City Research Online

City St George's, University of London

Citation: Noble, R., Kaltz, O., Nunney, L. & Hochberg, M. E. (2016). Overestimating the Role of Environment in Cancers. Cancer Prevention Research, 9(10), pp. 773-776. doi: 10.1158/1940-6207.capr-16-0126

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/24708/

Link to published version: https://doi.org/10.1158/1940-6207.capr-16-0126

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Overestimating the role of environment in cancers

Robert Noble[1], Oliver Kaltz[1], Leonard Nunney[2], Michael E Hochberg[1,3*]

[1] Institut des Sciences de l'Evolution, University of Montpellier, Montpellier, France

[2] Department of Biology, University of California, Riverside, CA 92521, USA

[3] Santa Fe Institute, Santa Fe, NM 87501, USA

* mhochber@univ-montp2.fr; +33 (0)4 67 14 36 67

The authors declare they have no conflicts of interest.

Running title: Overestimating the role of environment

Keywords: stem cells, carcinogenesis, causation, environment, intrinsic risk

Financial support: M.E. Hochberg acknowledges support from the Agence National de la Recherche (EvoCan ANR-13-BSV7-0003-01), the Institut multi-organismes cancer ('Physique Cancer' (CanEvolve PC201306), and the Institut National du Cancer (2014-1-PL BIO-12-IGR-1).

Article type: Commentary

1,750 words; 110 words in abstract; 1 table; 2 figures

Abstract:

In a recent article, Wu *et al.* (*Nature* 2016;**529**:43–47) review previous studies and present new estimates for the contribution of extrinsic factors to cancer development. The new estimates are generally close to 100%, even for bone and brain cancers that have no known associations with lifestyle and are typically not considered to be preventable. We find that the results of Wu and colleagues are incompatible with previous estimates derived from epidemiological and genetic data. We further argue that their methods are fundamentally flawed because they overlook important effects of tissue type on cancer risk. We therefore conclude that their results give a misleading view of cancer etiology and preventability.

Quantifying how much cancer risk may be avoided through changes in lifestyle or environment is important for guiding prevention efforts. In a recent article, Wu *et al.* [1] review previous studies and present new estimates for the contribution of extrinsic factors to cancer development. The authors define extrinsic factors as "environmental factors that affect mutagenesis rates (such as ultraviolet (UV) radiation, ionizing radiation and carcinogens)" that are "highly modifiable and thus preventable". They derive their new estimates by reanalyzing a previously published data set using two different methods: an "intrinsic risk line" (IRL) and a multistage model (MM). The estimates of extrinsic risk are generally close to 100%, even for bone and brain cancers that have no known associations with lifestyle and are typically not considered to be preventable. We question the reliability of these results for the following reasons: (i) the two methods produce contradictory estimates of extrinsic contributions for individual cancer types, which are also inconsistent with evidence from epidemiological studies and mutational signatures; (ii) the IRL method relies on the biologically implausible assumption that the intrinsic risk per stem cell division is a constant; and (iii) results from the MM model are undermined if one considers plausible, higher mutation rates. These considerations cast doubt on whether the new estimates of Wu *et al.* [1] accurately characterize cancer etiology or preventability.

If the IRL and MM methods of Wu *et al.* [1] were reliable, then we would expect them to produce consistent estimates when applied to the same cancer types. We might also expect them to be consistent with estimates derived from epidemiological studies (Table 1) and from an analysis of mutational signatures [2] cited by Wu *et al.* [1]. In fact, the data show that while the estimates derived from prior studies are correlated, the generally much higher IRL and MM estimates are uncorrelated both with each other and with the two other

data sets (Fig. 1). These inconsistencies have two potential implications. First, it is possible that the estimates derived from epidemiological studies and analysis of mutational signatures are unreliable, and the true contribution of extrinsic risk is close to 100% for almost all cancer types; however, given the significant correlation of these two independent data sets and our technical concerns over the other two methods (see below), this appears unlikely. Second, at least one, and potentially both, of the methods proposed by Wu *et al.* [1] are unreliable. The latter possibility suggests that a careful examination of the assumptions underlying the two novel methods is warranted.

For their IRL analysis, Wu *et al.* [1] used simple linear regression to estimate how cancer risk scales with the lifetime number of stem cell divisions across diverse tissue types, based on the data set of Tomasetti & Vogelstein [3]. Underlying their method (and prior analysis by Tomasetti & Vogelstein [3]) is the assumption that the intrinsic risk of a given cancer type over a lifetime can be expressed as

$$\text{risk per stem cell division} \times \text{lifetime number of stem cell divisions (lscd)}. \quad (1)$$

This equation is not an exact description (because risk must eventually saturate at 100%), but it is a reasonable approximation when the risk per stem cell division is much less than 1 and the lifetime risk is less than 10% (which holds for almost all of the data set). It follows that if the risk per stem cell division is the same for a subset of cancer types, then the lifetime risk of each type will simply be proportional to the lscd of the affected tissue. However, in any multistage model of carcinogenesis (i.e. any model requiring more than one mutational hit) the risk per stem cell division is not a constant, since it depends on the extent to which cancer-promoting mutations have accumulated. Given that obtaining a

cancer is a rare event, then a simple multistage model requiring M mutational hits predicts that [4]

$$\text{intrinsic risk} = C(uK)^M, \quad (2)$$

where C is the number of stem cells in the tissue, u is the somatic mutation rate, and K is the stem cell division rate. Equation 2 can be rewritten as

$$\text{intrinsic risk} = \text{lscd} \times (uK)^M/K, \quad (3)$$

so that the risk per division in Equation 1 is replaced by $(uK)^M$/K, which is the long-term average risk per division. If M is constant across a subset of tissues and they undergo division at similar rates, then this risk scales linearly with size. For example, since leg bones contain approximately twice as many cells as arm bones, we would expect osteosarcoma to be about twice as common in legs as in arms. Equation 3 implies that for such a subset,

$$\log(\text{lifetime intrinsic risk}) = \log(\text{lscd}) + \log(\text{long-term average risk/division}). \quad (4)$$

Therefore, assuming the long-term average risk is constant, we would predict an approximately linear relationship between log(lifetime intrinsic risk) and log(lscd) with slope approximately equal to 1. Any deviations above this line would represent the contributions of extrinsic factors. However, if extrinsic factors may affect all of the observed risks then how can we know where to draw the line? One approach is to position the line so that it intercepts the minimum observed risk, in which case deviations from the line correspond to minimum estimates of the contribution of extrinsic factors. What is essential is that, wherever the line is drawn, its slope must be equal to 1, so that it represents the assumed linear relationship between intrinsic cancer risk and lscd. The Wu *et al.* [1] IRL method

follows the above procedure except that it estimates the contributions of extrinsic factors to cancer risk by measuring deviations from a regression line that has slope much less than one (Fig 2A). The latter method has no theoretical basis and its results are therefore unreliable.

A more fundamental problem is that the IRL method (similar to that in Tomasetti & Vogelstein [3]) relies on the assumption that the long-term average risk per division is the same among all cancer types. This assumption is false, since the risk varies considerably depending on tissue organization [5], number of mutational "hits" necessary to initiate cancer (*M*), the rate of stem cell division, and other intrinsic biological factors [4]. As noted above, whenever the number of mutational hits (*M*) required to induce cancer is greater than one, then cancer risk scales linearly with cell number, but is more strongly related to the number of divisions (to the power *M*) [4]. As a result, any simple regression method cannot discriminate extrinsic from intrinsic effects across diverse cancer types, and estimates derived in this way are overestimates, since variation interpreted as extrinsic risk (Fig 2A) can disappear when the data are subdivided into comparable groups (Fig. 2B), as pointed out in recent critiques of Tomasetti & Vogelstein's analysis [6-10]. Unreliability of the IRL method would explain the apparently anomalous results of Wu *et al.* [1], such as the inconsistent estimates for bone cancer (for example, approximately 0% estimated extrinsic contribution to the risk of arm, head and pelvis osteosarcoma but >77% for leg osteosarcoma).

In their second analysis, Wu *et al.* [1] estimated extrinsic contributions to cancer risk using a mathematical model of multistage carcinogenesis and the same previously published data set [3]. They concluded that the risks predicted by the multistage model are much lower than observed cancer rates, and therefore extrinsic factors must make a very large

contribution. This analysis is based on two assumptions: first that all tissues require the same number of driver mutations to initiate cancer (i.e. *M* is constant across all tissues), and second that the effective somatic mutation rate is between 10^{-10} and 10^{-6} per stem cell division. The first assumption is known to be false (e.g. comparing retinoblastoma, where *M* = 2, to the cancers of larger tissues, where *M* is larger [11]). At first sight, the second assumption appears plausible; however, higher mutation rates may be more realistic given that (i) cancer-promoting mutations per driver gene are generally considered to be at the high end of this range (as in recent literature [12, 13]), and (ii) some driver mutations result in a burst of cell proliferation leading to clonal expansion [14]. Whereas more complex models are required to examine the effect of clonal expansion in detail (for example [15-17]), it can more simply be approximated as an increase in the "effective" mutation rate [4]. With these concerns in mind, Nunney & Muir [7] found the same data set to be consistent with a similar multistage model provided that *M* was allowed to vary among tissue types and was estimated from the data. As expected, their analysis indicated a higher effective driver mutation rate of about 8×10^{-6} per stem cell division. Of course, this higher mutation rate could be partly the result of extrinsic factors, but in its present form the data does not permit us to distinguish between intrinsic and extrinsic effects. As a result, we cannot assign percent causality between these two sources, or even exclude the possibility that the variation in risk is largely explained by intrinsic differences between tissues, such as the number of mutational hits required to initiate cancer [7].

Wu *et al.*'s [1] attribution of nearly all cancer risk to "highly modifiable and thus preventable extrinsic factors" demands scrutiny because of its implications for prevention strategies. We agree that epidemiological studies [18, 19] and analyses of mutational signatures [2] suggest

that extrinsic factors contribute substantially to the risk of many cancer types. We also share concerns about the methodology underlying the recent very different estimates of Tomasetti & Vogelstein [3]. However, the methods of Wu *et al.* [1] similarly do not account for the effects of intrinsic, tissue-dependent factors that could explain much of the observed variation in cancer risk, and the resulting estimates are therefore likewise questionable. Both Wu *et al.* [1] and Tomasetti & Vogelstein [3] also do not consider the realistic possibility that within a given cell or tumor, some mutational "hits" may be ascribed to the environment, whereas others may be due to bad luck, and extrinsic factors may further influence the selective value of mutations by changing the microenvironment [17]. Taken together this suggests that estimating relative contributions, even for particular cancer types, will be a daunting challenge. We argue that beyond an improved knowledge of transformation pathways towards specific cancers, progress will require a greater understanding of how mutational processes undermine evolved cancer prevention and suppression in different cell types, and how they interact with tumour microenvironments and our external environments and life-styles.

1. Wu S, Powers S, Zhu W, Hannun YA. Substantial contribution of extrinsic risk factors to cancer development. *Nature* 2016;**529**:43–7.

2. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, *et al.* Signatures of mutational processes in human cancer. *Nature* 2013;**500**:415–21.

3. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 2015;**347**:78–81.

4. Nunney L. Lineage selection and the evolution of multistage carcinogenesis. *Proc. Biol. Sci.* 1999;**266**:493–8.

5. Michor F, Frank SA, May RM, Iwasa Y, Nowak MA. Somatic selection for and against cancer. *J. Theor. Biol.* 2003;**225:**377–82.

6. Noble R, Kaltz O, Hochberg ME. Peto's paradox and human cancers. *Philos. Trans. R. Soc. B Biol. Sci.* 2015;**370**:20150104.

7. Nunney L, Muir B. Peto's paradox and the hallmarks of cancer: constructing an evolutionary framework for understanding the incidence of cancer. *Philos. Trans. R. Soc. B Biol. Sci.* 2015;**370**:20150161.

8. Rozhok AI, Wahl GM, DeGregori J. A critical examination of the 'bad luck' explanation of cancer risk. *Cancer Prev. Res.* 2015;**8**:762–4.

9. Little MP, Hendry JH, Puskin JS. Lack of Correlation between Stem-Cell Proliferation and Radiation- or Smoking-Associated Cancer Risk. *PLoS One* 2016;**11:**e0150335.

10. Belpomme D, Irigaray P. Replicative random mutations as an unproven cause of cancer: A technical comment. *Mol. Clin. Oncol.* 2016;**4:**497–9.

11. Berger AH, Knudson AG, Pandolfi PPA continuum model for tumour suppression. *Nature* 2011;**476**:163-9.

12. Bozic I, Nowak MA. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *Proc. Natl. Acad. Sci. USA.* 2014;**111**:15964–8.

13. Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* 2015;**525**:261–4.

14. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;**194**:23-8.

15. Luebeck EG, Moolgavkar SH. Multistage carcinogenesis and the incidence of colorectal cancer. *Proc. Natl. Acad. Sci.* 2002;**99**:15095–100.

16. Zhang X, Simon R. Estimating the number of rate limiting genomic changes for human breast cancer. *Breast Cancer Res. Treat* 2005;**91**:121–4.

17. Rozhok AI, DeGregori J. Toward an evolutionary model of cancer: Considering the mechanisms that govern the fate of somatic mutations. *Proc. Natl. Acad. Sci.* 2015;**112**:8914–21.

18. Doll R, Peto R. The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today. *J. Natl. Cancer Inst.* 1981;**66**:1191–308.

19. Parkin DM, Boyd L, Walker LC. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br. J. Cancer* 2011; **105** Suppl 2:S77–S81.

20. Savage SA, Mirabello L. Using Epidemiology and Genomics to Understand Osteosarcoma Etiology. *Sarcoma* 2011**;**548151.

21. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, *et al.* The epidemiology of glioma in adults: A state of the science review. *Neuro. Oncol.* 2014;**16**:896–913.

22. Gamer M, Lemon J, Fellows I, Sing P. irr: Various coefficients of interrater reliability and agreement. Version 0.84 [software]. 2012 Jul 16. Available from: http://CRAN.R-project.org/package=irr

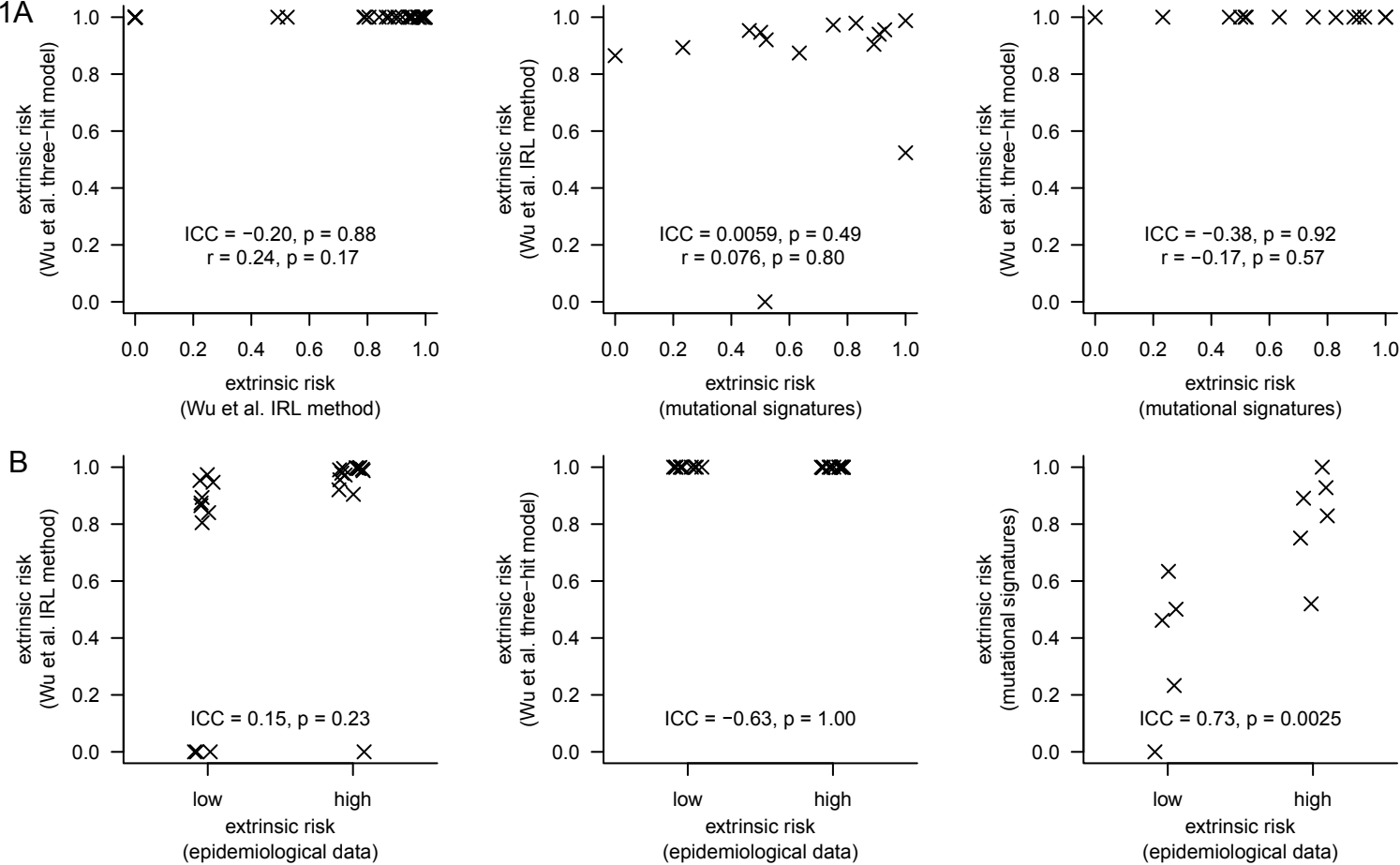
Figure 1 | Lack of consistency between different sets of estimates of extrinsic cancer risks. **A,** Comparison between results of the two new analyses by Wu *et al.* [1], and comparisons of these results with estimates derived from mutational signatures. **B,** Comparisons with estimates derived from epidemiological data, categorized as "low" (less than 50%) or "high" (more than 50%). Each point corresponds to a cancer type. ICC is the intraclass correlation coefficient; r is Pearson correlation coefficient. Epidemiological data is from Extended Data Table 2 in Wu *et al.* [1] and recent reviews [19-21] (see Table 1); estimates from mutational signatures are from Extended Data Table 3 in Wu *et al.* [1], derived from Alexandrov *et al.* [2]. Wu *et al.* [1] also present estimates derived from two-hit and four-hit models of carcinogenesis, which are likewise uncorrelated with the estimates from the IRL method, epidemiological data, and analysis of mutational signatures (data not shown). ICC values were calculated using the irr package for the R programming language [22]. In the calculation of ICC values, the "low" and "high" categories were assigned values of 0.25 and 0.75, respectively.

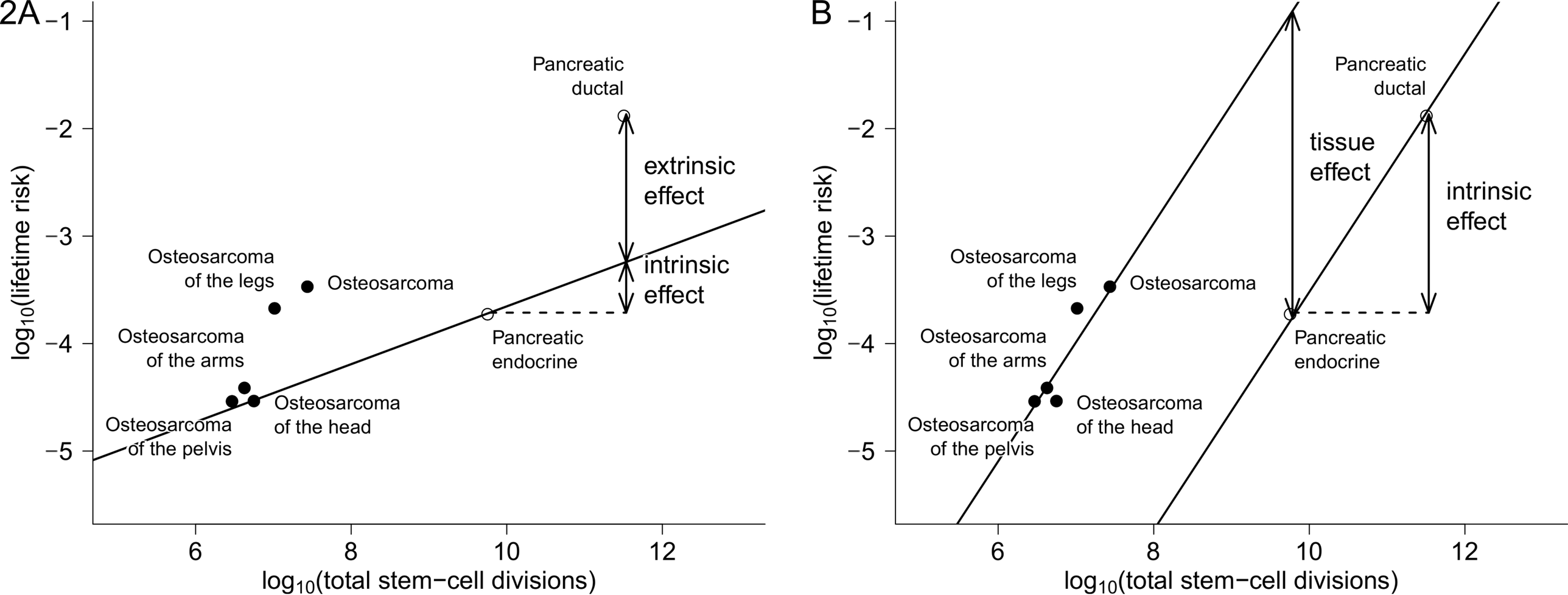
Figure 2 | Alternative models for explaining variation in cancer risk. For clarity, only part of the Tomasetti & Vogelstein [3] data set is shown. **A,** Wu *et al.* [1] assume that a single regression line can be used to quantify intrinsic cancer risk, and any additional risk must be

due to extrinsic factors. **B,** If we allow that the intrinsic cancer risk per stem cell division varies between tissues, then it is invalid to draw only a single regression line through the data for diverse cancer types. Instead, separate regression lines must be drawn for each cancer type or for sets of cancer types with similar features (such as anatomical site [6]). The vertical displacement of these regression lines corresponds to the tissue effect, which may be due to a combination of intrinsic factors (such as the number of mutational "hits" necessary to initiate cancer) and extrinsic factors (such as diet or smoking).

Cancer types with high (>50%) extrinsic risk	**Reference(s)**
Basal cell carcinoma	1
Colorectal adenocarcinoma	1, 19
Colorectal adenocarcinoma with FAP	*
Colorectal adenocarcinoma with Lynch syndrome	*
Duodenum adenocarcinoma with FAP	*
Esophageal squamous cell carcinoma	1, 19
Head & neck squamous cell carcinoma	1, 19
Head & neck squamous cell carcinoma with HPV-16	*
Hepatocellular carcinoma	1
Hepatocellular carcinoma with HCV	*
Lung adenocarcinoma (smokers)	*
Melanoma	1, 19
Small intestine adenocarcinoma	1
Thyroid papillary/follicular carcinoma	1
Cancer types with low (<50%) extrinsic risk	**Reference(s)**
Acute myeloid leukemia	19
Chronic lymphocytic leukemia	19
Gallbladder non papillary adenocarcinoma	19
Glioblastoma	21
Osteosarcoma	20
Osteosarcoma of the arms	20
Osteosarcoma of the head	20
Osteosarcoma of the legs	20
Osteosarcoma of the pelvis	20
Ovarian germ cell	19
Pancreatic ductal adenocarcinoma (acinar)	19

Table 1 | Categorization of cancer types by relative contributions of extrinsic factors to risk, according to epidemiological data. Risks for cancer types labeled * apply to subpopulations and are more than double the risks among the general population.


1A
extrinsic risk
(Wu et al. three−hit model)
ICC = −0.20, p = 0.88
r = 0.24, p = 0.17
extrinsic risk
(Wu et al. IRL method)
extrinsic risk
(Wu et al. IRL method)
ICC = 0.0059, p = 0.49
r = 0.076, p = 0.80
extrinsic risk
(mutational signatures)
extrinsic risk
(Wu et al. three−hit model)
ICC = −0.38, p = 0.92
r = −0.17, p = 0.57
extrinsic risk
(mutational signatures)
B
extrinsic risk
(Wu et al. IRL method)
ICC = 0.15, p = 0.23
low
high
extrinsic risk
(epidemiological data)
extrinsic risk
(Wu et al. three−hit model)
ICC = −0.63, p = 1.00
low
high
extrinsic risk
(epidemiological data)
extrinsic risk
(mutational signatures)
ICC = 0.73, p = 0.0025
low
high
extrinsic risk
(epidemiological data)


2A
B
−1
−2
−3
−4
−5
log10(lifetime risk)
6
8
10
12
log10(total stem−cell divisions)
Pancreatic ductal
extrinsic effect
intrinsic effect
Osteosarcoma of the legs
Osteosarcoma
Osteosarcoma of the arms
Pancreatic endocrine
Osteosarcoma of the head
Osteosarcoma of the pelvis
tissue effect
intrinsic effect