



Published in final edited form as:

Science. 2015 January 2; 347(6217): 78–81. doi:10.1126/science.1260825.

Variation in cancer risk among tissues can be explained by the number of stem cell divisions

Cristian Tomasetti^{1,*} and Bert Vogelstein^{2,*}

¹Division of Biostatistics and Bioinformatics, Department of Oncology, Sidney Kimmel Cancer Center, Johns Hopkins University School of Medicine and Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 550 North Broadway, Baltimore, MD 21205, USA

²Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, 1650 Orleans Street, Baltimore, MD 21205, USA

Abstract

Some tissue types give rise to human cancers millions of times more often than other tissue types. Although this has been recognized for more than a century, it has never been explained. Here, we show that the lifetime risk of cancers of many different types is strongly correlated (0.81) with the total number of divisions of the normal self-renewing cells maintaining that tissue's homeostasis. These results suggest that only a third of the variation in cancer risk among tissues is attributable to environmental factors or inherited predispositions. The majority is due to “bad luck,” that is, random mutations arising during DNA replication in normal, noncancerous stem cells. This is important not only for understanding the disease but also for designing strategies to limit the mortality it causes.

Extreme variation in cancer incidence across different tissues is well known; for example, the lifetime risk of being diagnosed with cancer is 6.9% for lung, 1.08% for thyroid, 0.6% for brain and the rest of the nervous system, 0.003% for pelvic bone and 0.00072% for laryngeal cartilage (1-3). Some of these differences are associated with well-known risk factors such as smoking, alcohol use, ultraviolet light, or human papilloma virus (HPV) (4,5), but this applies only to specific populations exposed to potent mutagens or viruses. And such exposures cannot explain why cancer risk in tissues within the alimentary tract can differ by as much as a factor of 24 [esophagus (0.51%), large intestine (4.82%), small intestine (0.20%), and stomach (0.86%)] (3). Moreover, cancers of the small intestinal epithelium are three times less common than brain tumors (3), even though small intestinal epithelial cells are exposed to much higher levels of environmental mutagens than are cells within the brain, which are protected by the blood-brain barrier.

Another well-studied contributor to cancer is inherited genetic variation. However, only 5 to 10% of cancers have a heritable component (6-8), and even when hereditary factors in predisposed individuals can be identified, the way in which these factors contribute to

*Corresponding author. ctomasetti@jhu.edu (C.T.); vogelbe@jhmi.edu (B.V.).

Authors' contributions: C.T. formulated the hypothesis. C.T. and B.V. designed the research. C.T. provided mathematical and statistical analysis. C.T. and B.V. performed research. C.T. and B.V. wrote the paper.

differences in cancer incidences among different organs is obscure. For example, the same, inherited mutant *APC* gene is responsible for both the predisposition to colorectal and small intestinal cancers in familial adenomatous polyposis (FAP) syndrome patients, yet cancers occur much more commonly in the large intestine than in the small intestine of these individuals.

If hereditary and environmental factors cannot fully explain the differences in organ-specific cancer risk, how else can these differences be explained? Here, we consider a third factor: the stochastic effects associated with the lifetime number of stem cell divisions within each tissue. In cancer epidemiology, the term “environmental” is generally used to denote anything not hereditary, and the stochastic processes involved in the development and homeostasis of tissues are grouped with external environmental influences in an uninformative way. We show here that the stochastic effects of DNA replication can be numerically estimated and distinguished from external environmental factors. Moreover, we show that these stochastic influences are in fact the major contributors to cancer overall, often more important than either hereditary or external environmental factors.

That cancer is largely the result of acquired genetic and epigenetic changes is based on the somatic mutation theory of cancer (9-13) and has been solidified by genome-wide analyses (14-16). The idea that the number of cells in a tissue and their cumulative number of divisions may be related to cancer risk, making them more vulnerable to carcinogenic factors, has been proposed but is controversial (17-19). Other insightful ideas relating to the nature of the factors underlying neoplasia are reviewed in (20-22).

The concept underlying the current work is that many genomic changes occur simply by chance during DNA replication rather than as a result of carcinogenic factors. Since the endogenous mutation rate of all human cell types appears to be nearly identical (23,24), this concept predicts that there should be a strong, quantitative correlation between the lifetime number of divisions among a particular class of cells within each organ (stem cells) and the lifetime risk of cancer arising in that organ.

To test this prediction, we attempted to identify tissues in which the number and dynamics of stem cells have been described. Most cells in tissues are partially or fully differentiated cells that are typically short-lived and unlikely to be able to initiate a tumor. Only the stem cells—those that can self-renew and are responsible for the development and maintenance of the tissue's architecture—have this capacity. Stem cells often make up a small proportion of the total number of cells in a tissue and, until recently, their nature, number, and hierarchical division patterns were not known (25-28). Tissues were not included in our analysis if the requisite parameters were not found in the literature or if their estimation was difficult to derive.

Through an extensive literature search, we identified 31 tissue types in which stem cells had been quantitatively assessed (see the supplementary materials). We then plotted the total number of stem cell divisions during the average lifetime of a human on the x axis and the lifetime risk for cancer of that tissue type on the y axis (Fig. 1) (table S1). The lifetime risk in the United States for all included cancer types has been evaluated in detail, such as in the

Surveillance, Epidemiology, and End Results (SEER) database (3). The correlation between these two very different parameters—number of stem cell divisions and lifetime risk—was striking, with a highly positive correlation (Spearman's $\rho = 0.81$; $P < 3.5 \times 10^{-8}$) (Fig. 1). Pearson's linear correlation 0.804 [0.63 to 0.90; 95% confidence interval (CI)] was equivalently significant ($P < 5.15 \times 10^{-8}$). One of the most impressive features of this correlation was that it extended across five orders of magnitude, thereby applying to cancers with enormous differences in incidence. No other environmental or inherited factors are known to be correlated in this way across tumor types. Moreover, these correlations were extremely robust; when the parameters used to construct Fig. 1 were varied over a broad range of plausible values, the tight correlation remained intact (see the supplementary materials).

A linear correlation equal to 0.804 suggests that 65% (39% to 81%; 95% CI) of the differences in cancer risk among different tissues can be explained by the total number of stem cell divisions in those tissues. Thus, the stochastic effects of DNA replication appear to be the major contributor to cancer in humans.

We next attempted to distinguish the effects of this stochastic, replicative component from other causative factors – that is, those due to the external environment and inherited mutations. For this purpose, we defined an “extra risk score” (ERS) as the product of the lifetime risk and the total number of stem cell divisions (\log_{10} values). Machine learning methods were employed to classify tumors based only on this score (see the supplementary materials). With the number of clusters set equal to two, the tumors were classified in an unsupervised manner into one cluster with high ERS (9 tumor types) and another with low ERS (22 tumor types) (Fig. 2).

The ERS provides a test of the approach described in this work. If the ERS for a tissue type is high – that is, if there is a high cancer risk of that tissue type relative to its number of stem cell divisions – then one would expect that environmental or inherited factors would play a relatively more important role in that cancer's risk (see the supplementary materials for a detailed explanation). It was therefore notable that the tumors with relatively high ERS were those with known links to specific environmental or hereditary risk factors (Fig. 2, blue cluster). We refer to the tumors with relatively high ERS as D-tumors (D for deterministic; blue cluster in Fig. 2) because deterministic factors such as environmental mutagens or hereditary predispositions strongly affect their risk. We refer to tumors with relatively low ERS as R-tumors (R for replicative; green cluster in Fig. 2) because stochastic factors, presumably related to errors during DNA replication, most strongly appear to affect their risk.

The incorporation of a replicative component as a third, quantitative determinant of cancer risk forces rethinking of our notions of cancer causation. The contribution of the classic determinants (external environment and heredity) to R-tumors is minimal (Fig. 1). Even for D-tumors, however, replicative effects are essential, and environmental and hereditary effects simply add to them. For example, patients with FAP are ~ 30 times as likely to develop colorectal cancer than duodenal cancer (Fig. 1). Our data suggest that this is because there are ~ 150 times as many stem cell divisions in the colon as in the duodenum.

The lifetime risk of colorectal cancer would be very low, even in the presence of an underlying *APC* gene mutation, if colonic epithelial stem cells were not constantly dividing. A related point is that mice with inherited *APC* mutations display the opposite pattern: Small intestinal tumors are more common than large intestinal tumors. Our analysis provides a plausible explanation for this striking difference between mice and men; namely, in mice the small intestine undergoes more stem cell divisions than the large intestine (see the supplementary materials for the estimates). Another example is provided by melanocytes and basal epidermal cells of the skin, which are both exposed to the same carcinogen (ultraviolet light) at the identical dose, yet melanomas are much less common than basal cell carcinomas. Our data suggest that this difference is attributable to the fact that basal epidermal cells undergo a higher number of divisions than melanocytes (see the supplementary materials for the estimates). The total number of stem cells in an organ and their proliferation rate may of course be influenced by genetic and environmental factors such as those that affect height or weight.

In formal terms, our analyses show only that there is some stochastic factor related to stem cell division that seems to play a major role in cancer risk. This situation is analogous to that of the classic studies of Nordling and of Armitage and Doll (10,29). These investigators showed that the relationship between age and the incidence of cancer was exponential, suggesting that many cellular changes, or stages, were required for carcinogenesis. On the basis of research since that time, these events are now interpreted as somatic mutations. Similarly, we interpret the stochastic factor underlying the importance of stem cell divisions to be somatic mutations. This interpretation is buttressed by the large number of somatic mutations known to exist in cancer cells (14-16,30).

Our analysis shows that stochastic effects associated with DNA replication contribute in a substantial way to human cancer incidence in the United States. These results could have important public health implications. One of the most promising avenues for reducing cancer deaths is through prevention. How successful can such approaches be? The maximum fraction of tumors that are preventable through primary prevention (such as vaccines against infectious agents or altered lifestyles) may be evaluated from their ERS. For nonhereditary D-tumors, this fraction is high and primary prevention could make a major impact (31). Secondary prevention, obtainable in principle through early detection, could further reduce nonhereditary D-tumor-related deaths and is also instrumental for reducing hereditary D-tumor-related deaths. For R-tumors, primary prevention measures are not likely to be as effective, and secondary prevention should be the major focus.

Supplementary Material

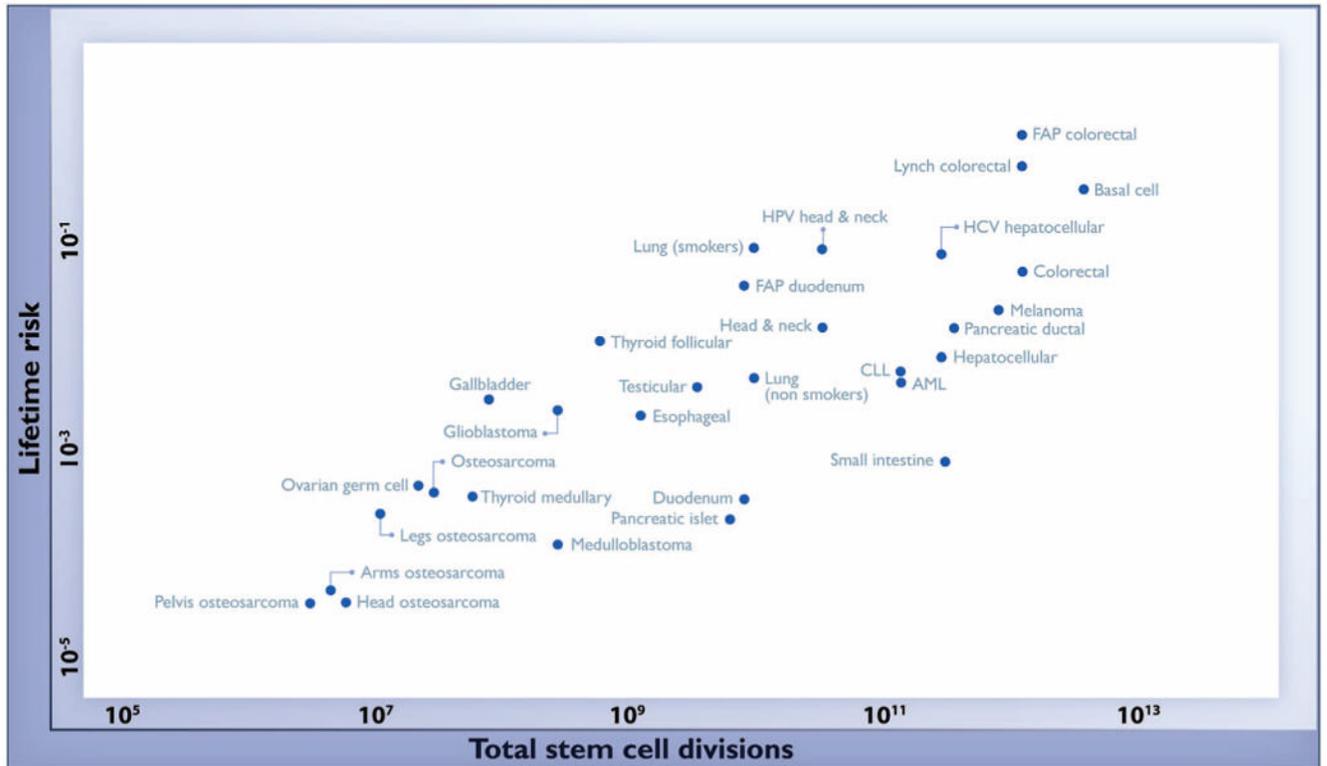
Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank E. Cook for artwork. This work was supported by the 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.

References and Notes

1. Dubai PM, et al. *Laryngoscope*. 2014; 124:1877–1881. [PubMed: 24474667]
2. Jaffe, N. *Pediatric and Adolescent Osteosarcoma*. Springer; New York: 2009.
3. National Cancer Institute, Surveillance, Epidemiology, and End Results Program. www.seer.cancer.gov
4. Danaei G, et al. *Lancet*. 2005; 366:1784–1793. [PubMed: 16298215]
5. Centers for Disease Control and Prevention. www.cdc.gov
6. Fearon ER. *Science*. 1997; 278:1043–1050. [PubMed: 9353177]
7. Lichtenstein P. *N Engl J Med*. 2000; 343:78–85. [PubMed: 10891514]
8. American Cancer Society. www.cancer.org
9. Armitage P, Doll R. *Br J Cancer*. 1957; 11:161–169. [PubMed: 13460138]
10. Armitage P, Doll R. *Br J Cancer*. 1954; 8:1–12. [PubMed: 13172380]
11. Boveri, T. *Zur Frage der Entstehung Maligner Tumoren*. G. Fischer; Jena, Germany: 1914.
12. Fearon ER, Vogelstein B. *Cell*. 1990; 61:759–767. [PubMed: 2188735]
13. Knudson AG Jr. *Proc Natl Acad Sci USA*. 1971; 68:820–823. [PubMed: 5279523]
14. Garraway LA, Lander ES. *Cell*. 2013; 153:17–37. [PubMed: 23540688]
15. Stratton MR, Campbell PJ, Futreal PA. *Nature*. 2009; 458:719–724. [PubMed: 19360079]
16. Vogelstein B, et al. *Science*. 2013; 339:1546–1558. [PubMed: 23539594]
17. Albanes D, Winick M. *J Natl Cancer Inst*. 1988; 80:772–775. [PubMed: 3385783]
18. Tomatis L. International Agency for Research on Cancer. *Environ Health Perspect*. 1993; 101(suppl 5):149–151. [PubMed: 8013403]
19. Ward JM, Uno H, Kurata Y, Weghorst CM, Jang JJ. *Environ Health Perspect*. 1993; 101(suppl 5):125–135. [PubMed: 8013399]
20. Sell S. *Environ Health Perspect*. 1993; 101(suppl 5):15–26. [PubMed: 7516873]
21. Frank, SA. *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. Princeton Univ. Press; Princeton, NJ: 2007. chap 4
22. Baker SG, Cappuccio A, Potter JD. *J Clin Oncol*. 2010; 28:3215–3218. [PubMed: 20547997]
23. Lynch M. *Proc Natl Acad Sci USA*. 2010; 107:961–968. [PubMed: 20080596]
24. Tomasetti C, Vogelstein B, Parmigiani G. *Proc Natl Acad Sci USA*. 2013; 110:1999–2004. [PubMed: 23345422]
25. Blanpain C, Fuchs E. *Nat Rev Mol Cell Biol*. 2009; 10:207–217. [PubMed: 19209183]
26. Booth C, Potten CS. *J Clin Invest*. 2000; 105:1493–1499. [PubMed: 10841502]
27. Reya T, Morrison SJ, Clarke MF, Weissman IL. *Nature*. 2001; 414:105–111. [PubMed: 11689955]
28. Snippet HJ, et al. *Cell*. 2010; 143:134–144. [PubMed: 20887898]
29. Nordling CO. *Br J Cancer*. 1953; 7:68–72. [PubMed: 13051507]
30. Kandoth C, et al. *Nature*. 2013; 502:333–339. [PubMed: 24132290]
31. Colditz GA, Wolin KY, Gehlert S. *Sci Transl Med*. 2012; 4:127rv4.



FAP = Familial Adenomatous Polyposis ♦ HCV = Hepatitis C virus ♦ HPV = Human papillomavirus ♦ CLL = Chronic lymphocytic leukemia ♦ AML = Acute myeloid leukemia

Fig. 1. The relationship between the number of stem cell divisions in the lifetime of a given tissue and the lifetime risk of cancer in that tissue

Values are from table S1, the derivation of which is discussed in the supplementary materials.

Clustering of cancer types

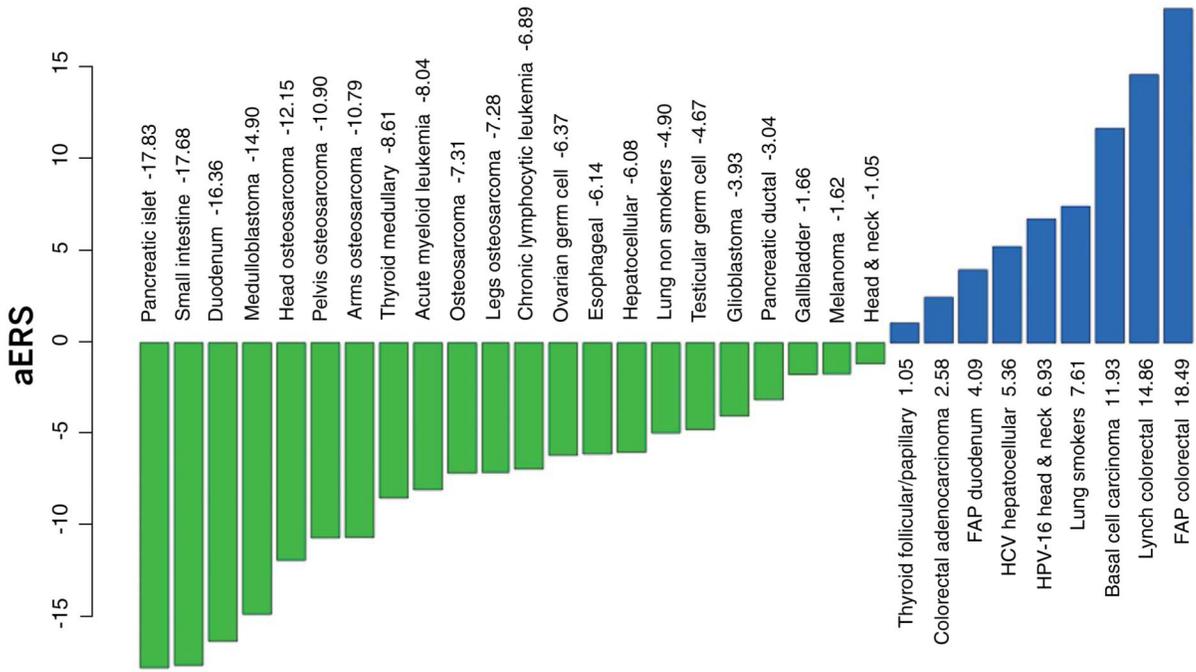


Fig. 2. Stochastic (replicative) factors versus environmental and inherited factors: R-tumor versus D-tumor classification

The adjusted ERS (aERS) is indicated next to the name of each cancer type. R-tumors (green) have negative aERS and appear to be mainly due to stochastic effects associated with DNA replication of the tissues' stem cells, whereas D-tumors (blue) have positive aERS. Importantly, although the aERS was calculated without any knowledge of the influence of environmental or inherited factors, tumors with high aERS proved to be precisely those known to be associated with these factors. For details of the derivation of aERS, see the supplementary materials.