Molecular Epidemiology of Human Cancer: Contribution of Mutation Spectra
Studies of Tumor Suppressor Genes

S. Perwez Hussain and Curtis C. Harris

Department of Pathology, University of Maryland School of Medicine, Baltimore, Maryland 21201 [S. P. H.], and Laboratory of Human Carcinogenesis, National Cancer Institute, NIH, Bethesda, Maryland 20892 [C. C. H.]

Molecular Epidemiology

The identification of individuals at high risk of cancer can offer promising avenues for the prevention of cancer. The majority of human cancer cases are caused by environmental, occupational, and recreational exposures to carcinogens (1). These carcinogens can affect one or multiple stages of carcinogenesis through both genetic and epigenetic mechanisms (2-4). Cancer risk assessment, a highly visible discipline in the field of public health, has historically relied on classical epidemiology, chronic animal bioassays of potential carcinogens, and the mathematical modeling of these epidemiological and laboratory findings (5). Public health regulators are forced to steer a prudent course of conservative risk assessment because of the limited knowledge of the complex pathological processes during carcinogenesis: differences in the metabolism of carcinogens and DNA repair capacities, variable genomic stability among animal species, and variation among individuals with inherited cancer predisposition make a definitive analysis of cancer risk extremely difficult. A molecular epidemiological approach integrates molecular biology, in vitro and in vivo laboratory models, and biochemistry and epidemiology to infer individual cancer risk (6-9). Achieving this goal is challenging both current molecular technologies and the epidemiological designs used to resolve bioethical dilemmas (10, 11).

The two major facets of the molecular epidemiology of human cancer risk are the assessment of carcinogen exposure, including biomarkers of effect, and the inherited or acquired host cancer susceptibility factors (reviewed in Refs. 12 and 13); thus, it is the gene-environment interaction that determines an individual's cancer risk (Fig. 1). Whereas the external environment is a major source of carcinogens, the cellular microenvironment also contains endogenous carcinogens including oxygen radicals generated by cancer-prone chronic inflammatory diseases. Kinzler and Vogelstein (14) have further classified cancer susceptibility genes as gatekeepers and caretakers (discussed below).

The gene-environment paradigm can improve cancer risk assessment. When combined with carcinogen bioassay in laboratory animals, laboratory studies of molecular carcinogenesis, and classical epidemiology, molecular epidemiology can contribute to the four traditional aspects of cancer risk assessment: (a) hazard identification; (b) dose-response assessment; (c) exposure assessment; and (d) risk characterization (Fig. 2). Improved cancer risk assessment has broad public health and economic implications (5).

The identification of individuals at high risk of cancer raises complex bioethical issues (Refs. 10 and 15-17; Fig. 2). One can argue that the knowledge of one's risk can be beneficial. However, more encompassing bioethical issues arise such as an individual's responsibility to family members and psychosocial concerns regarding the genetic testing of children. Therefore, the uncertainty of the current individual risk assessments and the limited availability of genetic counseling services dictate caution and, many argue, the restriction of genetic testing to those conditions amenable to either preventative or therapeutic intervention.

Cancer Susceptibility Genes

An intricately balanced control of cellular proliferation and death maintains normal tissue homeostasis and is accomplished by a network of genes. Many of these genes are implicated in the natural history of human cancer because of their consistent alteration in most types of human cancer. The p53 tumor suppressor gene is a remarkable example, because it is mutated in about half of all cancer types arising from a variety of tissues, and missense mutations occur at a high frequency (18-20). Other tumor suppressor genes that are important in human oncology, e.g., APC, WT1, or NFI, seem to have a more limited tissue contribution (Table 1). In addition to the consideration of single genes, recent data indicate the importance of molecular pathways involving cancer susceptibility genes, e.g., p16INK4 and Rb are involved in both the G1 checkpoint pathway (21) and the molecular pathogenesis of many types of cancer (22), and APC and β-catenin are involved in the initiating events in colon carcinogenesis (23).

Germline mutations in genes, e.g., p53, RB, p16INK4, and APC, have been identified in rare cancer-prone families (Table 1). Somatic mutations in these genes are also frequently found in common sporadic cancers, which attests to the value of studying rare familial cancer syndromes. These cancer susceptibility genes encode proteins that perform diverse cellular functions, including transcription, cell cycle control, DNA repair, and apoptosis. The increased cancer risk of an individual carrying one of these germline mutations can be extraordinarily high, i.e., more than 1000-fold in xeroderma pigmentosum (complementation group A-G; Fig. 3), where the "at risk" allele is infrequent in the general population. Whereas germline mutations in genes involved in carcinogen metabolism increase cancer risk only severalfold, an at risk allele can be very common, e.g., GSTM1 (occurring in about 50% in the Caucasian population), thus making the attributable cancer risk substantial.

Caretaker and Gatekeeper Genes

Recently, the concept of gatekeeper and caretaker genes characterized by their control of net cellular proliferation or maintenance of genomic integrity, respectively, has been introduced (14, 24). Examples of gatekeeper genes include APC and β-catenin in colon epithelial cells, Rb in retinal epithelial cells, NFI in Schwann cells, and VHL in kidney cells. The most prominent example of a gatekeeper is the APC gene in colorectal cancer. It is suggested that an alteration in
APC leads to a derangement of the cellular proliferation pathway that is important for maintaining a constant cell population. However, this function of APC has a high specificity for colonic epithelial cells but is lacking in most other organs. In murine as well as human colon cancer, APC mutation occurs early on in the process of carcinogenesis (25–27). Although other genes such as K-ras and p53 play important roles in the later stages of colorectal carcinogenesis, APC mutation and the less common β-catenin mutations are essential events in the initiation of neoplasia (reviewed in Ref. 24). If this concept of a gatekeeper pathway holds true for the initiation of neoplasia in general, then the identification of other gatekeeper genes can be anticipated.

Unlike gatekeeper genes, caretaker genes, generally maintain genomic stability and are not involved directly in the initiation of the neoplastic process. Genetic instability due to mutations in caretaker genes enhances the probability of mutation in other genes, including those in the gatekeeper pathway. Mismatch DNA repair genes, e.g., MSH2 and MLH1, are caretaker genes, and abnormalities in these genes enhance genomic instability and increase the risk of human colon cancer. Animal models based on this knowledge of human colon carcinogenesis have been developed (28–30). Similar in vivo as well as in vitro models should be developed for other tumor sites of carcinogenesis. Breast cancer susceptibility genes BRCA1 and BRCA2 have been recently included in the list of caretaker genes (14). In the same report, it is suggested that a predisposed individual with an inherited mutated allele of a caretaker gene is at a lower risk of cancer when compared with an individual with a mutated gatekeeper allele. This difference has been attributed to the finding that three or more additional somatic mutations are required to initiate neoplasia in the caretaker pathway, whereas only one additional somatic mutation is required to initiate neoplasia in the gatekeeper pathway. In addition to cancer as an endpoint, the molecular analysis of gatekeeper and caretaker genes in presumed preneoplastic morphological lesions could provide short-term and less expensive pathobiological endpoints for hazard identification and molecular epidemiological studies.

Mutator Concept

Multiple mutations are found in cancer cells. The presence of a mutator phenotype has been suggested as an important step in tumor development and forms the basis of the mutator hypothesis (Ref. 31, reviewed in Ref. 32). MIS arises from either deletions or insertions that are important for maintaining a constant cell population. However, this function of APC has a high specificity for colonic epithelial cells but is lacking in most other organs. In murine as well as human colon cancer, APC mutation occurs early on in the process of carcinogenesis (25–27). Although other genes such as K-ras and p53 play important roles in the later stages of colorectal carcinogenesis, APC mutation and the less common β-catenin mutations are essential events in the initiation of neoplasia (reviewed in Ref. 24). If this concept of a gatekeeper pathway holds true for the initiation of neoplasia in general, then the identification of other gatekeeper genes can be anticipated.

Unlike gatekeeper genes, caretaker genes, generally maintain genomic stability and are not involved directly in the initiation of the neoplastic process. Genetic instability due to mutations in caretaker genes enhances the probability of mutation in other genes, including those in the gatekeeper pathway. Mismatch DNA repair genes, e.g., MSH2 and MLH1, are caretaker genes, and abnormalities in these genes enhance genomic instability and increase the risk of human colon cancer. Animal models based on this knowledge of human colon carcinogenesis have been developed (28–30). Similar in vivo as well as in vitro models should be developed for other tumor sites of carcinogenesis. Breast cancer susceptibility genes BRCA1 and BRCA2 have been recently included in the list of caretaker genes (14). In the same report, it is suggested that a predisposed individual with an inherited mutated allele of a caretaker gene is at a lower risk of cancer when compared with an individual with a mutated gatekeeper allele. This difference has been attributed to the finding that three or more additional somatic mutations are required to initiate neoplasia in the caretaker pathway, whereas only one additional somatic mutation is required to initiate neoplasia in the gatekeeper pathway. In addition to cancer as an endpoint, the molecular analysis of gatekeeper and caretaker genes in presumed preneoplastic morphological lesions could provide short-term and less expensive pathobiological endpoints for hazard identification and molecular epidemiological studies.

Mutator Concept

Multiple mutations are found in cancer cells. The presence of a mutator phenotype has been suggested as an important step in tumor development and forms the basis of the mutator hypothesis (Ref. 31, reviewed in Ref. 32). MIS arises from either deletions or insertions that are important for maintaining a constant cell population. However, this function of APC has a high specificity for colonic epithelial cells but is lacking in most other organs. In murine as well as human colon cancer, APC mutation occurs early on in the process of carcinogenesis (25–27). Although other genes such as K-ras and p53 play important roles in the later stages of colorectal carcinogenesis, APC mutation and the less common β-catenin mutations are essential events in the initiation of neoplasia (reviewed in Ref. 24). If this concept of a gatekeeper pathway holds true for the initiation of neoplasia in general, then the identification of other gatekeeper genes can be anticipated.

Unlike gatekeeper genes, caretaker genes, generally maintain genomic stability and are not involved directly in the initiation of the neoplastic process. Genetic instability due to mutations in caretaker genes enhances the probability of mutation in other genes, including those in the gatekeeper pathway. Mismatch DNA repair genes, e.g., MSH2 and MLH1, are caretaker genes, and abnormalities in these genes enhance genomic instability and increase the risk of human colon cancer. Animal models based on this knowledge of human colon carcinogenesis have been developed (28–30). Similar in vivo as well as in vitro models should be developed for other tumor sites of carcinogenesis. Breast cancer susceptibility genes BRCA1 and BRCA2 have been recently included in the list of caretaker genes (14). In the same report, it is suggested that a predisposed individual with an inherited mutated allele of a caretaker gene is at a lower risk of cancer when compared with an individual with a mutated gatekeeper allele. This difference has been attributed to the finding that three or more additional somatic mutations are required to initiate neoplasia in the caretaker pathway, whereas only one additional somatic mutation is required to initiate neoplasia in the gatekeeper pathway. In addition to cancer as an endpoint, the molecular analysis of gatekeeper and caretaker genes in presumed preneoplastic morphological lesions could provide short-term and less expensive pathobiological endpoints for hazard identification and molecular epidemiological studies.

MIS2 arises from either deletions or insertions that are important for maintaining a constant cell population. However, this function of APC has a high specificity for colonic epithelial cells but is lacking in most other organs. In murine as well as human colon cancer, APC mutation occurs early on in the process of carcinogenesis (25–27). Although other genes such as K-ras and p53 play important roles in the later stages of colorectal carcinogenesis, APC mutation and the less common β-catenin mutations are essential events in the initiation of neoplasia (reviewed in Ref. 24). If this concept of a gatekeeper pathway holds true for the initiation of neoplasia in general, then the identification of other gatekeeper genes can be anticipated.

Unlike gatekeeper genes, caretaker genes, generally maintain genomic stability and are not involved directly in the initiation of the neoplastic process. Genetic instability due to mutations in caretaker genes enhances the probability of mutation in other genes, including those in the gatekeeper pathway. Mismatch DNA repair genes, e.g., MSH2 and MLH1, are caretaker genes, and abnormalities in these genes enhance genomic instability and increase the risk of human colon cancer. Animal models based on this knowledge of human colon carcinogenesis have been developed (28–30). Similar in vivo as well as in vitro models should be developed for other tumor sites of carcinogenesis. Breast cancer susceptibility genes BRCA1 and BRCA2 have been recently included in the list of caretaker genes (14). In the same report, it is suggested that a predisposed individual with an inherited mutated allele of a caretaker gene is at a lower risk of cancer when compared with an individual with a mutated gatekeeper allele. This difference has been attributed to the finding that three or more additional somatic mutations are required to initiate neoplasia in the caretaker pathway, whereas only one additional somatic mutation is required to initiate neoplasia in the gatekeeper pathway. In addition to cancer as an endpoint, the molecular analysis of gatekeeper and caretaker genes in presumed preneoplastic morphological lesions could provide short-term and less expensive pathobiological endpoints for hazard identification and molecular epidemiological studies.

MIS2 arises from either deletions or insertions that are important for maintaining a constant cell population. However, this function of APC has a high specificity for colonic epithelial cells but is lacking in most other organs. In murine as well as human colon cancer, APC mutation occurs early on in the process of carcinogenesis (25–27). Although other genes such as K-ras and p53 play important roles in the later stages of colorectal carcinogenesis, APC mutation and the less common β-catenin mutations are essential events in the initiation of neoplasia (reviewed in Ref. 24). If this concept of a gatekeeper pathway holds true for the initiation of neoplasia in general, then the identification of other gatekeeper genes can be anticipated.

Unlike gatekeeper genes, caretaker genes, generally maintain genomic stability and are not involved directly in the initiation of the neoplastic process. Genetic instability due to mutations in caretaker genes enhances the probability of mutation in other genes, including those in the gatekeeper pathway. Mismatch DNA repair genes, e.g., MSH2 and MLH1, are caretaker genes, and abnormalities in these genes enhance genomic instability and increase the risk of human colon cancer. Animal models based on this knowledge of human colon carcinogenesis have been developed (28–30). Similar in vivo as well as in vitro models should be developed for other tumor sites of carcinogenesis. Breast cancer susceptibility genes BRCA1 and BRCA2 have been recently included in the list of caretaker genes (14). In the same report, it is suggested that a predisposed individual with an inherited mutated allele of a caretaker gene is at a lower risk of cancer when compared with an individual with a mutated gatekeeper allele. This difference has been attributed to the finding that three or more additional somatic mutations are required to initiate neoplasia in the caretaker pathway, whereas only one additional somatic mutation is required to initiate neoplasia in the gatekeeper pathway. In addition to cancer as an endpoint, the molecular analysis of gatekeeper and caretaker genes in presumed preneoplastic morphological lesions could provide short-term and less expensive pathobiological endpoints for hazard identification and molecular epidemiological studies.
gene involved in the initiation of sporadic colorectal carcinoma (60). For example, there was a lack of any association between p53 status and MIS or loss of heterozygosity in a recent study analyzing 58 sporadic colorectal tumors (61). An analysis of 40 primary gastric carcinomas showed a negative correlation between p53 mutation and MIS and indicates two distinct pathways for their contributions in cancer development (62). In the same study, a positive association was found between MIS and mutation in the TGFß-RII gene. Treatment of wild-type and p53-null mouse embryo fibroblasts with 4-nitroquinoline 1-oxide did not show any difference in accumulation of point mutations in the lacI target gene (63). Also, in the same study, DNA from thymus and thymic tumors in both p53+/+ and p53−/− mice did not show any distinct difference in the mutation frequency in the target gene. However, the spontaneous mutation frequency of the lacI gene was high, i.e., 10−5, so that the modulating effects of p53, if any, might not be detected. In addition, the rate of nucleotide excision repair is substantially lower in rodent-derived cells than it is in human cells (64, 65), and the repair of UV damage in the nontranscribed strand of the active gene is markedly reduced in rodents when compared with that of humans (66).

MUTATIONAL SPECTRA OF TUMOR SUPPRESSOR GENES

Several endogenous and exogenous mutagens have been described as inducing a characteristic pattern of DNA alteration. Mutational spectra analyses are required to study the type, location, and frequency of DNA changes. Alterations of cancer-related genes found in tumors not only represent the interaction of a carcinogen with DNA and cellular DNA repair processes but also reflect the selection of those mutations that provide premalignant and malignant cells with a clonal growth and survival advantage. Study of the frequency, timing, and mutational spectra of p53 and other cancer-related genes provides insight into the etiology and molecular pathogenesis of cancer and generates hypotheses for future investigations. These include questions regarding carcinogen-DNA interactions, the function of the affected gene products, the mechanism of carcinogenesis in specific organs or tissues, and general cell biological processes such as DNA replication and repair.

Nonsense mutations, deletions, and insertions are the most frequent types of mutations in tumor suppressor genes that produce either an absentee or a truncated protein product. These mutations are clearly loss-of-function mutations. In contrast, the p53 tumor suppressor gene shows an unusual spectrum of mutations when compared with other suppressor genes, e.g., APC, BRCA1, or ATM, in which the mutations lead to a loss of function (Fig. 4). Missense mutations, in which the encoded protein contains amino acid substitutions, are commonly found in the p53 tumor suppressor gene. p53 missense mutations can cause both a loss of tumor suppressor function and a gain of oncogenic function by changing the repertoire of genes whose expression is controlled by this transcription factor (67–69). This functional duality may be one explanation for the high frequency of p53 mutations in human cancer.

p53

The p53 gene is well suited for mutational spectrum analysis for several reasons: (a) p53 mutations are common in many human cancers, and a sizeable database of more than 8000 entries has accrued, so that the analysis of this large database can yield statistically valid conclusions (70, 71) and can be readily accessed on the World Wide Web (http://www.iarc.fr/p53/homepage.html); (b) the modest size of the p53 gene (11 exons, 393 amino acids) permits the
study of the entire coding region, and it is highly conserved in vertebrates, allowing the extrapolation of data from animal models (72); and (c) the point mutations that alter p53 function are distributed over a large region of the molecule, especially in the hydrophobic midportion (Refs. 18, 19, and 73; Fig. 5). These numerous base substitutions alter p53 conformation and sequence-specific transactivation activity; thus, correlations between distinct mutants and functional changes are possible. Frameshift and nonsense mutations that truncate the protein are located outside of these regions (Fig. 5), and additional changes are possible. Frameshift and nonsense mutations that truncate the protein are located outside of these regions (Fig. 5), and additional changes are possible.

Exogenous mutagenic agents and/or endogenous mutagenic mechanisms have been implicated in the induction of mutations. These mutations are archived in the spectrum of p53 mutations found in human cancer (18, 19, 73, 76-78). Errors introduced during DNA replication, RNA splicing, DNA repair, and DNA demethylation are examples of endogenous mutagenic mechanisms. The DNA sequence context is an important factor that determines these events. Almost all short deletions and insertions occur at monotonic runs of two or more identical bases or at repeats of 2-8 bp DNA motifs, either in tandem or separated by a short intervening sequence (74). The mechanism that has been studied most is called slipped mispairing, a misalignment of the template DNA strands during replication that leads to either deletion, if the nucleotides excluded from pairing are on the template strand, or insertion, if the nucleotides excluded are on the primer strand (79). When direct repeat sequences mispair with a complementary motif nearby, the intervening oligonucleotide sequence may form a loop between the two repeat motifs and be deleted (80, 81). More lengthy runs and sequence repeats are more likely to generate frameshift mutations. The deletions and insertions in the p53 gene found in human tumors also may be biologically selected from the broad array of such mutations occurring in human cells. When compared with the distribution of missense mutations, these types of mutations occur more frequently in exons 2-4 (54%) and 9-11 (77%) than in exons 5-8 (20%; Fig. 5). The NH2 terminus of the p53 protein (encoded by exons 2-4; reviewed in Refs. 82-85) has an abundance of acidic amino acids that are involved in the transcriptional function of p53 (86, 87) and binds to transcription factors such as TATA-binding protein in transcription factor IID (82, 88-91). Experimental studies have shown that multiple point mutations in this domain are required to inactivate its transcriptional transactivation function (92). The COOH terminus (encoded by exons 9-11) of the p53 protein is enriched in basic amino acids.
that are important in: (a) the oligomerization and nuclear localization of the p53 protein (reviewed in Refs. 93–96); (b) the recognition of DNA damage (97, 98); (c) the negative regulation of p53 binding to promoter sequences; (d) the transcription of p53-transactivated genes (99); and (e) the induction of apoptosis (100). Laboratory studies have shown that at least two point mutations in the NH2 terminus of p53 are required to inhibit its transcriptional transactivity (92); therefore, deletions and insertions are a more detrimental mutagenic mechanism than single point mutations for disrupting these NH2-terminal and COOH-terminal functional domains.

Fig. 5. Schematic of the p53 molecule. The p53 protein consists of 393 amino acids with functional domains, evolutionarily conserved domains, and regions designated as mutational hotspots. Functional domains include the transactivation region (amino acids 20–42), the sequence-specific DNA-binding region (amino acids 100–293), the nuclear localization sequence (amino acids 316–325), and the oligomerization region (amino acids 319–360). Cellular or oncoviral proteins bind to specific areas of the p53 protein. Evolutionarily conserved domains (amino acids 17–29, 97–292, and 324–352; black areas) were determined using the Multiple Alignment Construction and Analysis Workbench (MACAW) program. Seven mutational hotspot regions within the large conserved domain are identified (amino acids 130–142, 151–164, 171–181, 193–200, 213–223, 234–258, and 270–286; checkered blocks). Vertical lines above the schematic, missense mutations.

Fig. 4. Class of mutations in p53, APC, BRCA1, and ATM genes in all human cancers. Missense mutations represent a high proportion of p53 mutations, whereas nonmissense mutations (e.g., frameshift, nonsense, and splice site mutations) are common in other tumor suppressor genes. Deletions and insertions that do not induce frameshifts have been referred to as in frame deletions.
Structure-Function Relationship of p53

The p53 mutation spectrum can also provide clues to the critical functional regions of the gene that, when mutated, contribute to the carcinogenic process. Because about 75% of the missense mutations are in the sequence-specific DNA-binding midregion of the protein (Refs. 18, 19, and 73; Fig. 5), investigators have focused on the transcription transactivator function of p53. However, these missense mutations and the resultant amino acid substitutions can cause aberrant protein conformations (101) that may alter other functional domains, including those in the COOH terminus of the p53 protein. This positively charged region contains the putative major nuclear localization signal (amino acids 316–325), the oligomerization domain (amino acids 319–360), and a DNA damage binding domain (amino acids 318–393; Refs. 102–105). p53 sequence-specific DNA binding and transcriptional transactivation can also be modulated by posttranslational mechanisms, including serine phosphorylation (103, 106) and the redox regulation of the cysteine residues responsible for binding zinc to p53 (107–109). Recently, it has been demonstrated that acetylation of the p53 COOH-terminal domain at amino acids 373 and 382 stimulates its sequence-specific DNA-binding activity (110). Furthermore, in another recent study, O-glycosylation of the p53 COOH terminus has been shown to activate DNA binding (111). In ML-1 cells exposed to Adriamycin or cisplatin, a COOH-terminal cleavage product of endogenous p53 has been shown to correlate with the up-regulation of p21 (112). The function-structure relationship revealed by the analysis of the p53 mutation spectrum (18, 73), its nuclear magnetic resonance and crystallographic three-dimensional structure (93, 94, 113), and functional studies of wild-type versus mutant p53 activity (reviewed in Ref. 83) have generated both hypotheses for further study and strategies for the development of rational cancer therapy.

Missense mutations in the p53 gene domains encoding loop 2 or 3 have been associated with more aggressive breast cancer (114) that was less responsive to chemotherapy with doxorubicin (115). Unlike mutations affecting interaction at the interface between the p53 protein and its consensus sequences in DNA, loop 2 or 3 mutations are less likely to be temperature sensitive. Additional studies are warranted to determine the importance of somatic p53 mutation in specific structural and functional domains and the success of cancer therapy. The structural and functional consequences of specific germline p53 mutations in Li-Fraumeni syndrome families should also be investigated. Mutation spectra studies of DNA repair genes such as XPD have shown dramatic gene phenotype effects (116). Germline mutations in XPD can exhibit either a phenotype of trichodystrophy without an increased risk of cancer or xeroderma pigmentosum with a >1000 increased risk of sunlight-induced skin cancer. Therefore, similar gene phenotype effects are likely to be found in Li-Fraumeni families.

DNA Methylation and Mutational Hot Spots

Methylated CpG sites of p53 harbor a strikingly high proportion of mutations that constitute major hotspots in human cancer (reviewed in Ref. 73). Deamination of 5-methylcytosine at CpG sites is considered to be one of the major endogenous mechanisms for the induction of these mutations (117). Deamination of 5-methylcytosine forms thymidine and generates a G-T mismatch which, if not repaired, produces a C to T transition. Deamination of cytosine can also generate a C to T transition if uracil glycosylase and a G-T mismatch repair are coupled repair and mismatch repair (133) implies a possible strand bias (123, 126). The positive correlation between cigarette smoking and G:C to T:A transversions on the nontranscribed strand of p53 in lung cancer (127) supports the above-mentioned hypothesis. Laboratory studies have also shown that the mutations induced by exogenous chemical carcinogens occur preferentially on the nontranscribed coding strand of p53 (128–130). C to T transitions at CpG sites are a common mutation produced by endogenous mechanisms such as deamination of 5-methylcytosine (117, 131, 132). It can be hypothesized that C to T transitions arising from the deamination of 5-methylcytosine at CpG dinucleotide sites in the p53 gene would not display a DNA strand bias, whereas C to T transitions arising from bulky chemical DNA adducts would occur primarily on the nontranscribed strand, due to the efficient repair of the transcribed strand. Alternatively, the reported association between transcription-coupled repair and mismatch repair (133) implies a possible strand bias for mismatch repair. Analysis of the p53 mutation database (70) for C to T transitions at CpG sites at mutational hotspot codons on the nontranscribed strand showed either bias for both transcribed and nontranscribed strands or no bias, depending on the specific codon and cancer type (Table 2). Although the total number of mutations is small, mutations at codon 175 (nucleotide residue 13,202) show a substantial bias for the transcribed strand in breast, colon, esophageal, head and neck, and lung cancers. In contrast, mutations at codon 282 (nucleotide residue 14,513) show a strong bias for the nontranscribed strand in brain, breast, colon, esophageal, head and neck, and lung cancers. Mutations at codon 248 (nucleotide residue 14,068) show no strand bias in breast, colon, and head and neck cancers, whereas esophageal and lung cancers show a bias for the nontranscribed strand. Mutations at codon 248 (nucleotide residue 14,068) only show...
Table 2 DNA strand bias at somatic and germline mutation hotspots in the p53 tumor suppressor gene

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>% codon 157</th>
<th>% codon 175</th>
<th>% codon 248</th>
<th>% codon 273</th>
<th>% codon 282</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Somatic mutation: C to T at CpG sites (nontranscribed strand)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>32</td>
<td>28</td>
<td>24</td>
</tr>
<tr>
<td>Breast</td>
<td>NA &lt;5</td>
<td>0</td>
<td>21</td>
<td>42</td>
<td>33</td>
</tr>
<tr>
<td>Colon</td>
<td>NA &lt;5</td>
<td>3</td>
<td>73</td>
<td>45</td>
<td>101</td>
</tr>
<tr>
<td>Esophageal</td>
<td>NA &lt;5</td>
<td>0</td>
<td>27</td>
<td>86</td>
<td>7</td>
</tr>
<tr>
<td>Head and neck</td>
<td>NA &lt;5</td>
<td>0</td>
<td>13</td>
<td>42</td>
<td>26</td>
</tr>
<tr>
<td>Liver</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>38</td>
<td>8</td>
<td>92</td>
</tr>
<tr>
<td>Lung</td>
<td>NA &lt;5</td>
<td>0</td>
<td>8</td>
<td>64</td>
<td>14</td>
</tr>
<tr>
<td>B. G to T at CpG sites (nontranscribed strand)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
</tr>
<tr>
<td>Breast</td>
<td>100 5</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
</tr>
<tr>
<td>Colon</td>
<td>NA &lt;5</td>
<td>80</td>
<td>5</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
</tr>
<tr>
<td>Esophageal</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>100</td>
<td>6</td>
<td>NA &lt;5</td>
</tr>
<tr>
<td>Head and neck</td>
<td>100 7</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>100</td>
<td>7</td>
</tr>
<tr>
<td>Liver</td>
<td>100 11</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>83</td>
<td>5</td>
</tr>
<tr>
<td>Lung</td>
<td>100 23</td>
<td>NA &lt;5</td>
<td>100</td>
<td>12</td>
<td>88</td>
</tr>
<tr>
<td>C. Germline mutation: C to T at CpG sites (nontranscribed strand)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germline</td>
<td>NA &lt;5</td>
<td>NA &lt;5</td>
<td>20</td>
<td>5</td>
<td>16</td>
</tr>
</tbody>
</table>

* The number of mutations less than five (<5) was not assessed (NA; Ref. 50).

** The germline data do not include multiple mutations from the same family.

a bias for the nontranscribed strand in brain and liver cancers. Mutations at codon 273 (nucleotide residue 2,896) show a transcribed strand bias in brain, esophageal, and head and neck cancers but show a nontranscribed strand bias in breast and liver cancers. No strand bias was observed at this codon in colon and lung cancers. One possibility of the occurrence of strand bias for C to T transitions at CpG sites could be due to the different rate or the fidelity of G-T mismatch repair, depending on either the surrounding DNA sequences of the mutational hotspot sites or differences in the etiological agents among cancer types. These findings are consistent with the hypothesis that chemical carcinogens found in the environment, diet, or tobacco smoke are responsible for this subset of p53 mutations.

The frequency of G to T transversions at CpG sites is found to be relatively low and exhibits a strong bias for the nontranscribed strand at a number of mutational hotspots. All G to T transversions at CpG sites at codon 157 (nucleotide residue 13,147) occur on the nontranscribed strand in breast, head, and neck, liver, and lung cancers. Similarly, at codon 248 (nucleotide residue 14,069), a strong bias for the nontranscribed strand exists in esophageal and lung cancers. G to T transversions at codon 175 (nucleotide residue 13,203) in colon cancer and codon 282 (nucleotide residue 14,514) in lung, liver, and head and neck cancers also show a substantial bias for the nontranscribed strand. The presence of strong bias for the nontranscribed strand in G to T transversions at CpG sites of hotspot codons suggests the involvement of exogenous chemical carcinogens that form bulky DNA adducts, rather than the endogenous mechanism such as the deamination of 5-methylcytosine.

Interestingly, analysis of germline C to T transitions at CpG sites of hotspot codons in Li-Fraumeni families shows a considerable bias for the transcribed strand (Table 2). G to T germline mutations at CpG sites are not reported in these families.

Fig. 6. Mutation spectra of p16INK4 in human cancers. Germline and somatic p16INK4 mutations show distinct patterns. Deletions and insertions constitute 33% of the somatic mutations, whereas only 5% of the germline mutations represent this class of mutation.

pl6\textsuperscript{INK4}

pl6\textsuperscript{INK4}, which encodes a cell cycle-inhibitory protein, has been found to be commonly mutated in a variety of human cancers (134–140). Epigenetic control of pl6\textsuperscript{INK4} expression by DNA methylation may also inactivate this cancer susceptibility gene (141, 142). Germ-line mutations in p16\textsuperscript{INK4} have been associated with familial cancers including melanoma, pancreatic cancer, and head and neck cancer (143). Because of the common occurrence of germline mutations in melanoma families, p16\textsuperscript{INK4} has been considered a melanoma susceptibility gene; however, half of the chromosome 9p-linked melanoma families do not show p16\textsuperscript{INK4} germline mutations. In vitro studies...
MOLECULAR EPIDEMIOLOGY OF HUMAN CANCER

Fig. 7. Germline mutational spectra in BRCA1 and BRCA2. Deletions and insertions represent the major types of mutations in both BRCA1 and BRCA2. Among the less common missense mutations, G:C to A:T and A:T to G:C transitions represent 17 and 12%, respectively, of the total mutations in BRCA1. G:C to A:T and A:T to G:C transitions represent only 6 and 5%, respectively, of the total mutations in BRCA2.

BRCA1 and BRCA2

BRCA1 on chromosome 17q12-21 and BRCA2 on chromosome 13q12-13 have been recently identified as cancer susceptibility genes that are involved in the familial predisposition to breast cancer (145–151). In addition to breast cancer, germline mutations in BRCA1 also predispose individuals to ovarian cancer (152–154), whereas BRCA2 mutations are implicated in male breast and pancreatic cancers (155–157). Mutations in these genes are considered to be responsible for about 80% of familial breast cancer cases. In contrast to other cancer susceptibility genes, somatic mutations in BRCA1 or BRCA2 are infrequent in breast and ovarian cancers (158–162). Mutation spectra studies of BRCA1 and BRCA2 have revealed that the majority of the mutations lead to a loss of function and include deletions and insertions leading to frameshift and nonsense mutations. Deletion and insertion mutations constitute about 60% of all mutations in BRCA1 and 80% of all mutations in BRCA2 (Fig. 7). However, a considerable number of missense mutations have been reported previously (163).

The major missense mutations reported in BRCA1 are G:C to A:T transitions (17%, with 12% at non-CpG sites) and A:T to G:C transitions (12%), and in BRCA2, only 6% of the total mutations are G:C to A:T transitions (with 2% at non-CpG sites), and 5% are A:T to G:C transitions. Although there are 100 different mutations reported throughout the BRCA1 gene, two mutations, i.e., 185delAG and 5382 insC, occur with considerably high frequencies and constitute about 20% of the total mutations reported so far. Interestingly, one of these mutations, 185delAG, has an ethnic commonality and segregates with early-onset breast cancers among 20% of Ashkenazi Jews (164). In a

Fig. 8. p53 mutational hotspots in human cancers. Most types of human cancers show the domination of specific p53 mutations at particular mutational hotspots. The characteristic patterns hypothesize molecular linkage between a particular cancer and a specific exogenous or endogenous carcinogen.
MOLECULAR EPIDEMIOLOGY OF HUMAN CANCER

LUNG (64%)
BREAST (15%)
HEAD & NECK (20%)

Fig. 9. p53 mutational hotspots of G to T transversions in human breast, head and neck, and lung cancers. Codon 157 is one of the mutational hotspots in lung cancer. A transversion of G to T at codon 157 in lung cancer is found more frequently among smokers than never-smokers and is a candidate early marker for identifying individuals at higher cancer risk.

recent study involving 5318 Ashkenazi Jews, the 185delAG and 5382insC mutations in BRCA1 and the 6174delT mutation in BRCA2 were analyzed. Among 120 carriers of BRCA1 and BRCA2 mutations, the estimated risk at the age of 70 years was reported to be 56% for breast cancer and 16% for both ovarian cancer and prostate cancer (165). Similar to BRCA1, a large percentage of known BRCA2 germline mutations (33%) is represented by two deletion mutations, 6174delT and 997del5. The 6174delT is found in about 8% of early-onset breast cancer cases in the Ashkenazi Jewish population (166), whereas 997del5 has been commonly reported in early-onset familial breast cancer cases from Iceland (153). Another interesting observation is the apparent occurrence of genotype-phenotype correlation with respect to a specific BRCA1 or BRCA2 mutation and the risk of breast and ovarian cancers. The presence of a truncating mutation in the first two-thirds of the BRCA1 gene instead of in the last third of the gene may significantly increase the risk of ovarian cancer in comparison with that of breast cancer (153). Likewise, the analysis of 25 families with multiple cases of breast and ovarian cancers with BRCA2 mutations also showed a significant genotype-phenotype correlation. Mutations leading to the truncation of BRCA2 in families with the highest risk of ovarian cancer were all found to be clustered in a region of approximately 3.3 kb in exon 11 (154).

Although the normal functions of BRCA1 and BRCA2 are not known, recent studies are providing interesting clues. For example, the transcriptional activation function of BRCA1 and BRCA2 has been reported, and the COOH-terminal region of BRCA1 (amino acids 1528–1863) showed significant transcriptional activation when fused to the GAL4 DNA-binding domain; this transactivation function was lost when a mutation was introduced in the COOH-terminal region (167). BRCA2 exon 3 has been found to have a sequence homology for the activation domain of c-Jun and showed transcriptional activation in yeast when linked to the lex-A DNA-binding domain as well as in two different mammalian cell lines, U2OS and NMuMG, when linked to the GAL4 DNA-binding domain (168). BRCA1 and BRCA2 have been suggested to play a role in embryonic cellular proliferation and development (169, 170). Mutations in BRCA1 or BRCA2 cause lethality at different stages of development in mouse embryos. Interestingly, BRCA2-5 to 6 mutants (deletion of the fifth and sixth exons) with a homozygous null p53 or p21 background showed an enhanced survival of embryos (171), suggesting a functional interaction between BRCA1, p53, and p21. An increased expression of p21 has been observed in the BRCA2-5 to 6 mutants (169). Furthermore, the formation of a complex between BRCA1/BRCA2 and Rad 51 and hypersensitivity of BRCA2 mutant mouse embryos to γ radiation has indicated a role for BRCA1 and BRCA2 genes in DNA repair pathways (170, 172, 173). These functions of BRCA1 and BRCA2, which have been suggested to play a role in maintaining genomic stability, give BRCA1 and BRCA2 an important place in the cancer susceptibility gene family. The finding of inherited mutations in the BRCA1 and BRCA2 genes in families with early-onset breast and ovarian cancers raises the significance of predictive testing or early diagnosis while prompting the debate of several technical and bioethical concerns (reviewed in Ref. 151).

Considering the variety of hereditary cancers and allelic deletions, one can predict the discovery of additional tumor suppressor genes, some of which may have a conspicuous role in carcinogenesis. The frequency of these cancer susceptibility genes and their attributable cancer risk are important considerations in the development of a
By the colony color of the yeast. Measuring p53 mutation load or the mutation frequency of p53 in non-tumorous liver in high HCC incidence geographic areas

Different public health and bioethical considerations apply to the genetic screening of family members of individuals carrying a high cancer risk allele in their germline. Different public health policy for genetic screening of the general population. Molecular Linkage between Carcinogen Exposure and Cancer

A number of specific p53 mutational hotspots have been recognized in different types of human cancer. The occurrence of an identical mutation that is experimentally induced by a carcinogen supports a causative role of a specific environmental carcinogen in certain tumor types. Molecular linkage between exposure to carcinogens and cancer is best exemplified by the p53 mutational spectra of hepatocellular carcinoma, skin cancer, and lung cancer. The most common techniques used for p53 mutation analysis include PCR-based assays like single-strand conformational polymorphism, denaturing gradient gel electrophoresis, and DNA sequencing. Recently, another method based on yeast functional assays was developed to detect p53 mutations. In this assay, loss of DNA binding and transcriptional transactivation function in mutant p53 is detected by the colony color of the yeast. Measuring p53 mutation load or the frequency of mutated alleles in non-tumorous tissue may indicate previous carcinogen exposure and identify individuals at increased cancer risk. However, the detection of rare cells with mutations in a proto-oncogene or tumor suppressor gene in normal-appearing human tissue represents a challenging task. The average spontaneous mutation per base pair in human cells is estimated to be in the range of $10^{-8}$ to $10^{-10}$, and these frequencies increase only 10- to 1000-fold with age. A dose-dependent relationship between diet and cancer is best exemplified by the p53 mutational spectra of non-tumorous liver also is positively correlated with dietary aflatoxin B₁ intake and codon 249p53 mutations is observed in hepatocellular carcinoma from Asia, Africa, and North America. The mutation load of 249p53 mutant cells in nontumorous liver also is positively correlated with dietary aflatoxin B₁ exposure. Exposure of aflatoxin B₁ to human liver cells in vitro produces 249p53 mutations. These results indicate that expression of the 249p53 mutant p53 protein provides a specific growth and/or survival advantage to liver cells and are consistent with the hypothesis that p53 mutations can occur early in liver carcinogenesis.

Sunlight exposure is a well-known risk factor for skin cancer. Tandem CC to TT transition mutations are frequently found in squamous and basal cell skin carcinoma. In vitro studies have shown the induction of the characteristic CC to TT mutations by UV exposure. Sunlight-exposed normal skin and pre-cancerous skin contain CC to TT tandem mutations. These results indicate that solar exposure may play a role in the occurrence of skin cancer.

Cigarette smoking has been established as a major risk factor for the incidence of lung cancer. Codons 157, 248, and 273 of the p53 gene have been designated as mutational hotspots in lung cancer. The majority of mutations found at these codons are G to T transversions. In addition to lung cancer, codon 157 also constitutes one of the hotspots for G to T transversions in breast and head and neck cancers. In smoking-associated lung cancer, the occurrence of G to T transversions has been linked to the presence of BP in cigarette smoke. Interestingly, codon 157 (GTC to TTC) mutations are not found in lung cancer from never-smokers. A dose-dependent increase in p53 G to T transversion mutations with cigarette smoking has been reported in lung cancer. Recently, it has been shown that BPDE, the metabolically activated form of BP, binds to guanosine residues in codons 157, 248, and 273, which are mutational hotspots in lung cancer. Cigarette smoke condensate or BP also neoplastically transforms human bronchial epithelial cells in vitro.

Assessment of Causation by the Bradford-Hill Criteria

Results obtained from molecular epidemiological studies can be used for the assessment of causation. Using the "weight of the evidence" approach, results from various studies can be integrated to provide a comprehensive assessment of the strength of association between a carcinogen and a cancer type.

### Table 3: Assessment of causation by the Bradford-Hill criteria

<table>
<thead>
<tr>
<th>Hypothesis: Dietary AFB₁ exposure can cause 249p53 (AGG → AGT) p53 mutations during human liver carcinogenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strength of association</strong></td>
</tr>
<tr>
<td>• Consistency</td>
</tr>
<tr>
<td>• Positive dose-response correlation between estimated dietary AFB₁ exposure and the frequency of 249p53 mutations in three different ethnic populations on three continents (179, 180, 206)</td>
</tr>
<tr>
<td><strong>Biological plausibility</strong></td>
</tr>
<tr>
<td>• AFB₁ is a potent mutagen and carcinogen in laboratory studies (207–209)</td>
</tr>
<tr>
<td>• AFB₁ is enzymatically activated by human hepatocytes (210, 211), and the 8,9-ABF₁ oxide binds to the third base (G) in codon 249 (212)</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
</tr>
<tr>
<td>• 249p53 mutations are observed in non-tumorous liver in high HCC incidence geographic areas (177)</td>
</tr>
<tr>
<td><strong>249p53 mutations are observed in non-tumorous liver in high HCC incidence geographic areas (177)</strong></td>
</tr>
</tbody>
</table>

### Table 4: Assessment of causation by the Bradford-Hill criteria

<table>
<thead>
<tr>
<th>Hypothesis: The chemical carcinogen BP in tobacco smoke can cause p53 hotspot mutations at codons 157, 248, and 273 in human lung carcinogenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strength of association</strong></td>
</tr>
<tr>
<td>• Consistency</td>
</tr>
<tr>
<td>• Cigarette smoking is associated with a dose-response increase in p53 mutations (G to T transversions) in human lung cancer (127)</td>
</tr>
<tr>
<td><strong>Biological plausibility</strong></td>
</tr>
<tr>
<td>• Tobacco smoke and BP are mutagens (223–225)</td>
</tr>
<tr>
<td>• BP is metabolically activated and forms BPDE DNA adducts in human bronchus in vitro (75-fold interindividual variation; Refs. 226 and 227)</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
</tr>
<tr>
<td>• Codon 157 (GTC → TTC) mutations are uncommon in other types of cancer, including lung cancer in never-smokers (73)</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
</tr>
<tr>
<td>• p53 mutations can be found in bronchial dysplasia (215–222)</td>
</tr>
</tbody>
</table>

MOLECULAR EPIDEMIOLOGY OF HUMAN CANCER

Table 3: Assessment of causation by the Bradford-Hill criteria

| • Biological plausibility |
| • BP diol epoxide binds to Gs in codons 157, 248, and 273, which are p53 mutations |
| • BP exposure to human cells in vitro produces 249p53 mutations |
| • 249p53 expression inhibits apoptosis (213) and p53-mediated transcription (118) and enhances liver cell growth in vitro (214) |

**MOLECULAR EPIDEMIOLOGY OF HUMAN CANCER**

Table 4: Assessment of causation by the Bradford-Hill criteria

| **Strength of association** |
| • Consistency |
| • Cigarette smoking is associated with a dose-response increase in p53 mutations (G to T transversions) in human lung cancer (127) |
| **Biological plausibility** |
| • Tobacco smoke and BP are mutagens (223–225) |
| • BP is metabolically activated and forms BPDE DNA adducts in human bronchus in vitro (75-fold interindividual variation; Refs. 226 and 227) |
| **Specificity** |
| • Codon 157 (GTC → TTC) mutations are uncommon in other types of cancer, including lung cancer in never-smokers (73) |
| **Temporal** |
| • p53 mutations can be found in bronchial dysplasia (215–222) |

**MOLECULAR EPIDEMIOLOGY OF HUMAN CANCER**

Table 4: Assessment of causation by the Bradford-Hill criteria

| **Strength of association** |
| • Consistency |
| • Cigarette smoking is associated with a dose-response increase in p53 mutations (G to T transversions) in human lung cancer (127) |
| **Biological plausibility** |
| • Tobacco smoke and BP are mutagens (223–225) |
| • BP is metabolically activated and forms BPDE DNA adducts in human bronchus in vitro (75-fold interindividual variation; Refs. 226 and 227) |
| **Specificity** |
| • Codon 157 (GTC → TTC) mutations are uncommon in other types of cancer, including lung cancer in never-smokers (73) |
| **Temporal** |
| • p53 mutations can be found in bronchial dysplasia (215–222) |
The difference in mutation spectra and cancer incidence within a different p53 mutational pattern in breast cancer is observed (199). Evidence" principle, Bradford-Hill (196) proposed criteria in the database analysis and Dorothea Dudek for editorial and graphic assistance.

Acknowledgments

We thank Mohammed Khan and Kamran Raja for help with the mutation database analysis and Dorothea Dudek for editorial and graphic assistance.

References


Molecular Epidemiology of Human Cancer: Contribution of Mutation Spectra Studies of Tumor Suppressor Genes

S. Perwez Hussain and Curtis C. Harris


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/58/18/4023.citation

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.