

CANCER ETIOLOGY

Mutational signatures associated with tobacco smoking in human cancer

Ludmil B. Alexandrov,^{1,2,3*} Young Seok Ju,⁴ Kerstin Haase,⁵ Peter Van Loo,^{5,6} Iñigo Martincorena,⁷ Serena Nik-Zainal,^{7,8} Yasushi Totoki,⁹ Akihiro Fujimoto,^{10,11} Hirowaki Nakagawa,¹⁰ Tatsuhiro Shibata,^{9,12} Peter J. Campbell,^{7,13} Paolo Vineis,^{14,15} David H. Phillips,¹⁶ Michael R. Stratton^{7*}

Tobacco smoking increases the risk of at least 17 classes of human cancer. We analyzed somatic mutations and DNA methylation in 5243 cancers of types for which tobacco smoking confers an elevated risk. Smoking is associated with increased mutation burdens of multiple distinct mutational signatures, which contribute to different extents in different cancers. One of these signatures, mainly found in cancers derived from tissues directly exposed to tobacco smoke, is attributable to misreplication of DNA damage caused by tobacco carcinogens. Others likely reflect indirect activation of DNA editing by APOBEC cytidine deaminases and of an endogenous clocklike mutational process. Smoking is associated with limited differences in methylation. The results are consistent with the proposition that smoking increases cancer risk by increasing the somatic mutation load, although direct evidence for this mechanism is lacking in some smoking-related cancer types.

Tobacco smoking has been associated with at least 17 types of human cancer (Table 1) and claims the lives of more than 6 million people every year (1–4). Tobacco smoke is a complex mixture of chemicals, among which at least 60 are carcinogens (5). Many of these are thought to cause cancer by inducing DNA damage that, if misreplicated, leads to an increased burden of somatic mutations and, hence, an elevated chance of acquiring driver mutations in cancer genes. Such damage often occurs in the form of covalent bonding of metabolically activated reactive species of the carcinogen to DNA bases, termed DNA adducts (6). Tissues directly exposed to tobacco smoke (e.g., lung), as well as some tissues not directly exposed (e.g., bladder), show elevated levels of DNA adducts in smokers and, thus, evidence of exposure to carcinogenic components of tobacco smoke (7, 8).

Each biological process causing mutations in somatic cells leaves a mutational signature (9). Many cancers have a somatic mutation in the *TP53* gene, and catalogs of *TP53* mutations compiled two decades ago enabled early exploration of these signatures (10), showing that lung cancers from smokers have more C>A transversions than lung cancers from nonsmokers (11–14). To investigate mutational signatures using the thousands of mutation catalogs generated by systematic

cancer genome sequencing, we recently described a framework in which each base substitution signature is characterized using a 96-mutation classification that includes the six substitution types together with the bases immediately 5' and 3' to the mutated base (15). The analysis extracts mutational signatures from mutation catalogs and estimates the number of mutations contributed by each signature to each cancer genome (15). Using this approach, more than 30 different base substitution signatures have been identified (16–18).

In this study, we examined 5243 cancer genome sequences (4633 exomes and 610 whole genomes) of cancer classes for which smoking increases risk, with the goal of identifying mutational signatures and methylation changes associated with tobacco smoking (table S1). Of the samples we studied, 2490 were reported to be from tobacco smokers and 1063 from never-smokers (Table 1). Thus, we were able to investigate the mutational consequences of smoking by comparing somatic mutations and methylation in smokers versus nonsmokers for lung, larynx, pharynx, oral cavity, esophageal, bladder, liver, cervical, kidney, and pancreatic cancers (Fig. 1 and table S2).

We first compared total numbers of base substitutions, small insertions and deletions (indels),

and genomic rearrangements. The total number of base substitutions was higher in smokers compared with nonsmokers for all cancer types together (q -value < 0.05) and, for individual cancer types, in lung adenocarcinoma, larynx, liver, and kidney cancers (table S2). Total numbers of indels were higher in smokers compared with nonsmokers in lung adenocarcinoma and liver cancer (table S2). The whole-genome-sequenced cases allowed comparison of genome rearrangements between smokers and nonsmokers in pancreatic and liver cancer, where no differences were found (table S2). However, subchromosomal copy-number changes entail genomic rearrangement and can serve as surrogates for rearrangements. Lung adenocarcinomas from smokers exhibited more copy-number aberrations than those from nonsmokers (table S2).

We then extracted mutational signatures, estimated the contributions of each signature to each cancer, and compared the numbers of mutations attributable to each signature in smokers and nonsmokers. Increases in smokers compared with nonsmokers were seen for signatures 2, 4, 5, 13, and 16 [the mutational signature nomenclature is that used in the Catalogue of Somatic Mutations in Cancer (COSMIC) and in (16–18)]. There was sufficient statistical power to show that these increases were of clonal mutations (mutations present in all cells of each cancer) for signatures 4 and 5 (q < 0.05), as expected if these mutations are due to cigarette smoke exposure before neoplastic change (supplementary text).

Signature 4 is characterized mainly by C>A mutations with smaller contributions from other base substitution classes (Fig. 2B and fig. S1). This signature was found only in cancer types in which tobacco smoking increases risk and mainly in those derived from epithelia directly exposed to tobacco smoke (figs. S2 and S3). Signature 4 is very similar to the mutational signature induced in vitro by exposing cells to benzo[*a*]pyrene (cosine similarity = 0.94) (Fig. 2B and fig. S3), a tobacco smoke carcinogen (19). The similarity extends to the presence of a transcriptional strand bias indicative of transcription-coupled nucleotide excision repair (NER) of bulky DNA adducts on guanine (fig. S1), the proposed mechanism of DNA damage by benzo[*a*]pyrene. Thus, signature 4 is likely the direct mutational consequence of misreplication of DNA damage induced by tobacco carcinogens.

Most lung and larynx cancers from smokers had many signature 4 mutations. Signature 4 mutations occurred more often in cancers from

¹Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ²Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ³University of New Mexico Comprehensive Cancer Center, Albuquerque, NM 87102, USA. ⁴Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea. ⁵The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. ⁶Department of Human Genetics, University of Leuven, 3000 Leuven, Belgium. ⁷Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, Cambridgeshire, UK. ⁸Department of Medical Genetics, Addenbrooke's Hospital National Health Service Trust, Cambridge, UK. ⁹Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. ¹⁰Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan. ¹¹Department of Drug Discovery Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. ¹²Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan. ¹³Department of Haematology, University of Cambridge, Cambridge CB2 0XY, UK. ¹⁴Human Genetics Foundation, 10126 Torino, Italy. ¹⁵Department of Epidemiology and Biostatistics, Medical Research Council (MRC)–Public Health England (PHE) Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. ¹⁶King's College London, MRC-PHE Centre for Environment and Health, Analytical and Environmental Sciences Division, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK.

*Corresponding author. Email: lba@lanl.gov (L.B.A.); mrs@sanger.ac.uk (M.R.S.)

smokers compared with nonsmokers in all cancer types together (table S2) and in lung squamous, lung adenocarcinoma, and larynx cancers (table S2). This finding largely accounts for differences in total numbers of base substitutions (Table 1). In nonsmokers, 13.8% of lung cancers showed many signature 4 mutations (Fig. 2A; >1 mutation per megabase), which may be due to passive smoking, misreporting of smoking habits, or annotation errors. Signature 4 mutations were also detected in cancers of the oral cavity, pharynx, and esophagus, albeit in much smaller numbers than in lung and larynx cancers, perhaps because of reduced exposure to tobacco smoke or more efficient clearance. Differences in mutation burden attributed to signature 4 between smokers and nonsmokers were not observed in these cancer types (Fig. 1). Signature 4 mutations were found at low levels in cancers of the liver, an organ not directly exposed to tobacco smoke, and were elevated in smokers versus nonsmokers (Fig. 1).

Signature 4 was not extracted from bladder, cervical, kidney, or pancreatic cancers, despite the known risks conferred by smoking and the presence of many smokers in these series. Additionally, this mutational signature was not extracted from cancers of the stomach, colorectum,

and ovary, nor from acute myeloid leukemia (in the analyzed series, the smoking status of patients with these cancers was unknown, but it is likely that many have been smokers). The tissues from which all of these cancer types are derived are not directly exposed to tobacco smoke. Simulations indicate that the lack of signature 4 is not due to statistical limitations (supplementary text and fig. S4). The absence of signature 4 suggests that misreplication of direct DNA damage due to tobacco smoke constituents does not contribute substantially to mutation burden in these cancers, even though DNA adducts indicative of tobacco-induced DNA damage are present in the tissues from which they arise (7).

Signatures 2 and 13 are characterized by C>T and C>G mutations, respectively, at TpC dinucleotides and have been attributed to overactive DNA editing by APOBEC deaminases (20, 21). The cause of the overactivity in most cancers has not been established, although APOBECs are implicated in the cellular response to the entrance of foreign DNA, retrotransposon movement, and local inflammation (22). Signatures 2 and 13 showed more mutations in smokers versus nonsmokers with lung adenocarcinoma (table S2). Because these signatures are found in many

other cancer types, where they are apparently unrelated to tobacco smoking, it seems unlikely that the signature 2 and 13 mutations associated with smoking in lung adenocarcinoma are direct consequences of misreplication of DNA damage induced by tobacco smoke. More plausibly, the cellular machinery underlying signatures 2 and 13 is activated by tobacco smoke, perhaps as a result of inflammation arising from the deposition of particulate matter or by indirect consequences of DNA damage.

Signature 5 is characterized by mutations distributed across all 96 subtypes of base substitution, with a predominance of T>C and C>T mutations (Fig. 2B) and evidence of transcriptional strand bias for T>C mutations (18). Signature 5 is found in all cancer types, including those unrelated to tobacco smoking, and in most cancer samples. It is “clocklike” in that the number of mutations attributable to this signature correlates with age at the time of diagnosis in many cancer types (17). Signature 5, together with signature 1, is thought to contribute to mutation accumulation in most normal somatic cells and in the germline (17, 23). The mechanisms underlying signature 5 are not well understood, although an enrichment of signature 5 mutations was found in bladder cancers harboring inactivating

Table 1. Mutational signatures and cancer types associated with tobacco smoking. Information about the age-adjusted odds ratios for current male smokers to develop cancer is taken from (2–4). Odds ratios for small cell lung cancer, squamous cell lung cancer, and lung adenocarcinoma are for an average daily dose of more than 30 cigarettes. Odds ratios for cervical and ovarian cancers are for current female smokers. Detailed information about all mutation types, all mutational signatures, and DNA methylation is provided in table S2. Nomenclature for signature identification numbers is consistent with the COSMIC database

(<http://cancer.sanger.ac.uk/cosmic/signatures>). The numbers of smokers and nonsmokers are unknown (i.e., not reported in the original studies) for acute myeloid leukemia, stomach, ovarian, and colorectal cancers. The patterns of all mutational signatures with elevated mutation burden in smokers are displayed in Fig. 2B. N/A denotes lack of smoking annotation for a given cancer type. Asterisks indicate that a signature correlates with pack years smoked in a cancer type. N.S. reflects cancer types without statistically significant elevation of mutational signatures. The odds ratio for all cancer types is not provided.

| Cancer type | Odds ratio | Nonsmokers | Smokers | Total number of mutational signatures found in the cancer type | Signature 4 found in cancer type | Mutational signatures with elevated mutation burden in smokers versus nonsmokers ($q < 0.05$) |
|--------------------------|------------|------------|---------|--|----------------------------------|---|
| All cancer types | ND | 1062 | 2490 | 26 | Y | 4, 5* |
| Small cell lung cancer | 111.3 | 3 | 145 | 6 | Y | N.S. |
| Lung squamous | 103.5 | 7 | 168 | 8 | Y | 4*, 5 |
| Lung adenocarcinoma | 21.9 | 120 | 558 | 7 | Y | 2*, 4*, 5*, 13* |
| Larynx | 13.2 | 6 | 117 | 5 | Y | 4*, 5 |
| Pharynx | 6.6 | 27 | 49 | 5 | Y | 5* |
| Oral cavity | 4.2 | 98 | 265 | 5 | Y | 5* |
| Esophagus squamous | 3.9 | 99 | 193 | 9 | Y | 5 |
| Esophagus adenocarcinoma | 3.9 | 67 | 175 | 9 | Y | N.S. |
| Bladder | 3.8 | 111 | 288 | 5 | N | 5* |
| Liver | 2.9 | 157 | 235 | 19 | Y | 4*, 5, 16 |
| Stomach | 2.1 | 472 | | 13 | N | N/A |
| Acute myeloid leukemia | 2.0 | 202 | | 2 | N | N/A |
| Ovary | 1.9 | 458 | | 3 | N | N/A |
| Cervix | 1.8 | 94 | 74 | 8 | N | N.S. |
| Kidney | 1.7 | 154 | 103 | 6 | N | 5 |
| Pancreas | 1.6 | 119 | 120 | 11 | N | N.S. |
| Colorectal | 1.3 | 559 | | 4 | N | N/A |

mutations in *ERCC2*, which encodes a component of NER (24).

Signature 5 (or a similar signature that is difficult to differentiate from signature 5 because of the relatively flat profiles of these signatures) was increased by a factor of 1.3 to 5.1 ($q < 0.05$; table S2) in smokers versus nonsmokers in all cancer types together and in lung squamous, lung adenocarcinoma, larynx, pharynx, oral cavity, esophageal squamous, bladder, liver, and kidney cancers. The association of smoking with signature 5 mutations across these nine cancer types therefore includes some for which the risks conferred by smoking are modest and for which normal progenitor cells are not directly exposed to cigarette smoke (Table 1). Given the clocklike nature of signature 5 (17), its presence in the human germline (23), its ubiquity in cancer types unrelated to tobacco smoking (18),

and its widespread occurrence in nonsmokers, it seems unlikely that signature 5 mutations associated with tobacco smoking are direct consequences of misreplication of DNA damaged by tobacco carcinogens. It is more plausible that smoking affects the machinery generating signature 5 mutations (24). Presumably as a consequence of the effects of smoking, signature 5 mutations correlated with age at the time of diagnosis in nonsmokers ($P = 0.001$) but not in smokers ($P = 0.59$).

Signature 16 is predominantly characterized by T>C mutations at ApT dinucleotides (Fig. 2B); exhibits a strong transcriptional strand bias consistent with almost all damage occurring on adenine (fig. S5); and, thus far, has been detected only in liver cancer. The underlying mutational process is currently unknown. Signature 16 exhibited a higher mutation burden

in smokers versus nonsmokers with liver cancer (table S2).

For smokers with lung, larynx, pharynx, oral cavity, esophageal, bladder, liver, cervical, kidney, and pancreatic cancers, quantitative data on cumulative exposure to tobacco smoke were available (table S1). Total numbers of base substitution mutations were positively correlated with pack years smoked (1 pack year is defined as smoking one pack per day for 1 year) for all cancer types together ($q < 0.05$) and for lung adenocarcinoma (table S3). For individual mutational signatures, correlations with pack years smoked were found in multiple cancer types for signatures 4 and 5 (table S3). Signature 4 correlated with pack years in lung squamous, lung adenocarcinoma, larynx, and liver cancers. Signature 5 correlated with pack years in all cancers together, as well as in lung adenocarcinoma, pharynx, oral cavity, and

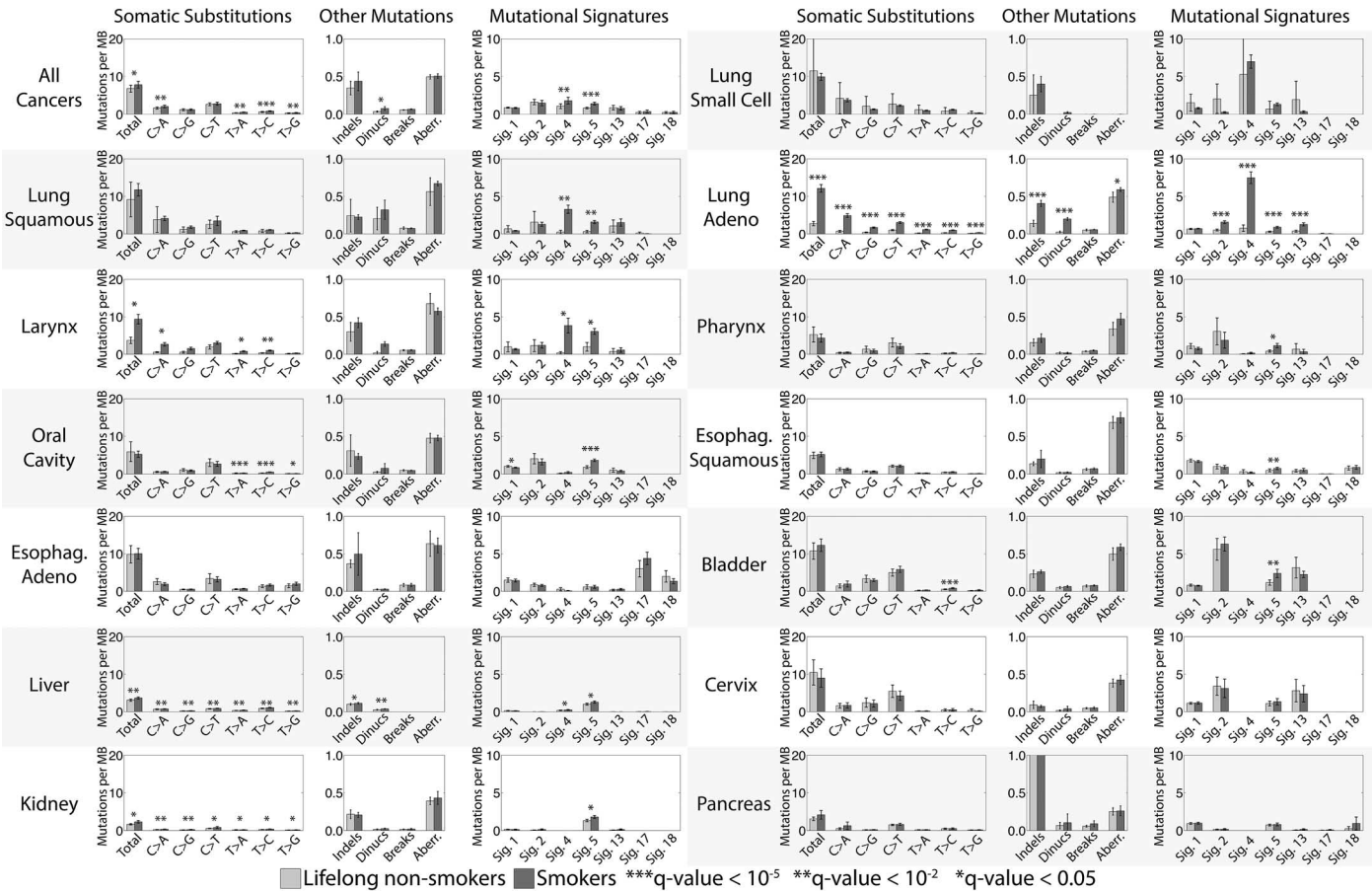


Fig. 1. Comparison between tobacco smokers and lifelong nonsmokers. Bars are used to display average values for numbers of somatic substitutions per megabase (MB), numbers of indels per megabase, numbers of dinucleotide mutations per megabase (Dinucs), numbers of breakpoints per megabase (Breaks), fraction of the genome that shows copy-number changes (Aberr.), and numbers of mutations per megabase attributed to mutational signatures found in multiple cancer types associated with tobacco smoking. Light gray bars represent nonsmokers, whereas dark gray bars are for smokers. Comparisons between smokers and nonsmokers for all features, including mutational signatures specific for a cancer type and overall DNA methylation, are

provided in table S2. Error bars correspond to 95% confidence intervals for each feature. Each q value is based on a two-sample Kolmogorov-Smirnov test corrected for multiple hypothesis testing for all features in a cancer type. Cancer types are ordered based on their age-adjusted odds ratios for smoking, as provided in Table 1. Data for numbers of breakpoints per megabase and fraction of the genome that shows copy-number changes were not available for liver cancer and small cell lung cancer. Adeno, adenocarcinoma; Esophag., esophagus. Note that the presented data include only a few cases (<10) of nonsmokers for small cell lung cancer, squamous cell lung cancer, and cancer of the larynx.

bladder cancers (table S3). In lung adenocarcinoma, correlations with pack years smoked were also observed for signatures 2 and 13. The rates of these correlations allow estimation of the approximate numbers of mutations accumulated in a normal cell of each tissue due to smoking a pack of cigarettes a day for a year: lung, 150 mutations; larynx, 97; pharynx, 39; oral cavity 23; bladder, 18; liver, 6 (table S3).

Consistent with our results, previous studies have reported higher numbers of total base substitutions in lung adenocarcinoma in smokers versus nonsmokers (mainly due to C>A substitutions) (25, 26). The same is true of signatures 4 and 5 in lung adenocarcinoma (18), signature 4 in liver cancer (27), and signature 5 in bladder cancer (24).

Differential methylation of the DNA of normal cells of smokers compared to nonsmokers has been reported (28). Using data from methylation arrays, each containing ~470,000 of the ~28 million CpG sites in the human genome, we evaluated whether differences in methyla-

tion are found in cancers. Overall levels of CpG methylation in DNA from cancers were similar in smokers and nonsmokers for all cancer types (fig. S6). Individual CpGs were differentially methylated (>5% difference) in only two cancer types: 369 CpGs were hypomethylated and 65 were hypermethylated in lung adenocarcinoma, with five hypomethylated and three hypermethylated in oral cancer (Fig. 3 and fig. S7). CpGs exhibiting differences in methylation clustered in certain genes but were not associated with known cancer genes more than expected by chance, nor with genes hypomethylated in normal blood or buccal cells of tobacco smokers (fig. S8 and tables S4 and S5) (28). Therefore, with the exception of lung cancer, CpG methylation showed limited differences between the cancers of smokers and nonsmokers (Fig. 3).

The genomes of smoking-associated cancers permit reassessment of our understanding of how tobacco smoke causes cancer. Consistent with the proposition that an increased mutation load caused by tobacco smoke contributes to in-

creased cancer risk, the total mutation burden is elevated in smokers versus nonsmokers with lung adenocarcinoma, larynx, liver, and kidney cancers. However, differences in total mutation burden were not observed in the other smoking-associated cancer types and, in some, there were no statistically significant smoking-associated differences in mutation load, signatures, or DNA methylation. Caution should be exercised in the interpretation of the latter observations. In addition to limitations of statistical power, multiple rounds of clonal expansion over many years are often required for development of a symptomatic cancer. It is thus conceivable that, in the normal tissues from which smoking-associated cancer types originate, there are more somatic mutations (or differences in methylation) in smokers than in nonsmokers but that these differences become obscured during the intervening clonal evolution. Moreover, some theoretical models predict that relatively small differences in mutation burden caused by smoking in preneoplastic cells could account for the observed increases in cancer

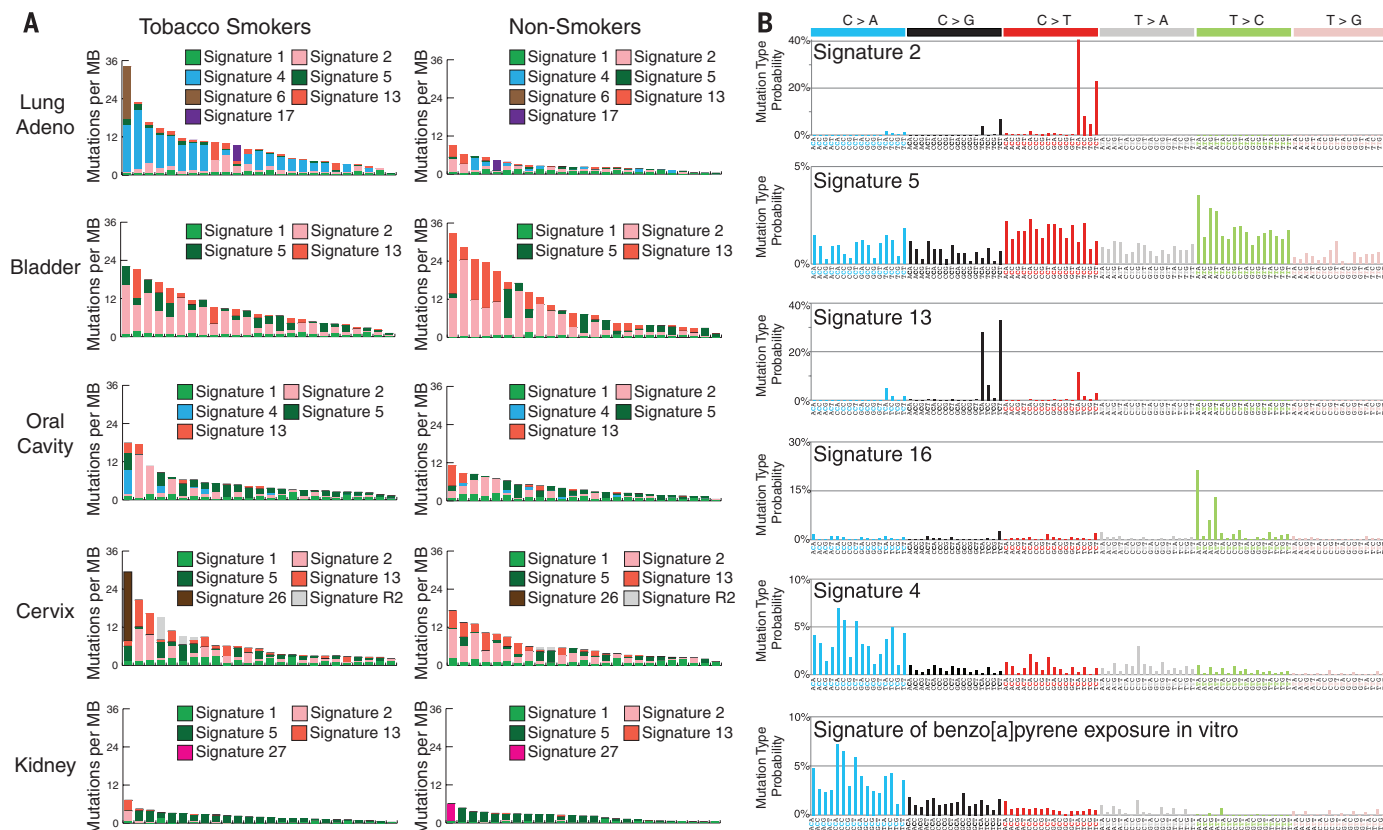


Fig. 2. Mutational signatures associated with tobacco smoking. (A) Each panel contains 25 randomly selected cancer genomes (represented by individual bars) from either smokers or nonsmokers in a given cancer type. The y axes reflect numbers of somatic mutations per megabase. Each bar is colored proportionately to the number of mutations per megabase attributed to the mutational signatures found in that sample. The naming of mutational signatures is consistent with previous reports (16–18). (B) Each panel contains the pattern of a mutational signature associated with tobacco smoking. Signatures are depicted using a 96-substitution classification defined by the

substitution type and sequence context immediately 5' and 3' to the mutated base. Different colors are used to display the various types of substitutions. The percentages of mutations attributed to specific substitution types are on the y axes, whereas the x axes display different types of substitutions. Mutational signatures are depicted based on the trinucleotide frequency of the whole human genome. Signatures 2, 4, 5, 13, and 16 are extracted from cancers associated with tobacco smoking. The signature of benzo[a]pyrene is based on in vitro experimental data (19). Numerical values for these mutational signatures are provided in table S6.

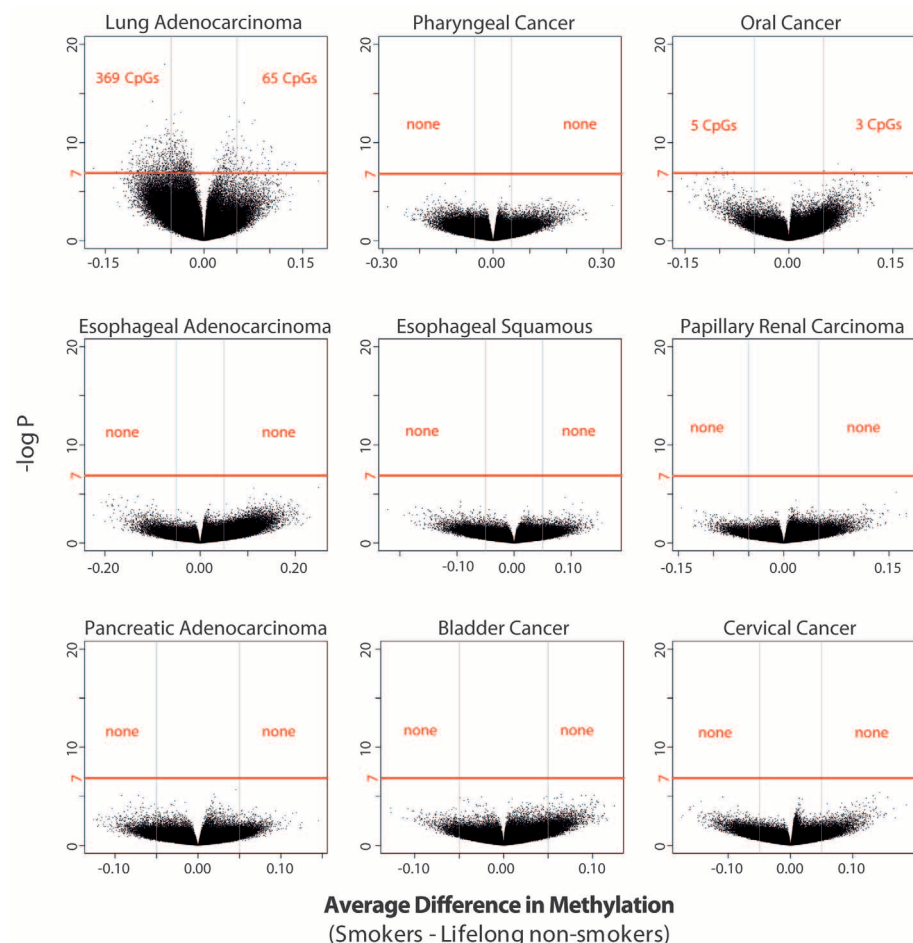


Fig. 3. Differentially methylated individual CpGs in tobacco smokers across cancers associated with tobacco smoking. Each dot represents an individual CpG. The x axes reflect differences in methylation between lifelong nonsmokers and smokers, where positive values correspond to hypermethylation and negative values to hypomethylation. The y axes depict levels of statistical significance. Results satisfying a Bonferroni threshold of 10^{-7} (above the red line) are considered statistically significant.

risks (29). Other models indicate that differences in mutation burden between smokers and non-smokers need not be observed in the final cancers (supplementary text and fig. S6). Thus, increased somatic mutation loads in precancerous tissues may still explain the smoking-induced risks of most cancers, although other mechanisms have been proposed (30, 31).

However, the generation of increased somatic mutation burden by tobacco smoking appears to be mechanistically complex. Smoking correlates with increases in base substitutions of multiple mutational signatures, together with increases in indels and copy-number changes. The extent to which these distinct mutational processes operate differs between tissue types (at least partially depending on the degree of direct exposure to tobacco smoke), and their mechanisms range from misreplication of DNA damage caused by tobacco smoke constituents to activation of more generally operative mutational processes. Although

we cannot exclude roles for covariate behaviors of smokers or differences in the biology of cancers arising in smokers compared with non-smokers, smoking itself is most plausibly the cause of these differences.

REFERENCES AND NOTES

1. B. Secretan et al., *Lancet Oncol.* **10**, 1033–1034 (2009).
2. S. S. Lim et al., *Lancet* **380**, 2224–2260 (2012).
3. B. Pesch et al., *Int. J. Cancer* **131**, 1210–1219 (2012).
4. A. Agudo et al., *J. Clin. Oncol.* **30**, 4550–4557 (2012).
5. S. S. Hecht, *Nat. Rev. Cancer* **3**, 733–744 (2003).
6. D. H. Phillips, in *The Cancer Handbook*, M. R. Allison, Ed. (Macmillan, 2002), pp. 293–306.
7. D. H. Phillips, *Carcinogenesis* **23**, 1979–2004 (2002).
8. D. H. Phillips, S. Venitt, *Int. J. Cancer* **131**, 2733–2753 (2012).
9. L. B. Alexandrov, M. R. Stratton, *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).
10. P. Hainaut, M. Hollstein, *Adv. Cancer Res.* **77**, 81–137 (1999).
11. M. F. Denissenko, A. Pao, M. Tang, G. P. Pfeifer, *Science* **274**, 430–432 (1996).

12. G. P. Pfeifer, M. F. Denissenko, *Environ. Mol. Mutagen.* **31**, 197–205 (1998).
13. L. E. Smith et al., *J. Natl. Cancer Inst.* **92**, 803–811 (2000).
14. F. Le Calvez et al., *Cancer Res.* **65**, 5076–5083 (2005).
15. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, P. J. Campbell, M. R. Stratton, *Cell Reports* **3**, 246–259 (2013).
16. L. B. Alexandrov, *Science* **350**, 1175–1177 (2015).
17. L. B. Alexandrov et al., *Nat. Genet.* **47**, 1402–1407 (2015).
18. L. B. Alexandrov et al., Australian Pancreatic Cancer Genome Initiative/CGC Breast Cancer Consortium/CGC MML-Seq Consortium/CGC PedBrain, *Nature* **500**, 415–421 (2013).
19. S. Nik-Zainal et al., *Mutagenesis* **30**, 763–770 (2015).
20. S. Nik-Zainal et al., *Cell* **149**, 979–993 (2012).
21. S. A. Roberts et al., *Mol. Cell* **46**, 424–435 (2012).
22. C. Swanton, N. McGranahan, G. J. Starrett, R. S. Harris, *Cancer Discov.* **5**, 704–712 (2015).
23. R. Rahbari et al., *Nat. Genet.* **48**, 126–133 (2016).
24. J. Kim et al., *Nat. Genet.* **48**, 600–606 (2016).
25. R. Govindan et al., *Cell* **150**, 1121–1134 (2012).
26. M. Imielinski et al., *Cell* **150**, 1107–1120 (2012).
27. A. Fujimoto et al., *Nat. Genet.* **48**, 500–509 (2016).
28. A. E. Teschendorff et al., *JAMA Oncol.* **1**, 476–485 (2015).
29. C. Tomasetti, L. Marchionni, M. A. Nowak, G. Parmigiani, B. Vogelstein, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 118–123 (2015).
30. M. Sopori, *Nat. Rev. Immunol.* **2**, 372–377 (2002).
31. H. Rubin, *Oncogene* **21**, 7392–7411 (2002).

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust (grant 098051). S.N.-Z. is a Wellcome-Brit Prize Fellow and is supported through a Wellcome Trust Intermediate Fellowship (grant WT100183MA). P.J.C. is personally funded through a Wellcome Trust Senior Clinical Research Fellowship (grant WT088340MA). M.R.S. is a paid advisor for GRAIL, a company developing technologies for sequencing of circulating tumor DNA for the purpose of early cancer detection. L.B.A. is personally supported through a J. Robert Oppenheimer Fellowship at Los Alamos National Laboratory. This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy (DOE) National Nuclear Security Administration under contract no. DE-AC52-06NA25396. Research performed at Los Alamos National Laboratory was carried out under the auspices of the National Nuclear Security Administration of the DOE. This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (grant FC001202), the UK MRC (grant FC001202), and the Wellcome Trust (grant FC001202). P.V.L. is a Winton Group Leader in recognition of the Winton Charitable Foundation's support toward the establishment of The Francis Crick Institute. D.H.P. is funded by Cancer Research UK (grant C313/A14329), the Wellcome Trust (grants 101126/Z/13/Z and 101126/B/13/Z), the National Institute for Health Research (NIHR) Health Protection Research Unit in Health Impact of Environmental Hazards at King's College London in partnership with PHE [the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, or PHE], and by the project EXPOSOMICS (grant agreement 308610-FP7) (European Commission). P.V. was partially supported by the project EXPOSOMICS (grant agreement 308610-FP7) (European Commission). Y.T. and T.S. are supported by the Practical Research for Innovative Cancer Control from Japan Agency for Medical Research and Development (grant 15ck0106094h0002) and National Cancer Center Research and Development Funds (26-A-5). We thank The Cancer Genome Atlas, the International Cancer Genome Consortium, and the authors of all studies cited in table S1 for providing free access to their somatic mutational data.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/618/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S6
References (32–54)

2 May 2016; accepted 23 September 2016
10.1126/science.aag0299

Mutational signatures associated with tobacco smoking in human cancer

Ludmil B. Alexandrov, Young Seok Ju, Kerstin Haase, Peter Van Loo, Iñigo Martincorena, Serena Nik-Zainal, Yasushi Totoki, Akihiro Fujimoto, Hidewaki Nakagawa, Tatsuhiro Shibata, Peter J. Campbell, Paolo Vineis, David H. Phillips and Michael R. Stratton

Science **354** (6312), 618-622.
DOI: 10.1126/science.aag0299

Assessing smoke damage in cancer genomes

We have known for over 60 years that smoking tobacco is one of the most avoidable risk factors for cancer. Yet the detailed mechanisms by which tobacco smoke damages the genome and creates the mutations that ultimately cause cancer are still not fully understood. Alexandrov *et al.* examined mutational signatures and DNA methylation changes in over 5000 genome sequences from 17 different cancer types linked to smoking (see the Perspective by Pfeifer). They found a complex pattern of mutational signatures. Only cancers originating in tissues directly exposed to smoke showed a signature characteristic of the known tobacco carcinogen benzo[*a*]pyrene. One mysterious signature was shared by all smoking-associated cancers but is of unknown origin. Smoking had only a modest effect on DNA methylation.

Science, this issue p. 618; see also p. 549

ARTICLE TOOLS

<http://science.sciencemag.org/content/354/6312/618>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2016/11/03/354.6312.618.DC1>

RELATED CONTENT

<http://stm.sciencemag.org/content/scitransmed/5/197/197ra101.full>
<http://stm.sciencemag.org/content/scitransmed/5/197/197ra102.full>
<http://science.sciencemag.org/content/sci/354/6312/549.full>

REFERENCES

This article cites 51 articles, 9 of which you can access for free
<http://science.sciencemag.org/content/354/6312/618#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)