

GENETIC PREDISPOSITION TO CANCER — INSIGHTS FROM POPULATION GENETICS

Steven A. Frank

Abstract | Individuals differ in their inherited tendency to develop cancer. Major single-gene defects that cause early cancer onset have been known for many years from their inheritance patterns, and inherited defects that have weaker effects on predisposition were also suspected to exist. Recent progress in cancer genetics has identified specific loci that are involved in cancer progression, many of which have key roles in DNA repair, cell-cycle control and cell-death pathways. Those loci, which are often mutated somatically during cancer progression, sometimes also contain inherited mutations. Recent genetic studies and quantitative population-genetic analyses provide a framework for understanding the frequency of inherited mutations and the consequences of these mutations for increased predisposition to cancer.

PENETRANCE

The frequency with which individuals that carry a given gene will show the manifestations associated with the gene. If a disease allele is 100% penetrant then all individuals carrying that allele will express the associated disorder.

POPULATION BOTTLENECK

Severe reduction in population size, during which previously rare genetic variants can become common by chance.

Department of Ecology
and Evolutionary Biology,
University of California,
Irvine, California
92697-2525, USA.
e-mail: safrank@uci.edu
doi:10.1038/nrg1450

Genetic factors affect the tendency to develop cancer. Predisposing mutations often influence DNA repair, cell-cycle regulation and cell-death pathways¹. Here, I address two broad questions. What processes determine whether predisposing mutations remain rare or become common within populations? And how do the biochemical consequences of particular mutations that affect DNA repair and cell-cycle pathways influence the accumulation of mutations for particular traits that predispose to cancer?

I begin by reviewing what is known about genetic predisposition to cancer. The key facts fall into four categories. First, major (that is, highly **PENETRANT**) mutations that show Mendelian inheritance cause strong genetic predisposition to nearly every type of cancer. Natural selection keeps these mutations rare because they cause severe disease early in life. Second, these major Mendelian mutations underlie only a small proportion of the genetic tendency to develop cancer, at least in the late-onset epithelial cancers such as breast and colon cancer. Polygenic inheritance — which involves many allelic variants, each of which has a small effect — seems to dominate genetic predisposition for the late-onset epithelial cancers. Third, conclusions about polygenic inheritance come mainly from statistical

studies of differences in the risk of developing cancer between twins, family members and unrelated individuals. The direct study of predisposing allelic variation is only just beginning, and the most promising line of investigation concerns variants that affect DNA repair. Fourth, families that show a high level of polygenic predisposition to breast cancer have a high level of constant risk of developing cancer later in life, which does not increase with age. By contrast, most individuals have a relatively low risk that does increase with age.

I devote the second part of this review to three promising topics for future study. First, polygenic variation could arise from many rare allelic variants or fewer common allelic variants. This raises issues concerning the age of variant alleles, which is measured as the time since the original mutation arose, and the forces — such as natural selection, **POPULATION BOTTLENECKS** and demographic expansions — that determine the frequencies of the variants. Second, the constant high level of risk that is seen in families that show strong predisposition to breast cancer provides clues about cancer progression. For example, if cancer arises as a result of progression through a series of stages^{2,3}, genetically predisposed families might progress rapidly through the early stages. Later in life, their risk of cancer is the constant risk of

passing the final stage in progression. Other individuals progress more slowly through the early stages. With more than one stage remaining later in life, their risk will increase with age. The final point I will discuss concerns the fact that DNA repair, cell-cycle regulation and apoptosis combine in a robust manner to regulate the balance between cell proliferation and cell death, thereby protecting against cancer. Because natural selection cannot purge mutations that are mostly hidden by robust pathways, mutations will continue to accumulate until their consequences become sufficiently deleterious that they are balanced by natural selection.

Genetic predisposition

Major genes that show Mendelian inheritance. Mutations with strong effects cause the early onset of cancer and the development of multiple tumours, are inherited in a simple Mendelian fashion and aggregate in families⁴. For example, most women who carry a highly penetrant mutation in *BRCA1* develop **breast or ovarian cancer**^{5,6}. Both copies of *BRCA1* are inactivated in tumours, indicating that this gene is a tumour suppressor and that the mutations involved are recessive. However, an individual needs to carry only one mutated copy to be at risk; the cancerous phenotype arises only after somatic mutation 'knocks out' the second copy in a small fraction of cells. So, although they are physiologically recessive, *BRCA1* mutations are inherited as dominant alleles.

In fact, most major cancer-associated mutations behave as physiologically recessive tumour-suppressor genes that are inherited dominantly⁴. So far, only three oncogenes (*RET*, *MET* and *CDK4*) have been identified among 31 cancer loci with Mendelian inheritance patterns that affect 20 different types of cancer⁴. Perhaps most mutations in physiologically dominant oncogenes are not viable because they would be expressed in all cells.

There are few direct estimates of the frequencies of predisposing cancer alleles that show Mendelian inheritance. Direct estimates count the numbers of mutated and unmutated alleles in the population by screening for molecular differences in genes between individuals. Some rough estimates have been made for the frequency of cases, from which crude calculations of allele frequencies and mutation rates can be made. For example, inherited cases of **retinoblastoma**, **Wilms' tumour** and skin cancer in **xeroderma pigmentosum** all occur at frequencies of $\sim 10^{-5}$ – 10^{-4} (REF 1). These diseases are inherited as dominant mutations, and most individuals who carry a highly penetrant mutation develop the disease during childhood or early life. Without treatment, carriers do not usually reproduce.

How can information from observed cases be used to calculate allele frequency? Suppose a mutation is expressed in all carriers, and those carriers die before they have reproduced. In this situation, each case must arise from a new mutation, and the frequency of mutated alleles is roughly equivalent to the probability of a new mutation arising ($q = u$, where q is the frequency of the mutant allele in the population and u is the mutation rate per generation).

Obtaining good estimates of mutation rates by direct methods is not easy. The commonly quoted values tend to be in the range of 10^{-6} – 10^{-5} per gene per generation⁷ — an order of magnitude lower than the frequency of cases. For this type of approximate calculation, a match within an order of magnitude suggests that we have roughly the right idea about the factors that influence allele frequencies.

Several other factors might influence the frequencies of mutations and observed cases. For example, some loci might be more mutable than others, some carriers might reproduce, a new mutation in the parental germline might be transmitted to several offspring, and the expression of some mutations (their penetrance) might depend on genetic background. In addition, the early onset of an apparently inherited form of cancer might be caused by somatic mutations early in development that occur in most cells⁸. High-throughput sequencing and other methods of analysing genetic polymorphisms will eventually provide more information about these possibilities. Until then, the rough correlation between case frequency and mutation rate provides a guide to the impact of dominantly inherited mutations that cause severe, early-onset disease.

Of course, some inherited mutations have low penetrance or cause later-onset disease. Natural selection removes a mutation from the population in proportion to both the probability that it causes disease and to the reduction in reproductive success of those individuals who express the disease. If the probability of expression in a carrier is the penetrance, p , the reduction in reproductive success is r , and q is the frequency of mutations, then qp is the frequency of cases, and the rate at which mutations are removed in each generation is qpr — the frequency of cases multiplied by the reduction in reproductive success in each case. Equilibrium occurs when mutations that are lost are matched by the influx of new mutations at rate u , so $qpr = u$ at equilibrium.

How does this theory apply to particular types of cancer? Dominantly inherited mutations of the *APC* gene cause the **colon cancer** syndrome familial adenomatous polyposis (FAP)⁹. Nearly all carriers develop cancer, with a median age of onset of about 40 years. The frequency of cases, qp , is of the order of 10^{-4} . We do not have historical data on the reduction in reproductive success that occurs in the absence of treatment. A reasonable value to use is $r \sim 10^{-1}$, which takes into account the fact that the age of reproduction in the past was probably somewhat lower than in modern societies. In this case, qpr is $\sim 10^{-5}$, which is again fairly close to the standard estimate for the mutation rate.

Mutations in the DNA mismatch repair (MMR) system lead to hereditary nonpolyposis colon cancer (HNPCC)¹⁰. Mutations in several MMR genes cause an increase in the somatic mutation rate, and more frequent somatic mutations lead to a high probability of early-onset cancer. The median age of diagnosis for HNPCC is about 42 years¹¹. The frequency of cases is at least of the order of 10^{-3} , but may be more because HNPCC can be difficult to distinguish from colon cancers that arise in the absence of MMR defects. Setting

the level of reproductive loss at $r = 10^{-1}$, the rate of removal of MMR mutations, qpr , is 10^{-4} or higher, which indicates a high mutation rate. Although mutations that increase the risk of developing HNPCC have been identified in five MMR loci so far¹⁰, mutations that influence HNPCC probably also occur in other MMR genes. There are 22 genes in the core MMR pathway¹². The effective mutation rate is nu , where n is the number of MMR loci and u is the mutation rate per locus. Using a value of $n \sim 10^1$, we again obtain a mutation rate per locus of $\sim 10^{-5}$.

These calculations show that we roughly understand the factors that influence the frequencies of cases and of mutant alleles. When the numbers do not match, another process must have a significant effect. *BRCA1* provides one example.

Mutations in *BRCA1*, which has an important function in the repair of double-strand DNA breaks, confer a high probability of developing breast or ovarian cancer⁶. Current estimates for the penetrance of breast cancer in carriers of *BRCA1* mutations range from 56% to 86%⁶. Lack of functional *BRCA1* leads to chromosomal abnormalities⁵ — a common feature of cancer cells. The median age of onset is ~ 50 years¹³, which is later than for most of the other cancers that show dominant Mendelian inheritance. The frequency of *BRCA1* mutant alleles and associated cases varies in different populations over the range 10^{-3} – 10^{-2} (REFS 6,14–16). To the best of my knowledge, there are no data that measure the decrease in reproduction in carriers of *BRCA1* mutations — a reasonable guess would be in the range 10^{-2} – 10^{-1} . These values give an estimate for $qpr = u$ of 10^{-5} – 10^{-3} , which is somewhat higher than the consistent calculation of $\sim 10^{-5}$ for other loci. In the case of *BRCA1*, this difference might be because of the high density of repetitive DNA elements in the gene⁵, which might cause a higher frequency of mutation.

I have used only the simplest concepts of mutation and selection from the theory of population genetics in the examples above. Many other forces can influence allele frequencies, such as recent expansions of the human population and subdivision into distinct ethnic groups. A rich mathematical theory provided by population genetics helps to study the diverse forces that shape allele-frequency patterns^{17,18}. However, because many of the current cancer-related data do not justify complex quantitative analyses, I have emphasized how the principles can be used in simple ways to think about easily observed patterns.

Other forms of genetic predisposition. High-penetrance mutations that show Mendelian inheritance are easy to identify, and most of the loci that are associated with a high level of risk might already have been identified. By contrast, it is more difficult to identify alleles with lower penetrance or later onset and to assess their relative importance.

In the absence of direct knowledge about many genes that predispose to cancer, statistical studies have analysed how environmental and genetic variation contribute to differences in cancer risk. For example, reflecting

environmental effects, immigrants take on the risk of colon cancer that is specific for their new home¹⁹. The risk of developing colon cancer for an individual in a specific geographical region is strongly associated with levels of meat consumption²⁰, so changes in diet might explain the altered risk of immigrants. Smoking^{21,22} and long-term exposure to certain carcinogens²³ also cause significant environmental risk.

To determine the genetic component of risk, statistical studies compare the frequencies of cancer occurrence between monozygotic twins, dizygotic twins, other family members and unrelated individuals²⁴. In principle, such studies could separate the relative contributions of shared genes, shared environment in the family, and differences in environment between unrelated individuals. However, the statistical power of such studies tends to be low, with wide CONFIDENCE INTERVALS for the relative roles of genes and environment. This problem is particularly severe for the rarer cancers because of low sample sizes in such studies.

A large study from the Swedish Family-Cancer database provided narrower confidence intervals for the proportions of cancer variance that are explained by genes and environment²⁵. The estimates for genetic contribution ranged from 1% to 53%, depending on the type of cancer. These may be lower limits, because certain types of genetic variation could not be separated from the effects of a shared environment. Confounding components include similar genotypes between parents, which would be classed as a shared environmental effect rather than a genetic effect. In this study, Mendelian loci explain only part of the total genetic contribution to cancer risk, indicating a significant role for polygenic variation.

An interesting analysis of the Anglian Breast Cancer Study Group study took a different approach to genetic predisposition²⁶. The authors first removed the two known Mendelian loci associated with breast cancer — *BRCA1* and *BRCA2* — from the analysis, and then fitted the remaining risk distribution to a polygenic model in which the small risks per variant allele are multiplied across loci. According to the fitted model, the 20% of the population that has the highest level of genetic predisposition has a 40-fold greater risk than the 20% of the population with the lowest level of predisposition. The model also predicted that more than 50% of breast cancers occur in the 12% of the population with the greatest predisposition. The known Mendelian loci account for only a small proportion of the total genetic risk, with the remainder being explained by polygenic variation.

It is difficult to tell how reliable these conclusions about polygenic inheritance are. Other models could be fitted to the same data, with different contributions of Mendelian loci, polygenic loci and environment. I favour the strong emphasis on polygenic inheritance, because most complex QUANTITATIVE TRAITS in nature show extensive polygenic variation^{27–29}. However, statistical models are hard to test directly, because it is difficult to obtain evidence that strongly supports one model and rules out other plausible models. One is often left with conclusions that are based as much on prior belief as on data.

CONFIDENCE INTERVAL
Range of values estimated from data that are likely to contain the true value of a parameter.

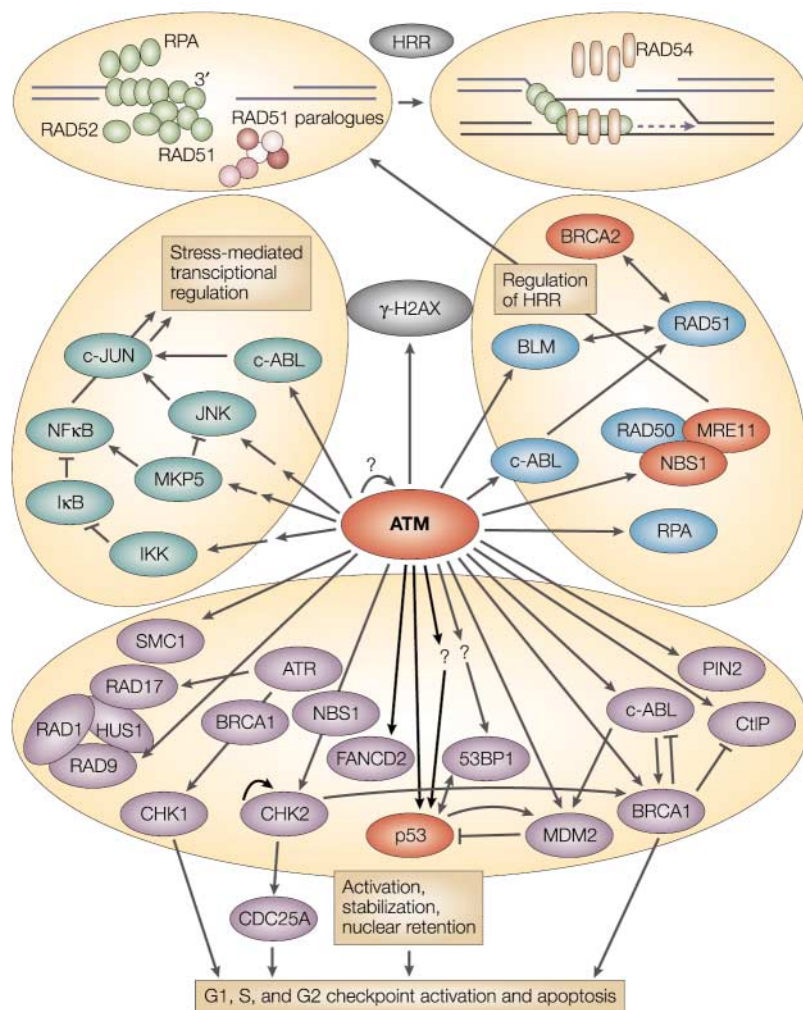


Figure 1 | Pathways involved in the response to DNA double-strand breaks. Double-strand breaks in DNA induce transcriptional changes, cell-cycle checkpoints and DNA-repair processes. Loss of fidelity in repairing double-strand breaks (DSBs) leads to chromosomal rearrangements and genomic instability, which are common attributes of cancer cells. Two of the five main DNA-repair pathways have key roles in responding to DSBs: the nonhomologous end-joining pathway and the homologous recombinational repair (HRR) pathway. Only some aspects of the HRR machinery are shown, in particular, those mediated by the ATM kinase, which is the master regulator in a signalling network that responds to DSBs. Note that the breast-cancer-related gene *BRCA1* is in the G2 checkpoint pathway, and another breast-cancer related gene, *BRCA2*, has a role in HRR regulation. Inherited mutations in *ATM* cause a rare disorder, ataxia telangiectasia, which involves progressive neuronal degeneration and cancer predisposition. At the cellular level, mutations in *ATM* lead to chromosome breakage, telomere instability, radiosensitivity, defective cell-cycle checkpoints, defective apoptosis and a reduced p53 response (reviewed in REF. 35). 53BP1, p53-binding protein 1; ATR, ATM- and RAD3-related; BLM, Bloom's syndrome protein; CHK1, checkpoint protein 1; CHK2, checkpoint protein 2; CtlP, CTBP-interacting protein; FANCD2, Fanconi anaemia, complementation group D2; IκB, inhibitor of κB; IKK, IκB kinase; JNK, c-JUN amino-terminal kinase; MKP5, MAP kinase phosphatase 5; NBS1, Nijmegen breakage syndrome 1; NFκB, nuclear factor-κB; RPA, replication protein A; SMC1, structural maintenance of chromosomes 1. Modified with permission from REF. 35 © (2002) Elsevier Science.

Direct study of polygenic inheritance. Ideally, one would like to know how particular genetic variants affect the biochemistry of cells, and how those biochemical effects influence progression to cancer. Although we are still a long way from this ideal, recent studies of DNA-repair genes provide hints about what could be learned¹².

Individuals vary in the ability of their cells to repair DNA damage³⁰. A relatively low repair efficiency is associated with a higher risk of cancer. Presumably, the association arises because higher rates of unrepaired somatic mutations and chromosomal aberrations contribute to faster progression to cancer. Most studies of repair capacity measure the effects of mutagens on DNA damage in lymphocytes. For example, a mutagen can be applied to cultures of lymphocytes; after a period of time, damage can be measured by the numbers of unrepaired single-strand or double-strand breaks, or by incorporation of a radioisotope. To study the role of DNA repair in cancer, measurements compare individuals with and without cancer. Berwick and Vineis³⁰ summarized 64 different studies that used a variety of methods to quantify repair. In these studies, a relatively low repair capacity was consistently associated with an ~2–10-fold increase in cancer risk.

Roughly speaking, repair efficiency has an inheritance pattern that is typical of a quantitative trait. A few rare Mendelian disorders cause severe deficiencies in repair capacity. Apart from these rare cases, repair capacity shows a continuous pattern of variation and has a significant heritable component^{31–33}. Measures of variability and heritability are statistical descriptions of the genetics of repair. Recent studies have made the first steps towards understanding the mechanistic relations between genetic variants and altered phenotypes.

Many genes in the five key repair pathways for different types of DNA damage^{12,34,35} are known, so genetic variants can be identified by sequencing the loci involved (FIG. 1). Specific variants can also be constructed, and their physiological consequences tested in cell-based assay systems. Mohrenweiser *et al.*¹² list 22 genes in the core pathway of the MMR system. This system primarily corrects mismatches and short insertion or deletion loops that arise during replication or recombination³⁶. The MMR system increases the accuracy of replication by a factor of 100–1000.

Eighty-five different variants have been found in seventeen different MMR genes that were screened in at least fifty unrelated individuals¹². Of these variants, 38% occurred at a frequency of 2% or more; 21% occurred at a frequency of 5% or more; and 12% occurred at a frequency of 20% or more. Similar results were obtained for the other DNA-repair pathways, as summarized by Mohrenweiser *et al.*¹² In 74 repair genes from various pathways, the average frequency of the wild-type allele is ~80%, with the remaining 20% being comprised of different allelic variants. Among the 148 alleles per person at the 74 repair loci, the average number of allelic variants is expected to be ~30. Presumably, each individual carries a very rare or unique genotype.

So, small variations in DNA repair are highly heritable, DNA repair efficiency is correlated with cancer risk and there are widespread amino-acid polymorphisms in the known repair genes. The next step will be to link these polymorphisms to variations in the biochemistry of repair, providing a mechanistic understanding of how genetic variation influences an important aspect of cancer predisposition³⁷.

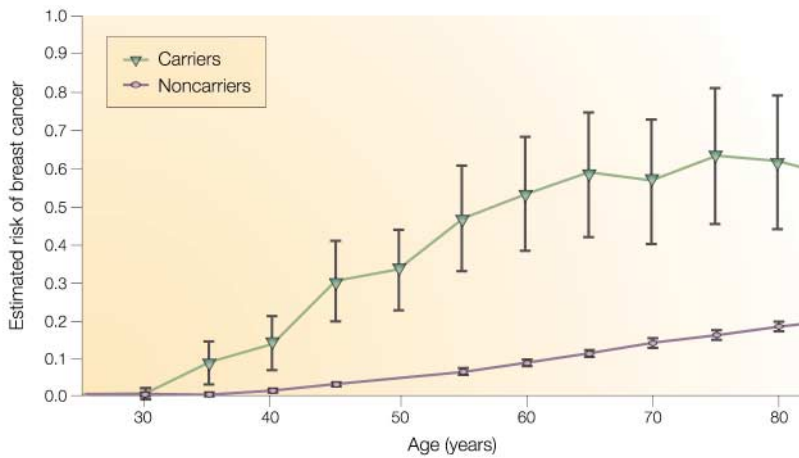


Figure 2 | Risk of breast cancer caused by *BRCA1* and *BRCA2* mutations. Risks for carriers versus non-carriers are shown. The curves show the estimated cumulative lifetime risk by a particular age, with 95% confidence intervals. Although cumulative risk does not decrease with age, the estimation method allows declines, which explains the decrease in cumulative risk for carriers in the 65–70 and 75–80 age intervals¹⁵. The data represented are from 5,318 Ashkenazi Jews. All individuals were genotyped for two specific mutations in each of *BRCA1* and *BRCA2*, yielding 120 mutants, which is 2.3% of the sample. There was no significant difference in risk between carriers of *BRCA1* and *BRCA2* mutations. Modified with permission from REF. 15 © (1997) Massachusetts Medical Society.

Age-specific incidence in predisposed individuals. The polymorphisms that occur in genes involved in DNA-repair hint at variations in cellular physiology that may be very common. The connection between DNA-repair efficiency and cancer seems plausible, because somatic mutations and chromosomal aberrations probably have a key role in cancer progression. However, at present, we cannot make a simple mechanistic connection between repair efficacy and the rate of progression to cancer, because the steps in cancer progression remain poorly understood. The most likely scenario is that each physiological risk factor affects the probability of cancer occurring at a particular age.

Women who carry *BRCA1* mutations develop cancer more frequently and at an earlier age than non-carriers^{6,15,16}. Such large differences in age-specific incidence probably occur only for the major Mendelian mutations. *BRCA1* mutations are unusually common for major Mendelian defects, so it is possible to collect enough data to draw an age-specific risk curve for carriers (FIG. 2). However, most Mendelian mutations are rare, so it is more difficult to obtain sufficient data on age-specific incidence. In addition, the role of genetic background cannot be assessed easily in studies of human cancer.

Mouse models with controlled genotypes provide a good way to evaluate the age-specific consequences of particular allelic variants³⁸. For example, mice show an accelerated onset of cancer when they carry mutations in nucleotide excision-repair genes^{39–41}. Such data provide a starting point for understanding the relations between genetic variants, physiological processes within cells and progression to cancer. However, because of the limited sample sizes that can be studied in the laboratory, experiments using mice do not usually provide

sufficient statistical power to measure the consequences of allelic variants that have small effects.

Peto and Mack⁴² presented the most provocative finding about the age-specific incidence of cancer. Women who are at high risk of developing breast cancer show an approximately constant incidence of cancer per year after a certain age, whereas in most individuals incidence rises significantly with age. This pattern appears in three different classes of susceptible individuals after the age at which a particular patient develops cancer. First, an individual with monolateral breast cancer has an annual risk of developing cancer in the contralateral breast of ~0.7% per year. Second, a monozygotic twin of a patient with breast cancer has an approximate risk of 1.3% per year, which is again ~0.7% per breast. Third, mothers and sisters of patients have a risk of ~0.3–0.4% per year. These patterns of age-specific risk indicate a strong role for genetic predisposition. Below, I discuss some hypotheses concerning cancer progression that could account for how genetic variation affects age-specific risk.

Promising directions for future research

Age and frequency of variant alleles. Dominant mutations that cause childhood cancers will be rapidly purged from the population. In lethal childhood diseases, such as untreated retinoblastoma, the ‘lifespan’ of mutant alleles must be relatively short. In other words, for a variant allele that causes severe childhood disease, the time to origin of the mutation must be short. By contrast, natural selection purges mutations that have a lower penetrance or confer a later age of onset more slowly. These weaker mutations can increase in frequency. Such weak alleles are often relatively old — for a weak variant allele, the time to origin of the mutation is generally longer than that of a highly penetrant, early-onset allele.

What are the expected ages and frequencies of variant alleles? To analyse this problem, it is necessary to make assumptions about mutation rates, selection against variant alleles, stochastic fluctuations in small populations and the demographic history of specific populations¹⁸. Demographic history matters, because bottlenecks in the past might have caused large, stochastic fluctuations in allele frequency. Recent population expansion causes most variant alleles to be relatively young, because most of the total births in the history of an expanding population will have been relatively recent. These theoretical considerations allow one to predict whether variants will be frequent or rare, young or old.

Given the observed frequencies of variant alleles and the estimates of their ages, what inferences can be made about mutation rate, selection, stochastic fluctuations and demographic history? The answer to this question helps to understand both why certain populations might have a higher level of genetic predisposition to cancer and the nature of the genetic predisposition in terms of the frequency, penetrance and age of onset of variant alleles.

Methods for estimating genetic risk in particular individuals depend on the frequency of variant alleles^{43,44}. If most genetic risk comes from a few relatively

QUANTITATIVE TRAIT
A measurable trait that depends on the cumulative action of many genes and that can vary among individuals over a given range to produce a continuous distribution of phenotypes. Common examples include height, weight and blood pressure.

common alleles that are relatively old, then those alleles will be associated with other polymorphisms in the genome that can be used as markers of risk. These associations arise because the original mutations will, by chance, occur in regions in which other SNPs are located nearby.

By contrast, most genetic risk might come from many rare, 'young' alleles. If so, then there will be no consistent association between known SNPs and genetic predisposition. Each particular mutation will have its own profile of linked marker polymorphisms, but those profiles will differ for each mutation. Because there may be many mutations, with each making only a small contribution to genetic risk, no overall association will occur between known marker polymorphisms and total genetic risk.

The data that are available at present do not allow a definitive distinction to be made between a few common, older variants and many rare, younger variants. What are perhaps the best data come once again from

the DNA-repair genes. Mohrenweiser *et al.*¹² partitioned the variance in allele frequency for particular mutations across 74 loci (FIG. 3). The fraction of the genetic variance is about the same for rare, intermediate and common alleles. To understand what this means, consider how to calculate genetic variance in allele frequencies for these data.

The contribution of a variant allele with frequency p_i to the variance at its locus is $p_i(1-p_i)$. For example, consider a case in which the allele frequencies at a locus are 0.01, 0.1 and 0.89. The rare allele contributes 0.01 to the variance and the intermediate allele contributes 0.09. The final allele is ignored, because given the values for the other alleles, its frequency is fixed by the condition that the total frequencies must add up to one. So, in this example, the rare allele contributes 10% and the intermediate allele contributes 90% of the total variance. If there were nine rare alleles with a frequency of 0.01 and one intermediate allele with frequency of 0.1, then the rare alleles together would contribute 50% of the total variance and the single intermediate variant would account for the other 50%.

FIGURE 3 shows that the total contribution of rare variants among the DNA-repair genes was found to be about the same as the intermediate variants and slightly less than the common variants. This occurred because there were more rare than common variants. By weighting the contribution of each variant only by its frequency, the total of all rare variants contributed about the same amount of genetic variance as the total of all intermediate or common alleles. Based on frequency alone, rare and common alleles make about the same contribution to variation.

These calculations provide information about the frequency of variant alleles. However, no data are available that connect the different variants to their consequences for disease. Inevitably, some of the variants will have little or no effect, whereas others may significantly increase risk. The common types are unlikely to be severely deleterious, but beyond that, no strong conclusions can be made about the effects of the variant alleles. As discussed previously, it is possible to engineer particular genetic variants and to look for physiological consequences in cell-based assays. However, it is difficult to measure small but biologically relevant variation in physiological performance and to determine how altered physiology affects disease progression.

Age-specific-incidence curves and multistage models of progression. We are some way from being able to connect widespread allelic variation to cancer risk. Many variant alleles no doubt have physiological consequences, and these consequences sometimes affect cancer progression. But few data exist to directly link particular allelic variants to their roles in disease progression. The main insights still come from epidemiological data and abstract mathematical models of cancer progression.

As described above, Peto and Mack⁴² showed that the risk of breast cancer remains approximately constant in each year for certain individuals. Close relatives of affected individuals also have high and nearly constant

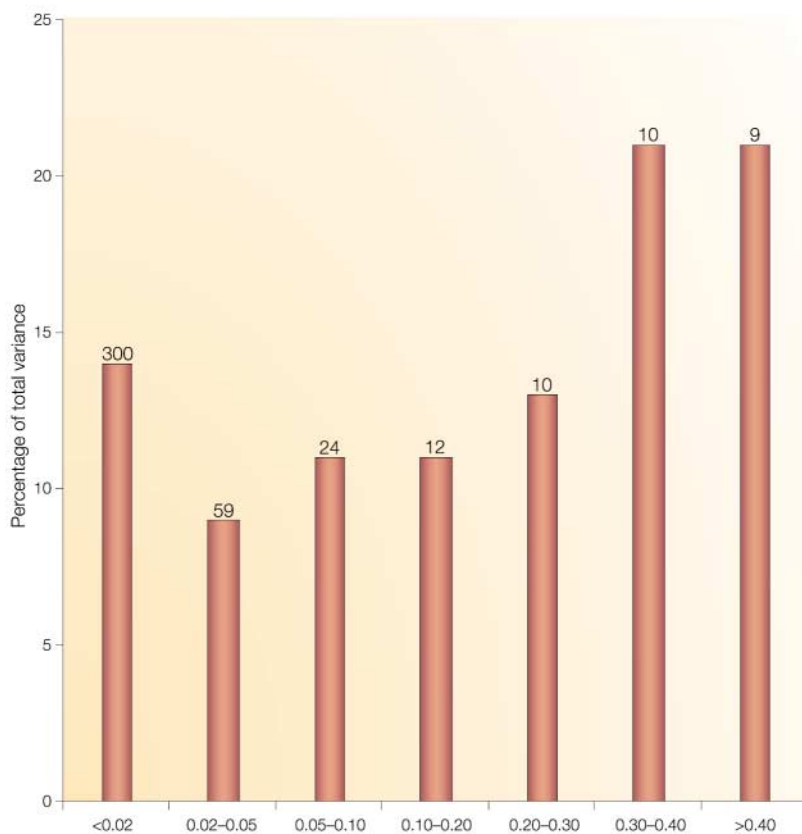


Figure 3 | Allele frequencies for DNA-repair genes. The relative variance in allele frequencies for rare and common alleles of 74 DNA-repair genes is shown. Each rare variant contributes a small fraction of the total variance, but there are many more rare than common variants. The total number of variants in each frequency category is shown above the bars. The variants are defined by changes in amino-acid sequence. The common variants probably have relatively mild phenotypic effects, allowing those variants to increase in frequency. Some of the rare variants might have more deleterious effects, but this has not been measured. To determine how rare and common variants affect disease, data would need to be obtained on the percentage of total variance in repair efficiency or disease outcome for each frequency class. The data shown here (taken from REF. 12) suggest that rare variants probably dominate in terms of contribution to disease.

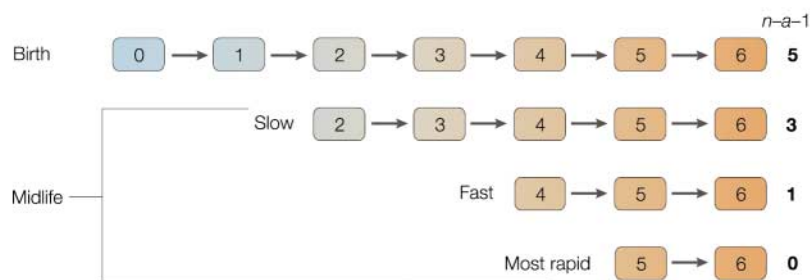


Figure 4 | Effects of multistage progression on age-related changes in cancer risk. A multistage model of progression⁴⁶ might explain why risk in highly predisposed families does not increase with age⁴². At birth, individuals must pass through n rate-limiting stages before progressing to cancer. The top line shows the six stages that a newborn must pass through in this case. As individuals grow older, many pass through the early stages. In midlife, most slow progressors have progressed to stage 2 and have four stages remaining. Faster progressors are in stage 4, and the most rapid group are in stage 5, or have passed all the way to stage 6, at which point cancer develops. Those rapid progressors in stage 5 only have to go through one remaining step. The probability that the final transition occurs determines the chance of progression in each year and remains constant over time. Individuals in stage 4 have two stages remaining. To progress to cancer, they must first pass to stage 5, and then to stage 6. Passing through two stages is more likely to happen in a future year than in the current year, because the first stage must be passed before the second stage can be breached. With two stages remaining, the rate of cancer increases linearly with age. For individuals in stage 2, four steps remain, and the rate of cancer per year increases with the third power of age. In general, the rate of cancer increases with age raised to the exponent $n-a-1$, where n is the initial number of stages and a is the number of stages passed by a particular age⁴⁶.

annual risks of developing cancer after the age at which the affected individual is diagnosed. The high rates of cancer in related individuals indicate a strong contribution of genetic predisposition. But why do predisposed individuals have a nearly constant risk over the later part of their life, whereas other individuals have an increasing risk as they age? Peto and Mack⁴² conclude: “A [...] model that may account for these peculiar temporal patterns is that many, and perhaps most, breast cancers arise in a susceptible minority whose incidence, at least on average, has increased to a high constant level at a predetermined age that varies between families.”

But why should predisposed individuals have constant annual risks after a certain age? Individuals who are not predisposed to breast cancer show an increasing risk with age, and the same is true for the other most common types of epithelial cancer when risk is measured in the absence of information about genetic predisposition⁴⁵. The classic multistage epidemiological model for cancer progression provides an intriguing hypothesis. In this model, there are n essential rate-limiting steps that must be completed before cancer develops². This does not mean that there are only n changes to a transformed cell, but that only n of those changes limit the rate of progression. The multistage model predicts that a log–log plot of annual incidence rate versus age will have a slope of $n-1$. Previous analyses of this model assumed that n is constant over all ages, so that the predicted slope of $n-1$ stays constant with age and the predicted plot is a straight line. Recently, I pointed out that as individuals age, they might pass some of the early stages⁴⁶. For example, if a typical individual has passed a stages by the age of 50, then the predicted slope at this age is $n-a-1$. In other words, the slope will

decline with age as individuals pass the early rate-limiting stages (FIG. 4).

If an individual has passed through all but the final stage in cancer progression, and has only one stage remaining, then their annual risk is constant — the risk is just the constant probability of passing the final stage. Families that have an increased predisposition may progress through the first $n-1$ stages quickly; subsequently, their annual risk is the constant probability of passing the final stage. Families with low genetic risk move through the early stages slowly, so that in mid-to-late life, members of those families typically have more than one stage to pass, and so continue to have an increasing rate of risk with advancing age.

If the early stages in cancer progression involve somatic mutations or chromosomal aberrations, impaired DNA-repair efficiency could explain why families with increased predisposition move quickly through the early stages. When they have progressed through the early stages, individuals from these families have a high constant risk later in life, while awaiting the final transition. By contrast, a better repair efficiency slows the transition through the early stages. Slow transitions early in life mean more stages to pass through later in life. With more stages remaining, individuals at low risk continue to show an increase in incidence with age.

Robust protection against cancer allows mutations to accumulate. Can we predict the kinds of traits that are likely to accumulate genetic variation and act as focal points for genetic predisposition to cancer? Do some cancers have a greater tendency for genetic predisposition than others? Detailed answers to these questions require data on how particular mutations and physiological variations affect progression to cancer. These data are not available. However, some general predictions can be made about how mutations accumulate in complex phenotypes. These predictions help to organize thinking about genetic predisposition and to guide future studies.

Late-onset cancers have less effect on reproduction than early-onset cancers. So, as outlined above, natural selection purges allelic variants that cause later expression of disease more slowly¹⁸. Data relating to the main Mendelian mutations that lead to cancer support the predicted association between age of onset and frequency. The age of cancer onset ranges from early to late for major mutations in *RB*, *APC* and *BRCA1*, which cause retinoblastoma, FAP and breast cancer, respectively. Estimates of the frequencies of mutations that cause rare cancers are not precise. However, it seems that higher penetrance and earlier onset of disease cause a lower frequency of dominant Mendelian mutations. Certainly, predisposing *RB* mutations occur at a much lower frequency (10^{-5} – 10^{-4}) than *BRCA1* mutations (10^{-3} – 10^{-2}).

A similar argument predicts that there is less polygenic variation in early-onset cancers compared to late-onset cancers. Most data on polygenic predisposition have been collected for late-onset cancers, so this prediction cannot be evaluated using existing data. It would be interesting to learn more about polygenic predisposition to childhood cancers, such as retinoblastoma.

Redundant DNA-repair pathways and multiple mechanisms that regulate the cell cycle and cell death can also influence the accumulation of mutations⁴⁷. For example, failure to repair double-strand DNA breaks increases the expression of p53, a transcription factor that in turn modulates the expression of many other genes⁴⁸. If DNA damage is moderate, p53 slows down the cell cycle so that the damage can be repaired and replication can be completed. If DNA damage is severe, p53 triggers an apoptotic pathway that leads to cell suicide. The function of p53 is mainly to protect against the damage that arises from environmental stress during the lifetime of the individual. However, a system that protects against the environment may also buffer against the negative effects of inherited mutations^{48–51}. For example, mutations that decrease the rate of DNA repair or allow more DNA damage to accumulate may have less effect because p53 compensates by adjusting the repair process and cell-cycle progression. So, the buffering effects of p53 can reduce the negative consequences of some inherited mutations, decreasing the rate at which natural selection removes those mutations from the population.

In general, the robustness provided by p53 and other regulatory controls of the cell cycle and DNA repair have obvious benefits. But robustness also allows some mutations to accumulate without having much effect on phenotype. Natural selection cannot purge mutations that are hidden by robust pathways. This relation between robustness and the accumulation of mutations provides a framework in which to understand how mutations accumulate. In particular, tissues that are protected against cancer by many mechanisms that regulate cell proliferation and cell death allow mutations to accumulate more readily than those tissues that require fewer genetic changes to progress to cancer^{47,53}.

The number of regulatory mechanisms that influence cancer development takes us back to the multistage

model of progression. A smaller number of rate-limiting steps means fewer checks on progression. In this case, less protection exists against the expression of each mutation; consequently, allelic variants accumulate more slowly, and less predisposing genetic variation occurs. An extra stage initially improves the survival and reproduction of individuals by pushing the onset of cancer to later ages. However, the additional buffering by an extra stage reduces the impact of any single deleterious mutation, and therefore allows the accumulation of more allelic variants in the population. Extra stages therefore lead to the evolution of later onset for cancer and significantly increased genetic predisposition to disease⁴⁷.

Conclusions

The roles of different stages and of robustness in cancer progression remain abstract at present. As information accumulates about specific DNA-repair and cell-cycle pathways and their robustness, it will be interesting to learn whether more robust pathways accumulate more allelic variants. Similarly, stages in progression may act as buffers in a robust system of protection against cancer. If so, then there may be less polygenic variation in early-onset cancers, such as retinoblastoma, that seem to have few stages, compared with polygenic variation in cancers with later onset and, apparently, more stages of progression, such as colon and breast cancer. In this regard, cancer provides a model for studying how the biological design of redundant protective and regulatory pathways influence genetic predisposition to disease⁴⁷.

Cancer may also provide insights into the general problem of how mutations accumulate in late-onset diseases, which is an important factor in ageing¹⁸. In particular, cancer studies may link perturbations in DNA repair and cell-cycle pathways to disease progression. The connection between biochemical perturbations and disease will clarify how ageing-related mutations accumulate.

- Vogelstein, B. & Kinzler, K. W. *The Genetic Basis of Human Cancer* (McGraw Hill, New York, 2002). **Provides good review chapters on all genetic aspects of cancer. There is emphasis on the biochemical roles of particular genes, but some information is given about polymorphisms in populations.**
- Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).
- Frank, S. A. Mathematical models of cancer progression and epidemiology in the age of high throughput genomics. *Int. J. Epidemiol.* (in the press).
- Marsh, D. & Zori, R. Genetic insights into familial cancers — update and recent discoveries. *Cancer Lett.* **181**, 125–164 (2002).
- Welch, P. L. & King, M. C. *BRCA1 and BRCA2* and the genetics of breast and ovarian cancer. *Hum. Mol. Genet.* **10**, 705–713 (2001).
- Couch, F. J. & Weber, B. L. In *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 549–581 (McGraw Hill, New York, 2002).
- Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
- Frank, S. A. & Nowak, M. A. Developmental predisposition to cancer. *Nature* **422**, 494 (2003).
- Kinzler, K. W. & Vogelstein, B. In *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 583–612 (McGraw Hill, New York, 2002).
- Boland, C. R. In *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 307–321 (McGraw Hill, New York, 2002).
- Lynch, H. T., Smyrk, T. & Jass, J. R. Hereditary nonpolyposis colorectal cancer and colonic adenomas: aggressive adenomas? *Semin. Surg. Oncol.* **11**, 406–410 (1995).
- Mohrenweiser, H. W., Wilson, D. M. & Jones, I. M. Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes. *Mutat. Res.* **526**, 93–125 (2003). **Reviews the evidence for significant genetic variation in DNA-repair capacity between families and the consequences for cancer susceptibility. Includes a good preview of future research in this area, combining biochemical analysis of genetic variants with frequencies of variants in populations.**
- Ford, D. *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.* **62**, 676–689 (1998).
- Tonin, P. *et al.* *BRCA1* mutations in Ashkenazi Jewish women. *Am. J. Hum. Genet.* **57**, 189 (1995).
- Struwing, J. P. *et al.* The risk of cancer associated with specific mutations of *BRCA1* and *BRCA2* among Ashkenazi Jews. *N. Engl. J. Med.* **336**, 1401–1408 (1997).
- Couch, F. J. & Weber, B. L. Mutations and polymorphisms in the familial early-onset breast cancer (*BRCA1*) gene. Breast Cancer Information Core. *Hum. Mutat.* **8**, 8–18 (1996).
- Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* 3rd edn (Sinauer, Sunderland, Massachusetts, 1997).
- Wright, A., Charlesworth, B., Rudan, I., Carothers, A. & Campbell, H. A polygenic basis for late-onset disease. *Trends Genet.* **19**, 97–106 (2003). **A review of population genetics theory regarding genetic susceptibility to disease, with emphasis on whether variant alleles that cause disease are likely to be common or rare in populations.**
- Haenszel, W. & Kurihara, M. Studies of Japanese migrants. I. Mortality from cancer and other diseases among Japanese in the United States. *J. Natl Cancer Inst.* **40**, 43–68 (1968).
- Armstrong, B. & Doll, R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int. J. Cancer* **15**, 617–631 (1975).
- Doll, R. Uncovering the effects of smoking: historical perspective. *Stat. Methods Med. Res.* **7**, 87–117 (1998).
- Vineis, P. *et al.* Tobacco and cancer: recent epidemiological evidence. *J. Natl Cancer Inst.* **96**, 99–106 (2004).
- Vineis, P. & Pirastu, R. Aromatic amines and cancer. *Cancer Causes Control* **8**, 346–355 (1997).
- Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
- Czene, K., Lichtenstein, P. & Hemminki, K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int. J. Cancer* **99**, 260–266 (2002).

26. Pharoah, P. D. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genet.* **31**, 33–36 (2002).
27. Barton, N. H. & Keightley, P. D. Understanding quantitative genetic variation. *Nature Rev. Genet.* **3**, 11–21 (2002).
28. Houle, D. Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195–204 (1992).
29. Mousseau, T. A. & Roff, D. A. Natural selection and the heritability of fitness components. *Heredity* **59**, 181–197 (1987).
30. Berwick, M. & Vineis, P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J. Natl Cancer Inst.* **92**, 874–897 (2000).
31. Grossman, L. *et al.* in *Advances in DNA Damage and Repair* (eds Dizdaroglu, M. & Karakaya, A. E.) 149–167 (Kluwer Academic/Plenum Publishers, New York, 1999).
32. Cloos, J. *et al.* Inherited susceptibility to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. *J. Natl Cancer Inst.* **91**, 1125–1130 (1999).
33. Roberts, S. A. *et al.* Heritability of cellular radiosensitivity: a marker of low-penetrance predisposition genes in breast cancer? *Am. J. Hum. Genet.* **65**, 784–794 (1999).
34. Bernstein, C., Bernstein, H., Payne, C. M. & Garewal, H. DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat. Res.* **511**, 145–178 (2002).
35. Thompson, L. H. & Schild, D. Recombinational DNA repair and human disease. *Mutat. Res.* **509**, 49–78 (2002).
36. Hsieh, P. Molecular mechanisms of DNA mismatch repair. *Mutat. Res.* **486**, 71–87 (2001).
37. de Boer, J. G. Polymorphisms in DNA repair and environmental interactions. *Mutat. Res.* **509**, 201–210 (2002).
38. Ishikawa, T. *et al.* DNA repair and cancer: lessons from mutant mouse models. *Cancer Sci.* **95**, 112–117 (2004).
39. Sands, A. T., Abuin, A., Sanchez, A., Conti, C. J. & Bradley, A. High susceptibility to ultraviolet-induced carcinogenesis in mice lacking XPC. *Nature* **377**, 162–165 (1995).
40. Nakane, H. *et al.* High incidence of ultraviolet-B- or chemical-carcinogen-induced skin tumours in mice lacking the xeroderma pigmentosum group A gene. *Nature* **377**, 165–168 (1995).
41. de Vries, A. *et al.* Increased susceptibility to ultraviolet-B and carcinogens of mice lacking the DNA excision repair gene XPA. *Nature* **377**, 169–173 (1995).
42. Peto, J. & Mack, T. M. High constant incidence in twins and other relatives of women with breast cancer. *Nature Genet.* **26**, 411–414 (2000).
- Shows different patterns of increase in breast cancer with age between individuals with and without a primary relative who had breast cancer. Suggests that families with a high level of genetic susceptibility have different age-specific patterns of disease from normal families, linking genetic variation to epidemiology.**
43. Weiss, K. M. & Terwilliger, J. D. How many diseases does it take to map a gene with SNPs? *Nature Genet.* **26**, 151–157 (2000).
44. Lee, C. Irresistible force meets immovable object: SNP mapping of complex diseases. *Trends Genet.* **18**, 67–69 (2002).
45. National Cancer Institute Surveillance, Epidemiology and End Results (SEER) Database <<http://seer.cancer.gov/>>
46. Frank, S. A. Age-specific acceleration of cancer. *Curr. Biol.* **14**, 242–246 (2004).
- Provides a theory to explain the different patterns of age-specific cancer onset in the genetically susceptible and normal families described in Ref. 42.**
47. Frank, S. A. Genetic variation in cancer predisposition: mutational decay of a robust genetic control network. *Proc. Natl Acad. Sci. USA* **101**, 8061–8065 (2004).
48. Malkin, D. in *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 387–401 (McGraw Hill, New York, 2002).
49. Bergman, A. & Siegal, M. L. Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**, 549–552 (2003).
50. de Visser, J. A. *et al.* Evolution and detection of genetic robustness. *Evolution Int. J. Org. Evolution* **57**, 1959–1972 (2003).
51. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
52. Sangster, T. A., Lindquist, S. & Queitsch, C. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. *Bioessays* **26**, 348–362 (2004).
53. Nunney, L. The population genetics of multistage carcinogenesis. *Proc. R. Soc. Lond. B* **270**, 1183–1191 (2003).

Acknowledgements
Work in the author's laboratory is supported by the National Science Foundation and the National Institutes of Health.

Competing interests statement
The author declares no competing financial interests

 **Online links**

DATABASES
The following terms in this article are linked online to:
Cancer.gov: http://www.cancer.gov/cancer_information/
Breast cancer | colorectal cancer | ovarian cancer | retinoblastoma | Wilms' tumour
Entrez Gene:
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>
APC | *BRCA1* | *BRCA2* | *CDK4* | *MET* | *p53* | *RET* | *RB*
OMIM: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>
Xeroderma pigmentosum

FURTHER INFORMATION
Steven Frank's web site: <http://stevfrank.org>
Access to this links box is available online.