

Review article

Detecting low penetrance genes in cancer: the way ahead

R S Houlston, I P M Tomlinson

Abstract

The search for the genes responsible for many complex genetic diseases is well under way and has already been successful in some cases. The study of cancer as a complex genetic disease has lagged behind other conditions, largely because of particular problems that are associated with malignant disease. Cancer also, however, presents specific opportunities for gene identification, which are not found in many other diseases. While the methods of genetic mapping and gene cloning used for other complex diseases will be applied to cancer, these must almost certainly be complemented by other methods, such as the study of somatic mutations, cancer associated phenotypes, and modifier genes for Mendelian cancers. Here, we review the strategies available for identifying cancer predisposition genes of low and moderate penetrance.

(J Med Genet 2000;37:161-167)

Keywords: low penetrance gene; cancer

Large scale genetic studies which aim to identify the moderate and low penetrance loci involved in many genetic but non-Mendelian diseases appear now almost to be a commonplace. Notable success in this field of research includes the recognition of a relationship between the *ApoE* genotype and Alzheimer's disease risk¹ and in the field of infectious diseases where variation in *TNF* and HLA have been shown to be associated with substantially different risks of TB and malaria.²⁻⁴ This is not to say that all these studies are near completion and in some cases, such as multiple sclerosis, localisation of predisposition genes (let alone gene identification) is proving very difficult.⁵⁻⁶ While it is easy to underestimate the difficulties inherent in studying diseases like diabetes, asthma, and rheumatoid arthritis, and indeed to underestimate the remaining problems, it is nevertheless surprising that the study of cancer has lagged behind that of other complex genetic diseases. Almost all cancer susceptibility alleles identified so far are rare and highly penetrant (for example, *APC*, *BRCA1*, *BRCA2*, *MSH2*, *LMH1*, *PTEN*, *CDNK2A*).⁷ They may cause a substantial proportion of cancers at young ages, but they are unlikely to be respon-

sible for a high proportion of all cancers, leaving a considerable potential contribution from less penetrant genes. It is possible to gain an insight into the potential impact of such genes on cancer incidence given that the relative risk of disease in first degree relatives is only of the order of 1.5-2.5 over all ages for most common cancers.⁷ Under a dominant low penetrance model, the ratio of cancer risk in susceptibles to that in the general population cannot exceed about 5, if 50% of all cancers occur in susceptibles, when about 10% of the population must be at increased risk. A dominant gene carried by 2% of the population will cause a risk 11 times that in the general population and will cause 22% of all cancers. Within this range, such genes will rarely produce striking multiple case families, except possibly for breast cancer, where an 11-fold increase in risk would correspond to a penetrance of 43% by the age of 70. In the case of colon cancer, for example, where the cumulative risk in the general population is only 1.2% by the age of 70, an 11-fold increase in risk in susceptibles would correspond to a penetrance of only 13%.

Why is cancer different from other common diseases?

Cancer has very few strictly unique features compared with other complex genetic diseases, but it has a combination of features that make it particularly problematical to study. First, there are very few cancers with non-Mendelian inheritance, but which have large sib relative risks (unless stratified by age). Second, common cancers (such as those of the colon, breast, or bronchus) are usually late onset and the parents of patients (even those who present "early") are often dead. Third, many cancers are fatal, making samples difficult to obtain retrospectively and even contemporaneously. Fourth, not only does cancer require particular combinations of genes and environment (like other diseases), but it also has a truly random component in that several somatic mutations must occur for a carcinoma or sarcoma to develop. Fifth, the challenge of discovering genes mutated somatically in cancer, but with no germline effects, has been a worthy distraction for the geneticist from the task of studying cancer as a complex genetic disease.

Nevertheless, despite the above problems and relatively low overall relative risks, the prospects for identifying low penetrance genes

Section of Cancer Genetics, Institute of Cancer Research, Cotswold Road, Sutton, Surrey SM2 5NG, UK
R S Houlston

Molecular and Population Genetics Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
I P M Tomlinson

Correspondence to:
Dr Houlston,
r.houlston@icr.ac.uk
or Dr Tomlinson,
i.tomlinson@icrf.icnet.uk

for cancer are far from bleak. The wide age range of presentation of most common cancers means that selection for early age of onset may be a more powerful way of enriching for disease with a genetic component in cancer than in many other diseases and for reducing environmental influences. In addition, the fact that some cancers progress in a stepwise fashion means that the more common benign precursor lesions can be studied instead of or in addition to the cancer itself. Moreover, as will be shown below, complementary strategies can be used to identify low penetrance cancer genes.

Types of low penetrance cancer predisposition genes

All Mendelian cancer predisposition genes appear to act in a cell autonomous fashion. Indeed, such a mechanism is more or less implicit in the action of tumour suppressor genes such as *APC*, *RB1*, and *TP53*. It is probable that non-Mendelian cancer predisposition loci will include many genes with cell autonomous effects, but will also include some genes with “global” effects (for example, carcinogen metabolism polymorphisms, behavioural differences reflected in different diets or tendencies to smoke, and anti-tumour immune response) and other genes with effects on the local tumour environment (for example, influencing stromal-epithelial interactions or production of paracrine hormones). Some potential low or moderate penetrance cancer genes have been characterised, but the only successful mechanism for identifying these genes has been the analysis of candidate loci. The most common experimental design has been the case-control study, comparing allele frequencies in cancer patients with those in healthy controls. A number of putative low penetrance genes have been described, conferring susceptibility to cancer through a variety of mechanisms. Table 1 provides a summary of loci reported to act as low penetrance genes, most

on the basis of more than one study. Of those which have known or probable modes of action, some, such as those involved in the metabolism of carcinogens, probably increase cancer risk by raising the “global” mutation rate (although the overall effect may appear to be site specific) while others such as the I1307K variant of *APC* may increase the mutation rate in a cell autonomous fashion.¹⁷

Recently, it has been proposed that a significant part of the non-Mendelian contribution to cancer might be derived from missense variants (or a restricted set of protein truncating variants) at classical tumour suppressor loci. This suggestion was prompted by the discovery of the I1307K variant at the *APC* locus, which confers about a two-fold increased risk of colorectal cancer in the Ashkenazi population.¹⁷⁻¹⁹ It has also been proposed that the E1317Q *APC* variant is associated with colorectal adenomas and hence with colon cancer.¹⁸ Individually, each of these types of variant would be at “sub-polymorphic” levels, contributing little to cancer risk and difficult to detect by usual gene mapping methods. Together, however, several rare variants at any locus such as *APC* could contribute significantly to cancer risk, although there is as yet little evidence to support this theory.

Linkage versus association analysis to detect low penetrance genes

The detection of high penetrance genes is generally through linkage studies using multiple case families. Occasional high penetrance cancer genes may remain to be identified. Moderate penetrance genes for cancer, if such genes exist for any particular tumour type, can also be mapped by linkage studies, although non-parametric methods are most likely to be of use given uncertainties regarding genetic heterogeneity and the certain occurrence of sporadic cases within families with a genetic predisposition. The optimal linkage design for detection of a

Table 1 Genes reported to act as low penetrance susceptibility loci

Class/locus	Cancer*	Putative mechanism	Reference
Metabolic polymorphisms			
CYP1A1	Lung, breast, colorectal, uterine, BCC	Altered metabolism (procarcinogen activation: polycyclic aromatic hydrocarbons)	8 9
CYP1A2	Bladder, colorectal	Altered metabolism (procarcinogen activation: nitrosamines and arylamines)	9
CYP2D6	Lung, liver	Altered metabolism (procarcinogen activation: nitrosamines)	8 9 10
GSTM1	Lung, bladder, breast, gastric, colon, head and neck, uterine	Altered metabolism (carcinogen detoxification: electrophilic compounds)	8 9 11
GSTT1	Colorectal, larynx, BCC, brain	Altered metabolism (carcinogen: electrophilic compounds)	9
NAT2	Bladder, colon, liver	Altered metabolism (carcinogen detoxification: aromatic amines, hydrazines)	8 9
Androgen receptor	Prostate	Altered metabolism (testosterone and dihydrotestosterone transactivation)	12 13
<i>MTHFR</i>	Colorectal , uterine	Methylation status	14 15 16
Tumour suppressor genes			
<i>APC</i> -I1307K	Colorectal	Hypermutability	17 18 19
DNA repair genes			
<i>ATM</i>	Breast	Genomic instability	20 21 22
Proto-oncogene polymorphisms			
<i>H-ras</i> -VNTR	Colorectal, breast, lung, bladder, leukaemia	Altered transcription/linkage disequilibrium	23

*Strongest associations in bold.

common low penetrance gene is not as simple as for a rare high penetrance gene. Families comprising three or more cases are likely to be more powerful for linkage than affected sib pairs or other two case families. For example, assuming a dominant model and gene frequency of 0.1 and risk ratio of 6, 473 affected sib pairs, 205 three affected sibs, 145 four affected sibs, 545 cousin pairs, or 550 avuncular pairs would be required to give a lod of 3.3. Increasing the number of affected subjects in a family does not necessarily represent an efficient strategy, since a large proportion of these large multiple case families are likely to be caused by highly penetrant genes or by chance. Furthermore, if the gene is common there is the possibility that a large family will be segregating two copies of the mutation. This leads to parents being homozygous at the disease locus with the consequence of the family being uninformative for linkage. While the composition of families used for linkage analysis will always depend on what can practicably be collected, affected sib trios probably provide the most efficient strategy for detecting low penetrance genes over a wide range of models.²⁴

A concern, given the low relative risks associated with most common cancers, is that few moderate penetrance cancer predisposition genes exist. Low penetrance genes, characterised by small genotypic risks or less (that is, <4) will only confer a sib relative risk of 1.7 or less and will therefore rarely give rise to multiple case families. It is difficult or impossible to identify such genes by linkage analysis, because the number of affected relative pairs required will be prohibitively large.²⁵ In contrast, the required sample size for a test based on allelic association can be vastly smaller, even allowing for multiple comparisons.²⁵⁻²⁶ For example, to detect by linkage a dominantly acting gene conferring a four-fold increase in risk with a frequency of 0.1 would require 1055 affected sib pairs. However, with certain caveats discussed below, only 200 cases and 200 controls (or 200 parent offspring trios using the transmission disequilibrium test) would be required if an association strategy was adopted, even if a high degree of significance was stipulated to allow for 100 000 comparisons (that is, 5×10^{-8}).

At present, association studies are limited to the evaluation of candidate gene loci and there is every expectation that such studies will continue and become more common. However, there is increasing interest in the possibility of systematic genome wide association studies. These are becoming increasingly feasible with the advent of high density marker maps of single nucleotide polymorphisms (SNPs) coupled with new methods of genotyping.²⁷ Furthermore, the availability of the complete human genome sequence opens up the possibility of conducting genome wide allelic association studies based on analysis of intragenic polymorphisms.

An essential issue in a genome wide association study is marker density since the power of any association test will decline rapidly as linkage disequilibrium diminishes, or if there are differences in their relative allele frequency.²⁸ A recent analysis of this problem

based on simulations suggested that a useful level of linkage disequilibrium is unlikely to extend beyond an average distance of roughly 3 kb in the general population.²⁹ This would mean that around half a million SNPs would be required for whole genome studies. Furthermore, the extent of linkage disequilibrium is similar in isolated populations unless the founding bottleneck is very narrow or the frequency of the variant is low (<5%).²⁹ The outlook for genome wide association studies from these findings appears grim. It is, however, not clear that the underlying premises on which these simulations were carried out truly reflect the real world and the outlook is likely not to be so negative. As pointed out by Risch and Merikangas,³⁰ the expectation with respect to linkage disequilibrium across the genome is unknown and studies of the *ApoE* gene and late onset Alzheimer's disease, and the insulin VNTR region and diabetes, show that significant linkage disequilibrium exists well outside these regions.

The issue of false positive results is of great concern in association studies. Their frequency can be reduced by imposing low p values (for example, 10^{-6} to 10^{-8}) or by replicating findings in independent samples, or ideally by both. The simplest and most efficient form of association study is the case-control approach based on a comparison of allele frequencies in unrelated cases and controls. A major problem inherent in this approach, however, is that spurious associations can arise from population stratification. Hidden stratification can be tested for. Provided that cases and controls are well matched, differences in the frequency of genotypes will only be seen at predisposition loci. Given that spurious associations result in a departure of genotypes from Hardy-Weinberg equilibrium, stratification can be detected by typing a series of unlinked markers chosen from a panel known to exhibit differences in allele frequency between populations.³¹ The appropriate Bonferroni correction would, of course, be required to assess the statistical significance of any putative association.

One method of circumventing the problem of occult population stratification is to use family based controls. The most common approach is the transmission disequilibrium test (TDT), which assesses the evidence for preferential transmission of one allele over the other from heterozygous parents.³² An attractive feature of the TDT is that it is a test of linkage and not merely of linkage disequilibrium, since only linkage disequilibrium can distort the distribution of marker genotypes among parents of affecteds. A problem, however, is that most cancers develop in later life and it may rarely be possible to determine parental genotypes directly. To obviate the requirement for parental genotypes, allied, although less powerful, statistics based on the use of sib genotypes have been devised.³³⁻³⁵

Given that the human genome project will lead to the identification of polymorphisms for all genes and the introduction of reliable, high density oligonucleotide arrays (or alternative methods) to detect the identified polymor-

phisms will allow genome wide searches to be conducted rapidly, the detection of low penetrance susceptibility genes will then be only restricted by the availability of large cohorts of well characterised patients and healthy controls. For a direct association study, based on the use of 100 000 SNPs, using a sample size of 1000 plus cases and controls would be sufficient to detect loci conferring fairly small genotypic risks (~ 1.01 - 3.00) under multiplicative and additive models. This number of cases would be similarly sufficient under a common recessive model provided that p (predisposition allele frequency) > 0.5 and under a common dominant model provided $p < 0.5$.²⁶

Thus there is, in summary, scope for identifying low or moderate penetrance cancer predisposition genes using association studies in particular, possibly selected for early onset or disease of specific histological type. Both of these approaches of subclassifying cancers are likely to prove useful in order to enrich for genetic homogeneity. It is, however, clear that the chances of successfully identifying new cancer genes with incomplete penetrance are uncertain, whether using association or linkage analysis. Cancer provides novel opportunities for gene identification using indirect methods and we believe that these should be exploited wherever possible.

Indirect approaches for identifying low penetrance cancer predisposition genes

As for any other genetic disease, cancer is amenable to specific techniques for enhancing the power of studies to detect genetic effects, for example, by studying genetically young or isolated populations, and selecting early onset or severe cases. There are, however, several techniques for detecting low penetrance predisposition genes, which have special application to cancer.

SOMATIC MUTATIONS

Although the proportion of diseases which results from somatic mutation may be under-recognised, cancer provides a near unique opportunity for identifying low penetrance predisposing genes by studying somatic mutations. There are precedents for carrying out such studies, which derive from Mendelian cancer syndromes. The *PTEN* gene causing Cowden's syndrome and the *DPC4/SMAD4* gene causing juvenile polyposis were both originally identified by mapping somatic homozygous deletions in cancers.^{36,37} In Mendelian cancer syndromes, the justification for extending results from somatic genetics to germline predispositions is that germline and somatic mutations are, in theory, functionally equivalent in a cell; one patient may inherit a variant and another may acquire it by somatic mutation, just as Knudson³⁸ predicted. For the identification of low or moderate penetrance genes, the use of somatic genetics is suitable for genes with cell autonomous action. An example comes from the study of Paget's disease, in which linkage analysis and the study of somatic mutations in osteosarcoma have independently identified a disease locus on chromosome

18q.^{39,40} Somatic and germline studies will be complementary; fine mapping information, for example, might come from allelic association in the germline or from minimal regions of deletion in the soma.

Identifying somatic mutations is not, of course, a straightforward task, although samples of most tumours are relatively easy to collect. There are also some doubts that the sort of variation present and selected in the soma (generally loss of tumour suppressor genes and oncogene amplification) would have the relatively subtle effects typical of a low penetrance gene; high penetrance, or cell lethality, or simply lack of suitable germline variation are arguably more likely. Nevertheless, mechanisms such as somatic hypermutability (as proposed for I1307K-*APC*) provide sufficient justification for identifying predisposition genes by studying somatic mutations; the existence of alternative and complex genetic pathways may also mean that a tumour following one pathway may still gain an advantage from a variant in another pathway even if that variant does not have such a profound effect as the somatic mutations which usually occur at that locus.

Many and varied methods of somatic mutation mapping and identification exist. One of the most interesting is allele specific loss of heterozygosity (LOH), which is suited to studying patients with multiple tumours, but without Mendelian family histories. This is an extension of the principle of differential LOH patterns in multiple tumours first advanced in studies of *MEN1*.⁴¹ The approach stems from the concept of using loss of heterozygosity information in tumours to increase the power of linkage since loss of the wild type allele occurs in cases of tumours caused by mutations in tumour suppressor recessive oncogenes.⁴² To date there have been few examples of the utility of this approach, although linkage of isolated hyperparathyroidism to *MEN1* has been reported.⁴³ The concept can clearly be extended to mapping a novel gene. Examples of multiple tumours, which might be amenable to analysis using this strategy, include colorectal adenomas, naevi, and solar keratoses. If the tumours have a genetic cause owing to a mutation in a tumour suppressor gene, every lesion from the same patient should show loss of the "wild type" allele. In this way, double somatic events can be distinguished from single events associated with the site of a predisposition allele. Using this strategy it is theoretically possible to identify a novel adenoma predisposition gene by the analysis of five adenomas from 200 subjects, given a rate of allelic loss in excess of 50%. This would apply even if the frequency of the deleterious gene and its associated risk were to vary widely (for example, 0.001-0.1 and 5-20 respectively). The utility of this approach is, however, dependent upon a low background rate of somatic events, a condition most likely to be fulfilled in early tumours.

ASSOCIATED NON-TUMOUR PHENOTYPES

Some inherited disease phenotypes (or, sometimes, normal variants) are associated with an

increased risk of tumours, without tumours being an integral part of the disease process (table 2). These non-tumour phenotypes are classically non-neoplastic, although there is some evidence that this classification is incorrect in some cases. Many of the associated phenotypes have high sib relative risks and may be relatively genetically homogeneous, so that these diseases are much more amenable to genetic study than the cancer itself. The genes involved in predisposition to the disease phenotype may either contribute directly to inherited cancer risk, or may harbour a set of alleles which increases cancer risk without leading to the associated disease phenotype. The usual methods of studying complex genetic diseases can be used to analyse families with these tumour associated phenotypes. In almost all cases, the phenotype is relatively common, usually more so than the associated cancer. In most cases, genetic studies of the diseases shown are in progress.

ASSOCIATED TUMOURS

Some inherited tumours are associated with an increased risk of tumours of a different site or type from the cancer of primary interest, without any good evidence that one tumour type can progress to the other. Like the associated non-tumour phenotypes, the associated tumours may have relatively high sib relative risks and genetic homogeneity, making them an attractive means of identifying genes for the cancer of interest. Examples include multiple cutaneous leiomyomas, which are associated with uterine leiomyomas and leiomyosarcomas,⁶⁵ melanocytic naevi which are associated with melanomas,⁵⁹ and palmar keratoses which are associated with bladder cancer.^{46 47} Such lesions are theoretically amenable to study by a combination of linkage analysis and somatic mutation mapping, although no study of this type has yet been completed.

MODIFIER GENES

Modifier genes for the severity of a Mendelian disease must be distinguished from minor susceptibility loci for that disease, on the grounds that the modifiers have no direct effect on disease susceptibility. In cancer, however, there is a requirement for somatic mutations and there is frequently an overlap, in theory and in practice, between the spectrum of germline and somatic mutations in a single tumour type. Thus, for example, a gene which modifies the number of adenomas in the Mendelian disease

familial adenomatous polyposis (FAP) has a good chance of also influencing the probability that a person develops a sporadic colorectal adenoma and thus carcinoma (by reason that it acts in a similar fashion on germline and somatic *APC* mutations). Such genes may therefore act as low penetrance susceptibility loci for cancer, as long as they harbour suitable genetic variation. Mendelian cancer modifier genes are more likely to be QTLs, but semi-quantitative or qualitative variation (for example, in progression of benign lesions to malignancy or the propensity to metastasise) may also exist; in attenuated FAP, for example, the modifier may have such a strong effect that no detectable tumour develops.

One strategy for identifying modifiers of Mendelian cancer syndromes is to use discordant sib pairs, both of whom are affected by the disease (or, at least, carry the disease causing mutation), but one of whom has mild disease and the other has severe disease. Alternatively, disease severity may be treated as a continuous variable. The best Mendelian cancer syndromes to study are those associated with multiple lesions, in order to minimise the effects of somatic mutations which occur randomly in time. It is not essential to exclude disease variation caused by genotype-phenotype correlations provided that studies are performed within families, but it is clear that the clinical data on which these studies are based must be collected in a uniform manner and that, in cases in which the phenotype may be influenced by exogenous factors, information on all possible confounding covariates is collected.

MOUSE MODELS

The increasingly complete synteny between the map of mouse and human genomes opens up the possibility of using mouse models for the identification of low penetrance cancer susceptibility genes. A complete discussion of such possibilities is beyond the scope of this article. Nevertheless, the ease with which mouse models can be used to investigate complex traits will undoubtedly lead to an increase in the use of animal models as a route into the human situation,⁶⁴ thus reversing the current trend to make mouse models of human disease. Although attractive, the use of animal models is not problem free. Specifically, the limited genetic variation in animal systems and the very different environments and reproductive strategies of mice and humans represent intrinsic problems, as evidenced by attempts to

Table 2 Diseases and phenotypes associated with cancer susceptibility

Disease or phenotype	Population prevalence (%)	Sib risk (%)	Cancer risk	Reference
Cryptorchidism	1	6	Testicular (6-fold)	44 45
Palmar keratoses	23	—	Bladder (12-fold), lung (4-fold)	46 47
Supernumerary nipples	1	—	Urogenital (8-fold)	48
Paget's disease of bone	2	12	Osteosarcoma (100-fold)	49 50
Multinodular goitre	1	4	Thyroid (4-fold, 20% of cases), breast (4-fold)	51
Coeliac disease	0.3	10	Lymphoproliferative disease (30 to 40-fold), oesophagus (8 to 12-fold), Mouth and pharynx (10-fold)	52 53 54
Inflammatory bowel disease	0.3	8	Colorectal (6-fold)	55 56
Endometriosis	10	70	Breast (1.3-fold), ovary (1.9-fold)	57 58
Naevi	NA	h ² = 84	Melanoma	59 60
Radiosensitivity	9	18	Breast (9-fold), lung (14-fold)	61 62

NA: not applicable; h²: heritability.

identify a human modifier gene for FAP using the Min mouse. In addition, the use of animal models to identify cancer genes is beset by the very different life span of humans and small mammals and hence the time frame within which a tumour must grow. Furthermore, one of the most important facets of tumourigenesis, namely organ or site specificity, can be difficult to explore using animal models, as seen in the Min mouse model of FAP⁶⁵⁻⁶⁷ and mouse models of HNPCC.^{65 68 69} Attempts to counter these problems in animal systems, for example by the use of extrinsic carcinogens as a strategy to accelerate tumourigenesis, may succeed, but also have potential to create additional problems.

RESPONSE TO TREATMENT

Intervention in cancer provides another potential way of identifying moderate or low penetrance predisposition genes. Genes influencing the response to therapy may also influence susceptibility. Naturally, the main role of these studies will be to modify therapeutic intervention and candidate susceptibility genes will emerge as a by product.

The fine mapping problem: a multidisciplinary approach

There is already a considerable body of publications and discussion of gene identification in complex diseases. Most conclude that good fortune, such as a prime candidate gene within a region of significant linkage, is the best recipe for success. In the absence of such good fortune, initial mapping screens for moderate or low penetrance cancer susceptibility genes may use any of the above techniques (although the indirect approaches must subsequently be confirmed using linkage or association studies or both). Subsequent steps must be taken to identify the gene(s) involved, and experience has shown that this can be a very troublesome task. Even when the human genome is sequenced, it may still be more efficient to proceed to fine mapping of these genes, rather than go directly to testing candidates within the minimal region of interest, both in terms of time, and as a means of overcoming the problem, resulting from linkage disequilibrium, of identifying the relevant functional variation within a genetic interval. There is considerable overlap between each of the indirect approaches for identifying cancer genes which we have detailed above. Thus, after initial mapping screens, if fine mapping is considered worthwhile, several of these methods can be used in parallel. It may even be the case that cancer has advantages over other complex diseases in this regard; minimal regions of allele loss do not, for example, exist in asthma (as far as is known). All the methods outlined above may be useful in specific circumstances, but the most useful are likely to include somatic mutation mapping, modifier identification, and the study of tumour associated phenotypes. Additional methods, such as *in vitro* studies of tumour suppressor genes or of gene expression in tumours may be used as supplementary techniques.

Establishing causality is clearly a major issue with respect to association studies. Conclusions about the relationship between a specific gene variant and cancer risk should be based upon the guidelines developed by Hill⁷⁰: (1) the strength of the association, weak associations being more likely to be attributable to bias or confounding; (2) reproducibility of the findings, based on the analysis of different cohorts and study designs; (3) biological plausibility and functional analyses; and (4) animal models, the ease with which "knock outs" can be made making suitable animal models more frequently available.

Conclusions

The identification of common, moderate, or low penetrance genes for cancer is potentially of great benefit, because it allows screening to be targeted to those at greatest risk. Cancer is therefore different from diseases such as type I diabetes or rheumatoid arthritis in which gene identification will provide benefits which are less direct. It is therefore ironic that cancer has lagged behind other diseases in the identification of non-Mendelian genetic predisposition. The strategies most commonly used for identifying the genes for other complex diseases can, in theory, be used for cancer, but some of these methods will be extremely difficult to use in practice, because of problems specific to cancer which are detailed above. Affected sib pairs who are alive with gastric or pancreatic cancer are extremely rare, for example, and few patients with ovarian cancer have living parents to provide optimal power for analysis using methods like the transmission disequilibrium test. It is likely to be necessary, therefore, to use supplementary, indirect methods for the identification of non-Mendelian cancer predisposition genes. The biology of cancer provides opportunities, through, for example, the study of somatic mutations, tumour associated phenotypes, and modifier genes which may not exist in other complex diseases. Most of these rely on the multistage mutation model of carcinogenesis, which provides both problems and opportunities for the identification of moderate and low penetrance cancer genes. Finally, one of the differences between cancer and many other multifactorial diseases is that we do actually know quite a lot of the genes that are associated with a high risk of cancer, and this clears the field for the clearly more difficult area of low and moderate penetrance genetics.

We are grateful to two anonymous reviewers for their comments.

- 1 Farrer LA, Cupples LA, Haines JL, *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* 1997;278:1349-56.
- 2 Knight JC, Udalova I, Hill AV, *et al.* A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet* 1999;22:145-50.
- 3 Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. *Proc Natl Acad Sci USA* 1997;94:3159-95.
- 4 Gilbert SC, Plebanski M, Gupta S, *et al.* Association of malaria parasite population structure, HLA, and immunological antagonism. *Science* 1998;279:1173-7.
- 5 Hogancamp WE, Rodriguez M, Weinschenker BG. Identification of multiple sclerosis-associated genes. *Mayo Clin Proc* 1997;72:965-76

- 6 Chataway J, Feakes R, Coraddu F, *et al*. The genetics of multiple sclerosis: principles, background and updated results of the United Kingdom systematic genome screen. *Brain* 1998;121:1869-87.
- 7 Houlston RS, Peto J. Genetics of common cancers. In: Eccles RA, Ponder B, Easton DE, Horwich A, eds. *Inherited predisposition to cancer*. Chapter 13. London: Chapman and Hall, 1996: 208-26.
- 8 d'Errico A, Taioli E, Chen X, Vineis P. Genetic metabolic polymorphisms and the risk of cancer: a review of the literature. *Biomarkers* 1996;1:149-73.
- 9 Vineis P, d'Errico A, Malats N, Boffetta P. Overall evaluation and research perspectives. In: Ryder W, ed. *Metabolic polymorphisms and susceptibility to cancer*. Chapter 25. IARC scientific publications No 148. Lyon: International Agency for Research on Cancer, 1999:403-8.
- 10 Rostamani-Hodjegan A, Lennard MS, Woods HF, Tucker GT. Meta-analysis of studies of the CYP2D6 polymorphism in relation to lung cancer and Parkinson's disease. *Pharmacogenetics* 1998;8:227-38.
- 11 McWilliams JE, Sanderson BJ, Harris EL, Richert-Boe KE, Henner WD. Glutathione S-transferase M1 (GSTM1) deficiency and lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 1995;4:589-94.
- 12 Stanford JL, Just JJ, Gibbs M, *et al*. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Res* 1997;57:1194-8.
- 13 Giovannucci E, Stampfer MJ, Krithivas K, *et al*. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc Natl Acad Sci USA* 1997;94:3320-23.
- 14 Ma J, Stampfer MJ, Giovannucci E, *et al*. Methylentetrahydrofolate reductase polymorphism, dietary interactions, and risk of colorectal cancer. *Cancer Res* 1997;57:1098-102.
- 15 Chen J, Giovannucci E, Kelsey K, *et al*. A methylentetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer Res* 1996;56:4862-4.
- 16 Esteller M, Garcia A, Martinez-Palones JM, Xercavins J, Reventos J. Germ line polymorphisms in cytochrome-P450 1A1 (C4887 CYP1A1) and methylentetrahydrofolate reductase (MTHFR) genes and endometrial cancer susceptibility. *Carcinogenesis* 1997;18:2307-11.
- 17 Laken SJ, Petersen GM, Gruber SB, *et al*. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet* 1997;17:79-83.
- 18 Frayling IM, Beck NE, Ilyas M, *et al*. The APC variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history. *Proc Natl Acad Sci USA* 1998;95:10722-7.
- 19 Woodage T, King SM, Wacholder S, *et al*. The APC1307K allele and cancer risk in a community-based study of Ashkenazi Jews. *Nat Genet* 1998;20:62-5.
- 20 Swift M, Morrell D, Massey RB, Chase CL. Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N Engl J Med* 1991;325:1831-6.
- 21 Easton DF. Cancer risks in A-T heterozygotes. *Int J Radiat Biol* 1994;66 (suppl 6):S177-82.
- 22 Athma P, Rappaport R, Swift M. Molecular genotyping shows that ataxia-telangiectasia heterozygotes are predisposed to breast cancer. *Cancer Genet Cytogenet* 1996;92:130-4.
- 23 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;329:517-23.
- 24 Goldgar DE, Easton DF. Optimal strategies for mapping complex diseases in the presence of multiple loci. *Am J Hum Genet* 1997;60:1222-32.
- 25 Risch N, Merikangas K. The future of genetic studies of complex human disease. *Science* 1996;273:1516-17.
- 26 Camp NJ. Genomewide transmission/disequilibrium testing-consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* 1997;61:1424-30.
- 27 Wang DG, Fan JB, Siao CJ, *et al*. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998;280:1077-82.
- 28 Muller-Myhsok B, Abel L. Genetic analysis of complex diseases. *Science* 1997;275:1328-9.
- 29 Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999;22:139-44.
- 30 Risch N, Merikangas K. Genetic analysis of complex diseases. *Science* 1997;275:1329-30.
- 31 Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;65:220-8.
- 32 Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;55:506-16.
- 33 Horvath S, Laird NM. A discordant-sibship test for disequilibrium and linkage: no need for parental data. *Am J Hum Genet* 1998;63:1886-97.
- 34 Knapp M. The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test. *Am J Hum Genet* 1999;64:861-70.
- 35 Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet* 1998;62:450-8.
- 36 Hahn SA, Schutte M, Hoque AT, *et al*. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* 1996;273:350-3.
- 37 Li J, Yen C, Liaw D, *et al*. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* 1997;275:1943-7.
- 38 Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971;68:820-3.
- 39 Cody JD, Singer FR, Roodman GD, *et al*. Genetic linkage of Paget disease of the bone to chromosome 18q. *Am J Hum Genet* 1997;61:1117-22.
- 40 Nellisery MJ, Padalecki SS, Brkanac Z, *et al*. Evidence for a novel osteosarcoma tumor-suppressor gene in the chromosome 18 region genetically linked with Paget disease of bone. *Am J Hum Genet* 1998;63:817-24.
- 41 Lubensky IA, Debelenko LV, Zhuang Z, *et al*. Allelic deletions on chromosome 11q13 in multiple tumors from individual MEN1 patients. *Cancer Res* 1996;56:5272-8.
- 42 Teare MD, Rohde K, Santibanez Koref MF. The use of loss of constitutional heterozygosity data to ascertain the location of predisposing genes in cancer families. *J Med Genet* 1994;31:448-52.
- 43 Teh BT, Esapa CT, Houlston R, *et al*. A family with isolated hyperparathyroidism segregating a missense MEN1 mutation and showing loss of the wild-type alleles in the parathyroid tumors. *Am J Hum Genet* 1998;63:1544-9.
- 44 Chilvers C, Pike MC. Cancer risk in the undescended testicle. *Eur Board Urol* 1992;1:74-9.
- 45 Czeizel A, Erosi E, Toth J. Genetics of undescended testis. *J Urol* 1981;126:528-9.
- 46 Cuzick J, Harris R, Mortimer PS. Palmar keratoses and cancers of the bladder and lung. *Lancet* 1984;i:530-3.
- 47 Cuzick J, Babiker A, De Stavola BL, McCance D, Cartwright R, Glashan RW. Palmar keratoses in family members of individuals with bladder cancer. *J Clin Epidemiol* 1990;43:1421-6.
- 48 Mehes K, Szule E, Torzsok F, Meggyessy V. Supernumerary nipples and urologic malignancies. *Cancer Genet Cytogenet* 1987;24:185-8.
- 49 Wick MR, Siegal GP, Unni KK, McLeod RA, Greditzer HG. Sarcomas of bone complicating osteitis deformans (Paget's disease): fifty years' experience. *Am J Surg Pathol* 1981;5:47-59.
- 50 Siris ES, Ottman R, Flaster E, Kelsey JL. Familial aggregation of Paget's disease of bone. *J Bone Miner Res* 1991;6:495-500.
- 51 Ron E, Kleinerman RA, Boice JD Jr, LiVolsi VA, Flannery JT, Fraumeni JF Jr. A population-based case-control study of thyroid cancer. *J Natl Cancer Inst* 1987;79:1-12.
- 52 Logan RF, Rifkin EA, Turner ID, Ferguson A. Mortality in celiac disease. *Gastroenterology* 1989;97:265-71.
- 53 Holmes GK, Prior P, Lane MR, Pope D, Allan RN. Malignancy in coeliac disease-effect of a gluten free diet. *Gut* 1989;30:333-8.
- 54 Houlston RS, Ford D. Genetics of coeliac disease. *Q J Med* 1996;89:737-43.
- 55 Kornfeld D, Ekbom A, Ihre T. Is there an excess risk for colorectal cancer in patients with ulcerative colitis and concomitant primary sclerosing cholangitis? A population based study. *Gut* 1997;41:522-5.
- 56 Satsangi J, Parkes M, Jewell DP, Bell JI. Genetics of inflammatory bowel disease. *Clin Sci* 1994;94:473-8.
- 57 Moen MH, Magnus P. The familial risk of endometriosis. *Acta Obstet Gynecol Scand* 1993;72:560-4.
- 58 Brinton LA, Gridley G, Persson I, Baron J, Bergqvist A. Cancer risk after a hospital discharge diagnosis of endometriosis. *Am J Obstet Gynecol* 1997;176:572-9.
- 59 MacKie RM. Incidence, risk factors and prevention of melanoma. *Eur J Cancer* 1998;34(suppl 3):S3-6.
- 60 Easton DF, Cox GM, Macdonald AM, Ponder BA. Genetic susceptibility to naevi - a twin study. *Br J Cancer* 1991;64:1164-7.
- 61 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;77:614-20.
- 62 Wu X, Zhao Y, Kemp BL, Amos CI, Siciliano MJ, Spitz MR. Chromosome 5 aberrations and genetic predisposition to lung cancer. *Int J Cancer* 1998;79:490-3.
- 63 Garcia Muret MP, Pujol RM, Alomar A, Calaf J, de Moragas JM. Familial leiomyomatosis cutis et uteri (Reed's syndrome). *Arch Dermatol Res* 1988;280(suppl):S29-32.
- 64 Silver LM. *Mouse genetics*. Chapter 7. Oxford: Oxford University Press, 1995:133-58.
- 65 Su LK, Kinzler KW, Vogelstein B, *et al*. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* 1992;256:668-70.
- 66 Heyer J, Yang K, Lipkin M, Edelman W, Kucherlapati R. Mouse models for colorectal cancer. *Oncogene* 1999;18:5325-33.
- 67 Taketo MM. Apc gene knockout mice as a model for familial adenomatous polyposis. *Prog Exp Tumor Res* 1999;35:109-19.
- 68 Prolla TA, Baker SM, Harris AC, *et al*. Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. *Nat Genet* 1998;18:276-9.
- 69 Andrew SE, McKinnon M, Cheng BS, *et al*. Tissues of MSH2-deficient mice demonstrate hypermutability on exposure to a DNA methylating agent. *Proc Natl Acad Sci USA* 1998;95:1126-30.
- 70 Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.