leads to Cdc48 recruitment for extraction and degradation of the incomplete translation product. Rqc2p, through specific binding to Ala(IGC) and Thr(IGU) tRNAs, directs the template-free and 40*S*-free elongation of the incomplete translation product with CAT tails. CAT tails induce a heat shock response through a mechanism that is yet to be determined.

Hypomorphic mutations in the mammalian homolog of *LTN1* cause neurodegeneration in mice (*21*). Similarly, mice with mutations in a central nervous system–specific isoform of $tRNA^{Arg}$ and GTPBP2, a homolog of yeast Hbs1 which works with PELOTA/Dom34 to dissociate stalled 80S ribosomes, suffer from neurodegeneration (*22*). These observations reveal the consequences that ribosome stalls impose on the cellular economy. Eubacteria rescue stalled ribosomes with the transfer-messenger RNA (tmRNA)–SmpB system, which appends nascent chains with a unique C-terminal tag that targets the incomplete protein product for proteolysis (*23*). The mechanisms used by eukaryotes, which lack tmRNA, to recognize and rescue stalled ribosomes and their incomplete translation products have been unclear. The RQC—and Rqc2p's CAT tail tagging mechanism in particular—bear both similarities and contrasts to the tmRNA *trans*-translation system. The evolutionary convergence upon distinct mechanisms for extending incomplete nascent chains at the C terminus argues for their importance in maintaining proteostasis. One advantage of tagging stalled chains is that it may distinguish them from normal translation products and facilitate their removal from the protein pool. An alternate, not mutually exclusive, possibility is that the extension serves to test the functional integrity of large ribosomal subunits, so that the cell can detect and dispose of defective large subunits that induce stalling.

REFERENCES AND NOTES

1. O. Brandman *et al.*, *Cell* **151**, 1042–1054 (2012).
2. Q. Defenouillère *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1221724110 (2013).
3. R. Verma, R. S. Oania, N. J. Kolawa, R. J. Deshaies, *eLife* **2**, e00308 (2013).
4. S. Shao, K. von der Malsburg, R. S. Hegde, *Mol. Cell* **50**, 637–648 (2013).
5. G.-W. Li, D. Burkhardt, C. Gross, J. S. Weissman, *Cell* **157**, 624–635 (2014).
6. D. Lyumkis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1702–1707 (2013).
7. S. Shao, R. S. Hegde, *Mol. Cell* **55**, 880–890 (2014).
8. R. Lill, J. M. Robertson, W. Wintermeyer, *Biochemistry* **25**, 3245–3255 (1986).
9. G. E. Katibah *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12025–12030 (2014).
10. D. Chu, D. J. Barnes, T. von der Haar, *Nucleic Acids Res.* **39**, 6705–6714 (2011).
11. P. F. Agris, F. A. P. Vendeix, W. D. Graham, *J. Mol. Biol.* **366**, 1–13 (2007).
12. F. H. C. Crick, *J. Mol. Biol.* **19**, 548–555 (1966).
13. A. P. Gerber, W. Keller, *Science* **286**, 1146–1149 (1999).
14. E. Delannoy *et al.*, *Plant Cell* **21**, 2058–2071 (2009).
15. M. A. T. Rubio, F. L. Ragone, K. W. Gaston, M. Ibba, J. D. Alfonzo, *J. Biol. Chem.* **281**, 115–120 (2006).
16. R. E. Monro, *Nature* **223**, 903–905 (1969).
17. D. P. Letzring, K. M. Dean, E. J. Grayhack, *RNA* **16**, 2516–2528 (2010).
18. S. Ito-Harashima, K. Kuroha, T. Tatematsu, T. Inada, *Genes Dev.* **21**, 519–524 (2007).
19. K. Kobayashi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17575–17579 (2010).
20. A. M. Burroughs, L. Aravind, *RNA Biol.* **11**, 360–372 (2014).
21. J. Chu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2097–2103 (2009).
22. R. Ishimura *et al.*, *Science* **345**, 455–459 (2014).
23. S. D. Moore, R. T. Sauer, *Annu. Rev. Biochem.* **76**, 101–124 (2007).
24. A. Gerber, H. Grosjean, T. Melcher, W. Keller, *EMBO J.* **17**, 4780–4789 (1998).
25. K. W. Gaston *et al.*, *Nucleic Acids Res.* **35**, 6740–6749 (2007).
26. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, *Genome Res.* **14**, 1188–1190 (2004).

ACKNOWLEDGMENTS

Electron microscopy was performed at the University of Utah and the University of California. We thank D. Belnap (University of Utah) and M. Braunfeld (University of California, San Francisco) for supervision of the electron microscopes; A. Orendt and the Utah Center for High Performance Computing and the NSF Extreme Science and Engineering Discovery Environment consortium for computational support; D. Sidote (University of Texas at Austin) for help processing RNA-seq data; and D. Herschlag and P. Harbury for helpful comments. Amino acid analysis was performed by J. Shulze at the University of California, Davis Proteomics Core. Edman sequencing was performed at Stanford University's Protein and Nucleic Acid Facility by D. Winant. This work was supported by the Searle Scholars Program (A.F.); Stanford University (O.B.); NIH grants 1DP2GM110772-01 (A.F.), GM37949, and GM37951 (A.M.L.); the Center for RNA Systems Biology grants P50 GM102706 (J.S.W.) and U01 GM098254 (J.S.W.); and the Howard Hughes Medical Institute (J.S.W.). The authors declare no competing financial interests. The cryo-EM structures have been deposited at the Electron Microscopy Data Bank (accession codes 2811, 2812, 6169, 6170, 6171, 6172, 6176, and 6201).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/347/6217/75/suppl/DC1
Materials and Methods
Figs. S1 to S13
Table S1
References (*27*–*41*)

7 August 2014; accepted 14 November 2014
10.1126/science.1259724

CANCER ETIOLOGY

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

Cristian Tomasetti[1]* and Bert Vogelstein[2]*

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 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

[1]Division 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. [2]Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, 1650 Orleans Street, Baltimore, MD 21205, USA.
***Corresponding author. E-mail: ctomasetti@jhu.edu (C.T.); vogelbe@jhmi.edu (B.V.)**

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.

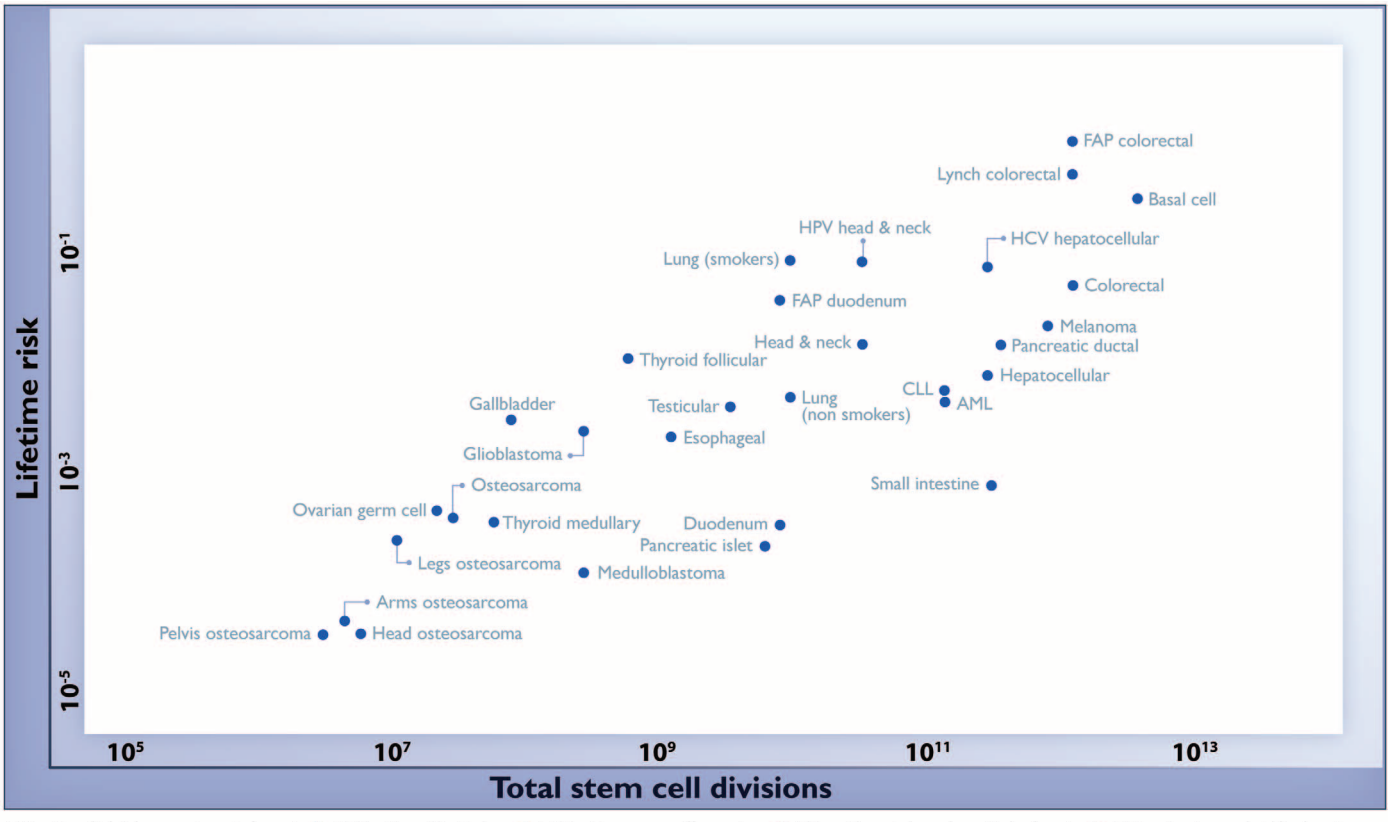
That cancer is largely the result of acquired genetic and epigenetic changes is based on the somatic mutation theory of cancer (*9*–*13*) and has been solidified by genome-wide analyses (*14*–*16*). The idea that the number of cells in a tissue and their cumulative number of divisions may be related 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



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.

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 very effective, and secondary prevention should be the major focus.

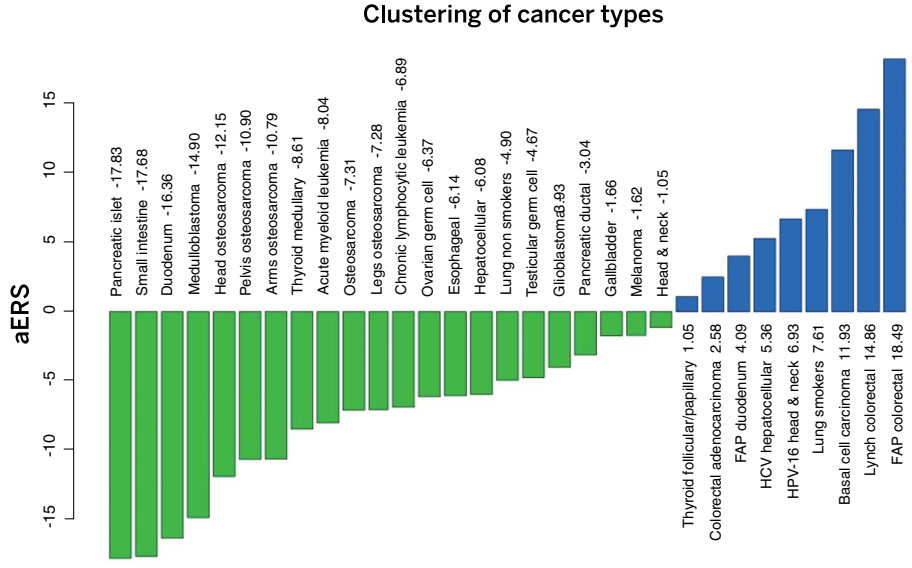


Fig. 2. Stochastic (replicative) factors versus environmental and inherited factors: R-tumor versus D-tumor classification. The adjusted ERS (aERS) is indicated next to the name of each cancer type. R-tumors (green) have negative aERS and appear to be mainly due to stochastic effects associated with 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.

REFERENCES AND NOTES

1. P. M. Dubal *et al.*, *Laryngoscope* **124**, 1877–1881 (2014).
2. N. Jaffe, *Pediatric and Adolescent Osteosarcoma* (Springer, New York, 2009).
3. National Cancer Institute, Surveillance, Epidemiology, and End Results Program; www.seer.cancer.gov.
4. G. Danaei *et al.*, *Lancet* **366**, 1784–1793 (2005).
5. Centers for Disease Control and Prevention; www.cdc.gov.
6. E. R. Fearon, *Science* **278**, 1043–1050 (1997).
7. P. Lichtenstein *et al.*, *N. Engl. J. Med.* **343**, 78–85 (2000).
8. American Cancer Society; www.cancer.org.
9. P. Armitage, R. Doll, *Br. J. Cancer* **11**, 161–169 (1957).
10. P. Armitage, R. Doll, *Br. J. Cancer* **8**, 1–12 (1954).
11. T. Boveri, *Zur Frage der Entstehung Maligner Tumoren* (G. Fischer, Jena, Germany, 1914).
12. E. R. Fearon, B. Vogelstein, *Cell* **61**, 759–767 (1990).
13. A. G. Knudson Jr., *Proc. Natl. Acad. Sci. U.S.A.* **68**, 820–823 (1971).
14. L. A. Garraway, E. S. Lander, *Cell* **153**, 17–37 (2013).
15. M. R. Stratton, P. J. Campbell, P. A. Futreal, *Nature* **458**, 719–724 (2009).
16. B. Vogelstein *et al.*, *Science* **339**, 1546–1558 (2013).
17. D. Albanes, M. Winick, *J. Natl. Cancer Inst.* **80**, 772–775 (1988).
18. L. Tomatis; International Agency for Research on Cancer, *Environ. Health Perspect.* **101** (suppl. 5), 149–151 (1993).
19. J. M. Ward, H. Uno, Y. Kurata, C. M. Weghorst, J. J. Jang, *Environ. Health Perspect.* **101** (suppl. 5), 125–135 (1993).
20. S. Sell, *Environ. Health Perspect.* **101** (suppl. 5), 15–26 (1993).
21. S. A. Frank, in *Dynamics of Cancer: Incidence, Inheritance, and Evolution* (Princeton Univ. Press, Princeton, NJ, 2007), chap. 4.
22. S. G. Baker, A. Cappuccio, J. D. Potter, *J. Clin. Oncol.* **28**, 3215–3218 (2010).
23. M. Lynch, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 961–968 (2010).
24. C. Tomasetti, B. Vogelstein, G. Parmigiani, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1999–2004 (2013).

25. C. Blanpain, E. Fuchs, *Nat. Rev. Mol. Cell Biol.* **10**, 207–217 (2009).
26. C. Booth, C. S. Potten, *J. Clin. Invest.* **105**, 1493–1499 (2000).
27. T. Reya, S. J. Morrison, M. F. Clarke, I. L. Weissman, *Nature* **414**, 105–111 (2001).
28. H. J. Snippert *et al.*, *Cell* **143**, 134–144 (2010).
29. C. O. Nordling, *Br. J. Cancer* **7**, 68–72 (1953).
30. C. Kandoth *et al.*, *Nature* **502**, 333–339 (2013).
31. G. A. Colditz, K. Y. Wolin, S. Gehlert, *Sci. Transl. Med.* **4**, 127rv4 (2012).

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, RO1-CA57345, and P50-CA62924. 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.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/347/6217/78/suppl/DC1
Materials and Methods
Fig. S1
Table S1
References (*32*–*146*)

4 September 2014; accepted 20 November 2014
10.1126/science.1260825

MUTAGENESIS

Smoking is associated with mosaic loss of chromosome Y

Jan P. Dumanski,[1,2]* Chiara Rasi,[1,2] Mikael Lönn,[3] Hanna Davies,[1,2] Martin Ingelsson,[4] Vilmantas Giedraitis,[4] Lars Lannfelt,[4] Patrik K. E. Magnusson,[5] Cecilia M. Lindgren,[6,7] Andrew P. Morris,[6,8] David Cesarini,[9] Magnus Johannesson,[10] Eva Tiensuu Janson,[11] Lars Lind,[11] Nancy L. Pedersen,[5] Erik Ingelsson,[2,11] Lars A. Forsberg[1,2]*

Tobacco smoking is a risk factor for numerous disorders, including cancers affecting organs outside the respiratory tract. Epidemiological data suggest that smoking is a greater risk factor for these cancers in males compared with females. This observation, together with the fact that males have a higher incidence of and mortality from most non–sex-specific cancers, remains unexplained. Loss of chromosome Y (LOY) in blood cells is associated with increased risk of nonhematological tumors. We demonstrate here that smoking is associated with LOY in blood cells in three independent cohorts [TwinGene: odds ratio (OR) = 4.3, 95% confidence interval (CI) = 2.8 to 6.7; Uppsala Longitudinal Study of Adult Men: OR = 2.4, 95% CI = 1.6 to 3.6; and Prospective Investigation of the Vasculature in Uppsala Seniors: OR = 3.5, 95% CI = 1.4 to 8.4] encompassing a total of 6014 men. The data also suggest that smoking has a transient and dose-dependent mutagenic effect on LOY status. The finding that smoking induces LOY thus links a preventable risk factor with the most common acquired human mutation.

Tobacco smoking killed ~100 million people during the 20th century and is projected to kill ~1 billion people during the current century, assuming that the current frequency of smoking is retained (*1*, *2*). Lung cancer is the prime cause of cancer-associated death in relation to smoking. However, smoking is also a risk factor for tumors outside the respiratory tract, and these are more common in males than females [hazard ratio in males: 2.2, 95% confidence interval (CI) = 1.7 to 2.8; in females: 1.7, 95% CI = 1.4–2.1] (*2*). Moreover, males have a higher incidence and mortality from most non–sex-specific cancers, disregarding smoking status, and this fact is largely unexplained by known risk factors (*3*, *4*). A recent analysis of noncancerous blood cells revealed that a male-specific chromosomal aberration, acquired mosaic loss of chromosome Y (LOY), is associated with an increased risk of nonhematological tumors among aging males (*5*).

Here, we analyzed possible causes of LOY by studying 6014 men from three independent prospective cohorts—TwinGene, n = 4373 (*6*, *7*); Uppsala Longitudinal Study of Adult Men (ULSAM), n = 1153 (*8*); and Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS), n = 488 (*9*)—from which comprehensive epidemiological records are available (tables S2 to S4). We included the following environmental, lifestyle, and clinical factors in the analyses: smoking, age, hypertension, exercise habits, diabetes, body mass index, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, education level, and alcohol intake. We also included genotyping quality as a confounder in the regression analyses, to adjust for possible influence of experimental noise. Similar definitions of factors were used in all cohorts, as outlined in tables S2 to S5 and described in detail in the materials and methods section of the supplementary materials. Estimation of LOY was based on single-nucleotide polymorphism (SNP)–array data using the 2.5MHumanOmni and HumanOmniExpress beadchips in the ULSAM and PIVUS/TwinGene studies, respectively (fig. S1). The estimation of the degree of mosaicism and scoring of LOY was undertaken using the continuous median logR ratio (mLRR-Y) estimate, calculated from SNP-array data as the median of the logR ratio of all SNP probes within the male-specific part of chromosome Y (MSY), as described previously (*5*). An mLRR-Y estimate close to zero indicates a normal chromosome Y state, whereas more negative mLRR-Y values denote an increasing level of blood cells with LOY. To facilitate comparisons between the three cohorts, we corrected the mLRR-Y values for all participants, using cohort-specific correction constants, as explained in the supplementary materials (figs. S1 and S2).

LOY was by far the most common postzygotic mutation found in the three cohorts. The age range at sampling in ULSAM and PIVUS was 70.7 to 83.6 years and 69.8 to 70.7 years, respectively, and we found LOY in 12.6% of ULSAM participants and 15.6% of PIVUS participants (figs. S3 and S4). The age range at sampling in TwinGene was 48 to 93 years, and the frequency of LOY in the entire cohort was 7.5% (fig. S5).

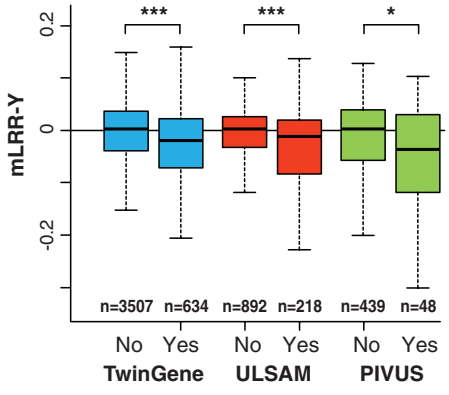


Fig. 1. The association between smoking status and the level of LOY (i.e., mLRR-Y) in three independent cohorts. In all cohorts, these unadjusted analyses indicate that the current smokers (Yes) (table S5) had a significantly higher degree of mosaic LOY in blood, compared with noncurrent smokers (No), composed of never-smokers and previous smokers. ***P < 0.001; *P < 0.05 (Kolmogorov-Smirnov tests: TwinGene, D = 0.15, $P = 1.131 \times 10^{-11}$; ULSAM, D = 0.15, P = 0.0006; PIVUS, D = 0.23, P = 0.0203). The definitions used for LOY scoring and the entire ranges of mLRR-Y data observed in each cohort are shown in figs. S3 to S5.

[1]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [2]Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [3]Södertörn University, School of Life Sciences, Biology, Huddinge, Sweden. [4]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. [5]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [6]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [7]Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA. [8]Department of Biostatistics, University of Liverpool, Liverpool, UK. [9]Center for Experimental Social Science, New York University, New York, NY 10012, USA. [10]Department of Economics, Stockholm School of Economics, Stockholm, Sweden. [11]Department of Medical Sciences, Uppsala University, Uppsala, Sweden.
***Corresponding Author. E-mail: jan.dumanski@igp.uu.se (J.P.D.); lars.forsberg@igp.uu.se (L.A.F.).**

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

Cristian Tomasetti and Bert Vogelstein

Science **347**, 78 (2015);
DOI: 10.1126/science.1260825

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.