PERGAMON

European Journal of Cancer 37 (2001) S88–S96

European Journal of Cancer

www.ejconline.com

Genetics and the common cancers

J. Peto[a,*], R.S. Houlston[b]

[a]*Section of Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK*
[b]*Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK*

Received 13 June 2001; received in revised form 10 July 2001; accepted 17 July 2001

Abstract

Rare highly penetrant genes cannot account for much of the familial risk for most common cancers, and there is increasing evidence that a high proportion of cancers arise in a susceptible minority who carry low-penetrance genes or gene combinations. The evidence for the existence of such genes and the prospects for identfying them are reviewed. © 2001 Published by Elsevier Science Ltd. All rights reserved.

Keywords: Cancer genetics; Low-penetration genes; Linkage analysis; Segregation analysis; Review

1. Introduction

Some of the mechanisms by which inherited mutations in certain genes cause a high cancer risk are already known. The same genes are often lost or mutated somatically in the multi-step evolution of cancer in non-susceptibles and may therefore be targets for novel approaches to screening, prevention or treatment. Most of the cancer susceptibility genes so far discovered are highly penetrant, but too rare to cause more than a few percent of most types of cancer. There is quite strong evidence that a considerably larger proportion of cancers are due to genetic susceptibility, but the relevant genes or gene combinations do not cause high enough risks to produce large multiple-case cancer families and may therefore prove difficult to identify by conventional linkage analysis. If a small number of common susceptibility genes account for these moderate cancer risks then many of them will probably be discovered quite rapidly by the methods discussed below. However, if large numbers of rare susceptibility genes are involved, whether acting independently or synergistically, they may be discoverable only through advances in cell and molecular biology.

2. Familial clustering of common cancers

2.1. Risks in first-degree relatives

First-degree relatives of patients with many common cancers are at increased risk for cancer at the same site. This has been recognised for many years, but for most cancers there are still too few systematic studies to provide precise estimates of these familial risks, particularly for relatives of younger patients. Table 1 gives pooled estimates of published risks in relatives of affected cases for a number of common cancers. (Data and references are given in Houlston and Peto [1]. For sites where numerous reports have been published, only the larger studies were included). For first-degree relatives (parents, siblings and children) of patients with most common cancers the risk of developing cancer at the same site is generally 2- to 3-fold higher than in the general population. Such apparently moderate relative risks could be due to the combined effects of shared environmental factors and interacting genes, but Table 2 shows that if single genes underlie these risks their effects must be substantial. For a dominant gene to cause a relative risk of 2 in patients' siblings, the risk in susceptible individuals has to be at least 10 times higher than in non-susceptibles, and for a recessive gene, the risk ratio must be over 20. Table 2 also shows that a moderate familial risk is consistent with a wide range of gene frequencies and genetic mechanisms, so the underlying genetic model cannot be inferred from the relative risks in first-degree relatives.

* Corresponding author. Also at: London School of Hygiene and Tropical Medicine, London WCIE 7HT, UK. Tel.: +44-20-8-643-8901; fax: +44-20-8-770-7876.
E-mail address: j.peto@icr.ac.uk (J. Peto).

PII: S0959-8049(01)00255-6

2.2. Age-specific familial risks

If tumours develop earlier in susceptibles, the relative risk in patients' relatives will be greatest in young relatives of young cases. This pattern of risk is exhibited by many common cancers, including breast, colon, prostate, bladder and melanoma. The distribution of age at diagnosis in susceptible individuals can differ from that in non-susceptibles for two reasons. Firstly, those susceptible may have a different underlying disease process. Ashley [2] suggested over 30 years ago that adenomatous polyposis coli (APC) patients have inherited one of the series of carcinogenic 'hits' that occur somatically in the development of sporadic colon cancer, based on the observation that the susceptible to non-susceptible incidence ratio declines with increasing age. In the same year, DeMars [3] suggested that apparently autosomal dominant syndromes such as APC and familial retinoblastoma should be interpreted as autosomal recessives at the cellular level, because the cancers "appear as a result of subsequent somatic mutations in which individual cells become homozygous for a recessive neoplasm-causing gene". Confirmation of this hypothesis has been one of the major achievements of the last two decades of cancer research, and it is now known that susceptibility to these and several other dominant cancer syndromes is due to germline mutation in one allele of a tumour suppressor gene, both alleles of which are frequently damaged or deleted somatically in the development of non-hereditary cancers of the same type. In a plausible extension of this model, inherited mutation in a DNA repair gene followed by somatic loss of the wild-type allele can cause genetic instability that usually kills the cell, but occasionally causes multiple damage leading immediately to cancer, so these subsequent events in carcinogenesis are no longer rate-limiting. This mechanism has recently been suggested for hereditary non-polyposis colon cancer (HNPCC) to explain the flat age–incidence curve in susceptible adults [4].

Secondly, progressive elimination of the most susceptible would result in earlier age at first cancer diagnosis in susceptibles even if the susceptible to non-susceptible incidence ratio were the same at each age. Incidence rates of second and subsequent skin tumours in mice treated with benzo(a)pyrene are consistent with such a model [5]. The marked difference between the distributions of age at diagnosis of patients' first tumours for hereditary and sporadic retinoblastomas is due almost entirely to the effects of elimination. In contrast to APC, where most somatic events occur later in life, the susceptible to non-susceptible incidence ratio will be virtually independent of age for embryonal tumours where sporadic cases involve an extra somatic mutation in a tumour suppressor gene that occurs in the rapidly growing target organ during embryogenesis or in infancy. The hypothesis of germline mutation among susceptibles in one allele of a tumour suppressor gene has proved to be the correct explanation of the 100 000-fold difference in incidence between hereditary and sporadic retinoblastomas; but as Hethcote and Knudson [6] noted, their data were also consistent with a model in which the number of somatic mutations, and hence the distribution of ages at diagnosis for all tumours (not first tumours, as most susceptibles develop multiple tumours) are the same for hereditary as for sporadic retinoblastoma.

Statistical models based on age-specific cancer rates have sometimes suggested unexpected mechanisms that were later shown to exist, but most remain untested. A novel hypothesis of this sort concerning breast cancer is suggested in Fig. 1, which shows the pattern of incidence in patients' contralateral breasts and in their MZ

Table 1
Summary estimates of the relative risk for the same type of cancer in patients' first-degree relatives (see Ref. [1] for references)

Site	Type of relative	Relative risk	95% CI
Breast	All	2.1	2.0–2.2
	Age < 50	2.3	2.2–2.5
	Age > 50	1.8	1.6–2.0
Ovary	All	3.1	2.5–3.7
Stomach	All	2.5	2.0–3.0
Colorectal	All	2.3	2.0–2.5
	Age < 45	3.9	2.4–6.2
Prostate	All	2.2	1.1–2.4
Renal	All	4.4	2.0–9.7
Testicular	Fathers	2.3	1.5–3.6
	Brothers	9.8	7.1–13.7
Uterus	All	2.1	1.3–3.4
Bladder	All	1.4	1.2–1.6
Melanoma	All	2.5	2.1–3.0

95% CI, 95% confidence interval.

Table 2
Dominant and recessive gene models that would cause relative risks of 5 or less in siblings of cancer patients

	Relative risk to sibling			
Gene frequency	1.5	2	3	5
Dominant gene model				
0.001	25	35	50	74
0.01	10	14	21	35
0.05	6	10	21	150
0.1	6	13	280	–
0.2	10	–	–	–
Recessive gene model				
0.01	143	203	289	414
0.05	30	44	64	96
0.1	16	24	36	60
0.5	8	35	–	–

Tabulated values are susceptible to non-susceptible risk ratios.

twins, sisters and mothers [7]. The high contralateral and MZ twin rates imply that a high proportion, and perhaps most, of all breast cancers arise in a susceptible minority of women; but the most surprising aspect of these data is the pattern of incidence in patients' mothers and sisters, which seems to suggest that the incidence rate rises abruptly to a constant level at different ages in different families. Such an effect would be difficult to reconcile with known mechanisms of carcinogenesis and susceptibility, but no equally simple model seems to fit the pattern shown in Fig. 1 [7]. Data on familial age-specific incidence rates for breast cancer are still very imprecise, however, particularly for twins and young relatives.

2.3. Familial associations between common cancers

A number of dominant susceptibility genes cause cancer at several sites. These include clustering of adenocarcinomas of the colorectum and endometrium in HNPCC families [8], the association of soft tissue sarcomas, leukaemia, brain and breast tumours in the Li–Fraumeni syndrome [9] and clustering in *BRCA1* and *BRCA2* families of cancers of the breast and ovary, with smaller risks for some other sites including colon, prostate and pancreas [10,11].

A systematic study of familial correlations between 28 different types of cancer using the Utah population database revealed a number of significant associations [12]. Environmental and behavioural risk factors which are shared within families contribute to some associations, such as the familial clustering for cancers of smoking-related sites. Clustering of cancers of the female genitalia, lip and oral cavity may reflect a common viral aetiology. Some associations are similar to those caused by rare penetrant genes. *BRCA1* and *BRCA2* presumably contribute to the associations between breast, colon and prostate and between ovary and pancreas, and the increased risk of soft tissue cancers among relatives of breast cancer probands may be due in part to the Li–Fraumeni syndrome. Truncating mutations in these genes are probably too rare to account for all of these associations, suggesting the existence of less penetrant variants, or perhaps other genes that affect some of the same pathways. Evidence of the HNPCC complex of cancers was seen in relatives of young uterine cancer probands. In addition to these expected associations between cancer sites, various significant associations such as thyroid cancer and non-Hodgkin's lymphoma with breast cancer and leukaemias with colorectal cancer suggest previously unrecognised hereditary effects.

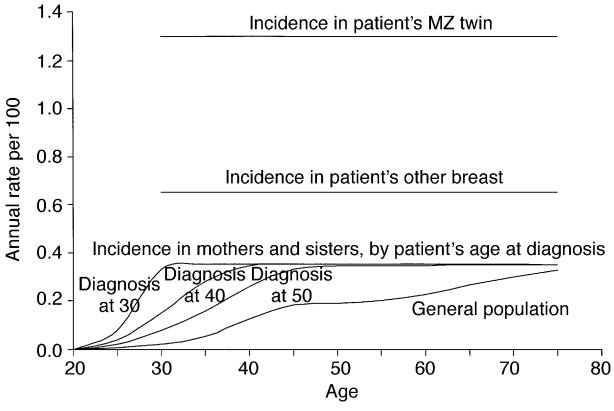

Fig. 1. The pattern of breast cancer incidence in patients' relatives and in the general population suggests that the incidence rises to a high constant level at different ages in different families (from Ref. [7]). MZ, monozygotic.

2.4. Benign lesions and other associated phenotypes

Houlston and Peto [1] list several benign lesions that are associated with an increased risk of cancer and have an inherited basis. In addition to the classical highly penetrant syndromes such as APC and xeroderma pigmentosum (XP), several such lesions are associated with moderate site-specific cancer risks. Melanoma risk is strongly correlated with numbers of melanocytic naevi [13]. Numbers of naevi are highly correlated in monozygotic twins, but not in dizygotic twins, suggesting a major genetic component in naevus prevalence and hence presumably in melanoma risk [14]. Further evidence of genetic susceptibility to naevi was found by Goldgar and colleagues in studies of families of melanoma cases [15]. Cuzick and colleagues [16] found that palmar keratoses, which are associated with bladder cancer, are more frequent in first-degree relatives of bladder cancer cases, particularly if the case also had keratoses. Spouses of cases also had an increased, but smaller, risk of bladder cancer, especially if the case had keratoses [16]. These results appear to implicate genetic, as well as environmental (possibly viral), factors in the aetiology of keratoses. Other associated phenotypes include cryptorchidism (testis), supernumerary nipples (urogenital), inflammatory bowel disease (colorectal) and endometriosis (breast and ovary) [1].

2.5. Segregation analysis

Genetic models of familial cancer can be formally tested using segregation analysis, in which the observed pattern of disease incidence in each pedigree is compared with that predicted under different models. Segregation analyses of breast, colorectal, lung, ovarian and testicular cancer pedigrees are summarised in Houlston and Peto [1]. The majority showed evidence for major dominant genes predisposing to these cancers. In principle, risks to second-degree and more distant relatives provide useful additional evidence on the mode of inheritance, since for a simple dominant gene the

excess relative risk (R−1) reduces by a factor of 2 for each degree of relationship [17]. In practice, however, unbiased data on second- and third-degree relatives are rarely available.

If a single gene accounts for the familial risk, segregation analysis of cancer families ascertained within population-based studies can, in principle, provide estimates of both gene frequency and age-specific cancer incidence for susceptibles and non-susceptibles. For most cancers, however, it is likely that several genes with varying penetrances contribute to the risk in relatives. For example, the pattern of familial breast cancer risk in the CASH study could be accounted for by a single dominant gene with a frequency of 0.0033 conferring a risk of 67% by age 70 years [18]; but it is now known that two major genes (*BRCA1* and *BRCA2*), the rare Li–Fraumeni (*TP53*) syndrome, and possibly ataxia telangectasia (AT) heterozygotes all contribute to the familial risk. Segregation analysis can be particularly misleading when there are several genes with different modes of inheritance. This is illustrated by testicular cancer, in which segregation analysis predicted an autosomal recessive gene [19]. The recently discovered testicular cancer locus *TCG1* on the X-chromosome [20] accounts for many fraternal pairs, but none of the father–son association.

In addition to the highly penetrant, but rare, *APC* and HNPCC genes, segregation studies suggest the presence of more common genes with lower penetrance predisposing to colorectal tumours in the general population [21]. Cannon-Albright and colleagues [22] fitted a dominant model with a gene frequency of 0.19 and a lifetime penetrance of 0.4 for colonic adenoma or colorectal cancer in susceptible individuals. This suggested that most adenomas arise in the susceptible minority, although the confidence limits were wide.

Breast cancer incidence is slightly higher in sisters of breast cancer cases than in their mothers [23]. This could be due to shared lifestyle factors among sisters, although various genetic mechanisms could account for the difference. Few segregation analyses of familial cancer have incorporated data on environmental risk factors. Sellers and colleagues [24] fitted a model in which lung cancer risk is increased both by genetic susceptibility and by smoking. The observed pattern of familial aggregation was compatible with the inheritance of a co-dominant gene conferring an increased susceptibility to lung cancer.

3. Highly penetrant cancer syndromes

3.1. Linkage analysis

A large number of rare hereditary syndromes in which carriers have a characteristic phenotype and are at high risk of developing cancer have been recognised clinically for many years. McKusick lists several hundred inherited disorders for which neoplasia is a major feature or complication, but the great majority are rare [25], and in total they are unlikely to be responsible for more than 1% of all cancers. There are also several hereditary syndromes with no associated phenotype that produce multiple-case families with one or more common types of cancer. Genetic linkage analysis of multiple case cancer families has led to the localisation of various highly penetrant rare genes. Common cancers for which susceptibility genes have been identified include breast and ovarian cancers (*BRCA1* and *BRCA2* on chromosomes 17 and 13) [26,27], colon cancer with adenomatous polyposis coli (*APC* on chromosome 5) [28], HNPCC (the *MSH2*, *MLH1*, *PMS1* and *PMS2* mismatch repair genes on chromosomes 2, 3 and 7) [29,30], melanoma (*CDNK2A* on chromosome 9) [31,32], testicular cancer (*TCG1* on the X chromosome) [20] and the Li–Fraumeni multiple cancer syndrome (*TP53* on chromosome 17) [9].

3.2. Gene frequency and penetrance

Once a gene has been identified, its frequency and penetrance can be estimated reliably only from very large population-based studies of unaffected individuals as well as cancer patients. Such studies have not yet been done. Families used for linkage cannot be used to estimate gene frequency because they have been selected for the occurrence of multiple cases. Penetrance estimates based on such families may also be misleading. For example, penetrance estimates for mutant *BRCA1* and *BRCA2* based on risks to relatives of unselected breast cancer patients in whom these genes were sequenced [33] were substantially lower than those derived from multiple case families [34]. Both genetic and shared environmental factors may contribute to this clinically important difference. This analysis indicated that about one woman in 400 has a germline mutation in *BRCA1* or *BRCA2* [33], but higher prevalence and even lower penetrance estimates will probably be obtained when these genes are sequenced in large numbers of unaffected women.

4. Genes of low penetrance

With the possible exception of HNPCC, highly penetrant genes that cause remarkable multiple case families are too rare to account for a substantial proportion of common cancers. There may also be predisposing genes of lower penetrance that account for a larger proportion of cancers, but they will rarely produce large numbers of cancers in a family. The relative risk in first-degree relatives is of the order of 2 at older ages for most common

cancers. The range of dominant and recessive models consistent with relative risks between 1.5 and 5 in siblings is shown in Table 2 [35]. If the relative risk were 2, for example, a dominant gene carried by 1 in 50 of the population 14 times that in non-carriers and would cause a risk in carriers (i.e. a gene frequency of 0.01 in Table 2) about 10 times that in the general population. This would rarely produce striking multiple case families. Even for lung cancer a 10-fold increased risk would correspond to a penetrance by age 70 years of less than 40%, and for colon cancer a 10-fold increased risk in susceptibles would correspond to a penetrance of less than 15% by age 70 years. Such a gene might be detectable by linkage analysis of affected sibling pairs, haplotypes in linkage disequilibrium with founder mutations, or an associated phenotype. Direct sequencing of candidate genes in cases and controls is perhaps the most promising approach, although so far it has proved rather unproductive [36].

If such low-penetrance genes cause a substantial proportion of all cancers their identification would be of great practical importance. There may also be susceptibility genes which are carried by the majority of the population, but these could not cause a detectably increased risk in relatives, and they would have such low penetrance that their discovery would be of scientific rather than immediate clinical value. In the extreme case of the polymorphic *CYP2D6* locus, where more than 90% of the population are at high risk, even if poor metabolisers were totally immune from lung cancer the resulting relative risk in first-degree relatives would be only 1.02.

4.1. Genetic polymorphisms

The only successful mechanism for identifying low-penetrance genes has been the analysis of polymorphisms at candidate loci. There have been many studies comparing the prevalence in cancer patients and unaffected controls of common polymorphisms in genes involved in the metabolism of external or endogenous mutagens or in the production or processing of sex hormones or their analogues [36,37]. Common variants reported to act as low-penetrance genes in more than one study are listed in Table 3 [1]. A few polymorphisms in such genes seem to alter the risk substantially, such as the N-acetyltransferase 2 (*NAT-2*) slow acetylator phenotype, which increases the risk of bladder cancer [38], particularly in workers heavily exposed to certain aromatic amines. But systematic meta-analysis reveals little or no effect for most such polymorphisms, and the pooled data for the minority that are statistically significant usually suggest odds ratios of less than 2, and often much less [38–44]. Thus, for example, early reports suggested a more than doubled lung cancer risk associated with glutathione S-transferase μ1 (*GSTM1*) deficiency, but the pooled results of subsequent genotyping studies give an odds ratio of only 1.14 (95% confidence interval (CI): 1.03–1.25) [39].

Polymorphisms in oncogenes or tumour suppressor genes may also confer a moderately increased cancer risk. An example is the I1307K single nucleotide polymorphism (SNP) in the *APC* gene, which is carried by about 1 in 20 Ashkenazi Jews and almost doubles their

Table 3
Common genetic polymorphisms that may act as low-penetrance susceptibility loci (from Ref. [1])

Class-Locus	Cancer*	Putative mechanism
Metabolic polymorphisms		
CYP1A1	Lung, breast, colorectal, uterine, BCC	Altered metabolism (procarcinogen activation: polycyclic aromatic hydrocarbons)
CYP1A2	Bladder, colorectal	Altered metabolism (procarcinogen activation: nitrosamines and arylamines)
CYP2D6	Lung, liver	Altered metabolism (procarcinogen activation: nitrosamines)
GSTM1	Lung, bladder, breast, gastric, colon, head and neck, uterine	Altered metabolism (carcinogen detoxification: electrophilic compounds)
GSTT1	Colorectal, larynx, BCC, brain	Altered metabolism (carcinogen: electrophilic compounds)
NAT2	Bladder, colon, liver	Altered metabolism (carcinogen detoxification: aromatic amines, hydrazines)
Androgen receptor	Prostate	Altered metabolism (testosterone and dihydrotestosterone transactivation)
MTHFR	Colorectal, uterine	Methylation status
Tumour suppressor genes		
APC-I1307K	Colorectal	Hypermutability
DNA repair genes		
ATM	Breast	Genomic instability
Proto-oncogene polymorphisms		
H-*ras*-VNTR	Colorectal, breast, lung, bladder, leukaemia	Altered transcription/linkage disequilibrium

BCC, basal cell carcinoma.

colon cancer risk [45]. To estimate the individual effects of rare polymorphisms will require very large studies, but their average effect can be observed. The increased cancer risk associated with rare alleles of the *HRAS1*-associated minisatellite was among the first such associations to be reported. Such alleles, which are carried by approximately 5% of the population, increase the risk of several common cancers by a factor of 1.5–2 [46].

There have been various reports of statistically significant gene–environment interactions, such as a much larger lung cancer risk due to passive smoking in women who were *GSTM1*-deficient [47], or an increased breast cancer risk due to smoking in post-menopausal women that was confined to *NAT2* slow acetylators [48]. In these examples, however, the estimates of the risk in susceptibles (although not their lower confidence limits) were inconsistent with the much lower overall effect of passive smoking on lung cancer [49] or of smoking on breast cancer (which is nil) in larger studies [50]. Many apparently significant gene–gene or gene–exposure interactions will arise by chance, but some will be real. The effects of such polymorphisms in combination with each other and with environmental risk factors could be substantial, but their total contribution to cancer incidence will not be known until data on risk factors and extensive genotyping are available for very large numbers of patients and controls. Molecular epidemiology is increasingly focusing on genes that may modify endogenous carcinogenic processes rather than those that affect susceptibility to environmental carcinogens [51].

4.2. DNA replication and repair genes

Some classical inherited diseases such as Fanconi anaemia, AT, XP and Bloom's syndrome which are associated with increased cancer risk involve chromosome instability or DNA replication or repair defects. Other important cancer susceptibility genes are also of this type. A normal function of the *TP53* gene is either to induce cell cycle arrest or to induce apoptosis in response to DNA damage, and inactivation of the HNPCC genes on chromosomes 2, 3 and 7 is associated with somatic instability of di- and tri-nucleotide repeats. The *p16* cell cycle regulator gene (*CDKN2A*) on chromosome 9, originally located by linkage in familial melanoma, is deleted or mutated in a high proportion of cancers at many sites [32]. If *p16* inactivation leads to uncontrolled cell division, it is also likely to increase the risk of other mutations. The possibility that other 'mutator' genes of this general class might be detectable phenotypically is suggested by the observation that more than a third of breast cancer patients exhibit lymphocyte radiation sensitivity, while only approximately 10% of the general population do [52]. Formal modelling of radiation sensitivity in breast cancer families is compatible with a dominant mechanism [53] that causes a relative risk of 1.5–2 in first-degree relatives of breast cancer patients and approximately 3 in their monozygotic twins. These figures are similar to the risks seen in relatives of breast cancer patients above the age of 50 years (Fig. 1). This effect, if real, is not due to a single gene. AT heterozygotes exhibit high radiation sensitivity in the test, but they do not constitute 10% of the population.

4.3. Linkage analysis of low-penetrance genes

Highly penetrant genes do not cause a high proportion of most cancers, but the preceding example shows that less penetrant genes carried by up to 20% of the population (Table 2) could well do so. If such genes cause a high proportion of cancers of a particular type they may be detectable by linkage analysis of affected pairs of siblings, particularly if the mode of inheritance is recessive, as the examples of Hodgkin's disease [54] and nasopharyngeal carcinoma [55] have demonstrated. For low penetrance genes that cause a small proportion of cancers, the number of affected relative pairs required will be prohibitively large [56], but the sample size for a test based on allelic association can be very much smaller, even allowing for multiple comparisons [57].

The observation that linkage disequilibrium can extend up to 100 kilobases or more among populations of European origin [58] lends some support to the idea that susceptibility genes might be found by discovering ancient DNA sequences that are commoner in cancer patients than in controls. Such disequilibrium studies are now feasible using the dense genome-wide map of over a million SNPs together with the human genome sequence [58]. This approach could detect any founder mutation that causes a substantial fraction of cancers of a particular type. *BRCA2* might, for example, have been discoverable through the higher prevalence among breast cancer patients of the sequences flanking the 6174delT mutation among Ashkenazi Jews or the 999del5 mutation in Icelanders. Less penetrant mutations would be more difficult to detect, however, particularly in genetically heterogeneous populations in which founder mutations are rarer. A similar idea underlies admixture analysis [59]. Most international differences in cancer rates are due to environmental or lifestyle rather than genetic effects, but some, such as the higher prostate cancer risk among African Americans than among white Americans, could be due to racial differences in the allelic spectrum of a particular gene. If so, in a population of mixed descent the high-risk African alleles would remain within long DNA sequences of African origin for many generations, and these sequences would be commoner in cancer patients of mixed race than in their siblings.

Another way in which such genes might be detected is by linkage analysis in large families of an associated

phenotypic or molecular marker of susceptibility such as the lymphocyte sensitivity in breast cancer patients described above or the susceptibility to adenomas in colon cancer modelled by Cannon-Albright and colleagues [22]. Other established or suspected susceptibility markers include clinically detectable lesions such as naevi, benign fibrocystic breast disease or palmar keratoses, metabolic markers such as slow acetylation, and various markers of DNA repair efficiency or chromosomal instability detectable by cellular assay or DNA analysis [37].

5. Conclusions

Studies of multiple case families have led to the identification of highly penetrant genes predisposing to colon, breast, ovarian and testicular cancers and melanoma. Such genes account for most striking multiple-case families and a substantial proportion of cancers diagnosed below age 40 years, although they cause a much smaller proportion of older cases. The large risks observed in relatives of young patients suggest the existence of such genes predisposing to other common cancers. With the possible exception of ovarian cancer, however, most of the familial risk in first-degree relatives of older cancer patients is probably not due to highly penetrant genes. A major unanswered question in cancer genetics is whether a substantial proportion of all cancers arise in susceptible individuals as a result of genes of lower penetrance. If so, some of these genes may be detectable by the various linkage approaches discussed above. Novel candidate genes will continue to be discovered through research on growth control and DNA repair pathways, and somatic mutations in oncogenes and tumour suppressor genes, some of which can also be inherited, will be discovered by powerful new methods such as the detection of all sequence differences between tumour DNA and normal DNA from the same patient.

The ability to identify a large number of individuals with a lifetime risk for a particular cancer of the order of 20–50% would present a number of practical opportunities and ethical difficulties. Phenotypic or genetic tests might be offered routinely as an adjunct to screening for cancers such as breast, colon and prostate, extending the ethical difficulties of obtaining fully informed consent for genetic testing to the general population. There would also be important implications for industrial and environmental exposure to carcinogens. Susceptible individuals can perhaps be excluded from certain occupations or persuaded to give up smoking, but if identified individuals in the general population suffer much higher than average risks from carcinogenic agents such as ionising radiation it could be argued that much more stringent environmental limits must be introduced. The discovery of mutant genes such as those predisposing to HNPCC has already raised these ethical quandaries, but they have yet to be resolved.

The study of cancer genetics has given rise to a succession of statistical fallacies during the 67 years since Cramer [60] noted that the overall cancer rate varies much less than the rates for individual sites between different countries. The correct explanation is the statistical formula for adding variances, but Cramer inferred that a certain fixed proportion of all populations must be 'cancer prone'. The low relative risks seen in relatives of cancer patients were mistakenly believed to be inconsistent with large effects due to mendelian genes. Familial clusters of common cancers which could not possibly have arisen by chance were often ascribed to ascertainment bias by those sceptical of the importance of genetic susceptibility in carcinogenesis. (A family in which four sisters develop breast cancer before age 45 years would be expected to occur by chance less than once every 1000 years in the whole of Britain.) The probability that one of the trillions of cells in a smoker's bronchus will become fully cancerous after 60 years of smoking is approximately 1 in 4, yet the illogical argument that genetics must play a dominant role because heavy smokers do not all develop lung cancer remains as popular as ever.

The relationship of early multi-stage models to subsequent discoveries of molecular mechanisms in carcinogenesis is also widely misunderstood. The most cited example, the hypothesis that hereditary retinoblastoma is due to an inherited or germline mutation in a tumour suppressor gene, provided a biologically plausible explanation both of the apparently dominant pattern of inheritance and of the 100 000-fold higher risk in susceptibles, and has proved to be correct; but the persistent belief [36] that differences between the distributions of age at diagnosis of hereditary and sporadic cases constituted an important part of the evidence for an extra rate-limiting step in sporadic retinoblastoma is wrong [6].

Such statistical misconceptions are not merely footnotes in the history of cancer research. Statisticians were right when they predicted more than 20 years ago that the discovery of penetrant susceptibility genes would also elucidate some of the fundamental processes of spontaneous carcinogenesis [2,6,35]. There is now a consensus among statisticians, again largely unsupported by laboratory evidence, that a substantial proportion of human cancers arise in carriers of less penetrant susceptibility genes. Modelling the relationship between the kinetics of specific genetic and cellular processes and age-specific cancer incidence rates may play a role in the discovery of these genes, and will certainly be an important goal once they have been identified [61]. The predictions of even the simplest genetic models are often counter-intuitive, and cancer researchers should be aware of familial incidence patterns and

the range of biological mechanisms that are consistent with them if this ultimate synthesis of cancer epidemiology and molecular biology is to be achieved.

References

1. Houlston RS, Peto J. Genetics of common cancers. In Eeles RA, Ponder B, Easton DE, Horwich A, eds. *Inherited Predisposition to Cancer*. Chapman Hall, in press.
2. Ashley DJB. Colonic cancer arising in polyposis coli. *J Med Genet* 1969, **6**, 376–378.
3. DeMars R. Contribution to discussion. In *Genetic Concepts and Neoplasia*. Twenty-third Annual Symposium on Fundamental Cancer Research, 1969 (University of Texas M.D. Anderson Hospital). Williams and Wilkins, Baltimore 1970, 105–106.
4. Janin N. A simple model for carcinogenesis of colorectal cancers with microsatellite instability. *Adv Cancer Res* 2000, **77**, 189–221.
5. Peto R, Parish SE, Gray RG. There is no such thing as ageing, and cancer is not related to it. In Likhachev A, Anisimov V, Montesano R, eds. *Age-Related Factors in Carcinogenesis*. IARC Scientific Publications no. 58. Lyon, IARC, 1986.
6. Hethcote HW, Knudson AG. Model for the incidence of embryonal cancers: application to retinoblastoma. *Proc Natl Acad Sci* 1978, **75**, 2453–2457.
7. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 2000, **26**, 411–414.
8. Olschwang S. Germline mutation and genome instability. *Eur J Cancer Prev* 1999, **8**(Suppl. 1), S33–S37.
9. Varley JM, Evans DG, Birch JM. Li-Fraumeni syndrome—a molecular and clinical review. *Br J Cancer* 1997, **76**, 1–14.
10. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. Risks of cancer in BRCA1 mutation carriers. *Lancet* 1994, **343**, 692–695.
11. Breast cancer linkage consortium. Cancer risks in BRCA2 carriers. *J Natl Cancer Inst* 1999, **91**, 1310–1316.
12. Goldgar DE, Easton DE, Cannon-Albright LA, Skolnick MH. A systematic population based- assesment of cancer risk in first degree relatives of cancer probands. *Journal of the National Cancer Institute* 1994, **86**, 200–209.
13. Swerdlow AJ, Green A. Melanocytic naevi and melanoma: an epidemiological perspective. *British Journal of Dermatology* 1987, **117**, 137–146.
14. Easton DF, Cox GM, Macdonald AM, Ponder BAJ. Genetic susceptibity to naevi—a twin study. *British Journal of Cancer* 1991, **64**, 1164–1167.
15. Goldgar DE, Cannon-Albright LA, Meyer LJ, Peipkorn MW, Zone JJ, Skolnick MH. Inheritance of nevus number and size in melanoma/DNS kindreds. *Cytogenetic Cell Genetics* 1992, **59**, 200–202.
16. Cuzick J, Babiker A, de Stavola BL, McCance D, Cartwright R, Glashan RW. Palmar keratoses in family members with bladder cancer. *Journal of Clinical Epidemiology* 1990, **43**, 1421–1426.
17. Risch N. Linkage strategies for genetically complex multilocus triats, I: multilocus models. *American Journal of Human Genetics* 1990, **46**, 222–228.
18. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. *American Journal of Human Genetics* 1991, **48**, 232–241.
19. Heimdal K, Olsson H, Tretli S, Fossa SD, Borresen AL, Bishop DT. A segregation analysis of testicular cancer based on Norwegian and Swedish families. *Br J Cancer* 1997, **75**, 1084–1087.
20. Rapley EA, Crockford GP, Teare D, *et al*. Localization to Xq27 of a susceptibility gene for testicular germ-cell tumours. *Nat Genet* 2000, **24**, 197–200.
21. Houlston RS, Collins A, Slack J, Morton NE. Dominant genes for colorectal cancer are not rare. *Annals of Human Genetics* 1992, **56**, 99–103.
22. Cannon-Albright LA, Skolnick MH, Bishop DT, Lee RG, Burt RW. Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers. *New England Journal of Medicine* 1988, **319**, 533–537.
23. Peto J, Easton DF, Matthews FE, Ford D, Swerdlow AJ. Cancer mortality in relatives of women with breast cancer: the OPCS study. *Int J Cancer* 1996, **65**, 275–283.
24. Sellers TA, Bailey-Wilson JE, Elston RC, *et al*. Evidence for mendelian inheritance in the pathogenesis of lung cancer. *Journal of the National Cancer Insitute* 1990, **82**, 1272–1279.
25. McKusick VA. *Mendelian Inheritance in Man: Catalogues of Autosomal Dominant Inheritance, Autosomal Recessive Inheritance and X-linked Phenotypes*. Johns Hopkins, 1994.
26. Hall JM, Lee MK, Morrow J, *et al*. Linkage analysis of early onset familial breast cancer to chromosome 17q21. *Science* 1990, **250**, 1684–1689.
27. Wooster R, Neuhausen SL, Mangion J, *et al*. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 1994, **265**, 2088–2090.
28. Bodmer W. Familial adenomatous polyposis (FAP) and its gene, APC. *Cytogenet Cell Genet* 1999, **88**, 99–104.
29. Lindbolm A, Tannergaard P, Werelius B, Nordenskjoid B. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nature Genetics* 1993, **5**, 279–282.
30. Peltomaki P, Aaltonen LA, Sistonen P, *et al*. Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 1993, **260**, 810–812.
31. Cannon-Albright LA, Goldgar DE, *et al*. Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. *Science* 1992, **258**, 1148–1152.
32. Piepkorn M. Melanoma genetics: an update with focus on the CDKN2A(p16)/ARF tumor suppressors. *J Am Acad Dermatol* 2000, **42**, 705–722.
33. Peto J, Collins N, Barfoot R, *et al*. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst* 1999, **91**, 943–949.
34. Ford D, Easton DF, Stratton M, *et al*. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* 1998, **62**, 676–689.
35. Peto J. Genetic predisposition to cancer. In Cairns J, Lyon JL, Skolnick M, eds. *Banbury Report No. 4: Cancer Incidence in Defined Populations*. Cold Spring Harbor, Cold Spring Harbor Laboratory, l980.
36. Ponder BAJ. Cancer genetics. *Nature* 2001, **411**, 336–341.
37. Houlston RS, Tomlinson IP. Detecting low penetrance genes in cancer: the way ahead. *J Med Genet* 2000, **37**, 161–167.
38. Hein DW, Doll MA, Fretland AJ, *et al*. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol Biomarkers Prev* 2000, **9**, 29–42.
39. Houlston RS. Glutathione S-Transferase M1 (GSTM1) status and lung cancer risk: a meta-analysis. *Cancer Epidemiology, Biomarkers and Prevention* 1999, **8**, 675–682.
40. Johns LE, Houlston RS. Glutathione S-Transferase M1 (GSTM1) status and bladder cancer risk: a meta-analysis. *Mutagenesis* 2000, **15**, 399–404.
41. Dunning AM, *et al*. A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 1999, **8**, 843–854.
42. Johns LE, Houlston RS. NAT2 status and bladder cancer risk: a meta-analysis. *Environmental and Molecular Mutagenesis* 2000, **36**, 221–227.
43. Houlston RS, Tomlinson IPM. Polymorphisms and colorectal tumor risk. *Gastroenterology* 2001, **121**, 282–301.
44. Houlston RS. CYP1A1 polymorphisms and lung cancer risk: a meta-analysis. *Pharmacogenetics* 1999, **9**, 221–224.

45. Woodage T, King SM, Wacholder S, *et al*. The APCI1307K allele and cancer risk in a community-based study of Ashkenazi Jews. *Nature Genet* 1998, **20**, 62–65.
46. Krontiris TG, Devlin B, Karp DD, Robert NJ, Risch N. An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N Engl J Med* 1993, **329**, 517–523.
47. Bennett WP, Alavanja MC, Blomete B, *et al*. Environmental tobacco smoke, genetic susceptibility, and risk of lung cancer in never-smoking women. *J Natl Cancer Inst* 1999, **91**, 2009–2014.
48. Ambrosone CB, Freudenheim JL, Graham S, *et al*. Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms, and breast cancer risk. *JAMA* 1996, **276**, 1494–1501.
49. Weinberg CR, Sandler DP. Gene-by-environment interaction for passive smoking and glutathione S- transferase M1? *J Natl Cancer Inst* 1999, **91**, 1985–1986.
50. IARC. *Tobacco Smoking*. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 38. Lyon, France, IARC, 1986.
51. Perera FP, Weinstein IB. Molecular epidemiology: recent advances and future directions. *Carcinogenesis* 2000, **21**, 517–524.
52. Scott D, Barber JB, Levine EL, Burrill W, Roberts SA. Radiation-induced micronucleus induction in lymphocytes identifies a high frequency of radiosensitive cases among breast cancer patients: a test for predisposition? *Br J Cancer* 1998, **77**, 614–620.
53. Roberts SA, Spreadborough AR, Bulman B, Barber JB, Evans DG, Scott D. Heritability of cellular radiosensitivity: a marker of low-penetrance predisposition genes in breast cancer? *Am J Hum Genet* 1999, **65**, 784–794.
54. Hors J, Dausset J. HLA and susceptibility to Hodgkin's disease. *Immunol Rev* 1983, **70**, 167–192.
55. Lu S, Day NE, Degos L, *et al*. Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature* 1990, **346**, 470–471.
56. Risch N, Merikangas K. The future of genetic studies of complex human disease. *Science* 1996, **273**, 1516–1517.
57. Camp NJ. Genomewide transmission/disequilibrium testing-consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* 1997, **61**, 1424–1430.
58. Reich DE, Cargill M, Bolk S, *et al*. Linkage disequilibrium in the human genome. *Nature* 2001, **411**, 199–204.
59. McKeigue PM. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet* 1998, **63**, 241–251.
60. Cramer W. The prevention of cancer. *Lancet* 1934, **1**, 1–5.
61. Peto J. Cancer epidemiology in the last century and the next decade. *Nature* 2001, **411**, 390–395.