# The genetics and genomics of cancer

Allan Balmain[1], Joe Gray[2] & Bruce Ponder[3]

doi:10.1038/ng1107

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

**The past decade has seen great strides in our understanding of the genetic basis of human disease. Arguably, the most profound impact has been in the area of cancer genetics, where the explosion of genomic sequence and molecular profiling data has illustrated the complexity of human malignancies. In a tumor cell, dozens of different genes may be aberrant in structure or copy number, and hundreds or thousands of genes may be differentially expressed. A number of familial cancer genes with high-penetrance mutations have been identified, but the contribution of low-penetrance genetic variants or polymorphisms to the risk of sporadic cancer development remains unclear. Studies of the complex somatic genetic events that take place in the emerging cancer cell may aid the search for the more elusive germline variants that confer increased susceptibility. Insights into the molecular pathogenesis of cancer have provided new strategies for treatment, but a deeper understanding of this disease will require new statistical and computational approaches for analysis of the genetic and signaling networks that orchestrate individual cancer susceptibility and tumor behavior.**

Over the past ten years the perception of the contribution of genetic susceptibility to the common cancers has changed. Knudson's hypothesis[1] and its molecular confirmation in retinoblastoma[2] focused attention on the role of genetic predisposition in certain rare cancers. But the origin of the common cancers was still predominantly viewed as environmental in the 1980s (refs. 3,4). This view was based on studies from the 1960s and 1970s that identified large differences in the incidence of specific cancers among populations and that showed that immigrants acquired the pattern of cancer risk of their new country[5].

Increased emphasis on the role of genetic predisposition in the common cancers began in the 1980s. Population-based epidemiological studies led to the implementation of genetic rather than environmental models to explain observed patterns of familial occurrence[6,7], and it was shown that genetic effects might account for a substantial fraction of cancer incidence without necessarily causing evident familial clustering[8]. Arguably, remaining doubts about the contribution of genetic susceptibility to the common cancers were dispelled by the demonstration in 1990 of genetic linkage in breast cancer families[9], which made use of the newly available DNA sequence polymorphisms.

## Cancer predisposition by rare, high-penetrance alleles

The first predisposing genes were identified as rare, mutated alleles that strongly increased the risk of cancer when inherited through the germ line. These mutated genes result in multiple cases of the disease in families and were identified using genetic linkage and positional cloning. The prototypic gene associated with familial cancer syndromes is the retinoblastoma gene (*RB1*), which has turned out to be one of the most important hubs of cellular signaling[10]. Other key signaling molecules such as p53 (encoded by *TP53*) were initially identified as important targets of viruses or somatic mutations in tumors[11–13] and were subsequently found to function as germline-inherited tumor predisposition genes[14].

High-penetrance alleles have provided many fundamental and unexpected insights into various aspects of cancer biology, including identification of the adenomatosis polyposis coli (APC), β-catenin and Tcf-4 pathway (reviewed in ref. 15) and the phosphatase PTEN, which is implicated in Cowden syndrome and in the development of a variety of tumor types[16,17]. The *VHL* gene product associated with Von Hippel Lindau syndrome[18]—a ubiquitin ligase that targets the hypoxia-inducible factor HIF-1 for degradation—is involved in angiogenesis. In addition, pathways have been identified that control important aspects of DNA repair and/or genomic stability, notably the DNA repair/checkpoint pathways that include products of the breast cancer–associated genes *BRCA1* and *BRCA2* (ref. 19) and those involved in DNA mismatch repair (see below)[20].

But this knowledge relates almost completely to events in the developing cancer cell. The explanation that it provides for how cancers develop is very incomplete—for example, we still have no mechanisms for the tissue specificity of many of the inherited cancer syndromes. This is perhaps not surprising: the regulation and breakdown of intracellular controls is only one part of understanding cancer. More essential and more challenging has been the attempt to understand the rules governing the organization of cells within a tissue and the nature of the cellular environment that restrains or promotes the emergence of a cancer cell.

Changes in the behavior of stromal cells from individuals with cancer were noted many years ago[21], and more recently epithelial–stromal interactions have been shown to influence tumor

[1]*UCSF Comprehensive Cancer Center and Department of Biochemistry and Biophysics, San Francisco, California 94143, USA.* [2]*UCSF Comprehensive Cancer Center and Dept. of Laboratory Medicine, San Francisco, California 94143, USA.* [3]*Cancer Research UK Department of Oncology, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK. Correspondence should be addressed to A.B. (e-mail: abalmain@cc.ucsf.edu), J.G (e-mail: jgray@cc.ucsf.edu) or B.P. (e-mail: bajp@mole.bio.cam.ac.uk).*

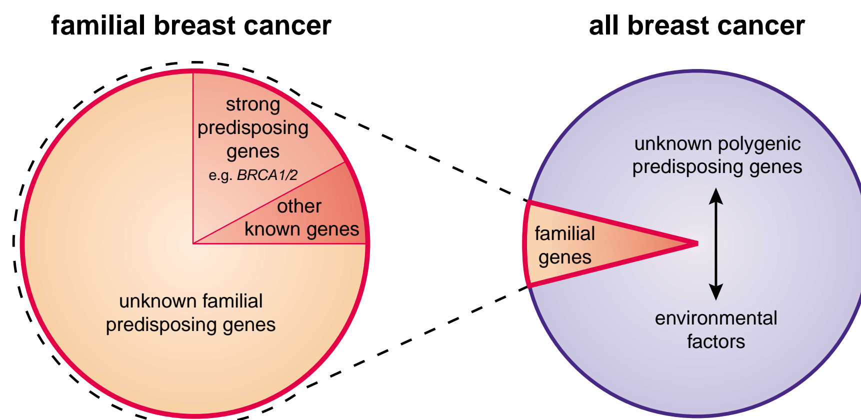


**Fig. 1** Breast cancer susceptibility genes. Familial breast cancer (left) constitutes only about 5–10% of total breast cancer (right). The genes known to be involved in familial breast cancer (*BRCA1* and *BRCA2*) account for only about 20% of the familial risk. Most of the genetic variants that contribute to the risk of developing sporadic breast cancer are unknown. Many of these may interact with environmental agents, such as radiation, that are known from epidemiological and experimental studies to cause cancer.

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

progression[22]. This co-dependence of cell types in tumors can be orchestrated through genetics: mutually exclusive mutations in the *PTEN* and *TP53* genes have been reported in stromal cells in breast cancer[23]. In this example, germline mutations in either of these genes can influence tumor development not only by their own effects on the tumor, but also by altering the behavior of accessory cells in the tumor microenvironment. Similar conclusions have been reached using mouse models for neurofibromatosis: haploinsufficiency for the tumor-suppressor gene *Nf1* in stromal and other 'normal' cells has been shown to be necessary for the development of neurofibroma[24].

## Cancer is a polygenic disorder

Predisposition by combinations of weak genetic variants may be of much greater significance to public health than the marked individual risks seen in the inherited cancer syndromes[25–27]. Such an argument is supported by the data on breast cancer[27]. Population-based epidemiological studies have shown that only 15–20% of the observed familial clustering of breast cancer occurs in families that carry a strongly predisposing *BRCA1* or *BRCA2* mutation (Fig. 1). In principle, the remaining 80–85% of familial risks might have a genetic or an environmental origin, but evidence from studies of breast cancer in twins[4,28], tumor incidence in the contralateral breast of an affected individual[4,28] and the pattern of inheritance in families[29,30] suggests that genetic factors predominate. The same evidence suggests, but cannot prove, that other, strong, *BRCA*-like genes are unlikely to account for much of the risk, which is explained better by a model that describes the combined action of several factors, each with an individually small effect[29,30]. In other words, the greater part of inherited predisposition to breast cancer, and therefore perhaps[31] of other common cancers, may be due to the effects of combinations of genetic variants at several (possibly a multitude) of different loci.

So what proportion of breast cancer can be explained by polygenic predisposition? This is difficult to answer. More helpful is to ask what predictions a polygenic model can make about the distribution of risk in the population[27]. This will determine the range of risks and the extent to which risk is concentrated in a predisposed minority of the population (Fig. 2). A model[27] based on the familial patterns of breast cancer in a population-based series from which *BRCA1* and *BRCA2* mutation carriers had been removed suggests that there is a wide distribution of risk, with up to a 40-fold difference between the top 20% and the bottom 20% of the population. The model also indicates that there is a marked concentration of risk, such that more than 50% of breast cancers occur in the most predisposed 12% of the population. Qualitatively similar conclusions have been suggested by other studies[4,29].

There are two general models for polygenic predisposition[32]. The first is the common variant–common disease model, in which common variants that have arisen only once, early in the history of the population, underlie disease predisposition in humans. In this model, the genes can be sought, in principle, by 'association studies' in which variant alleles of candidate genes are tested for significant differences in frequency between cancer cases and matched controls (ref. 33; see also Botstein and Risch[34], pp 228–237, this issue). Ideally, the variants to be tested should be those that are directly related to the mechanism of predisposition. Generally, however, these are not known. Instead, a set of polymorphisms is used to define haplotypes across the gene of interest and the frequency of each haplotype is compared between the cases and controls. The current interest in the haplotype structure of the human genome[35] is driven largely by the hope of using this information to design efficient whole-genome scans to search for disease associations and thereby to identify the loci of susceptibility alleles. The main limitation of this approach is its lack of power to detect alleles of weak effect without very large sample sizes.

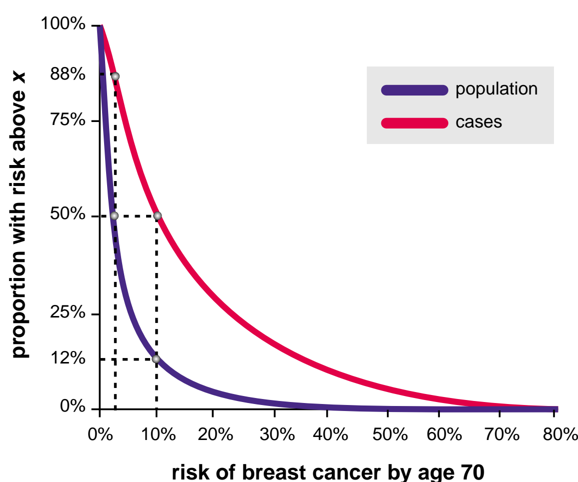


**Fig. 2** Risk distribution for breast cancer in the population. The blue curve shows that roughly 12% of the population have a risk of 10% or more of developing breast cancer by the age of 70; however, about 50% of all breast cancers develop in this subpopulation (red curve). By contrast, 50% of the population have a breast cancer risk of 3% or lower, and this subpopulation accounts for only 12% of all cancers. The conclusion[27] is that cancer risk is concentrated in a relatively small proportion of the female population.

The second model holds that the common variant–common disease theory is generally wrong[32,36] and that most significant variation underlying disease predisposition is in the form of rare alleles of recent origin, sometimes including several independent origins of the same allele. If this is true, the association approach outlined above will fail: first, because of a lack of statistical power to deal with rare, weak alleles; and second, because multiple independent copies of the same allele will probably lie on different haplotypes and therefore show much weaker haplotype association. Many hundreds of rare variant alleles may be involved and identifying them will be very difficult. To understand better the nature of genetic predisposition, we need much more detailed information about the nature and origins of variation in the human genome, and about the relationship between complex genotypes and phenotypic effect (see, for example, ref. 37).

The principle that combinations of common alleles can exert a profound influence on tumor susceptibility is clearly seen in mouse models of cancer susceptibility. The distribution of tumor number in a simple backcross population after carcinogen exposure shows that many mice have none or only a few tumors, but a small number have as many as 30 tumors or more and are clearly in the very high risk category[38]. By definition, the frequency of each allele in a backcross population between two inbred strains is 50–100%, and the number of different genes that account for differential strain susceptibility is relatively high[39,40]. Selection methods have also been used to enrich for the specific combinations of alleles that control a particular phenotype. For example, selective breeding from a mixture of eight parental inbred strains gave rise to two populations of mice that showed a more than 100-fold difference in susceptibility to inducers of skin cancer[41]. This seems precisely analogous to the model of polygenic susceptibility proposed in humans.

A main advantage of these mouse models is that it is possible to study genetic interactions in ways that are not feasible using data from human populations[42–44]. The model that emerges from these analyses is one of substantial heterogeneity: two individuals can show the same phenotype (cancer susceptibility) but for completely different genetic reasons. Identification of these different combinations of interacting alleles poses a significant obstacle to resolving the principal determinants of human cancer susceptibility. Simple association studies may detect ‘main effects’ at a single locus but will fail to identify interactions unless the appropriate interacting loci have been also tested.

Unless there are obvious candidates, either we must base our strategies on testing sets of interacting genes identified from mouse models and from tumor analysis (see below) or we must await the development of methods for carrying out complete genome scans in large numbers of individuals in well-characterized human populations. In the latter approach, the statistical difficulties in detecting true effects among the vast number of possible permutations of alleles will be severe, and methods for dealing with this complexity will be required.

## Somatic genetic changes in cancer cells

A cardinal feature of almost all cancer cells is genomic instability (reviewed in refs. 45–47), caused by either inherited mutations in genes that monitor genome integrity or mutations that are acquired in somatic cells during tumor development. The genetic alterations that result can occur at several levels, for example, in single nucleotides, small stretches of DNA (microsatellites), whole genes, structural components of chromosomes, or complete chromosomes.

Nucleotide, microsatellite or chromosome instability has been referred to as NIN, MIN or CIN, respectively[48]. Germline mutations leading to familial cancer syndromes have been identified that directly affect NIN; for example, individuals affected with xeroderma pigmentosum[49] develop multiple skin cancers because they are unable to repair ultraviolet-induced nucleotide mutations. MIN[50–52], caused by germline or somatic mutations in mis-match repair genes such as *MLH1*, *MSH2* or *MSH6* (ref. 53), can have serious consequences if the microsatellite target is located in an important growth controlling gene such as *BAX*[54], which mediates cell death, or the gene encoding the TGFβ type II receptor[55], which controls cell proliferation. Several events have been associated with CIN, including loss of telomere functions[47,56,57] and genetic alterations in genes that control chromosome segregation such as *BUB1*, *MAD2*, *BUB1R*[58,59], *APC*[60] or in the gene encoding the centrosome-associated kinase STK6 (refs. 61,62). In addition, mutations in *TP53* (refs. 63,64), the breast cancer susceptibility genes *BRCA1* and *BRCA2* (ref. 19) and the ataxia telangiectasia gene *ATM*[45] can affect gross genetic stability at several levels.

In classical models of multistage tumor development in both human[65,66] and mouse[67], it is assumed that each important event, whether it is a small-scale point mutation or a large-scale chromosomal change, confers a clonal selective advantage to the cell in which it occurs, setting the stage for the next advance leading to malignancy. But the relationship between early and late stages of tumor development and the role of genetic instability in progression remain unclear[68]. Some genetic alterations can be detected in tissue that appears normal by histological assessment[69,70], and extensive genetic changes can be seen at premalignant stages of human tumor development in many tissues[71–73]. The crucial determinant of progression may be the accumulation of specific combinations of genetic alterations or the occurrence of the mutations in a particular subset of target cells that has higher propensity for malignant progression[74]. Resolving such issues is an important goal for future research.

The consequences of genome instability in cancers are seen in the many aberrations that activate or inactivate the various genes that affect tumor cell behavior. Our ability to find these genes is increasing rapidly, owing to powerful analytical tools and the almost-complete genome sequence information from human and model organisms. Tools such as fluorescence *in situ* hybridization[75] allow the rapid detection and genomic localization of structural aberrations in metaphase spreads, the detection of individual aberrations, and the assessment of cell-to-cell variation in genome copy number as an indicator of genome instability[76]. Restriction landmark genome scanning[77], comparative genomic hybridization (CGH)[78], high-throughput quantitative PCR[79] and molecular ‘subtraction’ techniques such as representational display analysis[80] allow the rapid detection and genomic localization of aberrations in genome copy number and mutations throughout tumor genomes. Variations of these techniques also allow the assessment of methylation status[81].

Recently, the potential of high-throughput screening for detecting mutations in potential cancer genes has become apparent[82]. Transcriptional characteristics of tumors can be assessed with unprecedented completeness using large-scale microarray technologies[83] and/or high-throughput quantitative PCR[84]. Proteomic characteristics of tumors can be assessed by mass spectrometry[85] and by microarray techniques[86]. Combined applications of these techniques are providing ever-more detailed functional profiles of individual tumors.

The profiles even for clinically similar tumors are daunting in their diversity and complexity (W.-L. Kuo, manuscript in preparation). The types of aberration that develop include mutations in coding or regulatory sequences, changes in overall ploidy, small changes in genome copy number (such as gain or loss of a single genome copy), high amplification, structural rearrangement, homozygous deletion, and loss of

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

heterozygosity (LOH) resulting from the loss of one allele followed by reduplication. For example, Fig. 3 summarizes the frequencies of changes in gene copy number in ovarian cancer identified by CGH array analysis. The individual profiles show both small and large changes in copy number, as well as structural changes (data not shown), and clearly indicate the diversity that typically exists among clinically similar tumors. These regions of high copy number change may contain specific genetic variants that are preferentially amplified and/or act as germline tumor susceptibility genes. The challenge is to identify the specific genes that contribute to cancer progression when deregulated by these processes.

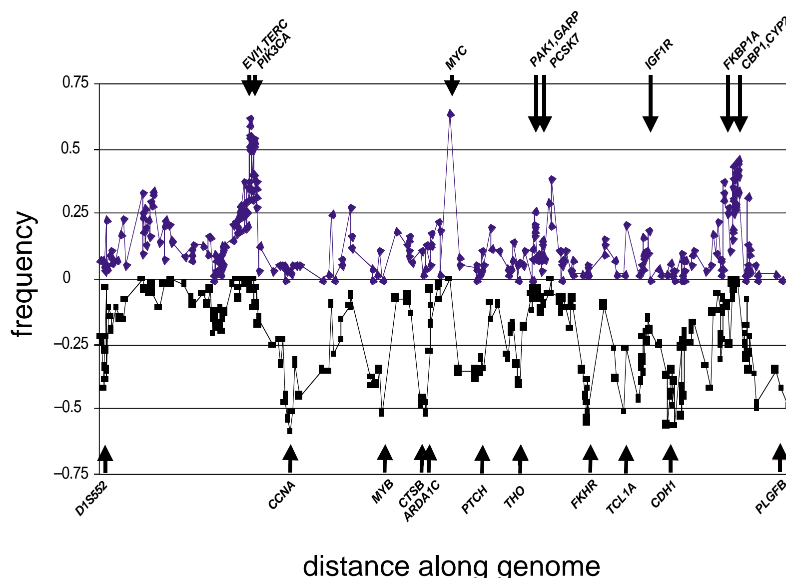


**Fig. 3** Genetic instability in ovarian cancer. Shown are composite array CGH profiles, with the frequency of changes in the relative gene copy number plotted as a function of genomic location, with chromosome 1pter to the left and 22qter and Xqter to the right. Peaks indicate gains (amplifications) and troughs indicate losses (deletions) in copy number on the different chromosomes. Some of the candidate genes in regions of copy number changes are indicated. Data are taken from W.-L. Kuo *et al.* (manuscript in preparation).

## Cancer genes as targets for therapy

Most efforts to identify potential therapeutic targets have focused on oncogenes that are activated recurrently or overexpressed in cancer cells. Genomic aberrations such as mutations, translocations and amplifications have guided the identification of such genes. Many growth factor receptors signal to the interior of the cell by phosphorylating their downstream interacting proteins at tyrosine residues. These tyrosine kinases have been the subject of much interest because the development of inhibitors seems straightforward. The tyrosine kinases encoded by *ABL* and *ERBB2* were among the earliest kinases identified in this class. The nuclear protein tyrosine kinase gene *ABL* was found to be activated by translocation of *BCR*[87]. This event occurs in 100% of chronic myelogeneous leukemias and at lower frequency in other leukemias.

The functional importance of this translocation has been confirmed by both model studies[88] and the effectiveness of the tyrosine kinase inhibitor, imatinib mesylate (Gleevec), against chronic myelogeneous leukemias in humans[89]. Amplification of the tyrosine kinase receptor gene *ERBB2* activates intracellular signaling in about 30% of breast cancers[90] and in a lower percentage of other solid tumor types. The functional importance of *ERBB2* activation, like that of *ABL*, has been shown both in model systems and by the therapeutic effectiveness of Herceptin, an antibody against ERBB2, in some tumors showing *ERBB2* amplification (ref. 91).

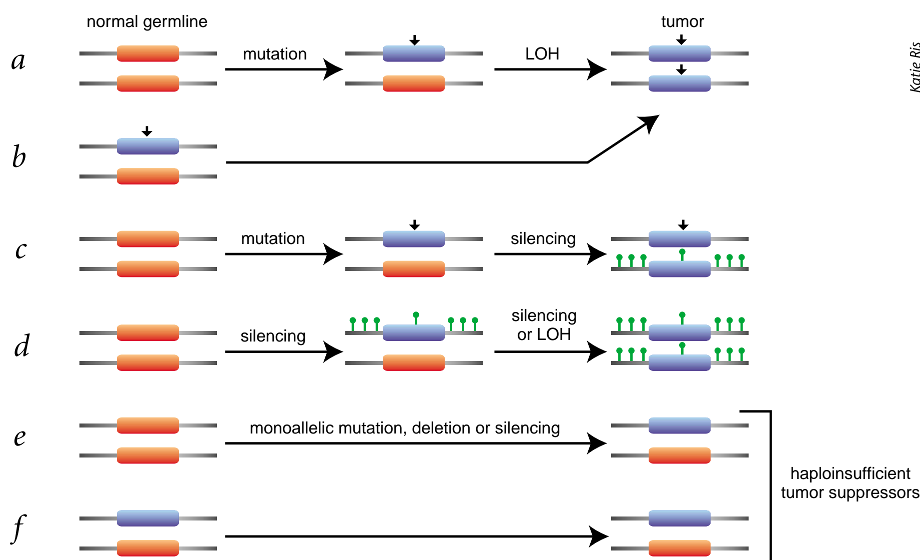


**Fig. 4** Loss of tumor-suppressor gene function in cancer. *a*,*b*, The classical Knudson two-hit model involves an initial mutational event (vertical arrowhead) that leads to gene inactivation during tumor development. Blue shaded bars indicate inactivated genes. LOH by non-disjunction, mitotic recombination or deletion results in the functional inactivation of both alleles. If the first mutation is inherited through the germ line, individuals carrying this mutation are often highly predisposed to tumor development. *c*, The mutational event can be followed by gene silencing through promoter methylation (vertical bars) without LOH. *d*, Biallelic silencing of both gene copies without LOH or gene mutation. *e*,*f*, Haploinsufficient tumor-suppressor genes do not need to lose both functional copies to confer increased risk. Loss of a single gene copy may occur by mutation, deletion or silencing, and the other functional allele may be retained. In some examples (*f*), a partially or completely nonfunctional allele may be inherited through the germ line, predisposing an individual to tumor development without requiring LOH or complete functional inactivation. Some 'low-penetrance' tumor susceptibility genes may be in this category.

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

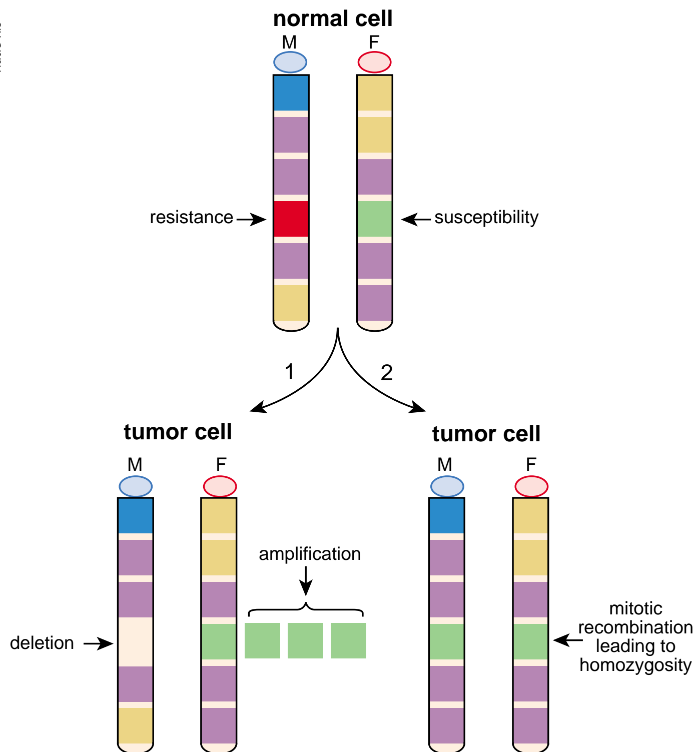


Fig. 5 Relationship between germline and somatic events involving weak tumor susceptibility genes. Some weak susceptibility genes may influence patterns of genetic events in tumors by causing preferential amplification of alleles that confer increased risk or loss of alleles that confer resistance (1). Mitotic recombination (2) may lead to homozygosity with respect to polymorphic (weak) tumor suppressor genes or hyperactive oncogenes, with no new somatic mutations. The letters M and F denote the chromosomes inherited from the mother and father, respectively. The multiplicity of genetic variants that influence susceptibility to sporadic cancers may contribute to the complexity of genetic aberrations in human cancers (see Fig. 3).

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

The recent identification of frequent and highly specific activating mutations in the gene *BRAF* in melanomas and other solid tumors[82] provides another opportunity to use specific kinase inhibitors[92] in therapeutic interventions directed at one of the main causal events of tumor development.

## Tumor-suppressor inactivation and haploinsufficiency

Early efforts to identify tumor-suppressor genes were guided by Knudson's hypothesis of biallelic gene inactivation (Fig. 4*a*,*b*) and typically focused on cooperation between a genomic aberration and an inactivating germline mutation. The large number of regions of recurrent LOH and physical deletion in human tumors identified in the past 10–15 years generated optimism that positional cloning would locate many other important tumor-suppressor genes. Progress has been slow, however, possibly because many of the important genes in those regions do not conform to the classical view that both alleles of the gene must be genetically inactivated to allow tumor progression[93].

A possibility is that tumors retain both copies of tumor-suppressor genes but effectively silence one or both alleles by methylation of the gene promoter (Fig. 4*c*,*d*; see Jaenisch and Bird[94], pp 245–254, this issue). In fact, changes in methylation may be as important as changes in DNA sequence or copy number in altering gene expression during the progression of cancer[95,96]. Alternatively, these 'new-generation' tumor-suppressor genes may influence tumor progression through functional haploinsufficiency—a process in which loss of one allele contributes to tumor progression even though a wild-type allele remains (Fig 4*e*,*f*). This has been shown in mice carrying single, inactivated alleles of the genes encoding p53 (ref. 97), TGF-β1 (ref. 98), 27Kip1 (ref. 99) and DMP1 (ref. 100). In each case, the genetically deficient mice are predisposed to tumor development, but the resulting tumors frequently retain a functional wild-type allele.

The prospect of compound haploinsufficiency, in which several contiguous genes that function in overlapping pathways to control cell growth are involved in the same deletion event, further complicates the search for the causal genetic changes in tumor progression. Results obtained from extensive studies of changes in human chromosome 3p in lung cancer show that several genes with tumor-suppressor activity are involved in the LOH event and that the remaining alleles are very rarely mutated[101]. Compound haploinsufficiency may involve single-copy deletions of interacting genes on different chromosomes, and if this is true it will be very difficult to identify the causal genetic changes in these chromosomal regions.

## Allele-specific genomic changes

If many common alleles affect susceptibility or resistance, it seems inevitable that selection pressures during tumor development will be reflected in patterns of allelic loss or gain in tumors, as well as in patterns of gene expression. It has been estimated that many cancers show LOH of at least 25%, and sometimes as much as 50%, of their alleles[48]. These changes are probably the result of a complex interplay between several inherited alleles of weak susceptibility genes and the somatic events responsible for tumor progression (Fig. 5).

Genetic background clearly affects the pattern of genetic changes seen in mouse tumors, because strain-specific mutations in oncogenes[102], allelic imbalance[103] and aberrations in gene copy number[104] are apparent in analyses of chemically induced or transgene-induced tumors. Observations of preferential allelic imbalance in mouse tumors have been correlated with the locations of germline susceptibility genes, as assessed by linkage analysis (A.E. Toland *et al.*, manuscript submitted). This suggests that, although studies of changes in gene copy number in human tumors have been highly informative, much more information remains buried in the same DNA samples. Analysis of allele-specific alterations in copy number and expression seems particularly promising for elucidating the relationships between germline polymorphisms and the development of sporadic cancers.

In summary, genomic, genetic and epigenetic events seem to be increasingly intertwined in tumor formation—even in tumors that seem to arise spontaneously. The combination of changes in gene dosage, methylation-based silencing and polymorphisms that cause diminished gene function greatly complicates the search for these important genes, because little trace of their action can be found by genetic analysis of the tumor DNA (Fig. 4*f*). Nevertheless, we can expect that the tools discussed above, coupled with population-based studies of the genetics of cancer incidence, will lead ultimately to their identification.

## Clinical applications and the future of cancer research

Although 'understanding the biology' is often given as the justification for genetic studies of cancer, prevention and treatment are also important goals. New and better treatments are certainly needed, but there are substantial gains to be made by applying genetic information to more effective use of the treatments that

we already have. Sensitive and non-invasive methods of early diagnosis may be developed from the detection of cancer-specific DNA mutations or abnormal DNA methylation patterns in readily accessible body fluids[105], or by using DNA replication proteins as markers of cycling cells exfoliated from cancers or preneoplastic lesions[106]. With improvements in specificity, or the use of sequential tests, such approaches may provide a cost-effective means of screening in the general population. It may also be possible to develop sensitive imaging techniques that will report specific molecular changes that are characteristic of cancer.

An important problem with current screening tests, such as prostate-specific antigen (PSA) screening for prostate cancer, is that the early lesions that are detected will not all progress to significant cancer. Large numbers of individuals can suffer considerable side-effects from interventions, such as prostatectomy or radical radiotherapy, that would not be needed if the potential of the supposed early cancerous change were accurately known. A similar situation can occur in the treatment of established cancers; for example, most women are recommended to have chemotherapy after surgery for early breast cancer to reduce the possibility of recurrence in the few individuals who have residual disease but cannot be identified. 'Molecular profiles' of individual cancers or preneoplastic lesions, based on either patterns of gene expression or patterns of genomic abnormality, are likely to find increasing application in predicting which individuals will benefit from additional treatment and which individuals do not need it[107].

The same principles will be applied to the selection of the treatment most likely to be effective in an individual case. Common cancers such as breast cancer will be reclassified into perhaps 10 or 20 subtypes on the basis of the patterns of the disrupted cellular pathways. Different patterns of molecular lesion will be found—empirically or by biological reasoning—to predict sensitivity or resistance to different agents, and the choice of treatment will be made accordingly. A possible downside is that the common cancers will be converted by molecular analysis into several individual uncommon subtypes with different sensitivities to treatment, which will be a less-attractive prospect for the companies on whom the development of new drugs currently depends.

The evaluation of new drugs will itself become quicker and more efficient: molecular readouts, perhaps in healthy volunteers, will be used to confirm that the drug hits the intended target at a minimum biologically effective dose; other readouts, biopsy or image-based, will provide rapid information about tumor response.

Finally, testing for predisposing genes will allow the identification of individuals at increased risk and sometimes the reassurance for those who are not. This has already entered standard clinical practice for dealing with high-penetrance familial cancer syndromes. The main outstanding problems are a lack of effective methods for non-surgical prevention or the early detection of most cancers, which means that the commonly recommended action after a positive gene test is surgical removal of the tissue at risk. The decision for or against surgery is made difficult by the variable and currently unpredictable age of cancer onset and the different organs that can be affected in many of the syndromes. In principle, effective prevention either may be based on the specific mechanism of the predisposition or (more probably) may be generic, as in the use of anti-estrogen treatment to prevent breast cancer. In either case, the rarity of the syndromes and the small numbers of at-risk individuals will mandate national and international cooperation to provide evidence of effect.

Recent modeling of the distribution of risk in the population that may result from the effects of common, weak genetic variation suggests the exciting possibility that a substantial fraction of sporadic cancers may occur in a predisposed minority of the population[27]. If confirmed, these models will allow the explicit planning of medical interventions in the population that are based both on medical and on economic criteria. The obstacles, which are still formidable, are first to identify most of the genetic and non-genetic factors that make up the risk distribution, and then to find the appropriate model that will combine them to give a single, individual prediction of risk.

The next decade should see the accumulation of enormous amounts of data on cancer genetics, epigenetics, genomics and gene expression. Progress will be made in mapping genome variation and understanding its origins, and this will be applied to the search for genes involved in predisposition. The emergence of a deeper understanding of the complex networks that control the altered behavior of the malignant cell will come only from the integration of knowledge obtained from these different sources. In this, we can learn a great deal from other systems for the study of complex interactions. The *TP53* gene has been referred to as one of the hubs of the cancer cell[64] and the same may be true for others such as *RB1*, *RAS* and *PTEN*, all of which occupy highly connected link positions in signal transduction.

A 'systems biology' approach to the cancer cell that is based on an understanding of scale-free networks[108] may be the only way to ensure disintegration of the support systems for cancer cells that have proved to be so resistant to conventional therapies. To do this, however, we must first establish the complex wiring diagrams of the cancer cell to identify the crucial, rate-limiting signals that pass through the principal hubs.


1. Knudson, A.G. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **68**, 820–823 (1971).
2. Friend, S.H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643–646 (1986).
3. Doll, R. & Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J. Natl. Cancer Inst.* **66**, 1191–1308 (1981).
4. Peto, J. Cancer epidemiology in the last century and the next decade. *Nature* **411**, 390–395 (2001).
5. Buell, P. & Dunn, J.E. Cancer mortality among Japanese Issei and Nisei of California. *Cancer* **18**, 656–664 (1965).
6. Newman, B., Austin, M.A., Lee, M. & King, M.C. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc. Natl. Acad. Sci. USA* **85**, 3044–3048 (1988).
7. Claus, E.B., Risch, N. & Thompson, W.D. Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am. J. Hum. Genet.* **48**, 232–242 (1991).
8. Peto, J. in *Cancer Incidence in Defined Populations* Vol. 4, Banbury Report (eds. Cairns, J., Lyon, J.L. & Skolnik, M.) 203–213 (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1980).
9. Hall, J.M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
10. Weinberg, R.A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
11. Linzer, D.I. & Levine, A.J. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* **17**, 43–52 (1979).
12. Lane, D.P. & Crawford, L.V. Tantigen is bound to a host protein in SV40-transformed cells. *Nature* **278**, 261–263 (1979).
13. Baker, S.J. *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217–221 (1989).
14. Malkin, D. *et al.* Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**, 1233–1238 (1990).
15. Kinzler, K.W. & Vogelstein, B. Colorectal tumors. In *The Genetic Basis of Human Cancer* (ed. Kinzler, K.W.) 583–612 (McGraw Hill, New York 2002).
16. Stambolic, V. *et al.* Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell* **95**, 29–39 (1998).
17. Eng, C. & Parsons, R. Cowden syndrome. In *The Genetic Basis of Human Cancer* (ed. Kinzler, K.W.) 527–537 (McGraw Hill, New York 2002).
18. Linehan, W.M., Zbar, B. & Klausner, R.D. Renal cancer. In *The Genetics Basis of Human Cancer* (ed. Kinzler, K.W.) 449–474 (McGraw Hill, New York 2002).
19. Venkitaraman, A.R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108**, 171–182 (2002).
20. Hoeijmakers, J.H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
21. Schor, S.L., Schor, A.M., Durning, P. & Rushton, G. Skin fibroblasts obtained from cancer patients display foetal-like migratory behaviour on collagen gels. *J. Cell Sci.* **73**, 235–244 (1985).
22. Olumi, A.F. *et al.* Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res.* **59**, 5002–5011 (1999).
23. Kurose, K. *et al.* Frequent somatic mutations in PTEN. and TP53 are mutually exclusive in the stroma of breast carcinomas. *Nat. Genet.* **32**, 355–357 (2002).
24. Zhu, Y., Ghosh, P., Charnay, P., Burns, D.K. & Parada, L.F. Neurofibromas in NF1: Schwann cell origin and role of tumor environment. *Science* **296**, 920–922 (2002).


© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

25. Holtzman, N.A. & Marteau, T.M. Will genetics revolutionize medicine? *N. Engl. J. Med.* **343**, 141–144 (2000).
26. Bell, J. The new genetics in clinical practice. *Br. Med. J.* **316**, 618–620 (1998).
27. Pharoah, P.D. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.* **31**, 33–36 (2002).
28. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
29. Cui, J. *et al.* After BRCA1 and BRCA2—what next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. *Am. J. Hum. Genet.* **68**, 420–31 (2001).
30. Antoniou, A.C. *et al.* A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br. J. Cancer* **86**, 76–83 (2002).
31. Peto, J. Breast cancer susceptibility—a new look at an old model. *Cancer Cell* **1**, 411–412 (2002).
32. Wright, A.F. & Hastie, N.D. Complex genetic diseases: controversy over the Croesus code. *Genome Biol.* **2**, COMMENT2007 (2001).
33. Cardon, L.R. & Bell, J.I. Association study designs for complex diseases. *Nat. Rev. Genet.* **2**, 91–99 (2001).
34. Botstein, D. & Rich, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.*, **33**,.228–237 (2003)
35. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
36. Couzin, J. Genomics. New mapping project splits the community. *Science* **296**, 1391–3139 (2002).
37. Horikawa, Y. *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.* **26**, 163–175 (2000).
38. Balmain, A. & Nagase, H. Cancer resistance genes in mice: models for the study of tumour modifiers. *Trends Genet.* **14**, 139–144 (1998).
39. Tripodis, N., Hart, A.A., Fijneman, R.J. & Demant, P. Complexity of lung cancer modifiers: mapping of thirty genes and twenty-five interactions in half of the mouse genome. *J. Natl. Cancer Inst.* **93**, 1484–1491 (2001).
40. Nagase, H., Mao, J.H. & Balmain, A. A subset of skin tumor modifier loci determines survival time of tumor-bearing mice. *Proc. Natl. Acad. Sci. USA* **96**, 15032–15037 (1999).
41. Saran, A. *et al.* Genetics of chemical carcinogenesis: analysis of bidirectional selective breeding inducing maximal resistance or maximal susceptibility to 2-stage skin tumorigenesis in the mouse. *Int. J. Cancer* **88**, 424–431 (2000).
42. Fijneman, R.J., de Vries, S.S., Jansen, R.C. & Demant, P. Complex interactions of new quantitative trait loci, Sluc1, Sluc2, Sluc3, and Sluc4, that influence the susceptibility to lung cancer in the mouse. *Nat. Genet.* **14**, 465–467 (1996).
43. van Wezel, T. *et al.* Gene interaction and single gene effects in colon tumour susceptibility in mice. *Nat. Genet.* **14**, 468–470 (1996).
44. Nagase, H., Mao, J.H., de Koning, J.P., Minami, T. & Balmain, A. Epistatic interactions between skin tumor modifier loci in interspecific (spretus/musculus) backcross mice. *Cancer Res.* **61**, 1305–1308 (2001).
45. Rouse, J. & Jackson, S.P. Interfaces between the detection, signaling, and repair of DNA damage. *Science* **297**, 547–551 (2002).
46. Kolodner, R.D., Putnam, C.D. & Myung, K. Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* **297**, 552–557 (2002).
47. Maser, R.S. & DePinho, R.A. Connecting chromosomes, crisis, and cancer. *Science* **297**, 565–569 (2002).
48. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
49. Friedberg, E.C. How nucleotide excision repair protects against cancer. *Nat. Rev. Cancer* **1**, 22–33 (2001).
50. Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**, 558–561 (1993).
51. Thibodeau, S.N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816–819 (1993).
52. Parsons, R. *et al.* Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**, 1227–1236 (1993).
53. Lynch, H.T. & de la Chapelle, A. Genetic susceptibility to non-polyposis colorectal cancer. *J. Med. Genet.* **36**, 801–818 (1999).
54. Rampino, N. *et al.* Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**, 967–969 (1997).
55. Parsons, R. *et al.* Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* **55**, 5548–5550 (1995).
56. Hastie, N.D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866–868 (1990).
57. Blackburn, E.H. & Challoner, P.B. Identification of a telomeric DNA sequence in *Trypanosoma brucei*. *Cell* **36**, 447–457 (1984).
58. Cahill, D.P. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303 (1998).
59. Jallepalli, P.V. & Lengauer, C. Chromosome segregation and cancer: cutting through the mystery. *Nat. Rev. Cancer* **1**, 109–117 (2001).
60. Fodde, R., Smits, R. & Clevers, H. APC, signal transduction and genetic instability in colorectal cancer. *Nat. Rev. Cancer* **1**, 55–67 (2001).
61. Bischoff, J.R. *et al.* A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J.* **17**, 3052–3065 (1998).
62. Zhou, H. *et al.* Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat. Genet.* **20**, 189–193 (1998).
63. Fukasawa, K., Choi, T., Kuriyama, R., Rulong, S. & Vande Woude, G.F. Abnormal centrosome amplification in the absence of p53. *Science* **271**, 1744–1747 (1996).
64. Vogelstein, B., Lane, D. & Levine, A.J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
65. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
66. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
67. Frame, S. & Balmain, A. Integration of positive and negative growth signals during ras pathway activation *in vivo*. *Curr. Opin. Genet. Dev.* **10**, 106–113 (2000).
68. Marx, J. Debate surges over the origins of genomic defects in cancer. *Science* **297**, 544–546 (2002).
69. Jonason, A.S. *et al.* Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc. Natl. Acad. Sci. USA* **93**, 14025–14029 (1996).
70. Deng, G., Lu, Y., Zlotnikov, G., Thor, A.D. & Smith, H.S. Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science* **274**, 2057–2059 (1996).
71. Rehman, I., Quinn, A.G., Healy, E. & Rees, J.L. High frequency of loss of heterozygosity in actinic keratoses, a usually benign disease. *Lancet* **344**, 788–789 (1994).
72. Gong, G. *et al.* Genetic changes in paired atypical and usual ductal hyperplasia of the breast by comparative genomic hybridization. *Clin. Cancer Res.* **7**, 2410–2414 (2001).
73. Shaaban, A.M. *et al.* Histopathologic types of benign breast lesions and the risk of breast cancer: case-control study. *Am J. Surg. Pathol.* **26**, 421–430 (2002).
74. Brown, K., Strathdee, D., Bryson, S., Lambie, W. & Balmain, A. The malignant capacity of skin tumours induced by expression of a mutant H-*ras* transgene depends on the cell type targeted. *Curr. Biol.* **8**, 516–524 (1998).
75. Pinkel, D. *et al.* Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc. Natl. Acad. Sci. USA* **85**, 9138–9142 (1988).
76. Gray, J.W. & Collins, C. Genome changes and gene expression in human solid tumors. *Carcinogenesis* **21**, 443–452 (2000).
77. Hayashizaki, Y. *et al.* Restriction landmark genomic scanning method and its various applications. *Electrophoresis* **14**, 251–258 (1993).
78. Kallioniemi, A. *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818–821 (1992).
79. Ginzinger, D.G. *et al.* Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. *Cancer Res.* **60**, 5405–5409 (2000).
80. Lisitsyn, N. & Wigler, M. Cloning the differences between two complex genomes. *Science* **259**, 946–951 (1993).
81. Zardo, G. *et al.* Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat. Genet.* **32**, 453–458 (2002).
82. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
83. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
84. Bernard, P.S. & Wittwer, C.T. Real-time PCR technology for cancer diagnostics. *Clin. Chem.* **48**, 1178–1185 (2002).
85. Conrads, T.P., Anderson, G.A., Veenstra, T.D., Pasa-Tolic, L. & Smith, R.D. Utility of accurate mass tags for proteome-wide protein identification. *Anal Chem.* **72**, 3349–3354 (2000).
86. Knezevic, V. *et al.* Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* **1**, 1271–1278 (2001).
87. Heisterkamp, N., Stam, K., Groffen, J., de Klein, A. & Grosveld, G. Structural organization of the *BCR* gene and its role in the Ph′ translocation. *Nature* **315**, 758–761 (1985).
88. Bernardi, R., Grisendi, S. & Pandolfi, P.P. Modelling haematopoietic malignancies in the mouse and therapeutical implications. *Oncogene* **21**, 3445–3458 (2002).
89. Druker, B.J. *et al.* Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N. Engl. J. Med.* **344**, 1038–1042 (2001).
90. Kallioniemi, O.P. *et al.* ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. USA* **89**, 5321–5325 (1992).
91. Vogel, C.L. *et al.* First-line Herceptin monotherapy in metastatic breast cancer. *Oncology* **61** (suppl. 2), 37–42 (2001).
92. Lyons, J.F., Wilhelm, S., Hibner, B. & Bollag, G. Discovery of a novel Raf kinase inhibitor. *Endocr. Relat. Cancer* **8**, 219–225 (2001).
93. Balmain, A. Cancer: new-age tumour suppressors. *Nature* **417**, 235–237 (2002).
94. Jaenisch R, & Adrian, B. Epigenetic regulation: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33**, 245–254 (2003)
95. Baylin, S. & Bestor, T.H. Altered methylation patterns in cancer cell genomes: cause or consequence? *Cancer Cell* **1**, 299–305 (2002).
96. Jones, P.A. & Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).
97. Venkatachalam, S. *et al.* Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *EMBO J.* **17**, 4657–4667 (1998).
98. Tang, B. *et al.* Transforming growth factor-β1 is a new form of tumor suppressor with true haploid insufficiency. *Nat. Med.* **4**, 802–807 (1998).
99. Fero, M.L., Randel, E., Gurley, K.E., Roberts, J.M. & Kemp, C.J. The murine gene p27Kip1 is haplo-insufficient for tumour suppression. *Nature* **396**, 177–180 (1998).
100. Inoue, K., Zindy, F., Randle, D.H., Rehg, J.E. & Sherr, C.J. Dmp1 is haplo-insufficient for tumor suppression and modifies the frequencies of Arf and p53 mutations in Myc-induced lymphomas. *Genes Dev.* **15**, 2934–2939 (2001).
101. Ji, L. *et al.* Expression of several genes in the human chromosome 3p21.3 homozygous deletion region by an adenovirus vector results in tumor suppressor activities *in vitro* and *in vivo*. *Cancer Res.* **62**, 2715–2720 (2002).
102. You, M. *et al.* Parental bias of Ki-ras oncogenes detected in lung tumors from mouse hybrids. *Proc. Natl. Acad. Sci. USA* **89**, 5804–5808 (1992).
103. Linardopoulos, S., Silva, S., Klein, G. & Balmain, A. Allele-specific loss or imbalance of chromosomes 9, 15, and 16 in B-cell tumors from interspecific F1 hybrid mice carrying Emu-c-myc or N-myc transgenes. *Int. J. Cancer* **88**, 920–927 (2000).
104. Hodgson, G. *et al.* Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nat. Genet.* **29**, 459–464 (2001).
105. Sidransky, D. Emerging molecular markers of cancer. *Nat. Rev. Cancer* **2**, 210–219 (2002).
106. Davies, R.J. *et al.* Analysis of minichromosome maintenance proteins as a novel method for detection of colorectal cancer in stool. *Lancet* **359**, 1917–1919 (2002).
107. van de Vijver, M.J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
108. Albert, R., Jeong, H. & Barabasi, A.L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).