Genetic pathways and mutation profiles of human cancers: site- and exposure-specific patterns

I.A. Lea, M.A. Jackson, X. Li, S. Bailey, S.D. Peddada1 and J.K. Dunnicke2*

Integrated Laboratory Systems, Inc., Research Triangle Park, NC 27709, USA and 1National Institutes of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

*To whom correspondence should be addressed. Tel: +1 919 541 4811; Fax: +1 919 541 4255; Email: dunnickj@niehs.nih.gov
Correspondence may also be addressed to I.A. Lea. Tel: +1 919 544 4589; Fax: +1 919 544 0380; Email: ilea@ils-inc.com

Cancer is a complex disease that involves the accumulation of both genetic and epigenetic alterations of numerous genes. Data in the Genetic Alterations in Cancer database for gene mutations and allelic loss [loss of heterozygosity (LOH)] in human tumors (e.g., lung, oral, esophagus, stomach and colon/rectum) were reviewed. Results for the genes and pathways implicated in tumor development at these sites are presented. Mutation incidence, spectra and codon specificity are described for lung, larynx and oral tumors. LOH occurred more frequently than gene mutations in tumors from all sites examined. The cell cycle gene, TP53 (all sites), and cell signaling gene, APC (colorectal and gastric cancers), were the only genes with similar incidences of LOH and mutation. Alterations of one or more cell cycle and cell signaling genes were reported for tumors from each site. Site-specific activation was apparent in the cell signaling mitogen-activated protein kinase oncogenes (KRAS in lung, HRAS in oral cancers and BRAF in esophageal and colorectal cancers). Analysis of genetic changes in lung tumors showed that the incidence of mutations in the TP53 and KRAS genes and the incidence of LOH in the FHIT gene were significantly greater in smokers versus non-smokers (P < 0.01). In lung and oral cancers, the TP53 GC → TA transversion frequency increased with tobacco smoke exposure (P < 0.05). Furthermore, the TP53 mutational hot spots for lung and laryngeal cancers from smokers included codons 157, 245 and 273, whereas for oral tumors included codons 280 and 281.

Introduction

One of the cornerstones to understanding cancer is knowledge of the genes and mutations that initiate tumor development and give cells a selective growth advantage that allows tumor progression to occur. It is now well established that a clone of cancerous cells is unlikely to be the result of a single mutational event in a stem cell but rather a series of accumulated genetic and epigenetic events induced or inhibited by the interplay of DNA and environment (1–3). Single nucleotide polymorphisms are one mechanism whereby the genetic background has been implicated in defining an individual’s susceptibility to cancer; the Tp53 Arg/Pro polymorphism at codon 72 is one example (4–7), yet external environmental exposures also play an important role. These can give rise to a number of different genetic alterations, the most common of which include: chromosomal changes such as loss of heterozygosity (LOH), rearrangements, deletions, gains, polyploidy and translocations; gene mutations such as base substitutions, small insertions and deletions, allelic loss, amplification and rearrangements; and epigenetic events such as alteration in DNA methylation.

Perhaps the best-documented external factor associated with human cancer is that of exposure to tobacco smoke and lung cancer. Fifty years ago, Doll and Hill (8) first demonstrated a dose–response relationship between the number of cigarettes smoked and lung cancer death rate. Subsequent studies have established several molecular targets for tobacco smoke including the TP53, KRAS, FHIT, RB1 and HPRT genes (9–12). Recent studies continue to show that smoking is associated with cancer risk including not only lung cancer risk but also risk for oral, larynx, esophageal, gastric cancers and colorectal hyperplasia/polyps (13–15).

To help advance the understanding of how different genes contribute to tumorigenesis, databases have been designed to collect and report gene mutation data for many types of cancers. One of the best-known databases is the International Agency for Research on Cancer TP53 Mutation Database (http://www-p53.iarc.fr/). Others include the CDKNA2A (https://biodesktop.uvm.edu/perl/p16) and androgen receptor gene databases (http://www.androgendb.mcgill.ca); the breast cancer gene (http://www.tumorgeone.org/Breast/index.html) and the oral cancer gene (http://www.tumorgeone.org/Oral/oral.html) databases; and the molecular cytogenetics database (http://www.ncbi.nlm.nih.gov/sky/). These resources provide important data for single genes and/or cancers at single tumor sites, but in order to better understand the similarities and differences in tumorigenesis at different sites and define the types of alterations and accumulated genetic events that contribute to the tumor development, databases with information for multiple genes, cancer topographies (anatomical sites of the tumors), alterations and exposures are critical. Data from one such source, the Genetic Alterations in Cancer (GAC) database (http://dir-apps.niehs.nih.gov/gac/), are evaluated and summarized in this paper.

The GAC database was developed by National Institute of Environmental Health Sciences as a web-based system for collecting and summarizing data reported in the web-sighted literature for genetic alterations in various types of human cancers and laboratory rodent tumors (16). It provides a unique tool for assessing groups of related genes for mutations, allelic loss and homozygous deletions based on species and tumor site. Tumor morphologies from different topographical sites (regions of the body) and exposures to environmental agents can also be distinguished. Collating these types of data allows a broader view of tumor development and facilitates comparative analysis of multiple genes in genetic pathways to cancer rather than focusing on single genes.

Data in the GAC database for genetic alterations in human tumors were evaluated with an emphasis on results from smokers compared with non-smokers. Tumors sites were selected based on their potential association with tobacco smoke exposure (13–15). The results are summarized to demonstrate how disruption of genetic pathways, such as cell cycle control and cell signaling, by mutations and allelic loss in one or more genes can contribute to tumor development. This analysis is designed to provide further insight into the modes of action of carcinogenesis. We illustrate how changes in specific genes, and different patterns of gene alterations in various topographical sites, may serve as markers of exposure.

Materials and methods

Data selection

Mutation data were obtained from the GAC database (http://dir-apps.niehs.nih.gov/gac/) which was last updated in April 2007 (containing ~3175 records). A complete description of GAC, including data acquisition, was described previously (16).

Tumor groups were defined as ‘Spontaneous’ (no known exposure reported), ‘Tobacco (smoke)’ (samples from smokers, either current or previous cigarette/pipe) and ‘Unexposed (Tobacco)’ (samples from non-smokers reported to have never smoked). The datasets were unrestricted with regard to tumor origin (primary, metastatic or recurrent tumors) and morphology.

Abbreviations: GAC, Genetic Alterations in Cancer; LOH, loss of heterozygosity; MAPK, mitogen-activated protein kinase.
Data from human tumors evaluated for somatic mutations (missense, nonsense and silent point mutations, frameshift insertion and deletions, in-frame insertions and deletions) or LOH (single allele loss of intragenic markers or markers tightly linked to a gene) were selected for evaluation. The results are presented as separate datasets; bi-allelic alterations in single tumors are not shown. Genes included in the analysis are shown in Table I and are grouped in the cell cycle or cell signaling pathway in which they play a role (this is not necessarily their only pathway). Lists of references from which datasets were obtained are given as supplementary data (available at Carcinogenesis Online).

Statistical analysis Gene mutation and LOH data for human lung, oral, esophageal, stomach and colorectal tumors were retrieved from the GAC database of peer-reviewed literature. To analyze the incidence of alterations, data were organized by tumor site and agent group [spontaneous, tobacco smoke, unexposed (tobacco)]. Data from detailed data lists showing the results for each tumor sample from all studies retrieved were generated by the program, captured and prepared for broad-based statistical analysis. This included data from a total of 181 peer-reviewed studies of lung tumors, 83 studies of oral tumors, 133 studies of esophageal tumors, 163 studies of stomach tumors and 233 studies of colon/rectum tumors from smokers, non-smokers or individuals for which smoking status was not reported (references to all studies are in the supplementary data available at Carcinogenesis Online).

The mutation incidence was compared for pairs of data groups (e.g. data for TP53 or KRAS or APC, etc. from lung versus oral tumors, lung versus esophageal tumors, oral versus stomach tumors, etc. or lung tumors from smokers versus non-smokers) using a standard two-sample Z-statistic for proportions. However, since sample proportions were derived from multiple studies and variables between studies (e.g. age, gender, exposure, and other environmental factors) were not specifically factored in, extra-variation between studies was taken into account by using a non-parametric estimator for extra binomial variability (17). The P values for the tests were derived using a non-parametric bootstrap methodology (18). TP53 mutational spectra were compared for smokers and non-smokers in lung and oral cancers by the Chi-square test.

Results

Altered genetic pathways

Data for the incidence of somatic point mutations, in-frame and frameshift insertions and deletions and LOH at gene-specific markers in human lung, oral, esophageal, gastric and colorectal cancers are summarized in Table I. The data are from studies of spontaneous tumors (exposure to potential carcinogens was not reported) and are organized by tumor site and genetic pathway. Gene mutations were universally studied more often than LOH for all of the genes and tumor sites, yet LOH generally occurred at higher frequencies.

Cell cycle genes Disruption of the cell cycle was a common event in all cancers studied (Table I). This was primarily the result of alterations in the TP53 gene, which occurred through both LOH and point mutations in approximately 40–50% of tumors studied. In other cell cycle genes, allelic loss was typically found more often than point mutations (see the CDKN2A gene for oral, esophageal and colon/rectum tumors and RB1 gene for lung, esophagus and colorectal cancers).

Allelic loss of BRCA1, CDKN2A, RB1 and TP53 was detected in colorectal cancer but mutations were only reported for TP53. LOH was also reported for these same four genes in esophageal tumors, but at a higher frequency than in colorectal cancer. Mutations were seen in TP53 as well as in the CDK4 inhibitor genes, CDKN2A and CDKN2B. In gastric tumors, genes that were studied for either allelic loss or mutations demonstrated low incidences of mutations in CDKN2A and CDKN2B. Moreover, the incidence of TP53 mutations was significantly lower at this site than at all the other topographic sites studied (P < 0.01). Fewer studies examined the cell cycle genes in oral cancer, but of those that did, CDKN2A mutations were detected in 7% of the cases, CDKN2B mutations were not found and LOH in RB1 occurred in 12% of cases. TP53 represented the majority of the data for this group with mutations in 39% and LOH in 48% of the tumors. Lung tumors showed a high incidence of TP53 and RB1 allelic loss and some evidence of CDKN2B loss but mutations only in TP53 (42%) and CDKN2A (9%).

Cell signaling: Wnt Activation of the Wnt pathway in tumor development is most often studied through the APC and CTNBB1 genes; however, at the topographical sites included in Table I, there was little evidence of CTNBB1 mutations reported in the studies reviewed. In lung and oral tumors, only two of >250 tumors analyzed had mutations and in digestive tract cancers, the incidence of mutation was ~5% for >2000 tumors. Studies of LOH at these tumor sites were not available.

Potential loss of APC function was indicated by the presence of both mutations and LOH in colon/rectum (36% mutations and 33% LOH) and gastric (18% mutations and 21% LOH) cancers. In contrast, for lung, oral and esophageal tumors APC alterations were reported most frequently as LOH rather than mutations. The incidences of allelic loss were 28% (lung) and 48% (oral and esophagus) and were at a level comparable with that in colorectal tumors but gene mutations occurred in <5% of the tumors analyzed.

Cell signaling: mitogen-activated protein kinase/extracellular signal-regulated kinase Activation of the mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase pathway oncogenes through mutation or allelic loss occurred most frequently in colorectal cancer, primarily through mutations of the KRAS (31%) and BRAF (14%) genes (Table I). In gastric cancers, the MAPK pathway was rarely activated by gene mutations but evidence from a low number of tumors suggests that HRAS allelic loss may occur. Similar allelic losses were detected in esophageal, oral and lung cancers but further studies are needed to verify this. Likewise, evidence based on a low number of tumors suggests that the BRAF gene may be activated more frequently in esophageal tumors than the RAS genes, which were only rarely activated.

The incidence of KRAS mutations was significantly higher (P < 0.01) in spontaneous lung tumors (19%) compared with oral tumors (4%). Overall, KRAS mutations occurred at a significantly higher incidence in colon/rectum tumors (31%) than at any other site examined. HRAS mutations were more prevalent in oral tumors than lung or digestive tract tumors, reaching significance in esophagus (P < 0.05) and stomach (P < 0.05); this has been linked to tobacco exposure, primarily chewing tobacco. (19). BRAF and NRAS alterations were uncommon in lung, oral and gastric tumors.

Cell signaling: apoptosis Other genes implicated in tumorigenesis include DCC and FHIT; both are thought to be tumor suppressor genes. DCC functions as a dependence receptor required for apoptosis induction (20). FHIT also induces apoptosis (21). In digestive tract tumors (esophagus, stomach and colon/rectum), allelic losses of DCC or FHIT were detected in 26–48% of all cases (Table I). In lung tumors, loss of a FHIT allele was detected in 56% of cases and evidence based on a low number of tumors suggested that DCC allele loss was also frequent (41%). Results from one study of lung tumors indicated that point mutations in FHIT are rare (22); mutations were also infrequent in studies of stomach tumors (6 of 121 tumors) but not esophageal cancer (34% of 89 tumors had mutations).

Gene pathways in lung cancer of smokers and non-smokers

One of the strongest etiological factors for the development of tumors in lung and head and neck cancers is tobacco smoke. Although tumors shown in Table I were considered spontaneous in nature, the group undoubtedly comprised smokers and non-smokers. Exposures to polycyclic aromatic hydrocarbons including benzo[a]pyrene, aromatic...
<table>
<thead>
<tr>
<th></th>
<th>Lung</th>
<th>Oral</th>
<th>Esophagus</th>
<th>Stomach</th>
<th>Colon/rectum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% tumors with mutations(^a)</td>
<td>% tumors with LOH</td>
<td>% tumors with mutations</td>
<td>% tumors with LOH</td>
<td>% tumors with mutations</td>
</tr>
<tr>
<td>Cell cycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRCA1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>9 (45/494)(^b)</td>
<td>0 (0/146)</td>
<td>0 (0/32)</td>
<td>0 (0/32)</td>
<td>0 (0/32)</td>
</tr>
<tr>
<td>CDKN2B</td>
<td>25 (2/8)</td>
<td>7 (24/339)</td>
<td>60 (3/5)</td>
<td>60 (3/5)</td>
<td>60 (3/5)</td>
</tr>
<tr>
<td>RB1</td>
<td>0 (0/21)</td>
<td>47 (73/155)</td>
<td>NA</td>
<td>12 (3/26)</td>
<td>0 (0/49)</td>
</tr>
<tr>
<td>Cell signaling: Wnt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTNNB1</td>
<td>0.8 (2/251)</td>
<td>0 (0/33)</td>
<td>0 (0/33)</td>
<td>0 (0/33)</td>
<td>0 (0/33)</td>
</tr>
<tr>
<td>Cell signaling: MAPK/ERK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAF</td>
<td>2 (7/411)</td>
<td>NA</td>
<td>0 (0/13)</td>
<td>11 (2/19)</td>
<td>11 (2/19)</td>
</tr>
<tr>
<td>HRAS</td>
<td>1 (4/454)</td>
<td>24 (21/89)</td>
<td>11 (19/170)</td>
<td>29 (7/24)</td>
<td>0 (0/176)</td>
</tr>
<tr>
<td>KRAS</td>
<td>19 (506/2714)</td>
<td>NA</td>
<td>4 (5/122)</td>
<td>2 (5/216)</td>
<td>2 (5/216)</td>
</tr>
<tr>
<td>NRAS</td>
<td>1 (7/505)</td>
<td>NA</td>
<td>0 (0/112)</td>
<td>0 (0/112)</td>
<td>0 (0/112)</td>
</tr>
<tr>
<td>Cell signaling: apoptosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCC</td>
<td>NA</td>
<td>41 (7/17)</td>
<td>NA</td>
<td>26 (37/143)</td>
<td>26 (37/143)</td>
</tr>
<tr>
<td>FHIT</td>
<td>2 (1/41)</td>
<td>56 (199/353)</td>
<td>0 (0/3)</td>
<td>34 (30/89)</td>
<td>34 (30/89)</td>
</tr>
</tbody>
</table>

NA, not available; ERK, extracellular signal-regulated kinase.

\(^a\)Mutations were somatic missense, nonsense and silent point mutations, frameshift deletions and insertions, in-frame deletions and insertions.

\(^b\)Values in parentheses show number of tumors with mutation/total number of tumors studied.

\(^c\)Values in parentheses show number of tumors with LOH/total number of informative samples.
amines, aldehydes and tobacco-specific nitrosamines from tobacco smoke cause DNA adduct formation and elicit DNA alterations (23–25). To analyze the extent of these alterations, we have compared incidences of mutations and LOH in lung tumors of smokers and non-smokers (Table II). While the list of genes studied was not exhaustive, both smokers and non-smokers demonstrated alterations in the same genes but with some significant differences in the incidence of alterations. The most frequently altered cell cycle gene was TP53 (Table II) where a statistically significant increase in the incidence of mutations was observed between smokers (39%) and non-smokers (26%; P < 0.01). Allele loss occurred in ~40% of cases but this did not vary with tobacco exposure. Mutations in CDKN2A were rare and although the incidence of LOH was higher it did not differ significantly between smokers and non-smokers. In the Wnt pathway, evidence based on a low number of tumors suggests that APC loss occurred less frequently in smokers than non-smokers. The number of KRAS mutations was significantly higher in smokers (20%) compared with non-smokers (3%; P < 0.001), yet no mutations were detectable in HRAS or NRAS genes (Table II). Data for cell signaling pathways involving FHIT and DCC genes were restricted to allelic loss and showed FHIT LOH occurred significantly more frequently in tumors from smokers compared with non-smokers (P < 0.01). DCC loss was evident in 16% of smokers but did not occur in non-smokers.

**TP53 spectra in tumors: smokers and non-smokers**

Evidence shows that the mutational specificity of tobacco smoke occurs at the level of the base rather than codon with GC → TA transversions indicative of exposure (26,27). To analyze the effect of tobacco smoke exposure on the incidence of single base substitutions at primary sites of smoke deposition: lung, larynx and oral cavity, we have compared TP53 mutational spectra from smokers and non-smokers (Figure 1). Laryngeal tumors from non-smokers were not included in these analyses due to insufficient data. When all single base substitutions were considered, a higher frequency of TC → AT and GC → TA substitutions occurred in lung, larynx and oral tumors compared with other substitutions. Substitutions of AT bases and GC → TA transversions occurred infrequently (<20% of cases) and did not vary significantly with tobacco smoke exposure. Figure 1 demonstrates the percentage of single base substitutions that occurred in different tumor groups. No difference was observed between the percentage of GC → AT transitions and GC → TA transversions in lung (25% versus 30%) and laryngeal (30% versus 26%) tumors of smokers in contrast to oral tumors where GC → AT substitutions in smokers (33%) were significantly higher than GC → TA transversions (15%; P < 0.001). Similarly, in oral tumors from non-smokers the number of GC → AT transitions (41%) was significantly higher than the number of GC → TA transversions (2%; P < 0.001).

When the frequency of substitutions was compared between smokers and non-smokers, it was evident that exposure to tobacco smoke significantly increased the number of GC → TA transversions. In lung tumors, the percentage of GC → TA mutations in non-smokers was 22%, whereas in smokers it was 30% (P < 0.05) with the increase occurring predominantly at CpG sites (data not shown). Oral cancers displayed a similar increase in GC → TA substitutions after exposure to tobacco smoke (2% in non-smokers compared with 15% in smokers; P < 0.05) and these occurred at both CpG and non-CpG sites. This increase was accompanied by a decrease in the frequency of GC → AT transitions following exposure to tobacco. In neither oral nor lung cancers did the changes in frequency of GC → AT substitutions between smokers and non-smokers reach significance. Substitutions of other bases did not differ significantly with tobacco smoke exposure (GC → CG, AT → CG, AT → GC and AT → TA).

Although the mutational specificity of tobacco smoke occurs at the nucleotide level rather than the codon level, in vitro evidence demonstrates that benzo[a]pyrene binds with greatest strength to codons 157, 248 and 273 (28–30). We have analyzed these codons for the presence of mutations. Positional spectra for tumors from smokers clearly reveal the presence of mutations at these codons in lung and larynx tumors (Figure 2). In oral tumors, mutations in these codons occurred most frequently at codon 248; codons 157 and 273 mutations were less numerous. Mutations at codon 175 were common in all cancers and also codon 245 in lung and larynx tumors. The latter codons are considered common mutational hot spots present in many types of cancer and are not necessarily associated with tobacco exposure (31).

To determine whether the base substitutions in the most frequently mutated codons in smokers (as reported in the literature; 28, 29, 30, 31) showed specificity for tobacco smoke exposure, the percentage of GC → TA transversions occurring at codons 157, 175, 245, 248 and 273 was determined (Table III). In tumors other than lung cancer, this type of substitution is an infrequent occurrence (31); however, in lung tumors from smokers all the frequently mutated codons demonstrated GC → TA transversions (25–88%) indicating a possible role for tobacco smoke in their formation. In comparison, tumors from

| Table II. Percentage of lung tumors from smokers and non-smokers with gene alterations |
|-----------------------------------------|-----------------|-----------------|-----------------|
| **Tobacco smoke** | **Unexposed** |
| % tumors with mutations | % tumors with LOH | % tumors with mutations | % tumors with LOH |
| **Cell cycle** | | | |
| CDKN2A | 4 (6/136) | 33 (8/24) | 0 (0/47) | 29 (5/17) |
| CDKN2B | NA | NA | NA | NA |
| RB1 | NA | 47 (47/99) | NA | 19 (10/53) |
| TP53 | 39\(c\) (627/1591) | 38 (47/125) | 26 (105/410) | 40 (14/35) |
| **Cell signaling: Wnt** | | | |
| APC | NA | 43 (31/72) | NA | 67 (4/6) |
| **Cell signaling: MAPK/ERK** | | | |
| Braf | NA | NA | NA | NA |
| Hras | 0 (0/122) | NA | 0 (0/42) | NA |
| Kras | 20\(c\) (616/3114) | NA | 3 (22/696) | NA |
| Nras | 0 (0/193) | NA | 0 (0/69) | NA |
| **Cell signaling: apoptosis** | | | |
| Dcc | NA | 16 (13/81) | NA | 0 (0/9) |
| Fhit | NA | 57\(c\) (262/460) | NA | 27 (34/126) |

**Notes:**

- ERK, extracellular signal-regulated kinase.
- CDKN2A, cyclin-dependent kinase inhibitor 2A.
- CDKN2B, cyclin-dependent kinase inhibitor 2B.
- RB1, retinoblastoma 1.
- TP53, tumor protein 53.
- APC, adenomatous polyposis coli.
- Braf, v-raf murine sarcoma virus Harvey murine sarcoma viral oncogene homolog.
- Hras, Harvey rat sarcoma viral oncogene homolog.
- Kras, Kirsten rat sarcoma viral oncogene homolog.
- Nras, nerve growth factor receptor.
- Dcc, deleted in colon cancer.
- Fhit, fragile histidine triad.
- LOH, loss of heterozygosity.

\(c\) Per cent tumors with alteration statistically higher in smokers compared with unexposed, non-smokers (P < 0.01).
Fig. 1. TP53 mutational spectra for lung, oral and laryngeal cancers showing the total percentage of single base pair substitutions that were GC → AT, GC → TA, GC → CG, AT → CG, AT → GC and AT → TA for the entire gene. TS indicates tumors arose in current or previous tobacco smokers; NS indicates tumors arose in non-smokers. Black bars designate oral cancers; dark gray designate lung and light gray designate larynx cancers.

Fig. 2. TP53 positional spectra showing the percentage of single base substitutions occurring at each codon in lung, larynx and oral tumors from smokers.
non-smokers rarely exhibited this substitution (Table III). Codons 157, 245 and 273 harbored GC → TA transversions in larynx cancers but of these, only codon 157 was a mutational hot spot in smokers. Codons 175 and 248 did not display GC → TA changes, instead GC → AT and GC → CG were observed. In oral cancers from smokers, codon 157 displayed 100% GC → TA transversions, yet this was not a frequently mutated codon. Mutational hot spots were evident at codons 175, 248, 281 and 288, yet none of these codons displayed GC → TA changes. At other codons, the percentage of GC → TA substitutions was 66% and 100% at codons 191 and 298, respectively (data not shown). In non-smokers, mutations were not observed at codons 157 and 175 and none of the substitutions at codons 245, 248 and 273 were GC → TA.

**Discussion**

In this paper, we have presented an analysis of published data collected from a large number of studies and analyzed using the GAC database. In performing this analysis, we have combined data from individual studies, some of which were small, to look for overall trends in gene alterations in tumors while minimizing the effect that single study anomalies have on the conclusions.

Current understanding of the molecular pathways of cancer development describes a multitude of genetic and epigenetic changes that accumulate as tumors develop (32). To understand the roles played by different genes in this process, we have characterized some of the genetic changes that occur in tumors from different topographical sites. Several mechanisms are known to activate or inactivate genes; we have presented data for two that are commonly studied: mutation and gene LOH. Other types of alterations are known to affect genes including homozygous deletion, methylation and chromosomal LOH; these undoubtedly also play a vital role in tumor development but have not been evaluated here.

Greeman et al. (33) recently analyzed 518 protein kinase genes from 210 diverse cancers for mutations. The authors concluded that the majority of the somatic mutations detected in these genes were ‘passenger’ mutations that did not directly effect tumor development. Similarly, Sjoblom et al. (34) sequenced 13 023 different genes in breast and colorectal cancers and showed that on average ~90 mutations per tumor were observed and that 11 genes per tumor were mutated at a significant frequency. However, these studies did not take into account other types of genetic alteration such as LOH that also play a significant role in tumor formation. The data we have presented in this paper demonstrates that in oral tumors and tumors of the digestive tract and lung, genes are more frequently affected by allele loss than by mutation. Gene LOH was reported in each tumor group analyzed for allelic loss and the incidence was >30% in over two-thirds of the studies. These data probably represent an underestimate of the total frequency of allelic loss because analysis was restricted to intragenic markers and markers tightly linked to a gene. Chromosome markers located in the vicinity of a gene or gene locus that could conceivably extend into the gene were not considered in these analyses of gene LOH.

**Table III.** Percentage of TP53 mutations at codons 157, 175, 245, 248 and 273 that were GC → TA transversions

<table>
<thead>
<tr>
<th>% GC → TA transversion mutations</th>
<th>Lung TS</th>
<th>Lung NS</th>
<th>Larynx TS</th>
<th>Oral TS</th>
<th>Oral NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codon 157</td>
<td>88</td>
<td>0</td>
<td>50</td>
<td>100</td>
<td>NA</td>
</tr>
<tr>
<td>Codon 175</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Codon 245</td>
<td>61</td>
<td>0</td>
<td>67</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Codon 248</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Codon 273</td>
<td>50</td>
<td>17</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

TS, tobacco smokers (current or previous); NS, non-smokers (never); NA, no mutation data available.

When gene mutations are considered in parallel with allele loss, it is evident that TP53 was one of the few genes inactivated at similar frequencies by allele loss and mutation; this occurred at all cancer sites examined. The only other example of this occurring was in the APC gene and this was restricted to gastric and colorectal cancers. Although the datasets defining the mutation and LOH data are distinct and cannot be used to infer bi-allelic inactivation, it is tempting to conclude that this data show TP53 and in some instances APC fulfill Knudson’s theory of gene inactivation in cancer development. Although several other tumor suppressor genes have been studied, they typically demonstrated very low levels of gene mutations in association with higher levels of LOH. Whether these genes perform a role in cancer development and what that role is remains unclear but possibly gene inactivation occurs by an alternative mechanism: gene methylation or germ line mutation. Alternatively, single allele inactivation may exert a dominant oncogenic effect or it may simply confer a predisposition to cancer. The number of genes and topographical sites where this occurred might suggest this as an important avenue of future research.

Data for alterations of oncogenes were less numerous and largely restricted to mutation studies of the RAS and BRAF genes of the MAPK/extracellular signal-regulated kinase pathway. No conclusions could be drawn concerning the role LOH plays in activation of these genes because of the lack of information but mutations appeared to play site-specific roles in the development of lung (KRAS), oral (HRAS), esophageal (BRAF) and colorectal (BRAF and KRAS) cancers. In the case of lung and oral cancers, it is possible that these were the result of site-specific exposures to environmental carcinogens (tobacco smoke or chewing tobacco). Still with the addition of allelic loss data for these genes, the specificity of gene activation for specific topographies may be negated.

When considered by topography, it is evident that alterations in cell signaling play a role in cancer formation: in colorectal and gastric cancer through genes of the Wnt pathway and the cell cycle genes, particularly TP53, at all sites of the digestive tract, oral and lung cancers. The potential for cell cycle disruption was particularly evident in esophageal cancers where LOH was detectable in a number of different genes. Constitutive transcriptional activation of genes involved in signal transduction in the MAPK/extracellular signal-regulated kinase pathway occurred most frequently in colorectal and lung cancers. All topographies but particularly lung cancers demonstrated alterations in FHIT and DCC, suggesting that inhibition of apoptosis was common to many tumors.

Some of the most well documented gene alterations in these datasets were KRAS mutations in lung cancers. The reason for this abundance of data is that it represents an instance where a common environmental exposure, tobacco smoke, evokes a specific genetic response (35). When data for tumors with known exposure to tobacco smoke are compared with those that were unexposed, the incidence of KRAS and TP53 mutations was significantly higher in smokers. Likewise, the incidence of allelic loss of the FHIT gene in smokers was significantly higher than in non-smokers. Other gene alterations have not been reported in sufficient numbers to draw conclusions about the effect of smoking on their incidence. This analysis does demonstrate that the range of genes activated in smokers and non-smokers was identical, suggesting that one mechanism by which tobacco can give rise to lung cancer is through an increased frequency of genetic alterations. Of course, these analyses are limited by the scope of the published literature; studies tend to focus on the same, well-characterized genes resulting in an absence of information for many genes and genetic pathways. In future, studies of a broader spectrum of genes would help to ascertain whether tumorigenesis in smoking-related lung cancers is the result of alterations in genetic pathways unique to tobacco smoke exposure and distinct from tumors with no exposure to tobacco.

There are >60 carcinogenic components of tobacco smoke which include polycyclic aromatic hydrocarbons, aromatic amines, aldehydes and nitrosamines (15); direct exposure to these carcinogens occurs not only in lung but also in the upper aerodigestive tract,
particularly the mouth and larynx. In the lung, it has been shown that preferential binding of benzo[a]pyrene and acrolein to guanine residues leads to an increased incidence of GC → TA transversions (25,36). From this it follows that GC → TA transversions do not frequently occur at sites not directly exposed to tobacco smoke (37,27). Indeed, our analysis of oral tumors from non-smokers demonstrated this; GC → TA substitutions were present as only 2% of the mutations. Surprisingly, this was not the case for lung cancers from non-smokers where 22% of alterations were GC → TA transversions. It is possible that this number was an overestimation of the true rate due to the inclusion of a potentially anomalous study. Gao et al. (38) reported a 66% mutation rate in lung tumors of non-smokers compared with <36% for all other studies. Each tumor in this study demonstrated multiple mutations such that 10 of 15 mutated tumors had 48 mutations comprising 31% of the total mutations for the group. This undoubtedly influenced the results for this group, yet even if this study is excluded, the percentage of mutations that were GC → TA in lung tumors (16% excluding the Gao study) is significantly higher than oral tumors (2%). Why this should occur is not clear but it is possible that lung exposure to polycyclic aromatic hydrocarbons, aromatic amines and nitrosoamines derives from alternative sources that are not pertinent to oral tumorigenesis. One of these might be sidestream tobacco smoke for passive/involuntary smokers. Qualitatively, sidestream smoke has the same chemical constituents as mainstream smoke and could result in exposure in non-smokers (39). Another possible exposure that might present an environment risk to non-smokers is the complex chemical mixtures of diesel fuel and vehicle exhaust fumes (40,41). Despite the differences in the lung and oral non-smoking groups, our data show an increase in the percentage of GC → TA transversions in tobacco smokers; this was most apparent in oral tumors but was also evident in lung tumors. The increase in GC → TA transversions was associated with an accompanying decrease in the percentage of GC → AT transitions; this was seen in both lung and oral tumors and was mirrored in laryngeal tumors of smokers. The fact that this decrease occurred particularly at CpG sites supports the view that it was the consequence of a chemical exposure (42).

Efforts made to analyze some of the individual chemical components of tobacco smoke in lung tumors of laboratory rodents have supported these findings; GC → TA transversions were prevalent in lung tumors of mice exposed to benzo[a]pyrene (16,43–45). Interestingly, 4-(N-nitrosomethylamino)-1-(3-pyridyl)-1-butanone also found in a variety of tobacco products (chewing tobacco, snuff, cigarettes and cigars) primarily causes GC → AT transitions in exposed mice (16,46–49). Although exposure to 4-(N-nitrosomethylamino)-1-(3-pyridyl)-1-butanone undoubtedly occurs in smokers, our analysis of the human data in the literature did not demonstrate any increase in the percentage of GC → AT transitions in lung or oral cancers.

Unlike other carcinogens such as aflatoxin B1, the mutagenic specificity of tobacco smoke exposure is defined at the base level by G → T transversions rather than at the codon level (50–52). Still comparison of positional spectra and incidences of GC → TA transversions for lung, larynx and oral tumors from smokers define several codons as hot spots of mutations in TP53, all of which occur in the DNA-binding domain of the protein. For both lung and larynx, these were codons 157, 245, 248 and 273; this agrees with in vitro studies that show benzo[a]pyrene derivatives bind with greatest strength to codons 157, 248 and 273 (28,51,52). In contrast, mutational hot spots in oral tumors were restricted to codons 175 and 248 when the types of base change occurring at these codons was considered codon 248 showed 0–25% incidence of GC → T substitutions. This would suggest that the benzo[a]pyrene component of tobacco smoke did not play a prominent role in mutation at this site.

Codons 175 and 245 are commonly mutated in different types of cancer and are not usually associated with specific exposures. Despite this, lung and larynx (codon 245 only) tumors frequently demonstrate G → T transversions at these sites, suggesting that tobacco smoke may play a role in their mutation in some instances. In oral cancer, this did not occur; codon 175 and 245 mutations were not G → T transversions. Of course, this does not preclude tobacco smoke from play-

Received May 16, 2007; revised July 16, 2007; accepted July 20, 2007.