

FORMAL COMMENT

On the slope of the regression between stem cell divisions and cancer risk, and the lack of correlation between stem cell divisions and environmental factors-associated cancer risk

Cristian Tomasetti^{1,2*}, Bert Vogelstein³

1 Division of Biostatistics and Bioinformatics, Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland, United States of America, **2** Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **3** Ludwig Center & Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland, United States of America

* ctomasetti@jhu.edu



OPEN ACCESS

Citation: Tomasetti C, Vogelstein B (2017) On the slope of the regression between stem cell divisions and cancer risk, and the lack of correlation between stem cell divisions and environmental factors-associated cancer risk. PLoS ONE 12(5): e0175535. <https://doi.org/10.1371/journal.pone.0175535>

Editor: Nancy Lan Guo, West Virginia University, UNITED STATES

Received: June 9, 2016

Accepted: March 20, 2017

Published: May 16, 2017

Copyright: © 2017 Tomasetti, Vogelstein. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by The John Templeton Foundation, The Virginia and D.K. Ludwig Fund for Cancer Research, The Lustgarten Foundation for Pancreatic Cancer Research, The Sol Goldman Center for Pancreatic Cancer Research, and NIH Grants P30-CA006973, R37-CA43460, R01-CA57345, and P50-CA62924. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funders had no role in study

We are pleased that the analyses on smoking and radiation performed by Little et al. support the idea that the effects of various environmental factors on cancer risk do not correlate significantly with the lifetime number of stem cell divisions. This provides further evidence that the correlation we obtained for the U.S., and similarly found across the rest of the world, is in general not due to environmental factors but rather is due to intrinsic ones. We also point out mathematical and conceptual mistakes in the analysis of Little et al. that led them to erroneous conclusions.

Introduction

We are delighted that many investigators are considering the issues raised by our previous study. The recent article by Little et al. highlights some of these important issues and gives us an opportunity to comment on them.

Results and discussion

Little et al. [1] misrepresent our findings [2] by making incorrect statements that are similar to those previously found in some media. The problem with these types of statements has been already pointed out by others [3]. Specifically, the authors at times confuse the concept of absolute risk with that of relative risk. We never suggested that “most cancers arose by chance”—as stated in the first sentence of the Abstract of [1]. Instead, we stated that the variation in cancer risk is largely explained by stem cell divisions [2]. We interpreted this correlation as indicating that replicative mutations associated with stem cell divisions (bad luck) was largely responsible for this variation. The difference between absolute risk and relative risk is fundamental in epidemiological statistics. In fact, the first word in the title of our paper is “Variation”. While some journalists caused confusion by not differentiating between these two very different concepts, we would hope that this fundamental mistake is not perpetuated in the scientific literature. A similar mistake is reiterated in the Discussion section of Little et al., where it is stated that we incorrectly “infer that most cancers are just due to bad luck”. No such inferences were made by us, as we were well aware of the differences between absolute and relative risks.

In their new analysis, Little et al. assumed a modified version of the standard multistage carcinogenesis model of Armitage and Doll [4]. The authors attempt to justify the choice of an Armitage-Doll model by stating that “much use has been made” (p. 2) of this model in the

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: We are the disputed authors of the paper: Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 2015, 347 (6217):78-81.

past. The same justification could be made if they were to use aether models to explain the transmission of electromagnetic and gravitational forces. Neither the aether model nor the original Armitage and Doll model are now considered appropriate for reasons that became apparent only after these models were posited. The Armitage and Doll model does not account for clonal expansions, the fundamental component of carcinogenesis. A related error lies in the statement that “the Armitage-Doll model has also been used to determine the number of driver-gene mutations associated with two types of cancer” (p. 2), referring to another of our recent papers [5]. This statement is incorrect, as we used a different mathematical model, partially based on that of Durrett et al. [6], to arrive at the conclusions in [5].

Little et al. attempted to incorporate clonal expansions into their analysis (their equations 2 to 7), stating that their equations are as “has been done by Tomasetti et al. [5], based on approximations of Durrett and Moseley [6]”. This is also incorrect, as their equations do not follow from the mathematical principles described in [5], [6]. Little et al. used a different model and did not provide a justification for it. Thus, the Little et al. claim that they use “the same modified Armitage-Doll model” that we used is also incorrect. We did not use an Armitage-Doll type model, or any modified version of it. Moreover, the model used by Little et al. did not include the developmental phase of tissues. While in some tissues this will not affect the results substantially, in others it is critical because the great majority of cell divisions within those tissues occurs during their development phases.

Little et al. also claim that the fact that the slope of the regression line relating stem cell divisions to cancer incidence is less than 1.0—something already observed by us [7]—provides evidence against our conclusions. Let’s review their approach: based on the flawed mathematical model described above and making the strong (as admitted by Little et al.) assumption that all cancers types require the same number of driver genes to be mutated, Little et al. estimate that the slopes should be >3 , and not the observed <1 . The first problem is that the slopes observed in cancer incidence data are not constant, as assumed by Little et al., but actually non-linear (for one of many examples, see Figure 2.3 of [8]). Second, the number of driver genes that must be mutated is clearly not constant across cancers, as mistakenly assumed by the authors. Persuasive evidence about this point is provided by considering the distribution of driver mutations found across different cancer types, and across patients, as shown in Fig 5 of [9]. Thus, the fact that the slope of their regressions is <1 is not a novel finding and more importantly does not provide evidence for or against our conclusions. And performing the analysis by leaving out leukemia, bone and thyroid cancers, and again assuming all other cancers require the same number of driver gene mutations, will not fix the problem. As pointed out in [7], a slope <1 can be easily explained by simple heterogeneity in the number of driver mutations across cancer types. If the theoretical cancer incidence, based on modeling, is plotted against the number of stem cell divisions, then the slope of the regression can vary markedly depending simply on the assumed heterogeneity in the number of driver genes. An example is shown in Fig 1. Such plots should not be confused with the observed (rather than modeled) cancer incidence plots in [1].

The Extra Risk Score (ERS), and related RBERS, were the first measures to assess the evidence of the role played by environmental and inherited factors in particular cancer types. As consistent with the goals of [2], these measures can reflect only relative risk with respect to other cancers, not absolute risk. A cancer with a higher ERS is evidence of a larger role of those factors in that cancer type than in the other cancers [2]. Interestingly, the authors use radiation and smoking data to test our ERS. Such tests are unjustified. Why should a tissue with a higher ERS than another tissue be more affected by one single environmental factor? For example, liver cancers have high ERS, but this doesn’t mean that smoking or radiation plays a larger role

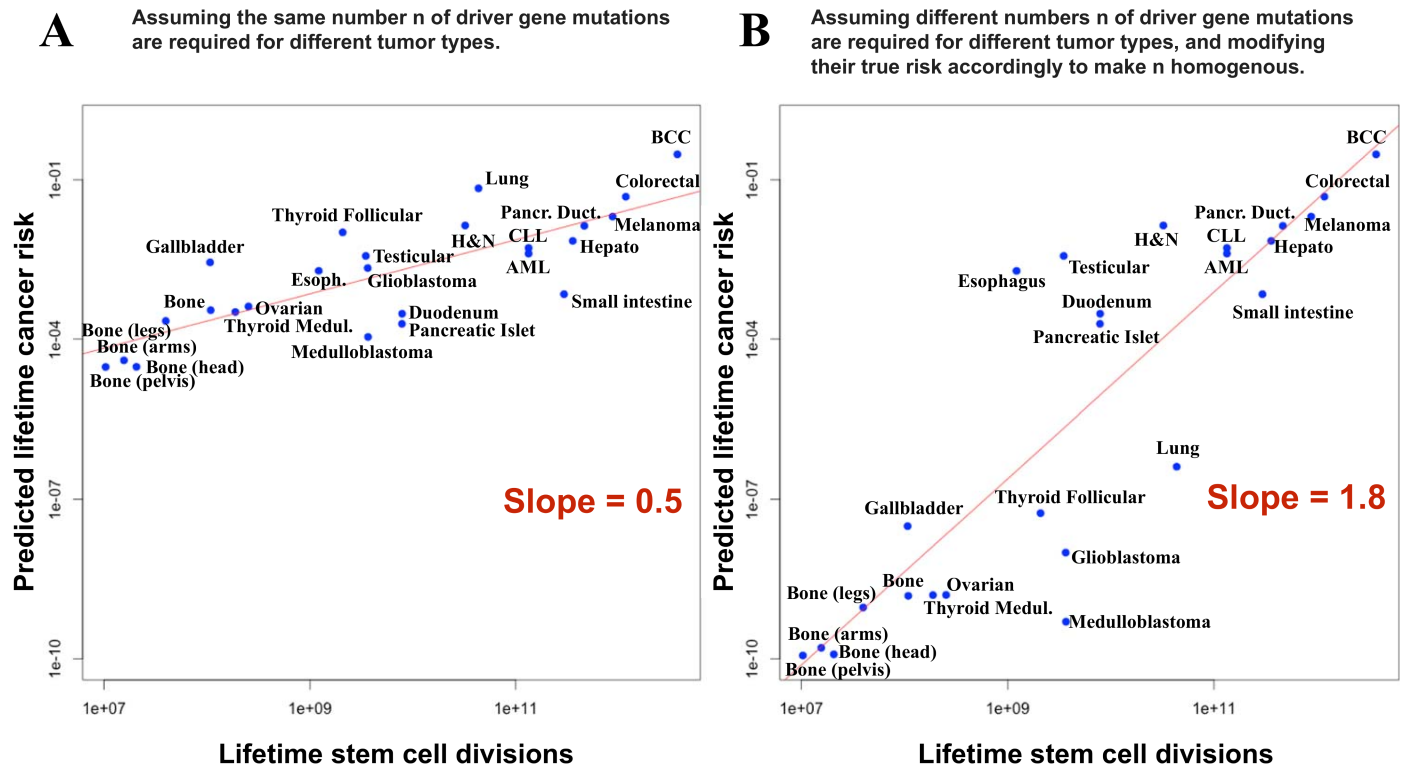


Fig 1. Dependence of the slope on the degree of heterogeneity in the number of required driver genes across tissues. (A) Regression line (red), for predicted cancer risk if it is assumed that all cancer types require exactly the same number of driver gene mutations, n , as assumed by Little et al. [1]. The slope of the regression line is 0.5. (B) Regression line (red), for predicted cancer risk if it is assumed that different cancer types require different numbers n of driver gene mutations (two in bone, ovarian, thyroid, gallbladder, brain, and lung cancer, and three drivers in all other cancer types.) To make the data points homogeneous in n , in order to regress on them, their risk was modified according to an Armitage and Doll model, for simplicity, with a mutation rate $u = 5 \times 10^{-7}$ per gene per cell division. All other variables were identical to those in A. The slope of the regression line is 1.8.

<https://doi.org/10.1371/journal.pone.0175535.g001>

in liver cancers than in other cancers. The lack of correlation between ERS and radiation- or smoking-induced cancer risk is expected from the data published in [2].

On the other hand, the lack of correlation found by Little et al. [1] is a relevant and informative result. It provides further evidence that substantially strengthens our preliminary analyses on the effects of atomic radiation [7, 10] and on the effects of stem cell numbers on the proportion of cancer cases explained by environmental factors [10]. In these studies, we showed a lack of correlation between the effects of the atomic bomb on cancer risk and the number of stem cell divisions in that tissue. Given that radiation and smoking are the strongest and most pervasive environmental carcinogens known today, the results of Little et al. provide important, independent evidence that environmental factors do not substantially contribute to the correlation discovered in [2].

Conclusions

There is also an important philosophical point that the paper by Little et al., and our response to it, brings up. The main objective of our paper [2] was to document a correlation between the lifetime number of stem cell divisions and cancer incidence among various tissue types. This correlation, graphically depicted in Fig 1 of [2] for the U.S. and similarly found across the rest of the world [10], is not based on any new assumptions or modeling. It simply points out a previously unrecognized association of data collected by investigators in two diverse fields:

epidemiology and developmental biology. Now that this correlation is documented, highly robust, and very statistically significant, it can be considered a reliable experimental result. If a particular theory, particularly one dependent on tenuous assumptions, leads to predictions that are inconsistent with the experimental result, it is more likely that the theory is flawed than the experimental result is flawed.

Author Contributions

Conceptualization: CT BV.

Investigation: CT BV.

Methodology: CT.

Writing – original draft: CT BV.

Writing – review & editing: CT BV.

References

1. Little MP, Hendry JH, Puskin JS. Lack of Correlation between Stem-Cell Proliferation and Radiation- or Smoking-Associated Cancer Risk. *PloS one*. 2016; 11(3):e0150335. <https://doi.org/10.1371/journal.pone.0150335> PMID: 27031507
2. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*. 2015; 347(6217):78–81. PMID: 25554788
3. <http://www.theguardian.com/science/grrlscientist/2015/jan/02/bad-luck-bad-journalism-and-cancer-rates>.
4. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British journal of cancer*. 1954; 8(1):1–12. PMID: 13172380
5. Tomasetti C, Marchionni L, Nowak MA, Parmigiani G, Vogelstein B. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112(1):118–23. <https://doi.org/10.1073/pnas.1421839112> PMID: 25535351
6. Durrett R, Moseley S. Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor Popul Biol*. 2010; 77(1):42–8. Epub 2009/11/10. <https://doi.org/10.1016/j.tpb.2009.10.008> PMID: 19896491
7. Tomasetti C, Vogelstein B. Technical Report. Musings on the theory that variation in cancer risk among tissues is explained by the number of divisions of the normal stem cells. *ArXiv.org* 2015. arXiv:1501.05035v4 [stat.AP].
8. Frank SA. *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. Princeton (NJ)2007.
9. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr., Kinzler KW. Cancer genome landscapes. *Science*. 2013; 339(6127):1546–58. <https://doi.org/10.1126/science.1235122> PMID: 23539594
10. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*. 2017; 355(6331):1330–4. <https://doi.org/10.1126/science.aaf9011> PMID: 28336671