ASIP Journal CME Program

The American Journal of Pathology, Vol. 172, No. 4, April 2008
Copyright © American Society for Investigative Pathology
DOI: 10.2353/ajpath.2008.070724

Tumorigenesis and Neoplastic Progression

The *in Vivo* Rate of Somatic *Adenomatous Polyposis Coli* Mutation

Chris Hornsby,* Karen M. Page,† and Ian Tomlinson‡

From the Departments of Computer Science, and Mathematics,† University College; and the Molecular and Population Genetics Laboratory,‡ London Research Institute, Cancer Research UK, London, United Kingdom*

Cancer arises through successive somatic mutations/epimutations of oncogenes and tumor-suppressor genes. Accurate estimates of the rates at which these (epi)mutations occur are a vital but missing link in our emerging quantitative understanding of tumorigenesis. Their absence has hindered arguments concerning the importance of genetic instability in tumorigenesis and the number of mutations that precede malignant conversion of healthy cell lineages. Herein, a novel method for calculating the *in vivo* mutation rate of the adenomatous polyposis coli (*APC*) tumor-suppressor gene is presented. The large majority of bowel cancers are thought to be initiated by a partial loss of APC function, with the age-onset pattern dramatically altered for the worse in familial adenomatous polyposis (FAP) because these patients harbor selected germline *APC* mutations. Colon cancer in the context of FAP can be thought of as occurring "one hit quicker" than in the sporadic setting. We were able to isolate and estimate the rate of the initiating *APC* mutation in sporadic cases using the age incidence of FAP to approximate the time taken for a cell lineage in a sporadic patient with one *APC* mutation to present clinically as a cancer. Our result of approximately 10^{-5} mutations per allele per year, although higher than previous estimates, appears to be consistent with the mutational spectrum of *APC*. The quality of fit provided by this method supports the theory that FAP and sporadic bowel cancer follow the same genetic pathway and are separated by only one mutation. ***(Am J Pathol 2008, 172:1062–1068; DOI: 10.2353/ajpath.2008.070724)***

The *in vivo* rate of somatic tumor suppressor or oncogene mutation cannot be measured directly. Often, it is not possible to isolate the precursor cells of a given cancer or, as a consequence, to probe their DNA for abnormalities. In cases where precursor cells can be isolated, the rarity of any specific mutation makes direct measurements of mutation rate impractical. Even in human cell cultures, determining mutation rate is very difficult[1] and has only been possible at a handful of loci. Estimates fall between 10^{-8} and 10^{-6} mutations per gene per cell generation.[2–4] Similar rates have been observed in yeast.[5,6]

These data provide only a blurred picture of the somatic mutation rate in humans. Aside from the 3 orders of magnitude over which they are spread, there is a question mark over how well the mutational characteristics of cultured cells mirror those of cells *in vivo*. Consequently, it is currently very difficult to ascribe a notional rate to tumor suppressor or oncogene mutation.

This is especially apparent when considering that the rate in a particular case will depend not only on microenvironmental factors that modify replication fidelity but also on the specific spectrum of genetic changes that lead to a selected mutant protein product.

Without resorting to cell lines, what alternative methods are available for measuring mutation rate? *In vitro*, the standard method is Luria-Delbruck fluctuation analysis. Many parallel clones are grown in culture from small parent populations containing no mutated genes. After a certain number of cell generations or when the clones have grown to a specific size, the prevalence of mutant cells within each population is recorded. The mean prevalence per clone is then used to estimate the underlying mutation rate. A simple *in vivo* analog of this experiment is to observe many patients (rather than clones) and measure the frequency with which neoplasia (rather than mutant colonies) arise, the neoplasia being markers of mutation. In other words, rather than trying to observe gene mutations at the microscopic level, a practical alternative is to observe malignancies at the population level (Figure 1). Luebeck and Moolgavkar[7] inferred a rate of 10^{-6} per gene per year for adenomatous polyposis coli (*APC*) mutation in patients with colorectal cancer via this approach.

Supported by a Medical Research, Council studentship (ref G75/47) through CoMPLEX.

Accepted for publication December 18, 2007.

Address reprint requests to Chris Hornsby, Department of Computer Science, University College London, WC1E 6BT. E-mail: c.hornsby@ucl.ac.uk.

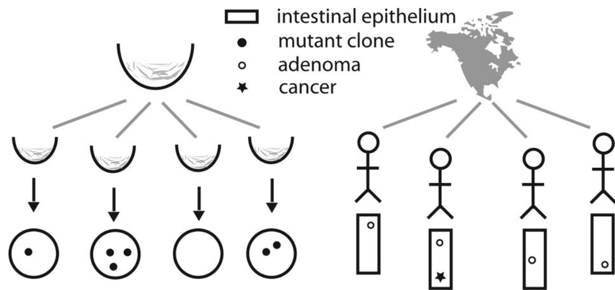

Figure 1. In Luria-Delbruck fluctuation analysis (**left**), the number of mutant colonies arising in plated clones can be used to estimate *in vitro* mutation rates. By analogy, *in vivo*, the number of tumors arising in individuals can be used.

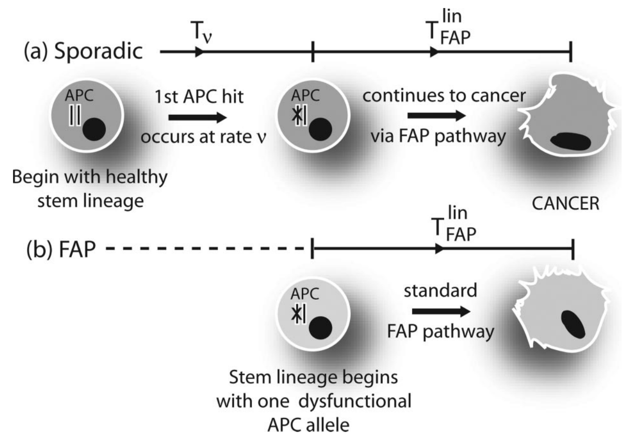

Figure 2. The pathways of FAP and sporadic bowel cancer are separated by only a single, truncating *APC* hit. In the case of FAP, the first hit already exists in the germline. In a sporadic patient, a given cell lineage takes T_ν years to acquire a truncating *APC* mutation in one allele and then a further $T_{\mathrm{FAP}}^{\mathrm{lin}}$ years to become malignant. A FAP lineage only takes $T_{\mathrm{FAP}}^{\mathrm{lin}}$ altogether to become malignant.

The primary advantage of using population data, in place of cell line data, is that they are indicative of *in vivo* rather than *in vitro* gene mutations. A second advantage is that neoplasms are markers of precisely the gene mutations whose rates of occurrence we would like to measure. Unfortunately, it is usually difficult to determine the number of precursor cells per patient at risk for a given cancer—a crucial factor in determining the mutation rate. Furthermore, besides the mutation of interest, there are often unknown additional pathogenic events required to produce a cancer. These confound the estimation procedure. For example, the estimate by Luebeck and Moolgavkar[7] of the *APC* mutation rate depends on an assumed number of progenitor cells per colorectum. It also depends on more difficult assumptions about the etiology of bowel cancer. Although it was reasonable for them to suppose that the initiating events are alterations in the two alleles of *APC*, assumptions about the events that follow from the second *APC* hit are more speculative. These compromise the accuracy of their estimate.

In cases where there is a well defined genetic and histological sequence, precursor legions can be used in place of cancers for estimating rates of gene mutation. The advantage of using neoplasia at an earlier stage of tumorigenesis is that they contain fewer genetic alterations in addition to the mutation of interest. Iwama[8] treated colonic adenomas as representative of two *APC* hits. From observed data on the incidence of adenomas in the bowel, he was able to infer a rate between 2×10^{-6} and 3×10^{-6} *APC* mutations per gene per year—similar to the estimate of Moolgavkar et al. Although Iwama's approach removes some of the uncertainty associated with etiology, it is still dependent on assumptions about the number of cells at risk of cancer per patient. Our approach removes this dependency, while still requiring only a limited knowledge of pathogenesis.

We infer the rate of *APC* mutation by comparing incidence of sporadic colon cancer with that of colon cancers arising in the context of familial adenomatous polyposis coli (FAP). FAP is a hereditary cancer syndrome caused by germline *APC* mutation. It is commonly assumed that colon cancers arising sporadically and in FAP patients proceed along similar pathways of somatic evolution with FAP cancers requiring one less mutation at the *APC* locus. If this is the case and if partial loss of APC function is initiating for the large majority of cancers of either type, then it is reasonable to use the age-onset pattern of FAP to estimate the time lapse between the first *APC* hit and clinical detection in sporadic bowel cancer (Figure 2). Using FAP incidence data in this manner removes the need for assumptions about the adenoma-carcinoma sequence after the first *APC* mutation. In addition, the method is insensitive to the assumed number of progenitor cells, because this information is implicit in the FAP data, provided that the number of progenitors is the same in FAP and sporadic patients.

Materials and Methods

Only one source of FAP data was available for the analysis. This was recorded in the UK over a period of several years ending in 1965.[9] To match the FAP data as closely as possible, sporadic data recorded in England and Wales between 1960 and 1962 by Doll et al[10] were used to estimate ν, the mutation rate per allele per year. It was assumed that sporadic bowel cancer emerges via the same genetic pathway as FAP but with an extra initiating *APC* "hit" (Figure 2). A quantitative model, describing the time taken for a sporadic bowel cancer to develop, was derived on the basis of this assumption. The initiating *APC* hit occurs at a rate of ν per allele per year in the model. The time between this hit and clinical detection was estimated from FAP incidence data. Fitting the model separately to sporadic male and female incidence data then allowed ν, the desired mutation rate, to be inferred in each case.

Time until Sporadic Cancer

It was assumed that the colonic epithelium of each patient is sustained by a population of stem cells. These stem cells form the target population for cancer. Periodically, each divides asymmetrically to give rise to one new stem cell and one non-stem cell daughter. A given stem cell, and its lineal stem cell descendants will be referred to collectively as a "stem cell lineage" (Figure 3). We assumed a fixed number of lineages, *N*, and each was treated as an independent entity. Before the first *APC*

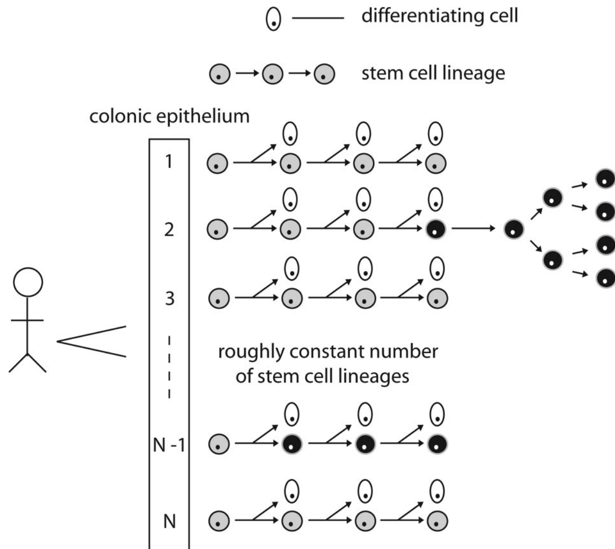

Figure 3. The epithelial sheet is replenished initially by a fixed number of independent stem cell lineages. These are shown in gray, and their differentiating progeny are shown in white. Black cells represent stem cell descendants that have at least one mutant *APC* allele.

mutation has occurred, each lineage retains only a single stem cell (undergoes no symmetric divisions). This restriction allows the time taken for the first *APC* mutation to be modeled with an exponential distribution. After the first *APC* mutation, however, the lineage can expand through symmetric divisions. A cancer is recorded when the first of the lineages has become malignant and grown to be of a detectable size.

The age at which any single lineage in a sporadic patient becomes malignant is denoted by $T^{\text{lin}}_{\text{spor}}$. It is the sum of two waiting times (Figure 2): 1) T_v, the time taken for the initial *APC* hit to occur; and 2) $T^{\text{lin}}_{\text{FAP}}$, the time taken for the events that comprise the FAP pathway to follow (see Figure 2). So $T^{\text{lin}}_{\text{spor}}$ can be expressed as

$$T^{\text{lin}}_{\text{spor}} = T_v + T^{\text{lin}}_{\text{FAP}} \tag{1}$$

T_v was assumed to follow an exponential distribution, with rate 2ν (because there are two alleles and ν is the mutation rate per allele per year). Two contrasting methods were then used to derive a distribution for $T^{\text{lin}}_{\text{FAP}}$ from the observed FAP data.

First of all, to avoid forcing a generic shape onto the distribution of $T^{\text{lin}}_{\text{FAP}}$, it was constructed as a survival curve directly from the FAP data. The construction proceeded in two stages. Initially, the observed FAP data, as summarized by Ashley,[11] were used to estimate the distribution of T_{FAP}, the time at which a patient (rather than a lineage) first develops a malignancy. So, $P[T_{\text{FAP}} \leq t_i]$ was inferred for $t_i = 10, 15, 20, 25, \ldots$ via the Actuarial method (see for example Parmar and Machin[12]). Second, we observed that if a patient has N independent lineages at risk of becoming cancerous, then there is a simple relationship between T_{FAP} and $T^{\text{lin}}_{\text{FAP}}$:

$$P[T^{\text{lin}}_{\text{FAP}} \leq t_i] = 1 - (1 - P[T_{\text{FAP}} \leq t_i])^{1/N}$$

Hence, a smooth approximation to $P[T^{\text{lin}}_{\text{FAP}} \leq t]$ was derived by interpolating the co-ordinates

$$(t_i, 1 - (1 - P[T_{\text{FAP}} \leq t_i])^{1/N}) \quad t_i = 10, 15, 20, \ldots$$

Using this approximation to $P[T^{\text{lin}}_{\text{FAP}} \leq t]$ and equation 1, it follows that:

$$P[T^{\text{lin}}_{\text{spor}} \leq t] = P[T_v + T^{\text{lin}}_{\text{FAP}} \leq t]$$

$$= \int_0^t P'[T_v \leq s]\, P[T^{\text{lin}}_{\text{FAP}} \leq t - s]\, ds$$

This gives the age at which cancer arises in a single lineage, $T^{\text{lin}}_{\text{spor}}$. For the patient, any of the N lineages has the potential to become a cancer. Consequently, the time taken for a patient to present with cancer, T_{spor}, is much quicker than for a single lineage. The two waiting times are related by

$$P[T_{\text{spor}} \leq t] = 1 - (1 - P[T^{\text{lin}}_{\text{spor}} \leq t])^N \tag{2}$$

Fit to Sporadic Data 1

With n fixed and $T^{\text{lin}}_{\text{FAP}}$ defined as above, $P[T_{\text{spor}} \leq t]$ (equation 2) was used to construct a likelihood function, $L(\nu)$, for the sporadic data (for details of this standard construction, see, for example, Ref. 7). In turn, $L(\nu)$ was used to calculate a posterior distribution, post(ν). Assuming a uniform prior for ν on 10^{-7} to 10^{-3} we took

$$\text{post}(\nu) = \frac{L(\nu)}{\int_{10^{-7}}^{10^{-3}} L(v)\, dv} \tag{3}$$

Equation 3 was evaluated using the Mathematica 5.2 software package by Wolfram Research Inc. (Champaign, IL).

The variance of this posterior is likely to be too narrow because uncertainty over the distribution of $T^{\text{lin}}_{\text{FAP}}$ was ignored in its derivation. To address this issue, in our second method for characterizing $T^{\text{lin}}_{\text{FAP}}$, we assume its distribution follows a predefined functional form whose parameters are to be inferred. Specifically, we used the formula of Armitage and Doll.[13] This gives the probability that a cancer requiring n successive mutations, each occurring at a rate u, has developed from an immortal lineage by time t:

$$P[T^{\text{lin}}_{\text{FAP}} \leq t] = 1 - \sum_{i=0}^{n-1} \frac{u^i t^i}{i!} e^{-ut} \tag{4}$$

so that

$$P[T_{\text{FAP}} \leq t] = 1 - (1 - P[T^{\text{lin}}_{\text{FAP}} \leq t])^N \tag{5}$$

Fit to Sporadic Data 2

Using equations 4 and 5, a likelihood function, $L(u, n, \nu)$, was constructed for the FAP and sporadic data together, given the approximated FAP curve (parameterized by u and n) and the model for the sporadic disease. This likelihood was the product of the likelihood for the spo-

radic data [$L(\nu)$, calculated as previously described] and a new likelihood, $L_{FAP}(u, n)$, for the FAP data.

The FAP data are drawn from a small study involving M patients, where M is ~100. The patients are observed at 5-year age intervals. During the ith age interval, a total of S_i patients are treated with surgery before being removed from the study, and D_i new cancers occur among the remaining patients. To account for censoring of patients through treatment, Pop_i, the effective number of patients at risk of a primary cancer during the ith interval, is defined by

$$Pop_i = M - \sum_{k=0}^{i-1}(D_k + S_k) - \frac{1}{2}S_i$$

ie, the number of healthy patients remaining in the study at the beginning of the interval less one-half those that are treated during the period. The likelihood, $L_{FAP}(u, n)$, of the FAP data can be expressed as a product of binomial probabilities. If the probability of a cancer arising in a given patient during the ith interval is p_i, then the probability that D_i cancers arise among the effective population Pop_i over the period is

$$\binom{Pop_i}{D_i} p_i^{D_i}(1 - p_i)^{Pop_i - D_i}$$

$L_{FAP}(u, n)$ is then equal to the probability of the given distribution of cases $\{D_1, D_2, \ldots\}$,

$$L_{FAP}(u, n) = \prod_i \binom{Pop_i}{D_i} p_i^{D_i}(1 - p_i)^{Pop_i - D_i}$$

If s_i and e_i are the start and end points of the ith age interval, then p_i can be evaluated, using equation 5, as

$$\begin{aligned} p_i &= P[T_{FAP} \leq e_i | T_{FAP} > s_i] \\ &= \frac{P[T_{FAP} \leq e_i] - P[T_{FAP} \leq s_i]}{1 - P[T_{FAP} \leq s_i]} \end{aligned} \qquad (6)$$

Ascertainment bias has been neglected in equation 6. Because the penetrance of FAP is 100%, this bias should be minimal. To fit both the sporadic and FAP data simultaneously, the combined likelihood function, $L(u, n, \nu)$, was used:

$$L(u, n, \nu) = L(\nu)\, L_{FAP}(u, n)$$

The posterior density at ν (Figure 4) was calculated assuming uniform priors for ν, u, and n, respectively, on 10^{-7} to 10^{-3}, 0 to 10^{-2}, and 2 to 7:

$$post(\nu) = \frac{\sum_{n=n_{min}}^{n_{max}} \int_{u_{min}}^{u_{max}} L(u, n, \nu)\, du}{\sum_{n=n_{min}}^{n_{max}} \int_{v_{min}}^{v_{max}} \int_{u_{min}}^{u_{max}} L(u, n, v)\, du dv}$$

Again, Mathematica was used to evaluate this expression.

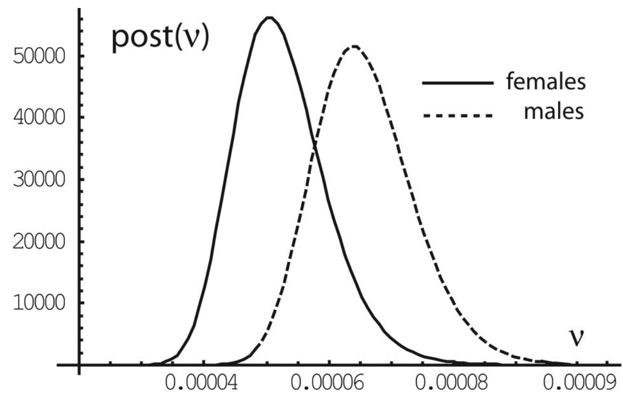

Figure 4. Posterior distribution of the *APC* mutation rate ν, measured in mutations per allele per year, calculated using a three parameter likelihood function. Sporadic colon cancer incidence data on cases diagnosed between 1960 and 1962 taken from Doll et al.[10]

Results

We assumed that sporadic bowel cancer emerges via the same genetic pathway as FAP but with an extra initiating *APC* hit (Figure 2). A stochastic model describing the time taken for a sporadic bowel cancer to develop was derived on the basis of this assumption. The intestinal cell lineages of a sporadic patient acquire initiating *APC* hits at a rate of ν per allele per year in the model. The remaining time taken for any such lineage to become malignant and present clinically was estimated from FAP incidence data. Fitting the resulting model back to the sporadic incidence curve allowed a posterior distribution for ν, the desired mutation rate, to be inferred.

This posterior distribution, post(ν), was derived via two separate methods. In the first, the time taken for an $APC^{+/-}$ cell lineage to become malignant and present clinically was assumed to be distributed according to a survival curve constructed from the FAP data. The shape of post(ν) (Figure 5) in this case implies that ν, the annual rate of *APC* mutation per allele, falls between 6×10^{-5} and 8×10^{-5} for both males and females. These estimates are insensitive to 2-order of magnitude changes in the assumed number of cell lineages per patient. However, substituting the optimum ν for males or females into the model gives only a rough approximation to the spo-

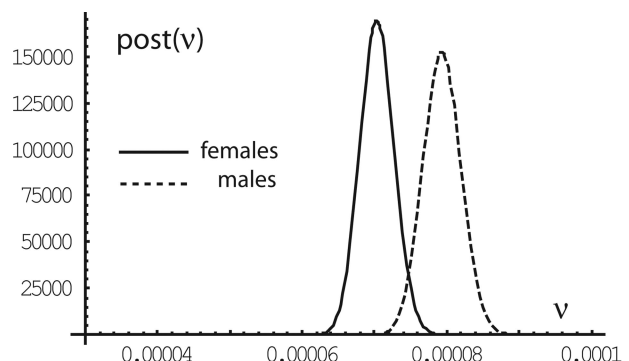

Figure 5. Posterior distribution of the *APC* mutation rate ν, measured in mutations per allele per year, calculated using a single parameter likelihood function. Sporadic colon cancer incidence data on cases diagnosed between 1960 and 1962 taken from Doll et al.[10]

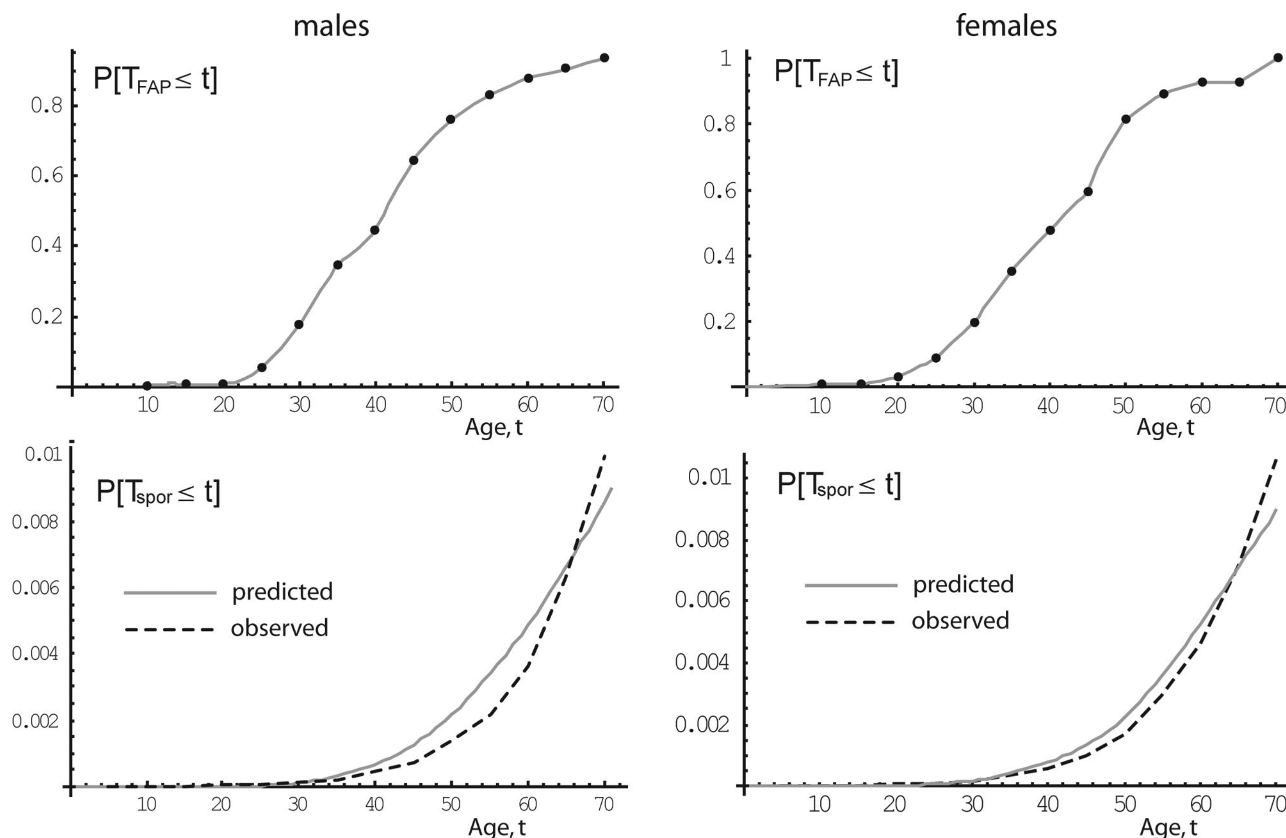

Figure 6. $P[T_{\text{FAP}} \leq t]$, the cumulative risk of FAP at age t, was calculated separately for males (left column) and females (right column) by interpolating observed FAP data. $P[T_{\text{spor}} \leq t]$, the cumulative risk of sporadic bowel cancer at age t, was constructed from $P[T_{\text{FAP}} \leq t]$, according to assumptions about the relationship between FAP and sporadic bowel cancer described in the text. Using the optimum ν, $P[T_{\text{spor}} \leq t]$ provides an adequate approximation to observed sporadic data.

radic incidence profiles (Figure 6). The discrepancy is systematic, in that for both males and females, we overestimate the risk before 65 years of age and underestimate it thereafter. It is likely that ascertainment bias in the FAP data is the underlying cause of this error. Ascertainment bias in the FAP data arises because only those patients who present clinically before death from other causes are included in the study. Consequently, risk of cancer at young ages, when mortality is low, is exaggerated whereas risk in older age as mortality increases is downplayed. Another problem is that the FAP data consist of only a small number of observations. By assuming the "real" distribution of FAP cases followed exactly a survival curve constructed from a small sample, the variances of the posterior distributions on ν were underestimated. To quantify the extent of this problem, as an alternative to using the exact survival curve suggested by observed data points, a parameterized class of possible approximations to the FAP curves was used. The posterior densities on ν in this case are only slightly wider (Figure 4). The optimized models naturally give a better fit to the sporadic data (Figure 7) and seem, crudely, to correct for the ascertainment bias described in the FAP data.

Discussion

Our calculations suggest that alleles of the *APC* tumor suppressor mutate to cause a selected protein change, about 6×10^{-5} times a year. This is 30 times faster than previous estimates.[8,7] A rough consistency check can be done by comparing our estimate against the mutational spectrum of the gene. More than 95% of *APC* mutations found in bowel cancers are nonsense or frameshift mutations. They take the form of small deletions/insertions or point mutations that result in truncation of the protein.[14] Assuming an error rate (insertions/deletions or mismatches) of 10^{-10} per base per cell generation[15] and assuming 100 stem cell divisions per annum, the number, BP, of base pairs through which *APC* can be truncated (either by insertion/deletion or point mutation) would be

$$BP = \nu \times 10^{8}$$

Substituting in order 10^{-5} for ν gives BP $\simeq$ 1000. This seems a sensible value for BP because more than 700 distinct, somatic *APC* mutations have been reported in bowel cancers to date.[16]

Approximations and Assumptions

Confidence in our estimates for ν should be based primarily, on the accuracy of the assumptions/approximations used in their calculation. The most crucial are as follows: 1) FAP and sporadic bowel cancer are both initiated by genetic alterations *in the APC* gene. 2) A lineage in a sporadic patient, with one *APC* hit, becomes malignant via the same mechanism as a FAP lineage with

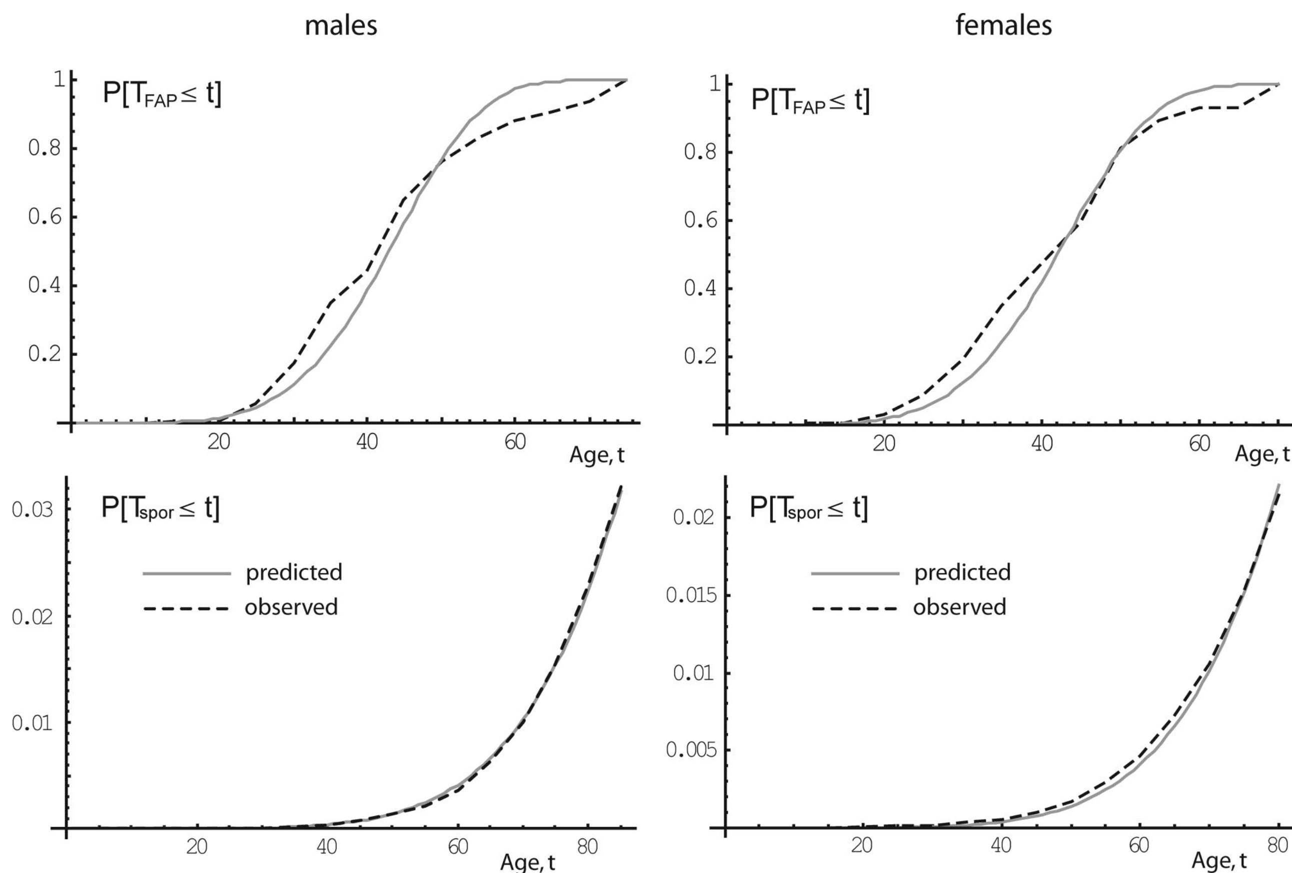

Figure 7. $P[T_{\mathrm{FAP}} \leq t]$, the cumulative risk of FAP at age t, is represented by a smooth function, parametrized by u and n. $P[T_{\mathrm{spor}} \leq t]$, the cumulative risk of sporadic bowel cancer at age t, was again constructed separately for males and females from $P[T_{\mathrm{FAP}} \leq t]$, as described in the text. Using the optimum parameter vector (u, ν, n), a good fit to the sporadic data can be made.

the germline genotype. 3) The target lineages for bowel cancer act independently. 4) Loss of the first *APC* allele in a sporadic patient follows an exponential distribution.

To evaluate the accuracy of these assumptions, a definition of "sporadic" colon cancer is necessary. We take sporadic to mean any case occurring in the general population, excluding those arising in the context of a known, highly penetrant germline mutation. Although initiating *APC* mutations appear to feature in the large majority of such cases,[17] there are alternative avenues available for a progenitor cell in the colonic epithelium.[18] Our method is also consistent with a model in which *APC* mutation appears as an early event, preceding any clonal growth, but is not necessarily initiating. However, cases with no *APC* mutation are a source of noise in the experiment. Because these cases form a minority of <30% and because their age relatedness is unlikely to be remarkably different from those initiated by *APC*, the impact should be minimal. A similar, but potentially more serious source of noise arises, because the registry data we use to represent sporadic cases are likely contaminated with FAP and other highly penetrant hereditary bowel cancers. However, these account for only 5% of all colorectal cancers,[19] so the effect is negligible. Concerning approximation 2, *APC* hits can broadly be categorized into two types: 1) truncating mutation and 2) LOH. It is well established that the type and the position of the two hits at *APC* are not independent in either FAP or sporadic bowel cancer. The first hit in sporadic cases and the germline mutations in FAP are both predominantly truncating mutations, but their spatial distribution is not identical. Consequently, due to nonindependence of the two hits, or otherwise, LOH may occur at a different frequency as a second hit in sporadic cancers. Fortunately, the difference in the spatial distributions is primarily due to certain germline mutations occurring at high frequencies relative to somatic mutations in sporadic cases. Specifically, germline mutations affecting codon 1061 and codon 1309 are relatively frequent in FAP. Because these tend to lead to different second hits, truncating mutation and LOH, respectively, the effect on the overall frequency of LOH as a second hit in FAP is limited. For example, Rowan et al[20] and Prall et al[21] detected LOH in 23 of 99 (23%) and 20 of 99 (20%) sporadic cancers, respectively, whereas Lamlum et al[22] found *APC* LOH in 42 of 210 (20%) FAP tumors.

Other factors that usually confound estimates taken from incidence data include the assumed number of cell lineages at risk and calendar year effects. Happily, the comparative nature of this study eliminates any sensitivity to the assumed number of lineages. It remains to be shown that the estimate is also robust under age-related changes in cell number. Calendar year effects are mild for bowel cancer before the 1980s[7] and are unlikely to

have a significant effect. Finally, a difficult problem, when modeling incidence data, is to decide how long it takes for a cancer to be detected once it has come into existence.[23] In this study, the FAP data are used to estimate the time taken for an $APC^{+/-}$ cell lineage to mutate into a cancer and present clinically. Assumptions about how this time is split between tumor formation and tumor progression are not required.

Given a small set of fairly conservative assumptions regarding sporadic colon cancer and its relationship to FAP, incidence data on the two diseases imply that the rate of truncating and disease-causing *APC* mutation is of order 10^{-5} mutations per allele per year. This estimate, although higher than previous estimates, seems to be consistent with the mutational spectrum of *APC*. Furthermore, the estimate neither requires accurate determination of the number of target cells in the colon nor depends on assumptions about time spans between the appearance of a cancer and its detection clinically. The quality of fit provided by the model supports the theory that FAP and sporadic bowel cancer follow the same genetic pathway and are separated by only one mutation.

References

1. Kendal WS, Frost P: Pitfalls and practice of Luria-Delbruck fluctuation analysis: a review. Cancer Res 1988, 48:1060–1065
2. DeMars R, Held KR: The spontaneous azaguanine-resistant mutants of diploid human fibroblasts. Humangenetik 1972, 16:87–110
3. Seshadri R, Kutlaca RJ, Trainor K, Matthews C, Morley AA: Mutation rate of normal and malignant human lymphocytes. Cancer Res 1987, 47:407–409
4. Araten DJ, Golde DW, Zhang RH, Thaler HT, Gargiulo L, Notaro R, Luzzatto L: A quantitative measurement of the human somatic mutation rate. Cancer Res 2005, 65:8111–8117
5. Lichten M, Haber JE: Position effects in ectopic and allelic mitotic recombination in Saccharomyces cerevisiae. Genetics 1989, 123:261–268
6. Yuan LW, Keil RL: Distance-independence of mitotic intrachromosomal recombination in Saccharomyces cerevisiae. Genetics 1990, 124:263–273
7. Luebeck G, Moolgavkar SH: Multistage carcinogenesis and the incidence of colorectal cancer. Proc Natl Acad Sci USA 2002, 99:15095–15100
8. Iwama T: Somatic mutation rate of the APC gene. Jpn J Clin Oncol 2001, 31:185–187
9. Veale AMO: Intestinal Polyposis. Eugenics Laboratory Memoirs, No. 40. Cambridge, UK, Cambridge University Press, 1965
10. Doll R, Payne P, Waterhouse J: Cancer Incidence in Five Continents, vol 1. Geneva, Switzerland, Union Internationale contre le Cancer, 1966
11. Ashley DJB: Colonic cancer arising in polyposis coli. J Med Genet 1969, 6:376–378
12. Parmar MKB, Machin D: Survival Analysis: A Practical Approach. Chichester, UK, Wiley 1995
13. Armitage P, Doll R: The age distribution of cancer and a multistage theory of carcinogenesis. Br J Cancer 2004, 91:1983–1989
14. Fearnhead NS, Wilding JL, Bodmer WF: Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumourigenesis. Br Med Bull 2002, 64:27–43
15. Kunkel TA, Bebenek K: DNA replication fidelity. Annu Rev Biochem 2000, 69:497–529
16. Laurent-Puig P, Beroud C, Soussi T: APC gene: database of germline and somatic mutations in human tumors and cell lines. Nucleic Acids Res 1998, 26:269–270
17. Rowan AJ, Lamlum H, Ilyas M, Wheeler J, Straub J, Papadopoulou A, Bicknell D, Bodmer WF, Tomlinson IPM: APC mutations in sporadic colorectal tumors: a mutational hotspot and interdependence of the two hits. Proc Natl Acad Sci USA 2000, 97:3352–3357
18. Segditsas S, Tomlinson I: Colorectal cancer and genetic alterations in the Wnt pathway. Oncogene 2006, 25:7531–7537
19. de la Chapelle A: Genetic predisposition to colorectal cancer. Nat Rev Cancer 2004, 4:769–780
20. Rowan A, Halford S, Gaasenbeek M, Kemp Z, Sieber O, Volikos E, Douglas E, Fiegler H, Carter N, Talbot I, Silver A, Tomlinson I: Refining molecular analysis in the pathways of colorectal carcinogenesis. Clin Gastroenterol Hepatol 2005, 3:1115–1123
21. Prall F, Weirich V, Ostwald C: Phenotypes of invasion in sporadic colorectal carcinomas related to aberrations of the adenomatous polyposis coli (APC) gene. Histopathology 2007, 50:318–330
22. Lamlum H, Ilyas M, Rowan A, Clark S, Johnson V, Bell J, Frayling I, Efstathiou J, Pack K, Payne S, Roylance R, Gorman P, Sheer D, Neale K, Phillips R, Talbot I, Bodmer W: The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. Nat Med 1999, 5:1071–1075
23. Michor F, Iwasa Y, Nowak MA: The age incidence of chronic myeloid leukemia can be explained by a one-mutation model. Proc Natl Acad Sci USA 2006, 103:14931–14934