in cancer research. Studies have shown that gene-expression data from tumors are clinically relevant in breast cancer and lymphoma prognosis (van't Veer et al. 2002; Dave et al. 2004) and are able to define cancer subtypes and response to therapies (Ramaswamy and Golub 2002). The use of

mutational profiling of tumor genomes has yielded important results during the past few years (Benvenuti et al. 2005). Large-scale exon resequencing of human tumors has been used to identify point mutations in candidate cancer genes in a variety of different tumors (Davies et al. 2002, 2005; Bardelli et al. 2003; Stephens et al. 2004; Sjoblom et al. 2006; Greenman et al. 2007; Wood et al. 2007; Jones et al. 2008; McLendon et al. 2008; Parsons et al. 2008), and new high-throughput methods for DNA methylation and histone modification profiling are being used to identify epigenomic alterations in cancer (American Association for Cancer Research and the European Union Network of Excellence Scientific Advisory Board 2008; Esteller 2008). In addition, several major projects that aim to identify all genetic alterations in common tumor types using genome-wide, high-throughput techniques are in progress, for example, The Cancer Genome Atlas (<http://cancergenome.nih.gov>) of the National Institutes of Health (NIH), the Cancer Genome Project (<http://www.sanger.ac.uk/genetics/CGP/>) at the Sanger Institute, and the International Cancer Genome Consortium (<http://www.icgc.org/>).

These genome-wide, high-throughput technologies have transformed the field of cancer research and have provided powerful ways to understand the mechanism of disease pathogenesis. They also have the potential to identify possible targets for therapy, discover molecular biomarkers that allow early detection of cancer, improve the diagnosis and prognosis of certain tumors, and predict the response to therapies (Baak et al. 2005; Chin and Gray 2008). However, these technologies also yield large volumes of data of multiple types. One of the main challenges is to distinguish between alterations that are causative (driver alterations) from those that are the consequences of the large number of cell divisions coupled with genome instability and checkpoint errors characteristic of cancer cells (passenger alterations) and are not directly involved in tumor development. New methods are needed to be able to prioritize the more promising candidates from genes that are unlikely to be contributing to tumorigenesis (Haber and Settleman 2007; Higgins et al. 2007; Furney et al. 2008a).

Analysis at the level of individual genes is informative, but it does not capture the full complexity of biological systems. Thus, it is also important to study the alterations identified in cancer cells at a more general level. One way to approach this is by embedding genes into functional or regulatory modules and focusing on the study of altered modules instead of single genes. Some of these approaches have been used in the analysis of microarray data, for example, the “module maps” (Ihmels et al. 2002; Segal et al. 2004; Tanay et al. 2004) and “molecular concept maps” (Tomlins et al. 2007).

Recently, more sophisticated studies exploiting data from different techniques and different types of alterations are becoming common in cancer research. A number of studies have revealed the effectiveness of integrative functional genomics in cancer research, in which information from complementary experimental data sources is combined to provide greater insight into the process of tumorigenesis (Rhodes et al. 2004; Bild et al. 2006; Carter et al. 2006; Liu et al. 2006; Stransky et al. 2006; Tomlins et al. 2007; Jones et al. 2008; McLendon et al. 2008; Parsons et al. 2008).

A SURVEY OF GENOMIC AND GENETIC ALTERATIONS IN CANCER

Analyzing Copy-Number Changes in Cancer

Aneuploidy in tumor cells, particularly in human cancers, is observed frequently. Cytogenetic methods such as karyotyping, fluorescence in situ hybridization (FISH), and CGH (Kallioniemi et al. 1992) have been used with great effect to analyze large structural chromosomal changes, gains, and losses of specific genes, and genome-wide gains and losses in cancer. In the last decade, array-based CGH (aCGH) (Pinkel et al. 1998) has become the technique of choice for investigating copy-number changes in cancer research and has been used to classify tumors, identify markers, and delineate the structure of chromosomal aneuploidies (Kallioniemi 2008). Recently, meta-analyses of CGH data from tumors have shown that tumors can be classified using these data (Baudis 2007; Jong et al. 2007). Access to the results of many CGH studies is provided in curated online databases such as the National Center for Biotechnology Information (NCBI)/National Cancer Institute (NCI)’s Cancer Chromosomes (Knutson et al. 2005) and Progenetix (Baudis and Cleary 2001).

The development of high-resolution single-nucleotide polymorphism (SNP) arrays has facilitated surveying of copy-number changes at a higher resolution and the detection of loss of heterozygosity (Mullighan et al. 2007; Weir et al. 2007). For instance, Mullighan et al. (2007) have applied this technology in more than 200 cases of pediatric acute lymphoblastic leukemia to identify a range of somatic deletions and amplifications.

Finding Genomic Rearrangements

Chromosomal translocations and subsequent gene fusion events have an important role in the initial steps of tumorigenesis. About 360 different gene fusion events have been identified (Mitelman et al. 2007). Translocations are recognized as a common mechanism of oncogenesis in leukemias and lymphomas (Mitelman et al. 1997; Rowley 1998), whereas relatively few translocations have been detected in solid tumors (Mitelman 2000). This is probably not because they are uncommon in solid tumors but because of technical and analytical limitations reflecting the complex genomic profiles and heterogeneous nature of these malignancies (Mitelman et al. 2007). Perhaps the most well-known chromosomal translocation in cancer is the Philadelphia chromosome discovered by Peter Nowell and David Hungerford in 1960 (Nowell 2007). Prior cytogenetic and molecular studies showed that it consisted of a translocation between chromosomes 9 and 22, resulting in a chimeric, constitutively active tyrosine kinase BCR–ABL fusion protein that is responsible for chronic myeloid leukemia (Groffen et al. 1984; Shtivelman et al. 1985).

Effects of Translocations in Cancer

At the molecular level, the effect of most of the translocations involved in cancer can be attributed to one of the following mechanisms: (1) Translocations can create chimeric proteins due to the fusion of parts of two genes, one in each breakpoint, as in the case of the BCR–ABL fusion protein (Groffen et al. 1984; Shtivelman et al. 1985). As a result of this fusion, the activity of the nonreceptor tyrosine kinase ABL is misregulated. This case is particularly relevant because of the effectiveness of the drug Gleevec (imatinib mesylate), which inhibits tyrosine kinase activity, in combating this type of cancer (Druker 2002). Numerous other translocations resulting in fusion proteins have also been described (Rabbitts 1994; Mitelman 2000; Rowley 2001). (2) Translocations can result in the misregulation of one of the genes involved in the fusion event by placing it close to the regulatory elements of another gene. This usually results in the ectopic expression of an apparently normal gene. Examples of these cases are common translocations (between chromosomes 8 and 2, 14, or 22) present in Burkitt's lymphoma that place the *MYC* gene close to an immunoglobulin gene, encoding either the heavy chain (IGH) or the kappa (IGK) or lambda (IGL) light chains. As a consequence of the translocation, the *MYC* gene becomes constitutively expressed because of the influence of regulatory elements of the immunoglobulins (Kuppers 2005).

The Mitelman Database of Chromosome Aberrations in Cancer (now part of the NCBI/NCI Cancer Chromosomes Database) catalogs chromosomal aberrations and relates them to tumor characteristics (Mitelman 2009). This database is manually curated from published literature by its authors.

Methods for Detecting Chromosomal Rearrangements

Numerous methods exist for the detection of chromosomal rearrangements (for review, see Morozova and Marra 2008). The earliest methods applied involved examination of chromosomes and chromosome banding patterns by microscopy. An important advance in molecular cytogenetics was the development of in situ hybridization techniques (Buongiorno-Nardelli and Amaldu 1970). This procedure is based on the hybridization of a labeled probe to a complementary target where probe copy number is assessed by microscopy. Some developments of the classical FISH methods are multiplex FISH (M-FISH) (Speicher et al. 1996; Speicher and Ward 1996), spectral karyotyping (SKY) (Schrock et al. 1996), and combined binary ratio labeling (Tanke et al. 1999), which allow the simultaneous

display of all chromosomes in 24 colors. FISH techniques are adequate to detect gross chromosomal aberrations; however, they are limited for smaller-scale chromosomal aberrations.

More recently, Arul Chinnaiyan and colleagues have applied a new integrative analytical methodology called cancer outlier profile analysis (COPA; MacDonald and Ghosh 2006). This method, which identifies associations between genomic and transcriptional abnormalities, allowed them to identify a family of common translocations in prostate cancer that brings ETS family genes under the control of TMPRSS2, in effect placing the expression of these genes under androgen-mediated regulation (Tomlins et al. 2005, 2006).

Sequencing approaches have also been developed for the detection of chromosomal aberrations. In this case, DNA from a tumor is cloned into a large insert, and the ends of the resultant clones are sequenced and then mapped onto the reference human DNA sequence. Paired ends that map farther apart than the maximum size tolerated by the clone indicate the presence of a structural aberration (Volik et al. 2003, 2006; Krzywinski et al. 2007). More recently, the combination of ultrafast DNA sequencing and bioinformatics allows high-resolution and massive paired-end mapping (PEM) (Korbel et al. 2007). This technique consists of the isolation of 3-kb sequence fragments and then end sequencing with 454/Roche technology, followed by mapping of paired-end reads back to the reference sequence using a computational algorithm developed by the authors. Campbell and colleagues have used this approach to identify structural variants in the genome of germ-line and lung cancer cells of two individuals. This analysis allowed the identification of 306 germ-line structural variants and 103 somatic rearrangements to the base-pair level of resolution (Campbell et al. 2008). In addition, Maher et al. (2009) have used a combination of high-throughput long- and short-read transcriptome sequencing to identify known and novel fusion transcripts in cancer cell lines and tumors.

Somatic Mutations in Cancer

Somatic mutations are alterations in the nucleotide sequence of a gene, such as single base-pair changes as well as those creating small insertions or deletions. Mutations can be classified in a variety of ways: (1) silent (no net effect on the amino acid code), missense (change of the original amino acid codon to another), or nonsense (change of the original amino acid codon to a stop codon); (2) loss of function (the function is lost or weakened) or gain of function (the protein becomes more active or gains a new or abnormal function); or (3) transition and transversion.

Mutational Patterns

Different types of mutations affect genes altered in cancers. However, one can draw some generalizations from the mutational patterns observed. For example, oncogenes usually undergo gain-of-function mutations. A typical example is BRAF. One of the most common changes observed in this kinase is the conversion of a valine to a glutamate at codon 599 within the activation loop of the kinase domain. This substitution leads to the constitutive activation of the protein product even in the absence of an activating signal. The “turned-on” BRAF kinase phosphorylates downstream targets leading to abnormal growth (Wan et al. 2004). On the other hand, tumor suppressor genes are usually rendered nonfunctional by loss-of-function mutations. A point mutation in TP53 inactivates its capacity to bind to the sequences it regulates transcriptionally (Vogelstein et al. 2000). “Disabled” TP53 cannot do its normal job of inhibiting cell growth and stimulating cell death in times of stress.

Databases

As data have accumulated, the results from mutational analysis studies have been stored in online databases. Some of these focus on a specific gene (p53 database, <http://www-p53.iarc.fr/>, Olivier et al. 2002); EGFR (<http://www.somaticmutations-egfr.org/>); whereas others are tissue-specific (Breast Cancer Mutations Database, <http://research.nhgri.nih.gov/bic/>). COSMIC (Catalogue of Somatic Mutations in Cancer, <http://www.sanger.ac.uk/genetics/CGP/cosmic>), on the other hand, stores somatic mutations that have been reported in the literature regarding many cancer types (Forbes et al. 2006).

Sequencing and Mutational Screens

Initial large-scale sequencing efforts focused on signaling pathways previously known to be mutated in at least one gene (Davies et al. 2002; Rajagopalan et al. 2002). In addition to well-known pathways, specific gene families have been scrutinized: the tyrosine kinases (Bardelli et al. 2003), lipid kinases (Samuels et al. 2004), tyrosine phosphatases (Wang et al. 2004), and tyrosine kinase receptors (Paez et al. 2004). These and similar studies pointed to the importance of kinase and phosphatase mutations and led to the identification of some important genes such as *PI3KCA*, *BRAF*, *EGFR*, and *JAK2* in many tumors.

The first report on the genomic landscape of somatic mutations focused on human breast and colorectal cancers (Sjjoblom et al. 2006). A two-stage strategy was followed in this study. In the discovery screen, the authors performed mutational screens for the consensus coding sequences (CCDS) in 11 breast and 11 colorectal tumors. The putative mutations were filtered to exclude silent changes, changes present in normal samples, known polymorphisms from dbSNP (Single Nucleotide Polymorphism Database, <http://www.ncbi.nlm.nih.gov/projects/SNP>), false-positive calls on visual inspection of sequence chromatograms, and confirmation by resequencing. The mutations passing all these criteria were sequenced again in a validation screen in 24 additional breast and colorectal tumors. After filtering as before, 921 and 751 mutations were identified in breast and colorectal cancers, respectively. In all, 92% of the mutations were single-base substitutions, the majority of which were missense. There were significant differences in the mutational spectra of the two tumor types at CG base pairs: Colorectal cancer samples were biased in TA transitions, whereas breast cancer samples were prone to GC transversions. A total of 44% and 11% of the colorectal mutations occurred in 5'-CG-3' and 5'-TpC-3' sites, respectively; these numbers were 17% and 31% for breast mutations. This result implies that there might be differences in the mechanisms of mutagenesis in the two tumor types.

To discriminate the “driver” mutations from the “passenger” mutations, a cancer mutation prevalence score was calculated as follows: Mutations were divided into different categories taking into account the type of the base mutated, the resulting base change, the 5' and 3' neighbors, and the codon usage. This resulted in the identification of 122 and 69 candidate genes for the breast and colorectal tumors, respectively. In these genes, some biological functions were overrepresented in the candidate genes, such as transcription factors and cell-adhesion- and signal-transduction-related genes.

Overall, this first large-scale sequencing effort revealed that the majority of the genes identified had not been previously known to have been mutated. In addition, different genes were mutated in breast and colorectal cancers. These genes also showed different biases in the type of nucleotide substitutions. Moreover, even the samples of the same cancer type were very heterogeneous, which might be the reason why gene sets related in a biologically meaningful way can explain prognosis, response to therapy, etc., better than individual genes.

Difficulties in Predicting Candidate Cancer Genes

The Sjjoblom et al. (2006) study, however, raised a lot of questions. Some were skeptical about the usefulness of the brute-force sequencing projects, emphasizing the importance of focusing on reverse engineering approaches (Loeb and Bielas 2007; Strauss 2007). Critics compared the high costs of such large-scale projects with the limited results obtained. It is true that high-throughput approaches cannot replace functional studies, but such bioinformatic screenings can guide experimental studies in more efficient directions, especially with the advent of more cost-efficient technologies. Other discussions centered around the robustness of the statistical methods, the background mutations rates, and the small sample sizes used (Forrest and Cavet 2007; Getz et al. 2007; Rubin and Green 2007). All of these are important factors that can affect the resulting genes found to be significantly mutated.

In another study, the Sjjoblom et al. (2006) analysis was extended to include all RefSeq genes (Wood et al. 2007). Using the same methods, 1718 genes with at least one nonsynonymous mutation

in either breast or colorectal cancer were identified. The mutation spectra of the two tumor types were similar to those of the previous analysis. Comparison of these with the sequencing of pancreas and brain tumors (Jones et al. 2008; Parsons et al. 2008) indicated that breast tumors have a somatic mutation spectrum different from that of the other three, with a relatively high number of mutations at 5'-TpG sites and a small number at 5'-CpG sites.

Of the 1718 genes with nonsynonymous mutations, 280 were predicted to be candidate cancer genes. One of the conclusions the authors reached was that very few genes are mutated at high frequencies in human cancers ("mountains" in the mutational landscape). These genes (e.g., *TP53*, *PTEN*, and *PIK3CA*) might have critical roles in tumorigenesis. On the other hand, a much larger number of genes are mutated at low frequencies. This indicates that a large number of the mutations confer only small advantages to the tumorigenic phenotype. However, this view also points to the difficulty of discrimination of driver mutations from passengers. Recently, Ding et al. (2008) sequenced 623 genes in 188 human lung adenocarcinomas, identifying 26 genes that were mutated at significantly high frequencies.

Common Variants in Cancer

The International HapMap project (International HapMap Consortium 2005) has facilitated the recent explosion of genome-wide association studies (GWAS) attempting to determine common variants (in general, SNPs) that contribute to common diseases. Many of these GWAS have identified SNPs associated with different tumor types. We mention only some of these studies below because it is not feasible to provide a comprehensive review of the field within the scope of this article.

In breast cancer, Cox et al. (2007) found a common coding variant in caspase 8 to be associated with an increased risk of the disease. Easton et al. (2007) identified five novel loci, including *FGFR2* and *TOX3*, showing genome-wide significant association with breast cancer. The *CASP8* and *TOX3* associations were independently confirmed by Tapper et al. (2008), who also identified SNPs in six genes associated with disease prognosis. Further recent studies have found associations at a number of genomic loci (Ahmed et al. 2009; Thomas et al. 2009; Zheng et al. 2009).

Studies of prostate cancer have also identified a number of loci associated with the disease, including independent replications of a risk locus at chromosome 8q24 (Gudmundsson et al. 2007, 2008; Haiman et al. 2007a,b; Yeager et al. 2007; Thomas et al. 2008). In addition, genome-wide associations have been identified in a number of other tumor types such as lung cancer (Amos et al. 2008; Y. Wang et al. 2008), chronic lymphocytic leukemia (Di Bernardo et al. 2008), colorectal cancer (Houlston et al. 2008; Tenesa et al. 2008), urinary bladder cancer (Kiemeny et al. 2008), diffuse cancer-type gastric cancer (Sakamoto et al. 2008), and basal cell carcinoma (Stacey et al. 2008), and also in multiple tumor types (Rafnar et al. 2009).

Epigenomic Alterations in Cancer

Epigenetic alterations are increasingly being recognized as central mechanisms of tumor development. Modifications of the DNA methylation landscape as well as of histone modifications seem to be a common feature of many tumor samples (Esteller 2007, 2008).

Types of Epigenetic Changes

The low level of DNA methylation in tumors compared to that in normal tissue counterparts was one of the first epigenetic alterations to be found in human cancer (Feinberg and Vogelstein 1983). This hypomethylation occurs mainly in gene-poor areas (Weber et al. 2005). The proposed mechanisms by which genome hypomethylation can contribute to the development of a cancer cell are generation of chromosomal instability (Eden et al. 2003), reactivation of transposable elements (Bestor 2005), and loss of imprinting (Feinberg 1999; Cui et al. 2003; Kaneda and Feinberg 2005).

In contrast, hypermethylation of CpG islands in promoter regions of certain genes (tumor suppressor genes) is an important event in many cancers. This is the case regarding the retinoblastoma tumor suppressor gene (*Rb*) (Greger et al. 1989; Sakai et al. 1991), *P16^{INK4a}* (Herman et al. 1994, 1995;

Gonzalez-Zulueta et al. 1995; Merlo et al. 1995), *hMLH1* (Herman and Baylin 2003), and *BRCA1* (breast cancer susceptibility gene 1) inactivation (Herman and Baylin 2003).

Histone modifications (such as acetylations or methylations) have direct effects on the regulation of gene transcription. Generally, histone acetylation is associated with transcriptional activation (Bernstein et al. 2007; Mikkelsen et al. 2007); however, the histone methylation effect depends on the residue modified (Bernstein et al. 2007; Mikkelsen et al. 2007). It is becoming clear that combinations of histone modifications have an effect on transcriptional regulation (Z. Wang et al. 2008).

Several lines of evidence point to the importance of alterations in histone modification as relevant steps in the transformation process. Examples include the association between CpG island hypermethylation in cancer and a particular combination of histones markers, namely deacetylation of histones H3 and H4, loss of histone H3 lysine K4 (H3K4) trimethylation, and gain of H3K9 methylation and H3K27 trimethylation (Fahrner et al. 2002; Ballestar et al. 2003; Vire et al. 2006). In addition, it has been observed that cancer cells undergo a general loss of monoacetylated and trimethylated forms of histone H4 (Fraga et al. 2005). However, it is thought that the main findings on the extent and implications of epigenomics in cancer are still to come in the future with the development of the international Human Epigenome Project (American Association for Cancer Research Human Epigenetic Task Force 2008) (<http://www.epigenome.org/>).

Methods for Detecting Epigenetic Modifications

Several approaches are available to study epigenetic modifications in normal and cancer cells. Some of these profile epigenetic alteration in a genome-wide manner, whereas others are centered in gene-specific alterations.

High-performance liquid chromatography and high-performance capillary electrophoresis allow the quantification of the total amount of 5-methylcytosine (Fraga and Esteller 2002; Esteller 2007). The study of DNA methylation at particular sequences has classically been based on the action of restriction enzymes that can distinguish between methylated and unmethylated recognition sites (Esteller 2007). Later, methods based on the use of bisulfite treatment of DNA, which changes unmethylated cytosines to uracil and leaves methylated cytosines unchanged, were developed (Clark et al. 1994; Herman et al. 1996). These methods can be coupled with polymerase chain reaction (PCR) and sequencing of candidate genes. They can also be combined with genomic approaches to detect genome-wide DNA methylation patterns, for example, by using promoter microarrays or arbitrary primed PCR, in which no prior sequence information is required for amplification.

In addition, techniques can be used that are based on chromatin immunoprecipitation (ChIP), with the ChIP-on-chip approach using antibodies against methyl-CpG-binding domain proteins (MBDs) (Lopez-Serra et al. 2006), which have a great affinity for binding to methylated cytosines. An antibody directly against 5-methylcytosine (methyl-DIP) can also be used (Weber et al. 2005; Keshet et al. 2006).

Another way of assessing genome-wide DNA methylation patterns is by using gene-expression profiling microarrays comparing mRNA levels from cancer cell lines before and after treatment with a demethylating drug (Suzuki et al. 2002; Yamashita et al. 2002). However, this method yields a significant amount of false positives, requiring confirmation by bisulfate genomic sequencing.

The profiling of histone modification marks is typically studied by ChIP using antibodies against specific histone modifications. The immunoprecipitated DNA is then analyzed by PCR with specific primers to investigate the presence of a candidate DNA sequence or on a microarray chip (ChIP-on-chip) to profile an extensive map of histone modifications (Azuara et al. 2006; Bernstein et al. 2007). More recently, ChIP has been combined with ultrasequencing techniques (ChIP-seq) to obtain higher resolution chromatin modification maps (Z. Wang et al. 2008).

Databases

Several databases have been created to collect and annotate alterations in DNA methylation (Table 1). DNA Methylation Database (MethDB, <http://www.methdb.de>) is a well-maintained resource

TABLE 1. Resources and databases for oncogenomics

Name	Description	Web address
Mutations in cancer		
IARC TP53 mutation database	Compiles all TP53 gene variations identified in human populations and tumor samples. Data compiled from peer-reviewed literature and generalist databases.	http://www-p53.iarc.fr
EGFR mutations database	Comprehensive compendium of all somatic mutations in EGFR that have been identified in human cancers and reported in peer-reviewed literature.	http://www.somaticmutations-egfr.org/
Breast cancer mutations database	Central repository for information regarding mutations and polymorphisms in breast cancer susceptibility genes.	http://research.nhgri.nih.gov/bic/
COSMIC	Catalogue of Somatic Mutations in Cancer.	http://www.sanger.ac.uk/genetics/CGP/cosmic
Structural alterations in cancer		
Progenetix	Overview of copy-number abnormalities in human cancer from CGH experiments. A curated database, it collects genomic gain/loss information of individual cancer and leukemia cases published in peer-reviewed journals.	http://www.progenetix.net/progenetix
Mitelman database of chromosome aberrations in cancer	Relates chromosomal aberrations to tumor characteristics, based on either individual cases or associations.	http://cgap.nci.nih.gov/Chromosomes/Mitelman
NCBI/NCI's cancer chromosomes	Integrates NCI/NCBI SKY/M-FISH and CGH database, NCI Mitelman database of chromosome aberrations in cancer, and NCI recurrent aberrations in cancer database.	http://www.ncbi.nlm.nih.gov/sites/entrez?db=cancerchromosomes
DNA methylation alterations in cancer		
MethyCancer	Database of human DNA methylation and cancer; it collects data from other public databases and resources and integrates this information with CpG island prediction and expression data.	http://methycancer.psych.ac.cn
Transcriptomic alterations in cancer		
GEO	Gene-expression/molecular abundance repository supporting MIAME-compliant data submissions and a curated, online resource for gene-expression data browsing, query, and retrieval. Not specific for cancer, but contains many cancer data sets.	http://www.ncbi.nlm.nih.gov/geo/
ArrayExpress	Public archive for transcriptomics data aimed at storing MIAME- and MINSEQE-compliant data in accordance with MGED recommendations. ArrayExpress Warehouse stores gene-indexed expression profiles from a curated subset of experiments in the archive. Like GEO, not specific for cancer but contains many cancer data sets.	http://www.ebi.ac.uk/microarray-as/ae/
Oncomine	Cancer microarray database and web-based data-mining platform aimed at facilitating discovery from genome-wide expression analyses.	http://www.oncomine.org/
Integrative projects		
The Cancer Genome Atlas	Comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through application of genome analysis technologies, including large-scale genome sequencing.	http://tcga.cancer.gov/
Cancer Genome Project	Sanger Center Project aimed at identifying somatically acquired sequence variants/mutations and, hence, genes critical in the development of human cancers.	http://www.sanger.ac.uk/genetics/CGP/
International Cancer Genome Consortium	International consortium with the goal of obtaining comprehensive description of genomic, transcriptomic, and epigenomic changes in 50 different tumor types and/or subtypes of clinical and societal importance across the globe.	http://www.icgc.org/
NCI-CGAP	Interdisciplinary program established and administered by the NCI to generate information and technological tools needed to decipher molecular anatomy of the cancer cell.	http://www.ncbi.nlm.nih.gov/ncicgap/
Integrative resources		
IntOGen	Discovery tool for cancer researchers; integrates multidimensional OncoGenomics Data for the identification of genes and groups of genes (biological modules) involved in cancer development.	http://www.intogen.org

Continued

TABLE 1. *Continued*

UCSC Cancer Genomics Browser	Suite of web-based tools to integrate, visualize, and analyze cancer genomics and clinical data. Displays a whole-genome and pathway-oriented view of genome-wide experimental measurements for individual and sets of samples alongside their associated clinical information.	http://genome-cancer.ucsc.edu
Other resources		
Cancer Gene Census	Ongoing effort to catalog those genes for which mutations have been causally implicated in cancer.	http://www.sanger.ac.uk/genetics/CGP/Census/
CancerGenes	Resource to simplify the process of gene selection and prioritization in large collaborative projects. Combines gene lists annotated by experts with information from key public databases.	http://cbio.mskcc.org/CancerGenes
CGPrio	Resource for the prioritization of candidate cancer genes after genomic experiments. Prioritization of oncogenes and tumor suppressor genes is based on computational classifiers that use different combinations of sequence and functional data, including sequence conservation, protein domains and interactions, and regulatory data.	http://bg.upf.edu/cgprio

Abbreviations: CGH: Comparative genomic hybridization; EGFR: Epidermal growth factor receptor; MGED: Microarray Gene Expression Database; MIAME: Minimum information about a microarray experiment; MINSEQE: Minimum information about a high-throughput sequencing experiment; NCBI: National Center for Biotechnology Information; NCI: National Cancer Institute

that stores DNA methylation data in a standard format (Grunau et al. 2001). In addition, specialized databases focus on methylation aberrations detected in cancer samples: PubMeth (<http://www.pubmeth.org>; Ongenaert et al. 2008), MeInfoText (<http://mit.lifescience.ntu.edu.tw/index.html>; Fang et al. 2008), and MethyCancer (<http://methycancer.psych.ac.cn>; He et al. 2008). MethyCancer collects data from other public databases and resources, including MethDB, and integrates this information with CpG island prediction and expression data. PubMeth and MeInfoText extract information from MedLine publications using text mining and manual curation.

TRANSCRIPTOMIC CHANGES IN TUMORS

The result of the cumulative effect of the different alteration types we have described is observed at the level of expression of the gene product. For example, genomic copy-number loss and epigenetic silencing may account for the down-regulation of the micro RNA (miRNA) gene expression, which further contributes to a genome-wide transcriptional deregulation at the level of mRNAs (Zhang et al. 2008). Therefore, to paint a complete picture of tumorigenesis, it is crucial to include changes at the expression level of both miRNAs and mRNAs. Actually, the use of high-throughput gene-expression profiling studies of tumorigenic cells has been used extensively and has changed cancer research substantially.

Methods for Detecting Transcriptomic Changes

Although it has long been known that tumor cells express some genes at abnormal levels, these large-scale expression studies showed that large numbers of genes are differentially expressed in cancer cells. Given that changes in expression are a reflection of the underlying complexity of different alterations, it is no surprise that high-throughput expression analysis is extremely difficult. How should long lists of deregulated genes be interpreted? How should one decide which of the transcriptionally deregulated genes are causally implicated in cancer?

Expression Analysis

One suggestion has come from “gene signature” studies. Instead of a single gene, tumorigenic phenotypes can be explained by the signature defined by the expression level of a list of genes. To identify groups of genes that change in expression, “unsupervised methods” have proved to be very useful.

Without any a priori information, these methods can help to discover patterns in the data. These methods led to the characterization of previously unknown, but clinically significant, subtypes of cancer in breast cancer (Perou et al. 2000; Sorlie et al. 2003), B-cell lymphoma (Alizadeh et al. 2000), Burkitt's lymphoma (Dave et al. 2006), prostate cancer (Lapointe et al. 2004), and lung cancer (Hayes et al. 2006). In addition to mRNA expression, even miRNA expression information has been proven to be helpful in dissecting cancer (He et al. 2005; Volinia et al. 2006).

There are also other methods used in expression analysis that make use of “supervised methods.” Using existing biological information as a guide, this approach has been successfully used to predict recurrence, metastasis, outcome, response to drugs, etc. (Beer et al. 2002; Pomeroy et al. 2002; Shipp et al. 2002; van de Vijver et al. 2002; van't Veer et al. 2002; Ramaswamy et al. 2003; Paik et al. 2004; Potti et al. 2006).

Databases

Accumulation of large amounts of expression data prompted the generation of public databases such as NCBI's Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>; Barrett et al. 2005), the European Bioinformatics Institute's ArrayExpress (<http://www.ebi.ac.uk/microarray-as/ae/>; Parkinson et al. 2007), the Stanford Microarray Database (<http://smd.stanford.edu/>; Marinelli et al. 2008), and Oncomine (<http://www.oncomine.org/>; Rhodes et al. 2007) (see Table 1). The first three mainly serve as data storage platforms and also provide data analysis options. Oncomine, on the other hand, is designed as a data-mining tool specific to cancer-related expression analysis. Such repositories make it possible to compare microarray results with one another. A higher level of information can be extracted by the meta-analysis of expression data from different studies.

Module Maps and Molecular Concept Maps

Given the heterogeneity of cancer and the noisy nature of expression data, however, being able to make discoveries at such a level necessitates the adoption of “gene-set-centered” approaches. An example of this is “module maps” (Ihmels et al. 2002; Tanay et al. 2004). This method was used in the meta-analysis of 2000 microarray experiments using 300 gene sets (Segal et al. 2004). A total of 456 gene modules were identified and were later used to compare different types of cancers. The authors found that previously unrelated tumor types could have similar expression patterns when analyzed at the level of modules. For example, a bone osteoblastic module (consisting of genes associated with proliferation and differentiation in bones) was found to be up-regulated in some breast cancers and down-regulated in lung cancer, hepatocellular carcinoma, and acute lymphoblastic leukemia.

Another example of integrative approaches to cancer expression data is the “molecular concept map” (Tomlins et al. 2007). Molecular concepts are sets of biologically related genes coming from gene annotations from external databases, computationally derived regulatory networks, and microarray gene-expression profiles coming from the Oncomine database. The gene signatures were obtained using COPA, mentioned above in Methods for Detecting Chromosomal Rearrangements. This method was developed to identify “outlier” gene sets, even if their expression level is low or a small number of samples show overexpression (Tomlins et al. 2005).

PRIORITIZATION OF CANDIDATE CANCER GENES

Cancer Gene Census

In 2004, Futreal et al. (2004) published a census of human cancer genes gleaned from published literature. Subsequent additions to the initial census of 291 genes have increased the total to more than 370 genes in 2008. A number of criteria were used for inclusion in the census. Only genes in which cancer-causing mutations have been reported were included, and a requirement for two independent reports of mutations in primary clinical samples was used. Genes involved in translocation or

copy-number change were included. However, genes for which there was only evidence of differential expression level or aberrant promoter DNA methylation in tumors were excluded.

The survey also included various data about each gene. For instance, the mutation type evident in the cancer gene (somatic, germ line, or both), neoplasm types associated with the gene (leukemias/lymphomas, mesenchymal, epithelial, etc.), the phenotypic nature of the mutated gene (dominant or recessive), and the mechanism of mutation affecting each gene (e.g., translocation, deletion, and frameshift) were recorded.

A number of general trends were highlighted in the analysis of the compiled list of genes. Approximately 90% of the genes had somatic mutations, 20% had germ-line mutations, and 10% were susceptible to both types of mutation. The most common somatic genetic changes seen were chromosomal translocations, with recurrent events frequently taking place in leukemias and lymphomas. A total of 90% of somatic mutations were phenotypically dominant in tumors, whereas 90% of germ-line mutations were found to be recessive.

In addition, the study examined the distribution of Pfam protein domains (Finn et al. 2006) in the proteins encoded by the cancer genes compared to the entire human proteome. Protein kinase domains, domains involved in transcriptional regulation, and DNA maintenance and repair-associated domains were overrepresented in the group of cancer genes.

Computational Prioritization of Cancer Genes

Many issues remain to be determined in understanding oncogenesis in different tumor types; for example, elucidation of candidate causative agents, distinguishing between driver and passenger alterations (Haber and Settleman 2007; Higgins et al. 2007), and characterization of the function of cancer genes in the oncogenic process (Hu et al. 2007). Oncogenomic experiments are now providing the cancer research community with numerous candidate causative genes. However, it is imperative to prioritize the more promising candidates from genes that are unlikely to be contributing to tumorigenesis. A number of previous computational studies have aimed at predicting cancer-associated missense mutations (Kaminker et al. 2007a,b).

Recently, we have described a number of different approaches for candidate cancer prioritization, irrespective of the oncogenic alteration. We have shown before that it is possible to develop an accurate classifier for distinguishing between Cancer Gene Census genes and other human genes (Furney et al. 2006). However, it is evident from cancer biology that altered proto-oncogenes and tumor suppressor genes promote oncogenesis in different ways. Furthermore, we have also shown that differences in sequence and regulatory properties exist between these two types of cancer genes (Furney et al. 2008b). These issues prompted us to devise separate classifiers for proto-oncogenes and tumor suppressor genes (Furney et al. 2008a). We constructed computational classifiers using different combinations of sequence and functional data including sequence conservation, protein domains and interactions, and regulatory data. We found that these classifiers are able to distinguish between known cancer genes and other human genes. Furthermore, the classifiers also discriminate candidate cancer genes from a recent mutational screen from other human genes. We have provided a web-based facility (CGPrio) through which cancer biologists may access our results (<http://bg.upf.edu/cgprio>).

INTEGRATION OF ONCOGENOMIC DATA TYPES

An integrative approach is necessary to obtain a more complete view of the deregulation of normal cellular processes that occurs during oncogenesis. During the past few years, a number of studies have revealed the effectiveness of integrative functional genomics in cancer research, whereby information from complementary experimental data sources is combined to provide greater insight into the process of tumorigenesis (Rhodes et al. 2004; Lu et al. 2005; Bild et al. 2006; Carter et al. 2006; Stransky et al. 2006; Tomlins et al. 2007).

Study Approaches

Integrative studies have combined data from different microarray experiments (Rhodes et al. 2004; Tomlins et al. 2007), expression and copy-number change data (Carter et al. 2006; Stransky et al. 2006), and expression of mRNAs and miRNAs (Lu et al. 2005). Other recent studies have used a comparative oncogenomic approach to identify genes contributing to oncogenesis and metastasis (Kim et al. 2006; Zender et al. 2006; Maser et al. 2007). For example, Kim et al. (2006) found an 850-kb amplicon from an array CGH analysis of a melanoma mouse model equivalent to a section of a much larger amplification observed in human melanoma. Using expression analysis, they were able to identify NEDD9 as the gene most likely to be responsible for driving metastasis.

Zender and colleagues (2006) identified syntenic amplifications in human liver carcinomas and a mouse model of hepatocellular carcinoma by array CGH of tumors from both species. A subset of candidate oncogenes was identified by excluding those genes absent in the amplified regions in either mouse or human tumors. RNA and protein expression analyses in both species of the remaining genes pinpointed *cIAP1* and *Yap* as oncogenes.

Maser et al. (2007) engineered murine lymphomas with destabilized genomes to mimic the far more prevalent chromosomal instability associated with human tumors. They generated a mouse lymphoma that was deficient for *Atm*, *Terc*, and *p53* and assessed these tumors and human T-cell acute lymphoblastic leukemias/lymphomas using array CGH. The authors found recurrent syntenic amplifications and deletions in the human and mouse lymphomas and, on targeted resequencing of candidate genes within syntenic regions, discovered frequent somatic mutations in *PTEN* and *FBXW7*.

Recently, three large-scale collaborative projects have resulted in the integrative analysis of human glioblastomas and pancreatic cancers (Jones et al. 2008; McLendon et al. 2008; Parsons et al. 2008). The Cancer Genome Atlas Research Network (<http://cancergenome.nih.gov>) presented an integrative analysis of DNA copy number, DNA methylation, and mRNA expression in more than 200 human glioblastomas (McLendon et al. 2008). In addition, they determined the nucleotide sequence in 91 of the tumors. This wealth of data allowed the authors to identify core signaling pathways that are affected in glioblastoma, including receptor tyrosine kinase (RTK) signaling and the *p53* and retinoblastoma tumor suppressor pathways.

Parsons et al. (2008) interrogated the same tumor type in 22 samples by sequencing >20,000 protein-coding genes, analyzing copy-number changes, and performing serial analysis of gene expression (SAGE) on 16 samples. This study found that the majority of tumors showed alterations in genes belonging to each of the *p53*, retinoblastoma, and *PI3K* pathways. In addition, the candidate cancer genes identified by the authors included several genes previously associated with glioblastoma (e.g., *p53*, *EGFR*, and *NF1*).

Jones et al. (2008) surveyed pancreatic tumors using a similar strategy of transcript nucleotide sequence determination, copy-number change evaluation, and gene-expression analysis. On average, they detected 63 alterations per tumor, most of which were point mutations. Through pathway analysis, they found a core set of 12 signaling pathways/processes in which at least one gene had a genetic alteration in 67%–100% of the tumors.

In addition, in a study by Parsons et al. (2008), *IDH1* (isocitrate dehydrogenase 1) was detected to be mutated in all secondary glioblastomas and was linked to a better prognosis. In a McLendon et al. (2008) study, on the other hand, integration of methylation profiling led to the identification of how *MGMT* (*O*⁶-methylguanine–DNA methyltransferase) promoter methylation status has substantial influence on the overall frequency and pattern of mutations in glioblastoma. This has clinical implications for alkylating agents, and one such agent is temozolomide, which is used in the clinical treatment of this cancer. The current standard practice for patients is surgical intervention followed by adjuvant radiation therapy or chemotherapy with temozolomide. However, this treatment only produces a median survival of 15 months.

These studies underline the need to investigate different types of aberrations in cancer and highlight how crucial the integration of these different methods is in understanding oncogenesis.

Integrative Oncogenomic Projects and Resources

The International Cancer Genome Consortium

The International Cancer Genome Consortium (ICGC; <http://www.icgc.org/>), launched in 2008, is a collaboration designed to produce high-quality genomic data in multiple cancer types. This international consortium has three primary goals: (1) to coordinate projects to generate comprehensive catalogs of somatic mutations in tumors in 50 different cancer types and/or subtypes that are of global, clinical, and societal significance; (2) to generate transcriptomic and epigenomic data sets from the same tumors; and (3) to ensure that these data are available to the research community at large as quickly as possible and with minimal restrictions.

The Cancer Genome Atlas

The Cancer Genome Atlas (TCGA; <http://tcga.cancer.gov/>) is a U.S. National Institutes of Health initiative involving the National Cancer Institute and the National Human Genome Research Institute. The goal of the project is to increase the understanding of cancer through the systematic use of various genome-wide technologies. The Cancer Genome Atlas Pilot Project (<http://cancergenome.nih.gov>) was undertaken as a feasibility study. Three tumor types—brain (glioblastoma multiforme), lung (squamous carcinoma), and ovarian (serous cystadenocarcinoma)—were selected for analysis in this pilot phase. The project entails collaboration among a central Biospecimen Core Resource, Cancer Genome Characterization Centers, Genome Sequencing Centers, and a Data Coordinating Center. Data produced by TCGA are available through the TCGA Data Portal. Initial fruit of TCGA's labor is the publication of their analysis to date of glioblastomas (see above for details; McLendon et al. 2008).

The Cancer Genome Project

The Cancer Genome Project (<http://www.sanger.ac.uk/genetics/CGP/>) comprises various endeavors at the Sanger Institute, including the Cancer Gene Census (<http://www.sanger.ac.uk/genetics/CGP/Census/>), COSMIC (<http://www.sanger.ac.uk/genetics/CGP/cosmic>), and a number of other cancer-related projects.

IntOGen

IntOGen (Integrative OncoGenomics, <http://www.intogen.org>) is a resource that integrates different types of oncogenomics data. At the time of this writing, this resource includes genomic alterations (amplifications and deletions), microarray expression profiles, and mutation screenings. The experiments are collected from different public databases or directly from the authors, and the type of cancer from the samples is annotated using the controlled vocabulary of the International Classification of Diseases (ICD-10 and ICD-O). All of the experiments are processed in a standard way and then analyzed statistically to identify genes that are significantly altered. Groups of experiments annotated with the same ICD term are combined to identify genes significantly altered in this cancer type.

IntOGen is designed to be a discovery tool for cancer researchers. Users interested in a particular gene can easily see whether their gene of interest has been found to be altered (e.g., overexpressed, mutated, or deleted) in different cancer types and subtypes. On the other hand, researchers interested in a particular tumor type are able to search for the genes that are more significantly altered (with mutations or genomic or transcriptomic alterations) in this type of cancer. Additionally, this resource is highly useful for prioritization of candidate cancer genes. The probabilities given by our prioritization method (CGPrio, described above) are integrated in IntOGen (Furney et al. 2008a). Users can upload a list of candidate cancer genes and prioritize them, taking into account evidence of oncogenic alterations detected in other experiments and the probabilities of being a cancer gene given by CGPrio.

IntOGen not only is focused on individual gene analysis, but also studies the implication of functional and regulatory modules in different cancer types. For example, users can search IntOGen for the biological pathways with a higher proportion of genes altered in a particular cancer type or have a wide view of the alterations of genes in a particular pathway in different tumor types.

Overall, the integration of a large compendium of oncogenomic experiments together with genomic data and statistical integrative analysis provides a powerful tool for online discovery of genes involved in different types of cancer.

SUMMARY

The last decade has witnessed profound changes in how cancer research is conducted. First, emerging technologies have allowed surveys of alterations in tumor cells on a genome-wide scale, giving rise to the field of oncogenomics. In tandem with this is the realization that, because of the complex nature of oncogenesis, methods that integrate multiple types of data are required to understand and elucidate the tumorigenic process. Recognition of this has led to the formation of the International Cancer Genome Consortium (ICGC), which will endeavor to use existing and improving technologies to perform integrative oncogenomic research in multiple cancer types. Projects under the aegis of the ICGC will occupy much of the cancer research field for years to come.

ACKNOWLEDGMENTS

We acknowledge funding from the International Human Frontier Science Program Organization (HFSP) and from the Spanish Ministerio de Educacion y Ciencia (MEC) grant number SAF2006-0459. N.L.-B. is the recipient of a Ramon y Cajal contract of the MEC and acknowledges support from Instituto Nacional de Bioinformatica.

REFERENCES

- Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, et al. 2009. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 41: 585–590.
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503–511.
- American Association for Cancer Research Human Epigenome Task Force; European Union, Network of Excellence, Scientific Advisory Board. 2008. Moving AHEAD with an international human epigenome project. *Nature* 454: 711–715.
- Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, et al. 2008. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 40: 616–622.
- Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, et al. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 8: 532–538.
- Baak JP, Janssen EA, Soreide K, Heikkilä R. 2005. Genomics and proteomics—The way forward. *Ann Oncol* 16: i30–ii44.
- Ballestar E, Paz MF, Valle L, Wei S, Fraga MF, Espada J, Cigudosa JC, Huang TH, Esteller M. 2003. Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J* 22: 6335–6345.
- Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, Saha S, Markowitz S, Willson JK, Parmigiani G, Kinzler KW, et al. 2003. Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* 300: 949.
- Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R. 2005. NCBI GEO: Mining millions of expression profiles—Database and tools. *Nucleic Acids Res* 33: D562–D566.
- Baudis M. 2007. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* 7: 226.
- Baudis M, Cleary ML. 2001. Progenetix.net: An online repository for molecular cytogenetic aberration data. *Bioinformatics* 17: 1228–1229.
- Beer DG, Kardias SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, et al. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8: 816–824.
- Benvenuti S, Arena S, Bardelli A. 2005. Identification of cancer genes by mutational profiling of tumor genomes. *FEBS Lett* 579: 1884–1890.
- Bernstein BE, Meissner A, Lander ES. 2007. The mammalian epigenome. *Cell* 128: 669–681.
- Bestor TH. 2005. Transposons reanimated in mice. *Cell* 122: 322–325.
- Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, et al. 2006. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439: 353–357.
- Buongiorno-Nardelli M, Amaldi F. 1970. Autoradiographic detection of molecular hybrids between RNA and DNA in tissue sections. *Nature* 225: 946–948.
- Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, et al. 2008. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 40: 722–729.

- Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38: 1043–1048.
- Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong LC, White RL. 1983. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305: 779–784.
- Chin L, Gray JW. 2008. Translating insights from the cancer genome into clinical practice. *Nature* 452: 553–563.
- Clark SJ, Harrison J, Paul CL, Frommer M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 22: 2990–2997.
- Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, et al. 2007. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 39: 352–358.
- Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S, Wu Y, He X, Powe NR, Feinberg AP. 2003. Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. *Science* 299: 1753–1755.
- Dave SS, Wright G, Tan B, Rosenwald A, Gascoyne RD, Chan WC, Fisher RI, Braziel RM, Rimsza LM, Grogan TM, et al. 2004. Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med* 351: 2159–2169.
- Dave SS, Fu K, Wright GW, Lam LT, Kluin P, Boerma EJ, Greiner TC, Weisenburger DD, Rosenwald A, Ott G, et al. 2006. Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med* 354: 2431–2442.
- Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, et al. 2002. Mutations of the BRAF gene in human cancer. *Nature* 417: 949–954.
- Davies H, Hunter C, Smith R, Stephens P, Greenman C, Bignell G, Teague J, Butler A, Edkins S, Stevens C, et al. 2005. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res* 65: 7591–7595.
- Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, Sullivan K, Vijayakrishnan J, Wang Y, Pittman AM, et al. 2008. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet* 40: 1204–1210.
- Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, et al. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455: 1069–1075.
- Doll R, Peto R. 1981. The causes of cancer: Quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 66: 1191–1308.
- Druker BJ. 2002. STI571 (Gleevec) as a paradigm for cancer therapy. *Trends Mol Med* 8: S14–S18.
- Dyrskjot L, Thykjaer T, Krühoffer M, Jensen JL, Marcussen N, Hamilton-Dutoit S, Wolf H, Orntoft TF. 2003. Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet* 33: 90–96.
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447: 1087–1093.
- Eden A, Gaudet F, Waghmare A, Jaenisch R. 2003. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 300: 455.
- Esteller M. 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8: 286–298.
- Esteller M. 2008. Epigenetics in cancer. *N Engl J Med* 358: 1148–1159.
- Fahrner JA, Eguchi S, Herman JG, Baylin SB. 2002. Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. *Cancer Res* 62: 7213–7218.
- Fang YC, Huang HC, Juan HF. 2008. MeInfoText: Associated gene methylation and cancer information from text mining. *BMC Bioinformatics* 9: 22.
- Feinberg AP. 1999. Imprinting of a genomic domain of 11p15 and loss of imprinting in cancer: An introduction. *Cancer Res* 59: 1743s–1746s.
- Feinberg AP, Vogelstein B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301: 89–92.
- Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, et al. 2006. Pfam: Clans, web tools and services. *Nucleic Acids Res* 34: D247–D251.
- Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, Flanagan A, Teague J, Wooster R, Futreal PA, et al. 2006. Cosmic 2005. *Br J Cancer* 94: 318–322.
- Forrest WF, Cavet G. 2007. Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 317: 1500; author reply 1500.
- Fraga MF, Esteller M. 2002. DNA methylation: A profile of methods and applications. *Biotechniques* 33: 632, 634, 636–649.
- Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, et al. 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of -histone H4 is a common hallmark of human cancer. *Nat Genet* 37: 391–400.
- Furney SJ, Higgins DG, Ouzounis CA, Lopez-Bigas N. 2006. Structural and functional properties of genes involved in human cancer. *BMC Genomics* 7: 3.
- Furney SJ, Calvo B, Larrañaga P, Lozano JA, Lopez-Bigas N. 2008a. Prioritization of candidate cancer genes—An aid to oncogenomic studies. *Nucleic Acids Res* 36: e115.
- Furney SJ, Madden SF, Higgins DG, Lopez-Bigas N. 2008b. Distinct patterns in the regulation and evolution of human cancer genes. *In Silico Biol* 8: 33–46.
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. *Nat Rev Cancer* 4: 177–183.
- Getz G, Hofling H, Mesirov JP, Golub TR, Meyerson M, Tibshirani R, Lander ES. 2007. Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 317: 1500.
- Gonzalez-Zuluea M, Bender CM, Yang AS, Nguyen T, Beart RW, Van Tornout JM, Jones PA. 1995. Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Res* 55: 4531–4535.
- Greenman C, Stephens P, Smith R, Dalgleish GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, et al. 2007. Patterns of somatic mutation in human cancer genomes. *Nature* 446: 153–158.
- Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B. 1989. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet* 83: 155–158.
- Groffen J, Stephenson JR, Heisterkamp N, Bartram C, de Klein A, Grosfeld G. 1984. The human *c-abl* oncogene in the Philadelphia translocation. *J Cell Physiol* 3: 179–191.
- Grunau C, Renault E, Rosenthal A, Roizes G. 2001. MethDB—A public database for DNA methylation data. *Nucleic Acids Res* 29: 270–274.
- Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Bake A, et al. 2007. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39: 631–637.
- Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Blondal T, et al. 2008. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 40: 281–283.
- Haber DA, Settleman J. 2007. Cancer: Drivers and passengers. *Nature* 446: 145–146.
- Haiman CA, Le Marchand L, Yamamoto J, Stram DO, Sheng X, Kolonel LN, Wu AH, Reich D, Henderson BE. 2007a. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* 39: 954–956.
- Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewski A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, et al. 2007b. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39: 638–644.
- Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* 100: 57–70.
- Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, Socinski MA, Perou C, Meyerson M. 2006. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J Clin Oncol* 24: 5079–5090.
- He L, Thomson JM, Hemann MT, Hernandez-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, et al. 2005. A microRNA polycistron as a potential human oncogene. *Nature* 435: 828–833.
- He X, Chang S, Zhang J, Zhao Q, Xiang H, Kusonmano K, Yang L, Sun ZS, Yang H, Wang J. 2008. MethyCancer: The database of human DNA methylation and cancer. *Nucleic Acids Res* 36: D836–D841.

- Herman JG, Baylin SB. 2003. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 349: 2042–2054.
- Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarr JR, Linehan WM, et al. 1994. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci* 91: 9700–9704.
- Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB. 1995. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res* 55: 4525–4530.
- Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. 1996. Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci* 93: 9821–9826.
- Higgins ME, Claremont M, Major JE, Sander C, Lash AE. 2007. Cancer-Genes: A gene selection resource for cancer genome projects. *Nucleic Acids Res* 35: D721–D726.
- Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, et al. 2008. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 40: 1426–1435.
- Hu P, Bader G, Wigle DA, Emili A. 2007. Computational prediction of cancer-gene function. *Nat Rev Cancer* 7: 23–34.
- Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. 2002. Revealing modular organization in the yeast transcriptional network. *Nat Genet* 31: 370–377.
- International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437: 1299–1320.
- Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, et al. 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321: 1801–1806.
- Jong K, Marchiori E, van der Vaart A, Chin SF, Carvalho B, Tijssen M, Eijk PP, van den Ijssel P, Grabsch H, Quirke P, et al. 2007. Cross-platform array comparative genomic hybridization meta-analysis separates hematopoietic and mesenchymal from epithelial tumors. *Oncogene* 26: 1499–1506.
- Kallioniemi A. 2008. CGH microarrays and cancer. *Curr Opin Biotechnol* 19: 36–40.
- Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258: 818–821.
- Kaminker JS, Zhang Y, Watanabe C, Zhang Z. 2007a. CanPredict: A computational tool for predicting cancer-associated missense mutations. *Nucleic Acids Res* 35: W595–W598.
- Kaminker JS, Zhang Y, Waugh A, Haverty PM, Peters B, Sebisanoic D, Stinson J, Forrest WF, Bazan JF, Seshagiri S, et al. 2007b. Distinguishing cancer-associated missense mutations from common polymorphisms. *Cancer Res* 67: 465–473.
- Kaneda A, Feinberg AP. 2005. Loss of imprinting of IGF2: A common epigenetic modifier of intestinal tumor risk. *Cancer Res* 65: 11236–11240.
- Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, Pikarski E, Young RA, Niveleau A, Cedar H, et al. 2006. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* 38: 149–153.
- Kiemeny LA, Thorlacius S, Sulem P, Geller F, Aben KK, Stacey SN, Gudmundsson J, Jakobsdottir M, Bergthorsson JT, Sigurdsson A, et al. 2008. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet* 40: 1307–1312.
- Kim M, Gans JD, Nogueira C, Wang A, Paik JH, Feng B, Brennan C, Hahn WC, Cordon-Cardo C, Wagner SN, et al. 2006. Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell* 125: 1269–1281.
- Knudson AG Jr. 1971. Mutation and cancer: Statistical study of retinoblastoma. *Proc Natl Acad Sci* 68: 820–823.
- Knutsen T, Gobu V, Knaus R, Padilla-Nash H, Augustus M, Strausberg RL, Kirsch IR, Sirotkin K, Ried T. 2005. The interactive online SKY/M-FISH & CGH database and the Entrez cancer chromosomes search database: Linkage of chromosomal aberrations with the genome sequence. *Genes Chromosomes Cancer* 44: 52–64.
- Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, et al. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318: 420–426.
- Krzywinski M, Bosdet I, Mathewson C, Wye N, Brebner J, Chiu R, Corbett R, Field M, Lee D, Pugh T, et al. 2007. A BAC clone fingerprinting approach to the detection of human genome rearrangements. *Genome Biol* 8: R224.
- Kuppers R. 2005. Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev* 5: 251–262.
- Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, et al. 2004. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci* 101: 811–816.
- Liu ET, Kuznetsov VA, Miller LD. 2006. In the pursuit of complexity: Systems medicine in cancer biology. *Cancer Cell* 9: 245–247.
- Loeb LA, Bielas JH. 2007. Limits to the Human Cancer Genome Project? *Science* 315: 762; author reply 764–765.
- Lopez-Serra L, Ballestar E, Fraga MF, Alaminos M, Setien F, Esteller M. 2006. A profile of methyl-CpG binding domain protein occupancy of hypermethylated promoter CpG islands of tumor suppressor genes in human cancer. *Cancer Res* 66: 8342–8346.
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, et al. 2005. MicroRNA expression profiles classify human cancers. 435: 834–838.
- MacDonald JW, Ghosh D. 2006. COPA—Cancer outlier profile analysis. *Bioinformatics* 22: 2950–2951.
- Mahe C, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM. 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458: 97–101.
- Marinelli RJ, Montgomery K, Liu CL, Shah NH, Prapong W, Nitzberg M, Zachariah ZK, Sherlock GJ, Natkunam Y, West RB, et al. 2008. The Stanford Tissue Microarray Database. *Nucleic Acids Res* 36: D871–D877.
- Maser RS, Choudhury B, Campbell PJ, Feng B, Wong KK, Protopopov A, O’Neil J, Gutierrez A, Ivanova E, Perna I, et al. 2007. Chromosomally unstable mouse tumors have genomic alterations similar to diverse human cancers. *Nature* 447: 966–971.
- McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogiannis M, Olson JJ, Mikkelsen T, Lehman N, Aldape K, et al. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455: 1061–1068.
- Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D. 1995. 5’ CpG island methylation is associated with transcriptional silencing of the tumor suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med* 1: 686–692.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553–560.
- Mitelman F. 2000. Recurrent chromosome aberrations in cancer. *Mutation Res* 462: 247–253.
- Mitelman F, Mertens F, Johansson B. 1997. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nat Genet* 15: 417–474.
- Mitelman F, Johansson B, Mertens F. 2007. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 7: 233–245.
- Mitelman F, Johansson B, Mertens F., eds. 2009. *Mitelman database of chromosome aberrations in cancer* The Cancer Genome Anatomy Project, Washington, D.C.
- Morozova O, Marra MA. 2008. From cytogenetics to next-generation sequencing technologies: Advances in the detection of genome rearrangements in tumors. *Biochem Cell Biol* 86: 81–91.
- Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, et al. 2007. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446: 758–764.
- Nowell PC. 2007. Discovery of the Philadelphia chromosome: A personal perspective. *J Clin Invest* 117: 2033–2035.
- Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. 2002. The IARC TP53 Database: New online mutation analysis and recommendations to users. *Human Mutation* 19: 607–614.
- Ongenaert M, Van Neste L, De Meyer T, Menschaert G, Bekaert S, Van Criekinge W. 2008. PubMeth: A cancer methylation database

- combining text-mining and expert annotation. *Nucleic Acids Res* 36: D842–D846.
- Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, et al. 2004. EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* 304: 1497–1500.
- Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, et al. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351: 2817–2826.
- Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, et al. 2007. ArrayExpress—A public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* 35: D747–D750.
- Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science* 321: 1807–1812.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. 2000. Molecular portraits of human breast tumors. *Nature* 406: 747–752.
- Peto J. 2001. Cancer epidemiology in the last century and the next decade. *Nature* 411: 390–395.
- Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, et al. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20: 207–211.
- Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, et al. 2002. Prediction of central nervous system embryonal tumor outcome based on gene expression. *Nature* 415: 436–442.
- Ponder BA. 2001. Cancer genetics. *Nature* 411: 336–341.
- Potti A, Mukherjee S, Petersen R, Dressman HK, Bild A, Koontz J, Kratzke R, Watson MA, Kelley M, Ginsburg GS, et al. 2006. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355: 570–580.
- Rabbitts TH. 1994. Chromosomal translocations in human cancer. *Nature* 372: 143–149.
- Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, Jakobsdottir M, Helgadóttir H, Thorlacius S, Aben KK, et al. 2009. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* 41: 221–227.
- Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. 2002. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* 418: 934.
- Ramaswamy S, Golub TR. 2002. DNA microarrays in clinical oncology. *J Clin Oncol* 20: 1932–1941.
- Ramaswamy S, Ross KN, Lander ES, Golub TR. 2003. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33: 49–54.
- Reddy EP, Reynolds RK, Santos E, Barbacid M. 1982. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* 300: 149–152.
- Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. 2004. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci* 101: 9309–9314.
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, et al. 2007. Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9: 166–180.
- Rowley JD. 1998. The critical role of chromosome translocations in human leukemias. *Annu Rev Genet* 32: 495–519.
- Rowley JD. 2001. Chromosome translocations: Dangerous liaisons revisited. *Nat Rev* 1: 245–250.
- Rubin AF, Green P. 2007. Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 317: 1500.
- Sakai T, Toguchida J, Ohtani N, Yandell DW, Rapaport JM, Dryja TP. 1991. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* 48: 880–888.
- Sakamoto H, Yoshimura K, Saeki N, Katai H, Shimoda T, Matsuno Y, Saito D, Sugimura H, Tanioka F, Kato S, et al. 2008. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet* 40: 730–740.
- Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ, et al. 2004. High frequency of mutations of the PIK3CA gene in human cancers. *Science* 304: 554.
- Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, et al. 1996. Multicolor spectral karyotyping of human chromosomes. *Science* 273: 494–497.
- Segal E, Friedman N, Koller D, Regev A. 2004. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 36: 1090–1098.
- Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, et al. 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 8: 68–74.
- Shtivelman E, Lifshitz B, Gale RP, Canaani E. 1985. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* 315: 550–554.
- Sjoberg T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, et al. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* 314: 268–274.
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, et al. 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci* 100: 8418–8423.
- Speicher MR, Ward DC. 1996. The coloring of cytogenetics. *Nat Med* 2: 1046–1048.
- Speicher MR, Gwyn Ballard S, Ward DC. 1996. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet* 12: 368–375.
- Stacey SN, Gudbjartsson DF, Sulem P, Bergthorsson JT, Kumar R, Thorleifsson G, Sigurdsson A, Jakobsdottir M, Sigurgeirsson B, Benediktsson KR, et al. 2008. Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat Genet* 40: 1313–1318.
- Stephens P, Hunter C, Bignell G, Edkins S, Davies H, Teague J, Stevens C, O'Meara S, Smith R, Parker A, et al. 2004. Lung cancer: Intragenic ERBB2 kinase mutations in tumors. *Nature* 431: 525–526.
- Stransky N, Vallot C, Rey F, Bernard-Pierrot I, de Medina SG, Seagraves R, de Rycke Y, Elvin P, Cassidy A, Spraggon C, et al. 2006. Regional copy number-independent deregulation of transcription in cancer. *Nat Genet* 38: 1386–1396.
- Strauss BS. 2007. Limits to the Human Cancer Genome Project? *Science* 315: 764–765.
- Suzuki H, Gabrielson E, Chen W, Anbazhagan R, van Engeland M, Weisenberg MP, Herman JG, Baylin SB. 2002. A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat Genet* 31: 141–149.
- Tabin CJ, Bradley SM, Bargmann CI, Weinberg RA, Papageorge AG, Scolnick EM, Dhar R, Lowy DR, Chang EH. 1982. Mechanism of activation of a human oncogene. *Nature* 300: 143–149.
- Tanay A, Sharan R, Kupiec M, Shamir R. 2004. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc Natl Acad Sci* 101: 2981–2986.
- Tanek HJ, Wiegant J, van Gijlswijk RP, Bezrookove V, Pattenier H, Heetebrij RJ, Talman EG, Raap AK, Vrolijk J. 1999. New strategy for multi-colour fluorescence in situ hybridisation: COBRA: COmbined Binary RAtio labelling. *Eur J Hum Genet* 7: 2–11.
- Tapper W, Hammond V, Gerty S, Ennis S, Simmonds P, Collins A, Eccles D. 2008. The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. *Breast Cancer Res* 10: R108.
- Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, et al. 2008. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 40: 631–637.
- Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, et al. 2008. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 40: 310–315.
- Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, et al. 2009. A multistage genome-

- wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**: 579–584.
- Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, Helgeson BE, Cao X, Wei JT, Rubin MA, Shah RB, et al. 2006. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res* **66**: 3396–3400.
- Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhana-sekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ, et al. 2007. Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* **39**: 41–51.
- Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**: 644–648.
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**: 1999–2009.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530–536.
- Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, et al. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**: 871–874.
- Vogelstein B, Kinzler KW. 1993. The multistep nature of cancer. *Trends Genet* **9**: 138–141.
- Vogelstein B, Kinzler KW. 2004. Cancer genes and the pathways they control. *Nat Med* **10**: 789–799.
- Vogelstein B, Lane D, Levine AJ. 2000. Surfing the p53 network. *Nature* **408**: 307–310.
- Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo WL, et al. 2003. End-sequence profiling: Sequence-based analysis of aberrant genomes. *Proc Natl Acad Sci* **100**: 7696–7701.
- Volik S, Raphael BJ, Huang G, Stratton MR, Bignel G, Murnane J, Brebner JH, Bajsarowicz K, Paris PL, Tao Q, et al. 2006. Decoding the fine-scale structure of a breast cancer genome and transcriptome. *Genome Res* **16**: 394–404.
- Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci* **103**: 2257–2261.
- Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, et al. 2004. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**: 855–867.
- Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, et al. 2008. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* **40**: 1407–1409.
- Wang Z, Shen D, Parsons DW, Bardelli A, Sager J, Szabo S, Ptak J, Silliman N, Peters BA, van der Heijden MS, et al. 2004. Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science* **304**: 1164–1166.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**: 897–903.
- Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37**: 853–862.
- Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhi R, Lin WM, Province MA, Kraja A, Johnson LA, et al. 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**: 893–898.
- Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science* **318**: 1108–1113.
- Yamashita K, Upadhyay S, Osada M, Hoque MO, Xiao Y, Mori M, Sato F, Meltzer SJ, Sidransky D. 2002. Pharmacologic unmasking of epigenetically silenced tumor suppressor genes in esophageal squamous cell carcinoma. *Cancer Cell* **2**: 485–495.
- Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, et al. 2007. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* **39**: 645–649.
- Zender L, Spector MS, Xue W, Flemming P, Cordon-Cardo C, Silke J, Fan ST, Luk JM, Wigler M, Hannon GJ, et al. 2006. Identification and validation of oncogenes in liver cancer using an integrative oncogenic approach. *Cell* **125**: 1253–1267.
- Zhang L, Volinia S, Bonome T, Calin GA, Greshock J, Yang N, Liu CG, Giannakakis A, Alexiou P, Hasegawa K, et al. 2008. Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc Natl Acad Sci* **105**: 7004–7009.
- Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, et al. 2009. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**: 324–328.

Cold Spring Harbor Protocols

Oncogenomics Methods and Resources

Simon J. Furney, Gunes Gundem and Nuria Lopez-Bigas

Cold Spring Harb Protoc; doi: 10.1101/pdb.top069229

Email Alerting Service

Receive free email alerts when new articles cite this article - [click here](#).

Subject Categories

Browse articles on similar topics from *Cold Spring Harbor Protocols*.

[Analysis of Gene Expression](#) (197 articles)
[Bioinformatics/Genomics, general](#) (192 articles)
[DNA Modification/Epigenetics](#) (42 articles)
[Genome Analysis](#) (191 articles)
[Genomic Analysis](#) (10 articles)

To subscribe to *Cold Spring Harbor Protocols* go to:
<http://cshprotocols.cshlp.org/subscriptions>