Human lung cancer and *p53*: The interplay between mutagenesis and selection

Sergei N. Rodin*†‡ and Andrew S. Rodin§

*Biology Department, Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010; †Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, 630090, Russia; and §Human Genetics Center, School of Public Health, University of Texas–Health Science Center, P.O. Box 20334, Houston, TX 77225

It is an almost consensus opinion that the major carcinogenic risk of tobacco smoke is in its direct mutagenic action on DNA of cancer-related genes. The key data supposedly linking smoke-induced mutations to lung cancer were obtained from the adduct spectrum of the *p53* tumor suppressor gene. Results of our analysis of p53 mutations compiled from the International Agency for Research on Cancer *p53* database (April 1999 update) and from the literature point to a different causative link. Our new analytical tests focused on complementary base substitutions and showed that it is strand-specific repair of primary lesions and site-specific selection of the resultant mutations that determine the lung cancer-specific hot spots of G:C to T:A transversions along the *p53* gene and also their increased abundance in lung tissues as compared with smoke-inaccessible tissues. However, on each of the two strands of p53 DNA, our tests revealed no significant difference between smokers and nonsmokers, either in the frequency of different types of mutations or in the frequency of their occurrence along the *p53* gene. Moreover, in both smokers and nonsmokers, there was the same frequency of lung tumors with silent p53 mutations. Accordingly, we offer here a selection-based explanation of why lung cancers with nonsilent p53 mutations are more common in smokers than in nonsmokers. We conclude that physiological stresses (not necessarily genotoxic) aggravated by smoking are the leading risk factor in the p53-associated etiology of lung cancer.

benzo[a]pyrene | repair | nongenotoxic stresses

Persistent smoking dramatically increases the risk of death from lung cancer (1). Of all of the carcinogenic effects of tobacco smoke, its mutagenic action is believed today to be the major cause of human lung malignancy (2, 3). The main data in support of that viewpoint came from the *p53* tumor suppressor gene mutations. *p53* is the "emergency brake" gene that expresses its tumor-preventing apoptotic and cell cycle checkpoint functions in physiologically stressful situations (4, 5). Mutational alteration of these functions often gives p53$^-$ mutant cells a proliferative advantage over wild-type p53$^+$ cells. Of the 3,800 potential mutagens contained in tobacco smoke (6), the best candidate for the role of a causative mediator between smoking, p53 mutations, and lung cancer is benzo[a]pyrene diol epoxide (BPDE), a metabolite of benzo[a]pyrene, one of the polycyclic aromatic hydrocarbons found in cigarette smoke. BPDE binds to DNA and forms bulky adducts at the N2 position of guanines (7). If the adducts remain unrepaired, their misreading by DNA polymerase during replication results predominantly ($\approx$70%) in G→T transversions (8, 9); this prevalence is, in fact, the mutational signature of BPDE. G→T transversions are frequent ($\approx$35%) in human lung cancer, but they are uncommon (<10%) in most other cancers, thus indicating a possible link between BPDE and lung cancer (2, 9). This hypothesis gained support (10) when the frequency distributions of *in vitro* BPDE adduct formation and *in vivo* p53 mutations of lung cancers along the human *p53* gene were compared. Strong BPDE adduct signals were detected at the guanines of codons 157, 248, and 273, the

mutational hot spots in human lung cancer. This coincidence led the authors (10) to conclude that "targeted adduct formation rather than phenotypic selection appears to shape the *p53* mutational spectrum in lung cancer." Whereas codons 248 and 273 are among major mutational hot spots in virtually all cancers, codon 157 was claimed to be a hot spot unique to lung cancer (10).

In mutationally characterizing the *p53* gene, one way is to plot the frequency of each of 12 different types of base substitutions in the gene as a whole (Fig. 1*A*), which we will call hereon a *p53* mutational pattern. Another way (Fig. 2) is to plot a frequency of the particular 1 of these 12 types as a function of a position in the *p53* gene, which we refer to as a *p53* mutational spectrum.

Thus, the previously proposed causative chain was as follows: (*i*) the major risk factor for lung cancer is smoking, (*ii*) smoke contains benzo[a]pyrene, (*iii*) BPDE forms bulky adducts at G bases of DNA, (*iv*) the adducts cause G→T transversions, (*v*) these transversions are a hallmark of mutant *p53* genes from lung cancers, and (*vi*) hot spots of G→T transversions coincide with preferential sites of p53 DNA-BPDE adducts. However, as noted in ref. 11, some pivotal epidemiological evidence required was missing. If BPDE is a major initiating mutagen that shapes the lung cancer *p53* mutational pattern and spectrum of smokers, then one would expect both the pattern and the spectrum to be essentially different in the lung tumors of confirmed nonsmokers. Until recently, published p53 mutational data in nonsmokers were sparse; consequently, reported differences (3, 12, 13) should be viewed skeptically. Furthermore, tumorigenesis is a multistep process, and the interaction of BPDE with *p53* DNA may represent only one step. Therefore, even if the *p53* pattern and spectral differences between smokers and nonsmokers were reliably demonstrated, other contributory causes, e.g., repair and selection, must be considered in addition to possible primary, BPDE-like adduct formation. With this in mind, we compared *p53* mutational patterns and spectra in contrasting epidemiological groups by using the information obtained from the *p53* mutation catalog (14) and updated by the recently available data.

First, we found a significant difference between smokers and nonsmokers in the frequency of mutations presumably originating in the transcribed and nontranscribed strands. However, within-strand *p53* mutational patterns of the two epidemiological groups are, in fact, identical. Second, the distribution of G→T transversions along the *p53* gene is also the same in lung cancers of smokers and in cancers of other, smoke-inaccessible tissues. Third, smokers do not differ from nonsmokers in the frequency of tumors with silent p53 mutations. Additionally,

Table 1. Tumor tissues classified in regard to the frequency of G:C→T:A transversions from the coding part of the *p53* gene as reported in the International Agency for Research on Cancer *p53* database (April 1999 update)*

Tissue	Total cases	No. of transversions		% G:C→T:A
		G→T	C→A	
Most accessible to smoke				
Lung total[†]	1,047	288	41	31.4
Ever smokers	344	108	16	36.0
Never smokers	75	12	4	21.3
Lung[‡]	985	256	38	29.8
Ever smokers	286	84	13	33.9
Never smokers	69	9	4	18.8
Esophagus	623	116	16	21.2
Oral cavity	726	106	34	19.3
Least accessible to smoke				
Rectum	85	8	1	10.6
Skin	461	34	14	10.4
Blood	589	43	16	10.0
Male genitals	213	15	6	9.9
Breast	1,033	79	22	9.8
Stomach	470	24	17	8.7
Lymph nodes	103	8	1	8.7
Colon	1,092	74	20	8.6
Sum	4,046	285	97	9.4

*Cell lines excluded.

[†]The data from ref. 18 are added. This lung p53 dataset does not include the "GAO" series of 107 p53 mutations reported in ref. 19 for 27 Chinese patients (10 smokers, 17 nonsmokers), as they may be the result of laboratory artifacts or represent a quite exotic, biased sample (see critique in ref. 3).

[‡]Cases of radon-, asbestos-, and mustard gas-associated p53 mutations excluded.

Fig. 1. p53 mutational patterns of lung adenocarcinomas from smokers (99 mutations) and nonsmokers (50 mutations). Because the strand with original lesions is identifiable only for (G→T, C→A), (G→C, C→G), and (C→T, G→A) pairs of complementary substitutions, we used only these three pairs to test the "mutagenesis" vs. "strand-asymmetric repair" alternative (see text). (*A*) In a standard (strand-nonspecific) representation, the patterns of smokers and nonsmokers show a difference, although statistically insignificant (*P* value 0.151). (*B*) The same patterns as in *A*, but organized as pairs of complementary transitions and transversions; the upper part may represent a nontranscribed strand, and the lower part, a transcribed strand.

silent G:C→T:A transversions in lung cancer do not correlate with the "silent spectrum" of BPDE adducts.

These results suggest the need to revise the tobacco smoke-associated etiology of lung cancers with p53 mutations. Although not entirely denying the implication of BPDE-like carcinogens as specific exogenous mutagens, the main purpose of this revision is to emphasize endogenous sources of mutations such as oxidative DNA damage (15, 16) and nongenotoxic stresses (4, 17). We propose that smoking aggravates these stresses and, thus, intensifies the processes of strand-asymmetric repair and selection of tumorigenic stem cells carrying mutations in stress-responsive genes such as *p53*. Although lung cancers with p53 mutations are more frequent in smokers than in nonsmokers, the increased selection rather than mutation rate may account for this difference in tissues insulted by smoking.

Materials and Methods

Table 1 introduces the main groups of p53 mutations analyzed. The mutations were retrieved from the International Agency for Research on Cancer p53 database (April 1999 update; 10,396 entries) and augmented by later data. The associated habits of smoking, snuffing, and chewing tobacco and drinking alcohol and hot beverages, etc., were confirmed from original papers according to the references in ref. 14. The relatively infrequent dubious cases such as "passive smokers" and "low drinkers" were not considered. The resulting lung cancer p53 data set was partitioned into three risk groups: smokers, nonsmokers, and patients with an unknown smoking history. To achieve a sufficient sample size, we pooled G→T transversions from tumors of several, supposedly smoke-inaccessible tissues (Table 1). The resulting sample of G→T transversions was large enough to be reliably compared with that of lung cancer.
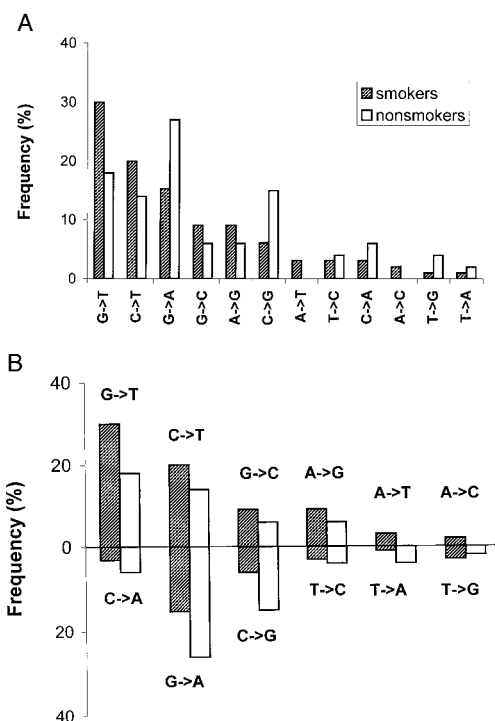
For statistical analysis of p53 mutational patterns and spectra, we used a number of tests. The popular Pearson χ^2 test, although nonparametric, is based on the expectation that, within any category, mutation frequencies are distributed normally. Therefore, if the observed mutation frequencies for some categories are very low, the results of the χ^2 test become invalid. This is the case with the p53 mutational spectra; it covers more than 100 sites but, except for the hot spots, many of these sites contain too few mutations for χ^2 analysis. The Adams and Skopek test (20) uses a Monte-Carlo simulation to estimate a *P* value of the hypergeometric bivariate (or multivariate) tabular analysis test and is recognized as superior to the χ^2 test by virtue of being able to cope with infrequent mutations. Lower *P* values indicate greater dissimilarity of the two spectra in question. For example, a *P* value of less than 0.05 indicates a statistically significant (at 5% level) difference between spectra. Similarly, high *P* values (e.g., 0.9) mean that both samples generated essentially the same spectrum.

We used HG-PUBL software (http://metalab.unc. edu/dnam/di-hypg.htm) with number of Monte-Carlo iterations set at 10,000 to make sure the sufficient test power was achieved (at least 1,700 iterations are recommended in ref. 21).

Results and Discussion

Complementary Base Substitutions in *p53* from Lung Cancer Distinguish Smokers from Nonsmokers. In Fig. 1, *p53* mutational pattern is plotted in two ways. A standard way (Fig. 1*A*) reveals a difference between smokers and nonsmokers that could be attributable to a direct action of BPDE-like mutagens on *p53*

MEDICAL SCIENCES

DNA. However, the second method (Fig. 1*B*) makes the same patterns point to a different cause. All 12 types of base substitutions are grouped into 6 pairs, each consisting of two complementary substitutions: (C→T, G→A) (G→T, C→A), etc. The overall p53 pattern (Fig. 1*A*) then is shown as two complementary subpatterns in the following fashion. In each pair, we select the substitution that is more frequent than the other in lung cancers of smokers. Six such substitutions plotted in decreasing order of magnitude form the first subpattern (above the line in Fig. 1*B*). Their complementary substitutions form the second subpattern (under the line in Fig. 1*B*). This new representation of the p53 mutational pattern shows that the excesses of G→T over C→A, G→C over C→G, A→G over T→C, etc., observed in smokers turned out to be either less pronounced in nonsmokers (in the case of G→T over C→A) or even reversed (excesses of C→G over G→C and G→A over C→T). At the same time, the differences in Fig. 1*A* disappear in Fig. 1*B*. In fact, the two parallel subpatterns (upper as well as lower in Fig. 1*B*) are almost perfectly correlated. The following six tests develop this surprising observation.

Testing the Strand-Specific p53 Patterns of Smokers and Nonsmokers.

The majority of p53 base substitutions in lung cancer involve G:C pairs, both in smokers (82.5%; 250 of 303) and in nonsmokers (82.9%: 58 of 70). Ninety-eight of 250 (39.2%) of these substitutions in smokers and 23 of 58 (39.6%) in nonsmokers occur at CpG sites. Two transitions in CpGs, C→T and G→A, are believed to originate from one event, namely, deamination of 5-mC, but in different strands (22, 23). If deamination occurs in the nontranscribed (coding) strand (NTS), the result is a **C**pG→**T**pG transition; if the mirror 5-mC is deaminated in the transcribed (noncoding) strand (TS), replication converts it to a complementary Cp**G**→Cp**A** transition in the coding strand. As to the complementary transversions (G→T, C→A and G→C, C→G), both presumed exogenous and endogenous sources, e.g., BPDE and oxidative DNA damage, respectively, point to modified Gs as primary lesions (reviewed in ref. 16). When such lesions occur in the NTS, they produce G→T and G→C transversions. Accordingly, the complementary C→A and C→G transversions most probably are derived from the similar lesions originating in the transcribed strand (24).

It thus appears that for more than 80% of p53 base substitutions in lung cancer (G:C→T:A, G:C→C:G, and C:G→T:A), we can identify the strand with a putative primary lesion (Fig. 1*B*). If so, the difference in the ratio of complementary patterns between smokers and nonsmokers with lung adenocarcinoma (Fig. 1*B*) reflects a strong, nontranscribed strand bias of p53 mutations in smokers. The NTS/TS ratio was calculated as (G→T + G→C + C→T)/(C→A + C→G + G→A) = 58/24 = 2.42 in smokers vs. 19/23 = 0.83 in nonsmokers. Similarly, the NTS/TS ratio = 74/35 = 2.1 for squamous cell (SC) and 24/7 = 3.42 for large cell (LC) carcinomas in smokers vs. 5/7 = 0.71 for the pooled (SC plus LC) data in nonsmokers.

This smoke-associated strand bias might result either from more active transcription-coupled repair or from the slower repair of the nontranscribed strand of *p53*. At any rate, to test the mutagenic role of BPDE-like carcinogens, it would seem rational to make the smoker vs. nonsmoker comparison of the p53 patterns not all-inclusive (Fig. 1*A*) but rather separately for each strand (Fig. 1*B*). If the lung cancer-specific p53 pattern is, indeed, formed mainly by BPDE-like mutagens, one would expect to find a significant difference between smokers and nonsmokers within each strand. However, these two epidemiological groups of patients with adenocarcinoma turned out to be absolutely indistinguishable in both the NTS and TS patterns formed by substitutions at G:C base pairs (*P* values for NTS and TS are 0.939 and 0.865, respectively). SC and LC carcinomas show the same negative result, but we cannot guarantee its
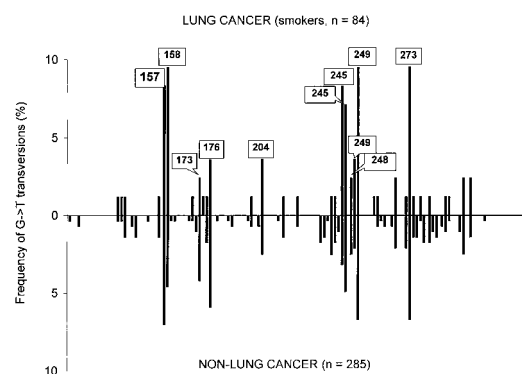

Fig. 2. Comparison of the p53 spectra of G→T transversions from lung cancer (radon-, asbestos-, and mustard gas-associated cases excluded) of ever smokers and cancers in non-lung tissues least accessible to smoke (see Table 1). Hot- and ''warm''-spot codons are indicated. Codon 248, one of the strongest BPDE targets, appears not to be a G→T transversion hot spot in either lung cancer or non-lung cancer spectra.

statistical significance because only 12 substitutions at G:C sites were available for nonsmokers with these malignancies.

Thus, whatever mechanisms might cause primary p53 lesions in lung, their pattern is reshaped by subsequent strand-biased repair in such a way that the proportion of G→T transversions in the standard representation of the p53 pattern (Fig. 1*A*) is notably higher in smokers than in nonsmokers. Consequently, it would be premature to attribute this prevalence to the mutational signature left by BPDE-like carcinogens.

Testing p53 Spectra of G→T Transversions in Lung Cancer of Smokers, Nonsmokers, and Non-Lung Cancers.

This surprising lack of within-strand differences prompted us to look closer at how the lung cancer p53 mutational spectrum of smokers was compared (and contrasted) so far to (*i*) the lung cancer p53 spectrum of nonsmokers and (*ii*) the p53 spectrum of cancers in smoke-inaccessible tissues. The BPDE hypothesis predicts that these two control groups will show different p53 spectra for G→T transversions whereas the alternative selection hypothesis implies no such difference. Therefore, to find out whether origin or selection of mutations is the major shaper of the lung-specific p53 spectrum, one should focus on G→T transversions for both smoker vs. nonsmoker and lung vs. non-lung comparisons because it is precisely G→T transversions that BPDE predominantly generates. Instead, all comparisons published so far (3, 12, 13) have been lumped together, thus making the spectrum-shaping roles of BPDE and selection indiscernible.

First, we compared the G→T spectra for lung cancers of smokers and nonsmokers. The Adam and Skopek test yielded a *P* value of 0.2968 (95% confidence limit in the 0.2878–0.3058 range), indicating that the spectra are statistically indistinguishable. However, only 12 G→Ts (including 3 associated with radon) represented the control group of nonsmokers. Therefore, we followed up the above test with an indirect but more powerful and ultimately more convincing one by contrasting the two distributions of G→T transversions along the *p53* gene, one for lung tissue tumors of smokers with the largest portion of G→Ts (30%; 84 of 286) and the other for the combined set of presumably smoke-inaccessible tissues with the lowest portion of G→Ts (7%; 285 of 4,046; Table 1). The result is shown in Fig. 2. The two spectra (84 vs. 285 G→Ts) are, in fact, identical. The Adams and Skopek test yielded a *P* value of 0.9793 (95% confidence limit in the 0.9765–0.9821 range). When we compared the non-lung spectrum of 285 G→Ts with that of 256 G→Ts from the total lung cancer dataset, we obtained a *P* value of 0.9072 (0.9015–0.9129 range), indicating no difference, as
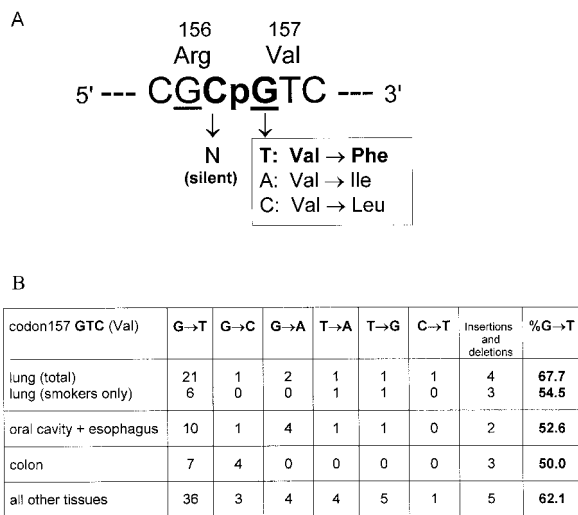
A

156   157
Arg   Val

5' --- CG**C**pG**TC** --- 3'

↓   ↓
N   T: **Val → Phe**
(silent) A: Val → Ile
          C: Val → Leu

B

codon157 GTC (Val)	G→T	G→C	G→A	T→A	T→G	C→T	Insertions and deletions	%G→T
lung (total)	21	1	2	1	1	1	4	67.7
lung (smokers only)	6	0	0	1	1	0	3	54.5
oral cavity + esophagus	10	1	4	1	1	0	2	52.6
colon	7	4	0	0	0	0	3	50.0
all other tissues	36	3	4	4	5	1	5	62.1

Fig. 3. (A) The scheme explaining high selectivity of G→T compared with all other changes at the CpG composed by codons 156 and 157. Underlined are guanines that show strong BPDE binding (10). (B) p53 mutational pattern of codon 157 showing cancer-common excess of G→T transversions; radon-associated cases are excluded.

well. Likewise, all other types of mutations separately mapped along *p53* (not shown) appeared to be statistically indistinguishable for these two groups. Thus, although the overall pattern of p53 mutations as a weighted combination of different mutation types can be tissue-specific, the p53 spectrum for each given mutation type emphatically is not. This unambiguously suggests that the major shaper of specific G→T spectra is not smoking-induced adducts but something else entirely, namely, selection.

The Mutational Pattern at Codon 157 Is Not Unique to Lung Cancer. In the p53 spectra limited to G→Ts, codon 157 is equally frequent in lung and non-lung cancers (Fig. 2). However, codon 157 appears to be a hot spot unique to lung cancer in previous reports (3, 12, 13) in which mutations of different types were pooled together along *p53*. Fig. 3 explains this "paradox." Two neighboring codons, 156 (CG**C**) and 157 (**G**TC), form a singular CpG between them (Fig. 3*A*). The guanine of this CpG is one of the most preferred BPDE-binding targets (10) and also is a G→T transversion hot spot in lung cancer. However, it was overlooked that in this particular codon, G→T prevails over any other type of mutations virtually in all tumor tissues (Fig. 3*B*). Colon tissue may serve as the most instructive example because this tissue rarely contacts benzo[*a*]pyrene and the overall p53 pattern of colon cancer is notable for having more than 50% of endogenous C→T/G→A transitions at CpG sites. However, in codon 157, both of these cancers show virtually the same excess of G→Ts: 54.5% in lung vs. 50% in colon (Fig. 3*B*). In fact, this excess is not lung-specific at all (Fig. 3*B*). The explanation suggests itself: compared with G→Ts, all other substitutions in codon 157 are much less tumorigenic.

p53 mutations prevalent in non-lung cancers are CpG transitions, including those in hot spots such as codons 248 (CGG) and 273 (CGT). Both C→T and G→A transitions at these CpGs lead to detrimental amino acid substitutions. The CpG formed by codons 156 and 157 (Fig. 3*A*) is different. Each substitution of the third base of codon 156, including C→T, is silent; G→A at the first position of codon 157 results in Val→Ile, a conservative substitution of one aliphatic amino acid (Val) for another, functionally similar one (Ile). This explains why C→T and G→A at this CpG are so rare, not only in lung tumors with dominating G→T transversions, but even in tumor tissues with dominating

CpG transitions (Fig. 3*B*). In contrast, G→T transversions result in an apparently more serious change: aromatic Phe instead of the original aliphatic Val. This is why G→T transversions "represent" codon 157 not only in lung but in most other tumor tissues.

Low tumorigenicity of G→A transition in codon 157 is evidenced by its frequency in tumors with multiple mutations. According to our theory of the origin and selection of multiple p53 mutations (25, 26), a p53 mutation of zero or low tumorigenicity can be carried in the tumor cell as a hitchhiker by another highly tumorigenic p53 mutation. The theory predicts that such zero- or low-tumorigenicity mutations will be more frequent in tumors when present in *p53* as "multiplets" rather than alone, as "singlets." This is exactly what we observed for Val157→Ile mutations caused by G→A; 40% of them (4 of 10) appear in tumors as p53 multiplets. By contrast, Val157→Phe derived from G→T is four times less frequent in multiplets at only 8.2% (6 of 73).

When p53 mutations of all types are lumped together in the same spectrum (see figure 3 in ref. 3), the roles of codon- and mutation-specific selection are slurred. Codon 157 is the most illustrative case. In lung cancer, G→Ts dominate over other mutations; therefore, codon 157 is as prominent as codon 273. But in other tumor tissues (such as colon), an overwhelming number of CpG transitions simply shadows infrequent G→T transversions so that the codon 157 mutation frequency sharply declines in comparison with codon 273 and other CpG hot spots. It happens not because CpG transitions do not originate in codon 157 but, rather, because they are not selectable in all cancers. The previous spectral analyses *en masse* (3, 12, 13) created the wrong impression that codon 157 gets changed only in lung cancer and that its high affinity to BPDE brings about its "uniqueness" in lung cancer.

Silent p53 Mutations in Tumors Do Not Correlate with BPDE Adducts. Most silent p53 mutations in human tumors seem to be selectively neutral somatic changes (27). Therefore, they turn up in tumors only as hitchhikers of nonsilent, tumor-driving mutations (25, 26). The hypothesis of BPDE-induced p53 mutations predicts that lung cancers should contain many more silent G→T transversions in smokers than in nonsmokers and more often in codons 248, 267, and 282 (all of CGG structure) because their second guanines show rather strong BPDE adduct signals (10). Free of interference by selection, silent G→Ts in the *p53* gene as a whole (and in these three codons in particular) are more suitable than nonsilent G→Ts for testing BPDE-induced mutagenesis in tumors. In reality, of 23 somatic silent p53 mutations in smokers (pooled for lung, oral cavity, and esophagus), there is only 1 G→T (in esophagus), and it is located in codon 173 (G**T**G), which is not a BPDE target (10).

The p53 database (14) contains 368 silent mutations. Thirty-four of these 368 are detected in lung cancers. Lung cancers demonstrate the highest proportion of silent G→T substitutions, 21% (7 of 34), in contrast to only 2.7% (9 of 334) in all non-lung cancers. However, these seven silent G→Ts, all from lung cancer patients of unspecified smoking status, occur in codons 113, 125, 173, 203, and 302. None of these codons has been reported as a BPDE target (10, 28).

The same logic is applicable to codon 156. Its internal CpG (Fig. 3*A*) is the most preferable target for BPDE in exon 5 (10). However, codon 156 is a mutational "cold spot" in all human cancers. The usual explanation is the relative unimportance of Arg156 in p53 functions (10, 28), meaning that any base substitution here is effectively a silent one. At any rate, none of the currently reported 13 mutations in codon 156 (including 10 in the CpG) for tumors in smoke-sensitive tissues (lung, esophagus, and oral cavity) happens to be a G→T or C→A, excess of either being a BPDE mutational signature.

Table 2. Frequency of tumors with silent p53 mutations in smokers and nonsmokers

Type	Total no. of tumors with p53 mutations	Tumors with a single p53 mutation		Tumors with multiple p53 mutations		Frequency[¶] of tumors with silent p53 mutations, %	
		No. of nonsilent mutations	No. of silent mutations	No. of nonsilent mutations	No. of silent mutations	Singlets	Total
Lung (63% vs. 31%)*							
Smokers total	328	306	8	26	4	1.54	2.30
Nonsmokers total	73	67	4	4	0	1.70	1.70
Smokers only[†]	278	263	8	15	1	1.81	2.04
Nonsmokers only[†]	68	63	4	2	0	1.82	1.82
Oral cavity + esophagus (57% vs. 33%)*							
Smokers total[‡]	316	287	6	43	5	1.08	1.98
Nonsmokers total	84	78	3	6	0	1.18	1.18
Smokers only[‡§]	170	154	6	20	1	2.01	2.35
Nonsmokers only[§]	46	41	3	4	0	2.15	2.15

*Shown in parentheses are the percentages of tumors with p53 mutations in smokers and nonsmokers, respectively.

[†]Cases of radon, asbestos, and mustard gas exposure excluded.

[‡]Include the cases of snuffing and chewing tobacco.

[§]Cases of drinking alcohol and hot beverages excluded.

[¶]For each group, the frequency was calculated as: $F = [(\text{no. of tumors with p53 mutations of interest})/(\text{total no. of tumors in the group})] \times 100\%$. The latter was estimated as $[(\text{no. of tumors with p53 mutations in the group})/\% \text{ of these tumors}] \times 100\%$. For example, the overall frequency of lung tumors with silent p53 mutations (single ones + those in p53 multiplets) in the group "smokers total," $F = [(8 + 4)/(328 \times 100\%/63\%] \times 100\%) = (12/328) \times 63\% = 2.3\%$.

The Hypothesis of Smoke-Induced Selection of p53 Mutations. Regarding the mutagenesis vs. selection alternative, it is important to distinguish the following two comparisons of p53 mutations in smokers and nonsmokers. The first comparison quantitatively sets off smokers against nonsmokers within the group of lung cancers carrying p53 mutations. The overall ratio is 344:75 = 4.6:1, tumors with multiple p53 mutations being particularly contrasting (Table 2). It may seem quite natural to relate this difference solely to direct mutagenic effects of smoking. However, this might be a simple effect of the difference in the number of cell divisions because smoking is known to increase the rate of cell proliferation in the bronchial epithelium. Moreover, the role of selection cannot be excluded, even in the case of p53 multiplets. The point is that cells use the *p53* gene nonconstitutively, only under stressful conditions (4, 5), in response to either DNA-damaging insults like ionizing radiation and drugs, or, more commonly, "nongenotoxic" hypoxia (4, 17). Wild-type p53 protein prevents tumor development either by committing stressed cells to apoptosis or by arresting them mainly at the G_1 stage of the cell cycle, before DNA replication (29, 30). Cells with mutant p53 often escape apoptosis and G_1 arrest and enter the S phase followed by cell division and proliferation. Our previous retrospective analysis of tumors with multiple p53 mutations (26) showed that the majority of their primary, mutation-prone changes also might originate in nondividing stem cells, during p53-regulated G_1 arrest, i.e., precisely when a potential carcinogenic advantage of mutant p53 protein over a normal one may be brought into effect by selection. If so, the higher frequency of p53 mutations in smokers (Table 1) may reflect not only the smoke-induced mutagenesis but also the increased selection pressure on the *p53* gene in tissues insulted by smoking.

The second difference between smokers and nonsmokers points more definitely to increased selection rather than mutation rate as the major driving force of lung tumorigenesis. It takes into account the frequency of lung cancers carrying p53 mutations in smokers compared with nonsmokers. From published data we found that, of all lung cancer cases, the ones possessing p53 mutations constitute 63% in smokers vs. only 31% of cases in nonsmokers (Table 2; see also ref. 31). The ratio is, therefore, 63:31 = 2.0. The similar ratio (57:33 = 1.73) distinguishes oral cavity and esophagus cancers (Table 2). Usually, this difference

also is interpreted as a strong evidence of smoke-induced mutagenesis. In fact, this evidence is invalid. Clearly, an elevated mutation rate may make the somatic malignant evolution faster, but is not necessary for malignancies to occur (32). Accordingly, it is not so clear why the increase of mutation rate as such should change the proportion of tumors with and without mutated *p53*. On the contrary, a smoke-raised selection pressure on stress-responsive genes such as *p53* readily explains the difference. Indeed, as soon as the p53⁻ mutant clone emerges, the probability of further malignant growth is determined by its proliferative advantage over p53⁺ cells around it. The more often stresses such as hypoxia (17) call on the *p53* gene, the higher the probability is for a p53 mutation to be selected, which, in turn, leads to the increase of the final portion of cancers with p53 mutation. Moreover, proliferative advantage depends not only on the tumorigenicity of a p53⁻ mutant cell *per se* but also on nongenotoxic damage inflicted on adjacent p53⁺ cells. Smoking may make the latter less competitive, and, as a result, the same p53 mutations are selected more often in lung tumors of smokers than nonsmokers.

Silent p53 Singlets in Tumors as Markers of Mutability. Finally, we address the key question of whether tobacco smoking raises a mutation rate in *p53* during lung tumorigenesis. All of the above comparisons were inconclusive in this regard because of direct or indirect effects of selection. The only class of p53 mutations that can be used for this purpose are single, silent base substitutions (Table 2). Because they are not accompanied by nonsilent p53 mutations, changes in other genes most likely "picked them up" during tumorigenesis. This means that silent p53 singlets represent tumors driven by p53-independent selection. Then, the frequency of tumors with silent p53 singlets reflects only the mutation rate in *p53*. Remarkably, smokers and nonsmokers turned out to be indistinguishable in this frequency (Table 2).

A total frequency of tumors with silent p53 mutations (i.e., singlets + multiplets) is higher in smokers than in nonsmokers (last column in Table 2). However, the synergistic effects of smoking and additional cancer risk factors (lung: exposure to radon/asbestos/mustard gas, oral cavity and esophagus: drinking alcohol and hot beverages) may account for the difference. At any rate, this difference between smokers and nonsmokers disappears when the additional factors are excluded (Table 2).

The question is whether these synergistic effects are mainly

genotoxic or/and caused by stress-induced selection? Available now p53 mutational data for two crucial control groups (such as nonsmokers exposed to radon and nonsmoking drinkers) are too small to examine this alternative for all p53 multiplets. However, we prefer the selection-based explanation because the frequency of silent p53 singlets is the same in smokers and in nonsmokers, independently of additional risk factors (Table 2).

Concluding Remarks. There is a tendency of current cancer research to downplay rather evident physiological insults caused by exogenous carcinogens in favor of their cancer-prone mutagenic action. Primarily nongenotoxic, these insults nevertheless change conditions for genetic processes such as strand- and sequence-specific repair and selection of tumorigenic mutations in genes such as *p53*. What if physiologically induced repair and selection may affect not only the base-to-base spectrum of p53 mutations, but even their overall pattern? Comparison of the standard (Fig. 1*A*) and our new graphic representation of p53 mutational patterns (Fig. 1*B*) gives a positive answer. Parallel patterns in Fig. 1*B* may reflect repair-associated differences between the two complementary DNA strands of *p53* that acquire mutations and how smoking, as many other stressors of cells, exacerbates this difference.

In all fairness, it should be noted that BPDE adducts on the transcribed strand are repaired faster than those on the non-transcribed strand. The guanines of codons 157, 248, and 273 are repaired two to four times slower when compared with other targets (28). However, closer examination of the BPDE repair rate profile for both DNA strands throughout *p53* exons 5, 7, and 8 (28) showed no correlation with specific transversions in the lung cancers of smokers, either with G→Ts representing the nontranscribed strand (correlation coefficient, $r = -0.13$) or with C→As representing the transcribed strand ($r = -0.53$). The strand-biased repair of targeted BPDEs seems to be largely irrelevant here, because the same bias distinguishes smokers from nonsmokers in two other pairs of complementary substitutions (G→C, C→G and C→T, G→A). Moreover, we have observed exactly a similar "big difference of *p53* patterns between strands; no difference within each strand" trend for esophagus and oral cavity cancers when comparing the patients with the history of alcohol and/or hot beverage consumption with the control group of nondrinkers. Needless to say, no

mutagenic action is suspected in this case. We will review these data in detail elsewhere.

Our results are consistent with two observations, namely, (*i*) no difference between smokers and nonsmokers in cancer-"neutral" *hprt* mutational spectra has been reported recently (33), and (*ii*) direct treatment with a carcinogenic dose of NMU (nitroso-methylurea), a suspected mammary tumor-specific strong mutagen, did not raise the mutational rate in the *Hras* gene in the experimentally NMU-induced rat mammary tumors (34). In fact, our hypothesis of a smoke-caused increase of physiological selection pressure on the stress-responsive *p53* gene develops further the ideas outlined in refs. 33 and 34.

Because even lung cancers of nonsmokers exhibit a significant excess of G:C→T:A transversions compared with most non-lung cancers, we assume that some primary causes, other than BPDE-adduct formation, exist that have a similar mutational signature (prevalence of transversions at Gs) but are lung-specific rather than smoking-specific. A logical candidate is oxidative DNA damage (15, 16). Consistent with this is a general view that the majority of somatic p53 mutations in tumors are of endogenous origin (11, 35, 36). Indeed, silent p53 mutations do not fit the hypothesis of smoke-raised mutation rate in general and the BPDE-adduct spectrum in particular. However, because the sample of available silent p53 mutations in lung cancer is still too small for reliable analysis, the approach outlined in refs. 25 and 37 can be used to reach the final verdict. It has three major stages: (*i*) measurement of the primary damage along *p53* caused by the carcinogen in question, (*ii*) multiplication of the primary damage frequency at each site by the repair rate value at that site, and (*iii*) multiplication of the postrepair damage frequencies by selection coefficients calculated for each site and the particular type of mutations the carcinogen in question causes. If, and only if, the product of these two multiplications is positively correlated with the real p53 mutational spectrum, it then will be possible to demonstrate a causal link between a tested mutagen and specific cancer. The third step (measuring selection) is a difficult but ultimately solvable task (25). The corresponding analysis is underway.

1. Doll, R. & Hill, A. B. (1999) *Bull. W. H. O.* **77,** 84–93.
2. Greenblatt, M. S., Bennett, W. P., Hollstein, M. & Harris, C. C. (1994) *Cancer Res.* **54,** 4855–4878.
3. Nernandez-Boussard, T. M. & Hainaut, P. (1998) *Environ. Health Perspect.* **106,** 385–391.
4. Kinzler, K. W. & Vogelstein, B. (1996) *Nature (London)* **379,** 19–20.
5. Levine, A. J. (1997) *Cell* **88,** 323–331.
6. Vineis, P. & Caporaso, N. (1995) *Environ. Health Perspect.* **103,** 156–160.
7. Friedberg, E. C., Walker, G. C. & Siede, W. (1995) in *DNA Repair and Mutagenesis* (Am. Soc. Microbiol., Washington, DC), pp. 39–41.
8. Maher, V. M., Yang, J. L. & McCormick, J. J. (1990) *Acta Biol. Hung.* **41,** 173–186.
9. Ruggeri, B., DiRado, M., Zhang, S. Y., Bauer, B., Goodrow, T. & Klein-Szanto, A. J. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1013–1017.
10. Denissenko, M. F., Pao, A., Tang, M. & Pfeifer, G. P. (1996) *Science* **274,** 430–432.
11. Krawczak, M. & Cooper, D. N. (1998) *Mutagenesis* **13,** 319–320.
12. Bennett, W. P., Hussain, S. P., Vahakangas, K. H., Khan, M. A., Shields, P. G. & Harris, C. C. (1999) *J. Pathol.* **187,** 8–18.
13. Pfeifer, G. P., Denissenko, M. F. & Tang, M. (1998) *Mutagenesis* **13,** 537–538.
14. Hernandez-Boussard, T., Rodriguez-Tome, P., Montesano, R. & Hainout, P. (1999) *Hum. Mutat.* **14,** 1–8.
15. Lindahl, T. (1996) *Philos. Trans. R. Soc. London B. Biol. Sci.* **351,** 1529–1538.
16. Loft, S. & Poulsen, H. E. (1996) *J. Mol. Med.* **74,** 297–312.
17. Graeber, T. G., Osmanian, C., Jack, T., Housman, D. E., Koch, C. J., Lowe, S. W. & Glaccia, A. J. (1996) *Nature (London)* **379,** 88–91.
18. Gealy, R., Zhang, L., Siegfried, J. M., Luketich, J. D. & Keohavong, P. (1999) *Cancer Epidemiol. Biomarkers Prevention* **8,** 292–302.
19. Gao, H.-G., Chen, J.-K., Stewart, J., Song, B., Rayappa, C., Whong, W.-Z. & Ong, T. (1997) *Carcinogenesis* **18,** 473–478.
20. Adams, W. T. & Skopek, T. R. (1987) *J. Mol. Biol.* **194,** 391–396.
21. Cariello, N. F. (1994) *Mutat. Res.* **312,** 173–185.
22. Rideout, W. M., Coetzee, G. A., Olumi, A. F. & Jones, P. A. (1990) *Science* **249,** 1288–1290.
23. Yang, A. S., Jones, P. A. & Shibata, A. (1996) in *Epigenetic Mechanisms of Gene Regulation* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 77–94.
24. Kuipers, G. K., Poldervaart, H. A., Slotman, B. J. & Lafleur, M. V. (1999) *Mutat. Res.* **435,** 141–150.
25. Rodin, S. N., Holmquist, G. P. & Rodin, A. S. (1998) *Int. J. Med. Med.* **1,** 191–199.
26. Rodin, S. N. & Rodin, A. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 11927–11932.
27. Strauss, B. S. (1998) *Genetics* **148,** 1619–1636.
28. Denissenko, M. F., Pao, A., Pfeifer, G. P. & Tang, M. (1998) *Oncogene* **16,** 1241–1247.
29. Linke, S. P., Clarkin, K. S. & Wahl, G. M. (1997) *Cancer Res.* **57,** 1171–1179.
30. Polyak, K., Xia, Y., Zweler, J. L., Kinzler, K. & Vogelstein, B. (1997) *Nature (London)* **389,** 300–305.
31. Takagi, Y., Osada, H., Kuroishi, T., Mitsudomi, T., Kondo, M., Niimi, T., Saji, S., Gazdar, A. F., Takahashi, T., Minna, J. D., *et al.* (1998) *Br. J. Cancer* **77,** 1568–1572.
32. Tomlinson, I. & Bodmer, W. (1999) *Nat. Med.* **5,** 11–12.
33. Curry, J., Karnaoukhova, L., Guenette, G. C. & Glickman, B. W. (1999) *Genetics* **152,** 1065–1077.
34. Cha, R. S., Thilly, W. G. & Zarbl, H. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3749–3753.
35. Krawczak, M., Smith-Sorensen, B., Schmidtke, J., Kakkar, V. V., Cooper, D. N. & Hovig, E. (1995) *Hum. Mutat.* **5,** 48–57.
36. Cairns, J. (1998) *Genetics* **148,** 1433–1440.
37. Holmquist, G. P. & Gao, S. (1997) *Mutat. Res.* **386,** 69–101.

MEDICAL SCIENCES