Cancer genetics

Bruce A. J. Ponder

CRC Department of Oncology, University of Cambridge, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK (bajp@mole.bio.cam.ac.uk)

Cancer genetics has for many years focused on mutational events that have their primary effect within the cancer cell. Recently that focus has widened, with evidence of the importance of epigenetic events and of cellular interactions in cancer development. The role of common genetic variation in determining the range of individual susceptibility within the population is increasingly recognized, and will be addressed using information from the Human Genome Project. These new research directions will highlight determinants of cancer that lie outside the cancer cell, suggest new targets for intervention, and inform the design of strategies for prevention in groups at increased risk.

With few exceptions, cancers are derived from single somatic cells and their progeny. The cells in the emerging neoplastic clone accumulate within them a series of genetic or epigenetic changes that lead to changes in gene activity, and so to altered phenotypes which are subject to selection[1]. Ultimately, a cell population evolves that can disregard the normal controls of proliferation and territory and become a cancer. Hanahan and Weinberg[2] identify six 'hallmark features' of the cancer cell phenotype: disregard of signals to stop proliferating and of signals to differentiate; capacity for sustained proliferation; evasion of apoptosis; invasion; and angiogenesis.

Several factors can influence the evolution of a cancer. They are summarized in Fig. 1. The bold horizontal arrows represent the pathway of successive genetic or epigenetic events through which the cell acquires the cancer phenotype. Mostly these are somatic events, but in many of the inherited cancer syndromes, discussed below, one of the events is inherited. The alternative pathways to the right signify that overtly similar cancers may contain different combinations of genetic events, which may confer different properties. (This is the basis of 'molecular profiling' of tumours to predict clinical behaviour[3].)

Influences on the pathway are represented as vertical arrows. One set of influences affects the probability that a pathway event will occur. Within the cancer cell these include acquired or inherited defects in DNA repair or in cell-cycle checkpoints (see articles in this issue by Evan and Vousden, pages 342–348, and Hoeijmakers, pages 366–374), and, possibly, defects in the regulation of epigenetic events[4]. The production and destruction of endogenous mutagens such as free radicals will also affect the probability of mutational events, and may be modified by genetic variation. External influences include environmental exposures, for example diet or cigarette smoke, the response to which again may be modified by genetic variation in metabolic systems acting inside or outside the cell (refs 5, 6, and see article in this issue by Peto, pages 390–395).

Other factors influence the outcome of pathway events once they have occurred. Within the cell, these might be any type of variation that modifies the effect of the pathway event on the cellular phenotype, or the response of the altered cells to signals from outside. Outside the cell, possible influences include paracrine interactions with neighbouring cells[7] and systemic effects such as the effectiveness of cellular defence mechanisms against the developing cancer, or levels of circulating hormones or growth factors[8,9]. Normal genetic variation in these factors is likely to be the source of much of the low-level predisposition to cancer, and of the genetic modifier effects seen in human and experimental tumours[10,11].

Before focusing on the factors that influence carcinogenesis, we should first consider the historical development of ideas surrounding events on the main pathway of cancer development.

Events on the cancer pathway

The idea that tumours arise from somatic genetic change originated in the early 1900s. It was not until the necessary technologies became available in the early 1970s that tumour formation was related to the action of specific genes. The concepts that developed were of course shaped by the assays on which they were based. The idea of gain-of-function genetic alterations came from experiments that involved gene transfer into recipient cells; these cells could then be assayed for 'transformation' — an approximation to a cancer phenotype. The idea of loss-of-function genetic change came from two different directions: from epidemiology and the study of inherited predisposition[12], and from cell hybridization experiments in which malignancy was found to be recessive to the normal phenotype[13]. This history is relevant because, even today, our partial knowledge of the development of cancer is necessarily constrained by the assays we have available.

Gain-of-function genetic events

The key concept in relation to gain-of-function events is the 'oncogene' (for review, see ref. 14). By the late 1960s, it had been shown that cells in culture could be transformed by several DNA viruses and retroviruses, and subsequently that a single gene from these viruses (the first example was *src*, from Rous sarcoma virus) could carry out this transformation. Genes related in sequence to those in the transforming retroviruses were found in the DNA of normal cells; these genes had functions in the control of normal cell growth or differentiation, but their inappropriate activation by a variety of mechanisms could lead to cancer. The normal cellular genes were termed 'proto-oncogenes'; their activated counterparts were 'oncogenes'[15].

In the late 1970s, fragmented DNA from human cancer cells was transferred into cultured non-neoplastic cells (mouse NIH-3T3 fibroblasts were used) by transfection. The aim was to see if transformation would result and, if so, to recover the active DNA sequences. The first transforming gene to be recovered from human cancer cells by this technique turned out to

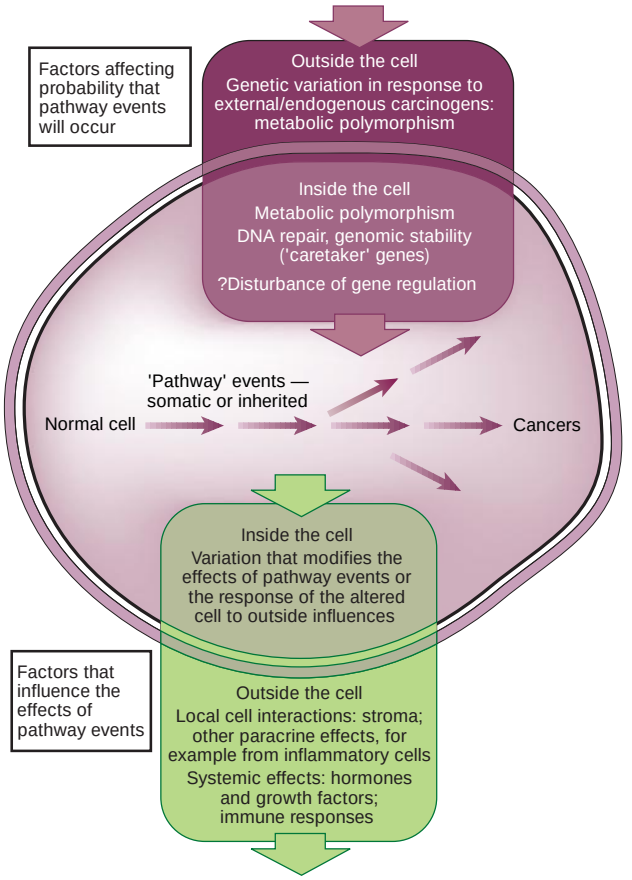


Figure 1 A framework for genetic effects on cancer development.

be a mutant form of *Ha-ras*, a proto-oncogene already familiar from retroviral studies[16]. Similar experiments have since identified many more transforming oncogenes[14], although these probably reflect only a subset of all gain-of-function genetic changes in cancer cells. For example, not all cells are good recipients for transfection, and the predominant use of rodent fibroblasts and of assays for 'transformation', rather than other aspects of the cellular phenotype that might be relevant, may have restricted the range of genes that could be found.

A further line of evidence for the role of activation of specific genes in cancer came from better techniques of chromosome analysis, starting with chromosome banding in the 1970s. In some tumours there were chromosomal translocations with consistent breakpoints, and some of these breakpoints proved to be in, or near to, already described proto-oncogenes — for example, c-*myc* in Burkitts lymphoma and c-*abl* in chronic myelogenous leukaemia[17]. In others, there were consistent regions of chromosomal amplification[18]. The inference, borne out by experiment, was that these specific chromosomal events could result in increased expression or activity of the related genes. Many more examples have been found[19], predominantly in haematological cancers and sarcomas where chromosomal identification is technically straightforward. A current question is whether the recently introduced techniques of chromosome analysis by molecular hybridization[20,21] will reveal similar mechanisms among the more complex chromosomal changes in epithelial malignancies, or whether perhaps epithelial cancers have different genetic mechanisms of development[22].

Loss-of-function genetic events

The impression given by the gene-transfer studies is of a single-step, gain-of-function mechanism for carcinogenesis. But this is a bias imposed by the methods used. The first evidence for loss-of-function genetic changes came from studies of children's cancers, in particular retinoblastoma. Like many cancers, retinoblastoma occurs in an inherited and a sporadic form. Knudson[12] described the distribution of age at diagnosis in inherited and sporadic cases. In inherited cases the distribution was consistent with a requirement for one further event for tumour formation. This event occurred with constant probability over time. In sporadic cases, the age distribution was more complex, and consistent with a need for two events. The inference was that in either case, two rate-limiting events were needed to form the tumour, and that in inherited cases one of these was already present in the germline. Comings[23] suggested that the two events might affect the two alleles of a single gene, implying that their effects would be recessive at the cellular level. Subsequently, in some inherited cases, a germline deletion was found on chromosome 13, implying that loss of a gene in that region might be the first event. This led to biochemical and molecular studies which showed that tumour development did indeed require loss of both copies of that region of chromosome 13 (ref. 24); using the chromosomal deletions as signposts, the *Rb* gene was ultimately cloned and found to be mutated in both copies in the tumours. *Rb* is thus the prototype of the class of 'tumour-suppressor genes'[25] where, in distinction to oncogenes, loss of function is required for tumorigenesis.

Linkage and positional cloning in inherited cancer syndromes has identified many more tumour-suppressor genes (for review, see ref. 26). Loss-of-function mutations are much more common than gain-of-function mutations in inherited predisposition, presumably because the loss of function is masked by the remaining normal allele during development (except in the recessive DNA-repair deficiencies), whereas a gain-of-function cancer-promoting mutation might well be lethal. In most inherited cancers, the germline loss-of-function allele represents one step on the pathway shown in Fig 1, and in most cases, as in retinoblastoma, the same genes are involved by somatic mutation in non-hereditary forms of the same cancer.

If the definition of a 'tumour-suppressor gene' were only that loss of function should contribute to cancer, then a list of potential genes could include not only genes such as *Rb*, but also a wider variety of genes acting at different points in Fig. 1. One might, for example, include genes that determine skin pigmentation as suppressors, on the grounds that fair-skinned individuals have a higher risk of skin cancer. Used as broadly as this, the term is perhaps of little help. Haber and Harlow[27] suggested a tighter definition which required the unequivocal demonstration of inactivating mutations of the gene. This had a practical rather than conceptual purpose — to lay down some unambiguous criteria by which the validity of the numerous candidates proposed as new suppressor genes could be judged. But four years later, we might be concerned that the requirement for mutation excludes genes where the predominant mechanism of loss of function is epigenetic[4]. If the term 'suppressor' is restricted to genes whose action lies within the cancer cell, two categories may usefully be distinguished. The first contains genes like *Rb* whose loss of function (by whatever mechanism) is rate limiting for cancer development and which lie on the direct pathway shown in Fig. 1 — the 'classical' tumour suppressors, termed 'gatekeepers' by Kinzler and Vogelstein[28]. Cancer predisposition due to these genes is tissue specific, although the mechanism of the specificity is generally unclear. The second group contains genes whose loss of function accelerates the acquisition of pathway events, but whose loss is not essential, and whose action lies outside the pathway itself. These are genes involved in DNA repair and genome integrity, which have been termed 'caretakers'[28]. (For details of DNA-repair genes, see review in this issue by Hoeijmakers, pages 366–374).

Somatic loss of a suppressor gene allele often involves a loss of chromosomal material, ranging in extent from a sub-band to the whole chromosome. Such events are conveniently assayed by 'loss of heterozygosity' (LOH), which is a comparison of polymorphic loci in DNA from blood and tumour in the same individual, and the finding of contiguous regions of tumour DNA where one allele is absent.

These regions might be expected to contain suppressor genes. LOH analysis has identified large numbers of regions of chromosomal loss in many of the common cancers[29], but the number of suppressor genes that have been identified convincingly, by the criterion of somatic mutation in the remaining allele, is small. There are several possible explanations: most LOH are noise; they reflect haploinsufficiency[30]; or perhaps the mutational criterion for identifying a suppressor gene is too stringent. In particular, there is growing evidence that epigenetic silencing rather than mutation is a common mechanism for loss of suppressor gene function.

Epigenetic pathway events

Epigenetic regulation of gene expression by methylation is an important mechanism of the determination of cell fate in embryogenesis. Disturbance of epigenetic mechanisms in the special case of genomic imprinting are responsible, for example, for loss of imprinting (LOI) and hence overexpression of the gene encoding insulin-like growth factor (IGF)-2 in the pathogenesis of Wilms tumour in Beckwith–Weidemann syndrome[31], and in some epithelial cancers, including colonic cancer[32]. It has been shown that methylation of regions rich in cytosine–guanine doublets ('CpG islands') in the promoter region in somatic cells is a common mechanism of epigenetic silencing of one or sometimes both alleles of tumour-suppressor genes such as *VHL*, *mlh1*, *p16* (*CDKN4*/*p16*^{INK4A}) and possibly *BRCA1* (ref. 4). It is not clear whether the epigenetic silencing of particular genes in cancer occurs through a stochastic process followed by selection, or whether certain promoters are predisposed (and if so, what might be the mechanisms involved) (reviewed in ref. 4). It is also unclear what determines whether a particular gene will lose function by an epigenetic or a mutational mechanism. Loss of function of the cyclin-dependent kinase inhibitor *p16* may occur through deletion, point mutation or promoter hypermethylation, but the frequency of each mechanism differs between tumour types[4]. Within the same tumour type, the mechanism may differ in different contexts. Thus, germline mutation of the *MLH1* gene is frequent in familial colon cancers with the microsatellite instability phenotype; but in sporadic cancers with this phenotype, promoter hypermethylation and loss of expression of *MLH1* (and, interestingly, LOI of the IGF-2 gene) is more common[33].

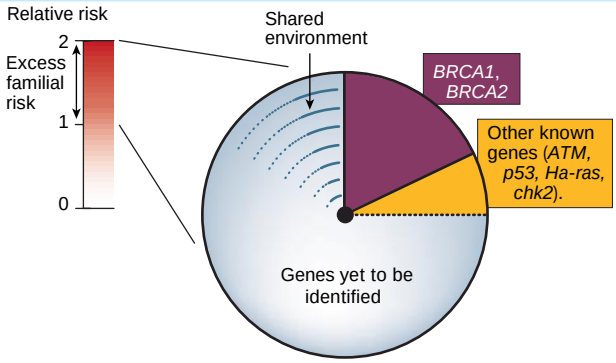
Although promoter hypermethylation has clearly been implicated in silencing of suppressor genes, there are other mechanisms by which changes in methylation might contribute to tumorigenesis. Examples come from the experimental manipulation of the activity of the maintenance DNA methylase Dnmt1 in mice. Thus, there is a reported increase in somatic mutation in mice heterozygous for loss of function of *Dnmt1*[34], and for widespread changes in gene expression in *Dnmt1*$^{-/-}$ mouse embryo fibroblasts rescued from apoptosis by inactivation of *p53* (ref. 35). The reduced incidence of intestinal adenomas in *Min* mice heterogeneous for a *Dnmt1*-null allele[36] (which seems counter to the increase in somatic mutation reported above) indicates that changes in genomic methylation may modify the phenotypic expression of a strong predisposing gene. The mechanisms of these effects, and their relevance to human cancer, require further investigation.

Epigenetic mechanisms can lead to a progressive, although patchy, silencing of some genes with age[37]. It is interesting to speculate to what extent our tissues may be a progressive mosaic either of gene silencing (or in the case of the IGF-2 gene, for example, of loss of imprinting), and what factors might influence this process[38]. The progressive silencing with age of the expression of β-galactosidase reporter genes in transgenic mice is well known. This is a highly artificial experimental situation which may have no relevance at all to endogenous genes in human tissues. Nevertheless, it is intriguing that histochemical staining of tissue sections showed the *β-gal* expression often to be strongly mosaic in intensity; the size of the positive patches diminished with age, and both the grain of the mosaic and its rate of disappearance differed on different genetic backgrounds[39]. Crosses between the relevant strains mapped a controlling locus to mouse chromosome 4 (ref. 40). Inheritance of methylation patterns in human DNA has also been described[41]. It is possible that susceptibility to cancer may be influenced by inherited variation in genes that regulate epigenetic silencing.

Patterns of pathway events

It has been estimated that between four and seven rate-limiting genetic events are needed for the development of the common epithelial cancers[42]. Because, presumably, the constraints to be overcome vary in importance between tissues, and can be evaded in different ways (for example, a signalling pathway may be disrupted at different points; see articles in this issue by Blume-Jensen and Hunter, pages 355–365, and Taipale and Beachy, pages 349–354), it is not surprising that the precise pattern of genetic alterations differs between cancers of different types, and of the same type[3,43]. But the patterns are not random. Specific associations of events are seen within individual tumours, and these presumably reflect the evolution of the tumours along particular pathways, as suggested in Fig. 1. Such patterns might potentially be important in several practical ways. They are the basis for the current optimism that 'molecular

Box 1
What genes might account for familial breast cancer?



Averaged across all ages, the risk of breast cancer to the sister, mother or daughter of a case is increased about twofold, as illustrated in the figure above. This excess familial risk provides an upper estimate (assuming all the risk is genetic) of the genetic effect that must be explained.

Modelling of genes that might be involved

The relationship between the familial relative risk (FRR) and the frequency and strength of predisposition of any predisposing allele is given by $FRR = [1 + p(1-p)(RR-1)^2]/[1 + p(RR-1)]^2$ where p is the allele frequency, and RR is the cancer risk in a carrier versus a non-carrier of the allele.

Assuming for purposes of illustration that the predisposing alleles are dominant, the table below shows some worked examples of the types of genetic effect that might explain the remaining familial clustering of breast cancer once *BRCA1* and *BRCA2* are accounted for. (Note that the real situation is quite unknown.)

	RR	Frequency of predisposing alleles in population	Contribution to excess familial risk	Number of such genes needed to account for all the observed familial risk*
'*BRCA3*'-like	10	0.002	0.16	4–5
	1.5	0.01	0.0025	240–320
Common, low-penetrance genes	1.5	0.1	0.020	30–40
	1.5	0.3	0.045	13–18

*Lower figure assumes that genes combine multiplicatively, upper figure assumes that genes combine additively.

Table 1 Inherited predisposition to cancer

	Contribution to overall cancer incidence	Clinical features	Frequency of predisposing alleles	Effect on individual risk
Inherited cancer syndromes	1–2% at most	Rare/unusual cancers or combinations of cancers. Sometimes with associated developmental defects or non-neoplastic phenotype. Mendelian dominant inheritance	Rare (~1:1,000 or less)	Strong: lifetime risks of cancer up to 50–80%
Familial cancers	?Up to 10% depending on definition	Families with several cases of common cancers that fall into a recognized pattern of cancer types (for example, breast and ovary; colon+endometrium+ urinary). Spectrum from families with multiple cases at young age (strongest evidence of predisposition) to two or three cases at older ages: many of the latter will be due to chance or to combinations of weaker genes. Generally show pattern consistent with dominant inheritance.	Uncommon to common	Moderate to weak
Predisposition without evident family clustering	No precise figure possible. Distribution of risk within population may result in substantial fraction of cancer incidence within predisposed minority	Single cases of cancer at any site, some with one or two affected relatives. The distribution of these cases in the population is probably determined by the combined effects of multiple genetic and non-genetic risk factors.	Multiple common alleles	Weak

profiling' of tumours by genomic or expression changes will provide information of clinical value[3,43]. If (which is not clear) the genetic pathway adopted by a given tumour is influenced either by genetic background or by environmental exposures, the 'molecular phenotype' may also define groups of tumours that aetiologically are more homogeneous, which would be valuable information in studies of genetic or environmental predisposition. Finally, adoption of a particular pathway of progression may constrain the possibilities for evolution of the cancer in the future. Clinical experience suggests that there are categories of pre-invasive change in, for example, prostate or breast epithelium which are, at the stage they are recognized, already largely determined in their potential for future malignancy. This implies that chance subsequent events in the evolution of these lesions cannot lead to a more malignant phenotype. If so, it will be important to find out whether molecular phenotypes can predict future malignant potential more accurately than current histological methods and, if they can, to use this information to judge strategies for intervention. A topical example is provided by the controversies surrounding radical treatment of early prostatic cancer detected by screening[44].

Genetic events outside the cancer pathway

So far, our focus has been on the developing cancer cell, and on the pathway genetic events and the deficiencies in DNA repair and genomic stability which may drive them. Productive though this focus has been and will continue to be, it provides only part of the picture. It is likely that genetic variation at other sites, both inside and outside the cancer cell, may substantially affect cancer development. This is illustrated by the following brief examples.

Gene–environment interaction

Genetic variation acting either within or outside the cancer cell may determine the outcome of interaction with exogenous carcinogens. A clear example is provided by the greater risk of cutaneous melanoma as a result of sun exposure in individuals with a fair skin, or who have many naevi (a phenotype which is genetically determined). Polymorphisms at the interleukin-1 locus, which are associated with increased production of interleukin-1β, are associated with both an increased risk of hypochlorhydria induced by the gastric pathogen *Helicobacter pylori*, and gastric cancer[45]. Analogous interactions are to be expected between chemical exposures and genetic variations in metabolic pathways, although well-attested examples are still rather few[5] (see ref. 5 and the article in this issue by Peto, pages 390–395). Such variation may in principle account for substantial differences in cancer susceptibility within the population, and knowledge of gene–environment interaction may indicate strategies for prevention in those at risk. Information about relevant genetic variation may also help in the design of epidemiological studies: categorization of subpopulations in terms of genetic risk may reduce heterogeneity and so increase power to detect causative exposures. Finally, tissue-specific patterns of gene expression may indicate which genes, and therefore which exposures, are likely to be relevant[46].

Local factors affecting the developing cancer cell

Wounding and chronic inflammation have long been known to be associated with cancer. Their effects may be mediated either through increased mitogenesis, which may be associated with increased mutation[47], or through paracrine effects, for example from inflammatory cells. Thus, production of the matrix metalloproteinase MMP9 by inflammatory cells has been implicated in the development of squamous cell carcinomas in an HPV-16 transgenic model, and various inflammatory cytokines have been shown to affect *p53* transcriptional regulation and apoptosis in epithelial cells (reviewed in ref. 48). Such processes presumably underlie the increased cancer risk in diseases such as ulcerative colitis and hereditary pancreatitis[49], which have an inherited component. It is also likely that there will be genetically determined variation in the wounding and inflammatory responses themselves, which will affect cancer initiation and progression.

There is accumulating evidence for an important role of paracrine interactions between epithelium and stroma in epithelial carcinogenesis[7]. Reciprocal 'conditioning' between cancer and adjacent stromal cells has been shown in tissue recombination experiments[50]. Irradiation of mammary gland stroma can promote the expression of tumorigenic potential by unirradiated epithelial cells[51]. Several studies provide evidence for a role of matrix metalloproteinases in the early as well as late stages of cancer development[7,52]. In general, transgenic mice that overexpress MMPs develop more cancers in response to oncogenic stimuli, whereas those that lack different MMPs or overexpress inhibitors develop fewer (but more malignant) cancers[53]. Although no data are currently available, it seems plausible that there will be polymorphic variation in MMP activity in human tissues, and that this may affect both the development of cancer and the behaviour of the cancers that result. Similar genetically determined variation may be expected in processes later in cancer development; such as angiogenic responses (see article in this issue by Liotta and Kohn, pages 375–379).

Systemic factors

Variations in circulating levels of hormones or growth factors show significant association with cancer risk. In one population-based study of oestradiol levels in post-menopausal women, there was an almost fivefold difference in risk of breast cancer between the upper and lower tertiles of circulating oestradiol level[54]. High levels of oestrogen might have carcinogenic effects either through direct stimulation of growth or as a by-product of mutagenic metabolites. Similar effects have been reported for several common cancers in

relation to the IGF family[9], and there is some evidence that a significant proportion of the variance in circulating IGF-1 levels is genetically determined[55]. Such genetic variation is a further plausible mechanism for a significant component of individual cancer susceptibility.

Inherited predisposition

The cardinal feature by which inherited predisposition is recognized clinically is family history. Cancer is common, so some families will contain several cases by chance. There is a spectrum of probability that a given family history reflects inherited predisposition from near-certainty of strong predisposition in the rare inherited cancer syndromes, to the possibility of weak effects in familial clusters (Table 1). Paradoxically, the largest category of inherited predisposition, in terms of expected fraction of cancer incidence, is the one with the weakest genetic effects — 'predisposition without evident family clustering'[56,57]. The combined contribution to overall breast cancer incidence of strongly predisposing mutations in *BRCA1* and *BRCA2*, which confer individual risks of around 60% by age 70, is less than 5%. By contrast, a predisposing allele with a relative risk of 2 and frequency of 20% could account for up to 20% of breast cancer incidence.

Strong predisposition

The human inherited cancer syndromes and their transgenic mouse counterparts have been reviewed extensively[58,59]. In the cases described so far, strong predisposition to cancer results either through inheritance of one of the events on the cancer 'pathway', or through effects on DNA repair or genome stability. Studies of the mechanisms of predisposition in these syndromes have led to substantial insights into cancer biology. Genetic testing for risk is now part of the standard of clinical care for families, although its value may be controversial when the practical benefits of the actions open to someone at risk are not clear[60].

Two features of these syndromes merit brief comment, because if we could explain them, we would know more about the development of cancer. They are tissue specificity and variability of expression. All inherited predisposition to cancer seems to show a considerable degree of tissue specificity, even in the case of predisposition by defective DNA repair. In most cases, there is no obvious lineage or physiological explanation for the patterns and the mechanisms are unknown. There may also be considerable variation in the age at onset of cancer and in the specific types of cancer that predominate not only within a given syndrome, but also within a single family. Some of this variation is due to different germline alleles of the main predisposing gene (for example, in Von Hippel Lindau disease[61], familial adenomatous polyposis[62] and multiple endocrine neoplasia type 2[63]) and some is environmental or chance. But much of the within-family variation is probably attributable to the effects of genetic modifiers. This has been clearly shown in a number of mouse cancer models[64], and by the demonstration that concordance of phenotype in neurofibromatosis type 1 is greatest in monozygotic twins and decays with increasing distance of relationship[10]. Many of these modifiers are likely to overlap with the low-penetrance predisposing genes described in the next section. One practical implication of modifier effects is that the quoting of risks for individuals who carry genes such as *BRCA1* is an uncertain business. This is relevant to insurance, where the uncertainties are perhaps not sufficiently recognized. Inappropriately high-risk figures may be used, which derive from reports of the extreme set of families that are usually the first to be studied. A more speculative implication is that if we knew the mechanisms of modification, we might exploit this knowledge for treatment[11] or prevention.

Weak predisposition

Weak predisposition to cancer may in principle result from weak alleles of the pathway or caretaker genes described in the last section, or from genetic variation at the other sites indicated in Fig. 1. The study of weak predisposition is of interest both for its possible public-health implications[56] and because just as the study of inherited cancer syndromes identified 'pathway' genes, so weak predisposition may point to a wider range of processes that are relevant to cancer development, and to interactions between them. The search for these genes is just beginning and as yet there are few data. The principles can be illustrated from studies of breast cancer.

In breast cancer, the risk to close relatives of a case, averaged across all ages, is about twofold. Most of this familial risk is probably genetic in origin (see article in this issue by Peto, pages 390–395). The risk is about the same for the mother, sisters or daughters of a case, suggesting dominant rather than recessive effects. Large population-based studies indicate that only 15–20% of overall familial risk is attributable to mutations in *BRCA1* and *BRCA2*[65]. The possibilities for the remaining 80% are some combination of a small number of moderately strong genes, and a larger number (possibly a hundred or more) of weaker genes (Box 1). If moderately strong genes exist, it should in principle be possible to identify them by linkage in families. The weaker genes will not, on the whole, result in multiple case families and so must be sought by a different approach: a comparison of the frequency of candidate genetic variants between a large series of cancer cases and controls (an 'association study'). The candidate genes might lie anywhere in the scheme outlined in Fig. 1. Of the first 40 or so candidates tested for association with breast cancer, a few show evidence of weak effects, most of which require independent confirmation. They include genes encoding steroid hormone receptors and paracrine growth factors, and genes involved in metabolism of exogenous chemicals, and in DNA repair[66]. The variant alleles are associated with risks of around 1.5-fold and are predicted to account for only a few per cent of breast cancer incidence. Collectively they account for only a very small fraction of the familial risk. Almost certainly there are many more genes to be identified, which together will account for a much higher fraction of cancer incidence than the genes in the inherited cancer syndromes.

The identification of these genes will be greatly accelerated by the data from the Human Genome Project[67]. The search relies on cataloguing the DNA sequence variation within the population, and showing (currently on a 'candidate' gene-by-gene basis) that particular variants are significantly associated either with disease susceptibility or with some other aspect of disease phenotype such as treatment response or survival[68]. The most readily assayed form of genomic variation is the single nucleotide polymorphism or 'SNP': of the order of one million SNPs have been identified and are available from genomic databases[69]. Comparable data from the mouse genome project will support similar studies in mice. Here, the availability of cancer models, and the possibilities of experimental manipulation on a defined genetic background, allow an empirical search for genetic modifiers and low-penetrance genes on a genome-wide basis, which may provide valuable candidates to test in human populations[11,70]. Lessons from the much longer history of quantitative genetic analysis in lower organisms are also likely to be valuable[71]. There are, of course, many problems still to be addressed (for review, see ref. 68), but possibly the most pressing is the lack of sufficiently large and well-documented human case-control sets to analyse. This, rather than the genetic or statistical technology, is currently the limiting factor. In general, funding agencies have in the past been curiously unwilling to face up to this; now when they may be changing, there is the potential threat from 'the new ethics' discussed in the article by Peto, pages 390–395, which may put further costs and difficulties in the way. Despite this, it seems certain that the next decade will see significant advances in understanding the polygenic basis of many diseases, including cancer.

The future

Some have hailed the approaching era of the polygenic basis of disease as a new dawn[72]; others are sceptical[73]. The sceptics argue, in relation to cancer predisposition, that the genes are weak in comparison to lifestyle and environmental causes or risk, and it will be difficult to use this type of genetic information to practical effect. The

numbers relating to avoidable cancer risks presented in the article by Peto seem to support this. However, as also discussed by Peto, the picture may be different if the aggregate effect of several genes and other non-genetic predisposing factors can define a spectrum of risk across the population which is sufficiently wide. In that case, these factors might be used to construct 'risk profiles' that would identify either small groups of people at high risk who account for a substantial fraction of cancer incidence, or large groups who are at very low risk (and who can therefore be discouraged from taking up costly and perhaps risky interventions). Our modelling of the distribution of breast cancer risk in a UK population (Pharoah *et al.*, unpublished data) predicts that there may be as much as a 40-fold difference in relative risk between the highest and lowest quintiles of the distribution that could be defined by a genotypic profile. As genes are identified, the predictive power of the available profiles can be tested in the large population cohorts that are being followed for cancer incidence. The goal of genotypic profiling is probably distant, because it may require that a majority of the tens or even hundreds of predisposing alleles be identified; and if it does become possible, there will be social and ethical issues to address. Nevertheless, it seems an attainable goal. □

1. Nowell, P. C. The clonal evolution of tumour cell populations. *Science* **194,** 23–28 (1974).
2. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).
3. Liotta, L. & Petricoin, E. Molecular profiling of human cancers. *Nature Rev. Genet.* **1,** 48–56 (2000).
4. Baylin, S. B. & Herman, J. G. DNA methylation in tumorigenesis. *Trends Genet.* **16,** 168–174 (2000).
5. Rothman, N. *et al.* The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochem. Biophys. Acta.* **1471,** C1–C10 (2001).
6. Olden, K. & Wilson, S. Environmental health and genomics: visions and implications. *Nature Rev. Genet.* **1,** 149–153 (2000).
7. Tlsty, T. D. & Hein, P. W. Know thy neighbor: stromal cells can contribute oncogenic signals. *Curr. Opin. Genet. Dev.* **11,** 54–59 (1998).
8. Risch, A. Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. *J. Natl Cancer Inst.* **90,** 1774–1768 (1998).
9. Yu, H. & Rohan, T. Role of the insulin-like growth factor family in cancer development and progression. *J. Natl Cancer Inst.* **92,** 1472–1489 (2000).
10. Easton, D. F., Ponder, M. A., Huson, S. M. & Ponder, B. A. J. An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. *Am. J. Hum. Genet.* **53,** 305–313 (1993).
11. Nadeau, J. H. Modifier genes in mice and humans. *Nature Rev. Genet.* **2,** 165–174 (2001).
12. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA* **68,** 820–823 (1971).
13. Harris, H. The analysis of malignancy by cell fusion: the position in 1988. *Cancer Res.* **48,** 3302–3306 (1988).
14. Park, M. in *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 205–228 (McGraw Hill, New York, 1998).
15. Bishop, J. M. Enemies within: the genesis of retrovirus oncogenes. *Cell* **23,** 5–6 (1981).
16. Parada, L. P., Tabin, C. J., Shih, C. & Weinberg, R. A. Human EJ bladder carcinoma oncogene is a homologue of Harvey Sarcoma virus ras gene. *Nature* **297,** 474–477 (1982).
17. Rabbitts, T. H. Chromosomal translocations in human cancer. *Nature* **372,** 143–149 (1994).
18. Brodeur, G. M. & Hogarty, M. D. in *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 161–172 (McGraw Hill, New York, 1998).
19. Mitelman, F., Mertens, F. & Johansson, B. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genet.* **13,** 417–474 (1997).
20. Gray, J. W. & Collins, C. Genome changes and gene expression in human solid tumours. *Carcinogenesis* **21,** 443–452 (2000).
21. The BAC Resource Consortium. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409,** 953–957 (2001).
22. DePinho, R. A. The age of cancer. *Nature* **408,** 248–254 (2000).
23. Comings, D. A general theory of carcinogenesis. *Proc. Natl Cancer Inst.* **70,** 3324–3328 (1973).
24. Cavenee, W. K. *et al.* Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305,** 779–781 (1983).
25. Weinberg, R. A. Tumor suppressor genes. *Science* **254,** 1138–1146 (1991).
26. Fearon, E. R. in: *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 229–236 (McGraw Hill, New York, 1998).
27. Haber, D. & Harlow, E. Tumor suppressor genes: evolving definitions in the genomic age. *Nature Genet.* **16,** 320–322 (1997).
28. Kinzler, K. W. & Vogelstein, B. Gatekeepers and caretakers. *Nature* **386,** 761–763 (1997).
29. Orborne, R. J. & Hamshere, M. G. A genome-wide map showing common regions of loss of heterozygosity/allelic imbalance in breast cancer. *Cancer Res.* **60,** 3706–3712 (2000).
30. Faro, M. L. *et al.* The murine gene $p27^{KIP1}$ is haploinsufficient for tumour suppression. *Nature* **396,** 177–180 (1998).
31. Rainier, S. *et al.* Relaxation of imprinted genes in human cancer. *Nature* **362,** 747–749 (1993).
32. Cui, H. *et al.* Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability. *Nature Med.* **4,** 1276–1280 (1998).
33. Herman, J. G. *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl Cancer Inst.* **95,** 6870–6875 (1998).
34. Chen, R. Z., Pettersson, U., Beard, C., Jackson-Crushy, L. & Jaenisch, R. DNA hypomethylation leads to elevated mutation rates. *Nature* **395,** 89–93 (1998).
35. Jackson-Crushy, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genet.* **27,** 31–39 (2001).
36. Cormier, R. J. & Dove, W. F. $Dnmt^{n/+}$ reduces the net growth rate and multiplicity of intestinal adenomas in C57BL/6-multiple intestinal neoplasia (Min)/+ mice. *Cancer Res.* **60,** 3965–3970 (2000).
37. Issa, J.-P. J. *et al.* Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nature Genet.* **7,** 536–540 (1994).
38. Feinberg, A. P. Methylation meets genomics. *Nature Genet.* **27,** 9–10 (2001).
39. Sapienza, C. Genome imprinting, cellular mosaicism and carcinogenesis. *Mol. Carcinogenesis* **3,** 118–124 (1990).
40. Engler, P. *et al.* A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. *Cell* **65,** 939–947 (1991).
41. Silver, A. J. & White, R. Inheritance of allelic blueprints for methylation patterns. *Cell* **54,** 145–152 (1988).
42. Renan, M. J. How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol. Carcinogenesis* **7,** 139–146 (1993).
43. Suzuki, S. *et al.* An approach to analysis of large scale correlations between genome changes and clinical endpoints in ovarian cancer. *Cancer Res.* **60,** 5382–5385 (2000).
44. Neal, D. E. & Donovan, J. L. Prostate cancer: to screen or not to screen? *Lancet Oncol.* **1,** 17–24 (2000).
45. El-Omar, E. M. *et al.* Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* **404,** 398–402 (2000).
46. Williams, J. A. Single nucleotide polymorphisms, metabolic activation and environmental carcinogenesis: why molecular epidemiologists should think about enzyme expression. *Carcinogenesis* **22,** 209–214 (2001).
47. Ames, B. N. & Gold, L. S. Too many rodent carcinogens: mitogenesis increases mutagenesis. *Science* **249,** 970–971 (1990).
48. Cordon-Carlo C. & Prives C. At the crossroads of inflammation and tumorigenesis. *J. Exp. Med.* **190,** 1367–1370 (1999).
49. Lowenfils, A. B. *et al.* Hereditary pancreatitis and the risk of pancreatic cancer. *J. Natl Cancer Inst.* **89,** 442–446 (1997).
50. Olumi, A. F. *et al.* Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res.* **59,** 5002–5011 (1999).
51. Barcellos-Hoff, M. H. & Ravani, S. A. Irradiated mammary gland stroma promotes the expression of tumorigenic potential by unirradiated epithelial cells. *Cancer Res.* **60,** 1254–1260 (2000).
52. Sternlicht, M. D. *et al.* The stromal protease MMP3/Stromelysin-1 promotes mammary carcinogenesis. *Cell* **96,** 137–146 (1999).
53. Coussens, L. M., Tinhl, C. C., Hanahan, D. & Werb, Z. MMP-9 supplied by bone marrow-derived cells contributes to skin carcinogenesis. *Cell* **103,** 481–490 (2000).
54. Thomas, M. V. *et al.* A prospective study of endogenous serum hormone concentrations in post-menopausal women on the island of Guernsey. *Br. J. Cancer* **76,** 401–405 (1997).
55. Jernstrom, H. *et al.* Genetic and non-genetic factors associated with variation of plasma levels of insulin-like growth factor-1 and insulin-like growth factor binding protein-3 in healthy pre-menopausal women. *Cancer Epidemiol. Biomarkers Prev.* (in the press).
56. Peto, J. in *Cancer Incidence in Defined Populations* Banbury Report 4 (eds Cairns, J., Lyon, J. L. & Skolnick, M.) (Cold Spring Harbor, New York, 1980).
57. Ponder, B. A. J. Inherited predisposition to cancer. *Trends Genet.* **6,** 213–218 (1990).
58. Vogelstein, B. & Kinzler, K. W. (eds) *The Genetic Basis of Human Cancer* (McGraw Hill, New York, 1998).
59. Klausner, R. D. Studying cancer in the mouse. *Oncogene* **18,** 5249–5252 (1999).
60. Ponder, B. A. J. Genetic testing for cancer risk. *Science* **278,** 1050–1054 (1997).
61. Chen, F. *et al.* Germline mutations in the Von Hippel Lindau disease tumour suppressor gene: correlation with phenotype. *Hum. Mutat.* **5,** 66–75 (1995).
62. Kinzler, K. W. & Vogelstein, B. in *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 565–587 (McGraw Hill, New York, 1998).
63. Eng. C. & The International RET Mutation Consortium. The relationship between specific RET protooncogene mutations and disease phenotype in multiple endocrine neoplasia type 2. *J. Am. Med. Assoc.* **276,** 1575–1579 (1996).
64. Halberg, R. B. *et al.* Tumorigenesis in the multiple intestinal neoplasia mouse: redundancy of negative regulations and specificity of modifiers. *Proc. Natl Acad. Sci. USA* **97,** 3461–3466 (2000).
65. The Anglian Breast Cancer Study Group. Prevalence of BRCA1 and BRCA2 mutations in a large population based series of breast cancer cases. *Br. J Cancer* **83,** 1301–1308 (2000).
66. Dunning, A. M. *et al.* A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **8,** 843–854 (1999).
67. Chakravarti, A. … to a future of genetic medicine. *Nature* **409,** 822–823 (2001).
68. Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nature Rev. Genet.* **2,** 91–99 (2001).
69. The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409,** 928–933 (2001).
70. Balmain, A. & Nagase, H. Cancer resistance genes in mice: models for the study of tumour modifiers. *Trends Genet.* **14,** 139–144 (1998).
71. Mackay, T. F. C. Quantitation trait loci in *Drosophila*. *Nature Rev. Genet.* **2,** 11–20 (2001).
72. Bell, J. The new genetics in clinical practice. *Br. Med. J.* **316,** 618–620 (1998).
73. Holtzman, N. A. & Marteau, T. M. Will genetics revolutionize medicine? *New Engl. J. Med.* **343,** 141–144 (2000).

Acknowledgements

I apologize that space has restricted the topics and the references that could be included. I thank D. Easton for advice on the figure in Box 1, and several colleagues for helpful suggestions. B.A.J.P. is a Gibb Fellow of The Cancer Research Campaign (CRC).