1 **Reviewer: Dr. Simo Schwartz Jr.**

1. We have changed the wording as suggested ("...numbers of oncogenic alterations (genetic mutations or epigenetic alterations)...")

2. We have changed "rare humans" to "familial cancer syndromes".

3. We favor a multistage model because it can be applied to the data with reasonable fit. We further explain our choice in a new final paragraph in the Discussion.

4. The observation that fewer mutations are needed for lower stage cancers is puzzling. Also surprising, for neither the Finnish MSI- patients nor the SEER patients do the different stages of cancer have mean ages of diagnosis in the expected rank order (Tables 1 and 2). This is perhaps due to ascertainment bias. We emphasize that our model provides rough estimates rather than absolute values for the number of mutations to cancer with a new sentence on page 7 in the Discussion. ("Such estimates should be considered rough guides rather than absolute values because our model does not account for all factors.").

2 **Reviewer: Dr. Steven Frank**

1. In the Methods section (page 4), we have provided a URL for the raw data used in the paper.

2. We have added a new sentence on page 7 of the Discussion that emphasizes that the estimated numbers of mutations should be viewed as rough guides rather than absolute values. ("Such estimates should be considered rough guides rather than absolute values because our model does not account for all factors.").

3. We have added a comment that your analysis was more consistent with a single mutation difference between FAP and sporadic colon cancer (page 7 of the Discussion).

3 **Reviewer: Dr. Suresh H. Moolgavkar**

1. We are aware of the references the reviewer has mentioned, and have now added them to a new last paragraph of the Discussion. We note there are many approaches that model cancer, and explicitly state that one of the reasons our approach differs is that we do not assume substantial growth until after the last required mutation has been acquired.

2. It is true that we ignore temporal trends in colo-rectal cancer. For the Finnish data set we now adjust for the ascertainment bias with a Finnish life table; as before for the United States data set we use a life table from
the United States population. We have added a sentence in the Methods section (page 5) to make this clear. This change produced a slight change in the numbers in Table 1.

3. Equation (13) in our American Journal of Pathology paper (reference (11)) is not in error. In this paper we model the time it takes until the first cell accumulates \( k \) mutations, after which we assume a cancer quickly develops. Since the data we consider is the age of cancer patients when they were diagnosed with cancer, there is an ascertainment bias. We consider two random variables: \( X \) which is the time until the first cell accumulates \( k \) mutations and which we model by the density function \( dF_{\gamma,k}(z) \), and \( Y \) which is the time an individual dies and which we model by the survival function \( S(z) \). \( S(z) \) is the probability an individual is alive at age \( z \), and was derived from a census life table. We assume \( X \) and \( Y \) are independent.

In equation (13) we calculate the probability that an individual develops cancer in the time interval \((t_1, t_2)\), conditioned on this happening before he dies:

\[
H_{\gamma,k}(t_1, t_2) = \frac{P(X \in (t_1, t_2)|X < Y)}{P(X < Y)} = \frac{\int_{t_1}^{t_2} S(z) dF_{\gamma,k}(z)}{\int_{0}^{\infty} S(z) dF_{\gamma,k}(z)}
\]

The reviewer claims that we should integrate with respect to the hazard function rather than the density function. This is not true. Perhaps we can best demonstrate this fallacy by supposing that the independent random variables \( X \) and \( Y \) do not have the distributions specified above but rather have distributions uniform on \([0, 1]\). Consider the random variable \( W \) equal to \( X \) if \( X < Y \) (if \( X \geq Y \) then continue to sample independent realizations of \( X \) and \( Y \) until \( X < Y \) and then set \( W = X \)). The density of \( W \) must therefore decrease from 0 to 1: if \( X \) is small there is better chance that \( X < Y \) and therefore \( W \) will be small then if \( X \) is large. If we take our approach and integrate the survival function \((S(z) = 1 - z)\) for a uniform random variable) by the density function \((dF(z) = dz\) on \([0, 1]\) for a uniform random variable) then:

\[
P(X \in (t_1, t_2)|X < Y) = \frac{\int_{t_1}^{t_2} (1- z) dz}{\int_{0}^{1} (1- z) dz} = 2t_2 - 2t_1 - t_2^2 + t_1^2
\]

This implies \( W \) has density function \( 2 - 2t \) on \([0, 1]\), which is decreasing as it must. If instead we integrate with respect to the hazard function \((h(z) = 1/(1 - z)\) for a uniform random variable) then:
\[
\frac{\int_{t_1}^{t_2} (1 - z)^{1/2} dz}{\int_0^1 (1 - z)^{1/2} dz} = t_2 - t_1
\]

And this implies \( W \) is uniform on \([0, 1]\), which is clearly wrong.