Molecular Epidemiology of Human Cancer: Contribution of Mutation Spectra

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 pathobiological 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 oxiradicals 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 NF1, 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., p16

Germline mutations in genes, e.g., p53, RB, p16

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, NF1 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 tissue 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 neoplastic 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). MIS² arises from either deletions or insertions in short repetitive sequences in the human genome known as microsatellites, and it is used as a measure of genetic instability. The occurrence of MIS was initially reported in colorectal cancer, particularly HNPPC (33–35). The presence of MIS as a biomarker of a mutator phenotype has been linked to the defect in mismatch repair genes, e.g., hMSH2, hMLH1, hPMS1, and hPMS2 (36–39). Tissue of MSH2-null mice shows hypermutability in a transgenic lacI-reporter system on exposure to N-methyl-N-nitrosourea (40). A proportion of HNPPC with MIS does not show a mutation in mismatch repair genes, which indicates the possible involvement of other mutator genes. In a recent study of five Japanese HNPPC families without a germline mutation in MSH2 or MLH1, one family showed MIS and a germline mutation in MSH6/GTF (41). However, tumors in mice with null mutations in MSH6 do not show MIS (42), which indicates possible subtle differences in DNA mismatch repair among animal species.

Candidate mutator genes are involved in multiple cellular functions that are important for maintaining genetic instability, such as DNA repair, DNA replication, chromosomal segregation, cell cycle regulation, and apoptosis. One example is the p53 tumor suppressor gene. Cells lacking expression of wild-type p53 displayed a high frequency of gene amplification, but the expression of wild-type p53 in Li-Fraumeni syndrome cells with mutant p53 reduced the frequency of PALA (a uridine biosynthesis inhibitor)-selected gene amplification (43, 44). In murine cells, p53 mutations lead to genomic instability and the amplification of drug-resistant genes (43). The role of p53 in DNA repair is less certain. A number of studies have shown p53 modulation of nucleotide excision repair in which the loss of p53 functions leads to a decrease in DNA repair and thus an increase in genomic instability (45–51). Suppression of spontaneous homologous recombination is another suggested function of p53 that contributes to the maintenance of genetic stability. Isogenic cell pairs from tumor cells or primary fibroblasts showed a 100-fold increase in the recombination rate when p53 is inactivated (52). Transfection of normal cells with either the dominant-negative p53alavariant expression vector or the human papilloma virus E6 gene showed a 10–80-fold higher spontaneous recombination rate (53). p53 is also involved in the fidelity of centrosomal duplication, which ensures balanced segregation of the chromosomes during cell division. Mouse embryonic fibroblasts lacking p53 protein produce multiple copies of centrosomes (54), resulting in unequal segregation of the chromosomes and genetic instability. Mutant p53 may increase the sensitivity of cells to mutagens. A 5-fold increase in the mutation frequency, as determined by the appearance of hypoxanthine guanine phosphoribosyl transferase-deficient (6-thioguanine-resistant) cells (induced by X-rays), was observed in human fibroblast SUSM-L/p53 cells containing a mutant p53 gene (codon 273 Arg-His) when compared with the parental SUSM-I cells (55). Human lymphoblastoid cell lines TK6 and WTK1 containing wild-type and mutant p53, respectively, showed different levels of (X-ray-induced) mutagenicity and apoptosis (56). A 28-fold increase in the spontaneous mutation frequency at the heterozygous thymidine locus was observed in WTK1 cells (p53-mutated) when compared with TK6 (normal) cells (57). TK6 cells were also more sensitive at the tk locus to the mutagenetic effects of X-rays, ethylenemethane sulfonate, methylmethane sulfonate, and mitomycin C. Mutant p53 induced an increase in the frequency of chromosomal breaks in dividing LIM12 cells (58). A study of the replication of the pZ402 shuttle vector containing the supF gene in Li-Fraumeni fibroblasts with mutated p53 allele 0 showed a frequent deletion of a portion of SV40 T antigen (58). A 20-fold increase in silent mutations has been observed in the mutated p53 gene in human cancer, which indicates its hypermutability (59). These silent mutations were not found to be selected or represented as genetic polymorphisms in the normal population. However, the fact that the alteration in p53 has been described as a relatively late event in colorectal carcinogenesis argues against its candidacy as a mutator.

---

² The abbreviations used are: MIS, microsatellite instability; HNPPC, hereditary nonpolyposis colorectal cancer; BP, benzo(a)pyrene; BPDE, BP-7,8-diol 9,10-epoxide.
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 TGF$\beta$-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 tumor cells 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 deletions and insertions are most commonly found in exons 2–4 and 9–11 (74), so evaluation of the entire DNA sequence yields relevant data. This situation differs from that of the ras oncoproteins, whose transforming mutations occur primarily in three codons, a few sequence-specific motifs, and a critical functional domain (75). The diversity of p53 mutational events permits more extensive inferences of the mechanism of DNA damage involved.

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 deamination 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 mononucleotide 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.

---

**Table 1: Examples of tumor suppressor genes involved in human cancers**

<table>
<thead>
<tr>
<th>Tumor suppressor gene</th>
<th>Chromosomal locus</th>
<th>Location/proposed function</th>
<th>Major types of somatic mutation</th>
<th>Examples of neoplasms with somatic mutations</th>
<th>Syndrome</th>
<th>Heterozygote carrier rate/10^5 births^a</th>
<th>Typical neoplasms</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53</td>
<td>17p13.1</td>
<td>Nuclear/transcription factor</td>
<td>Missense</td>
<td>Most human cancer types examined to date</td>
<td>Li-Fraumeni</td>
<td>~2</td>
<td>Carcinomas of the breast and adrenal cortex, sarcomas, leukemia, and brain tumors</td>
</tr>
<tr>
<td>RB1</td>
<td>13q14</td>
<td>Nuclear/transcription modifier</td>
<td>Deletion and nonsense</td>
<td>Retinoblastoma, osteosarcoma, and carcinomas of the breast, prostate, bladder, and lung</td>
<td>Retinoblastoma</td>
<td>~2</td>
<td>Retinoblastoma, and osteosarcoma</td>
</tr>
<tr>
<td>APC</td>
<td>5q21</td>
<td>Cytoplasmic/signal transduction</td>
<td>Deletion and nonsense</td>
<td>Carcinoma of the colon, stomach, and pancreas</td>
<td>Familial adenomatous polyposis</td>
<td>~10</td>
<td>Carcinomas of the colon, thyroid, and stomach</td>
</tr>
<tr>
<td>ATM</td>
<td>11q22</td>
<td>Nuclear/kinase Ser/Thr</td>
<td>Deletion and nonsense</td>
<td>Leukemia</td>
<td>Ataxia telangiectasia</td>
<td>~2</td>
<td>Leukemia and lymphoma</td>
</tr>
<tr>
<td>WT1</td>
<td>11p13</td>
<td>Nuclear/transcription factor</td>
<td>Missense</td>
<td>Wilms' tumor</td>
<td>Wilms' tumor</td>
<td>~0.5–1</td>
<td>Wilms' tumor</td>
</tr>
<tr>
<td>BRCA1</td>
<td>17q21</td>
<td>Nuclear/DNA repair</td>
<td>Deletion and nonsense</td>
<td>Breast</td>
<td>Familial breast/ovarian</td>
<td>~2–10</td>
<td>Carcinoma of the breast and ovary</td>
</tr>
<tr>
<td>BRCA2</td>
<td>13q12–13</td>
<td>Nuclear/DNA repair</td>
<td>Deletion and nonsense</td>
<td>Breast</td>
<td>Familial breast/ovarian</td>
<td>~2–10</td>
<td>Carcinoma of the breast and ovary</td>
</tr>
<tr>
<td>NF1</td>
<td>17q11</td>
<td>Cytoplasmic/GTPase activating protein</td>
<td>Deletion and nonsense</td>
<td>Schwanommas</td>
<td>Neurofibromatosis type 1</td>
<td>~30</td>
<td>Neural tumors</td>
</tr>
<tr>
<td>NF2</td>
<td>22q</td>
<td>Cytoplasmic/cytoskeletal-membrane linkage</td>
<td>Deletion and nonsense</td>
<td>Schwanommas and meningiomas</td>
<td>Neurofibromatosis type 2</td>
<td>~3</td>
<td>Central schwanommas and meningiomas</td>
</tr>
<tr>
<td>p16INK4</td>
<td>9p21</td>
<td>Nuclear/cyclin-dependent kinase inhibitor</td>
<td>Deletion and nonsense</td>
<td>Mesothelioma, pancreas melanoma, and glioblastoma</td>
<td>Familial melanoma</td>
<td>?</td>
<td>Melanoma</td>
</tr>
<tr>
<td>VHL</td>
<td>3p25</td>
<td>Nuclear/adaptor</td>
<td>Deletion</td>
<td>Unknown</td>
<td>von Hippel-Lindau</td>
<td>~3</td>
<td>Hemangioblastoma and renal cell carcinoma</td>
</tr>
</tbody>
</table>

^a Reviewed in Ref. 201 (see also Refs. 202–204).  
^b Ref. 205 and A. Knudson, personal communication.
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.
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 mutants 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). Germine 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 inefficient. The presence of a high frequency of C to T transitions at CpG dinucleotides in colon carcinomas (18) suggests the involvement of an endogenous deamination mechanism. However, a considerable variation also has been reported in mutational type at CpG sites among different types of cancer. Compared to colon carcinoma, low frequencies of transition mutations are observed in liver and lung cancer at CpG dinucleotide sites, and, in contrast, high frequencies of G to T transitions are reported in these cancer sites (73). A distinct pattern of mutation at CpG sites in different tissue types indicates either the possible involvement of exogenous agents or the selection of p53 mutants that vary among tissues in their pathobiological activities (118).

Recently, a role for cytosine methylation at CpG sites has been suggested in determining the preferential site of DNA damage by exogenous carcinogen. Interestingly, treatment of HeLa cells and bronchial epithelial cells with BPDE selectively induced guanine adduct formation at CpG sites in p53 codons 157, 248, and 273 (119). In contrast to nonmethylated cloned p53 gene sequences containing unmethylated cytosines, treatment of methylated sequences with BPDE preferentially produced BPDE-guanine adducts at CpG sites (120). In the same study, a similar trend also is observed in the PGK1 gene, with different methylation status, derived from an active or inactive X chromosome. BPDE, aflatoxin B1, 8,9-epoxide, N-acetoxy-2-acetylaminofluorene, and benz(a)chrysene diol epoxide, which interact with different moieties within guanine, showed more binding at methylated CpG sites than at unmethylated CpG sites in the hotspot codons of p53 (121). The precise role by which methylation results in the preferential interaction of BPDE is not clear; however, its influence in the creation of an intercalation site has been suggested (120). Rates of DNA repair may vary among p53 exons (122).

DNA Strand Bias of p53 Mutation

The transcribed strand of an active gene is usually repaired at a more rapid rate than the nontranscribed (DNA coding) strand (123). For example, DNA damage on the transcribed strand of the p53 gene is repaired faster compared with that of the nontranscribing or coding strand (124, 125). This preferential DNA repair is suggested as an important factor in determining the mutation pattern leading to 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:G to T:A 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:G to T:A 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>&lt;5</td>
<td>NA</td>
<td>&lt;5</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>Breast</td>
<td>&lt;5</td>
<td>0</td>
<td>21</td>
<td>42</td>
<td>33</td>
</tr>
<tr>
<td>Colon</td>
<td>&lt;5</td>
<td>3</td>
<td>73</td>
<td>45</td>
<td>101</td>
</tr>
<tr>
<td>Esophageal</td>
<td>&lt;5</td>
<td>0</td>
<td>27</td>
<td>86</td>
<td>7</td>
</tr>
<tr>
<td>Head and neck</td>
<td>&lt;5</td>
<td>0</td>
<td>13</td>
<td>42</td>
<td>26</td>
</tr>
<tr>
<td>Liver</td>
<td>&lt;5</td>
<td>NA</td>
<td>&lt;5</td>
<td>38</td>
<td>8</td>
</tr>
<tr>
<td>Lung</td>
<td>&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</td>
<td>&lt;5</td>
<td>NA</td>
<td>&lt;5</td>
<td>NA</td>
</tr>
<tr>
<td>Breast</td>
<td>100</td>
<td>5</td>
<td>NA</td>
<td>&lt;5</td>
<td>NA</td>
</tr>
<tr>
<td>Colon</td>
<td>NA</td>
<td>80</td>
<td>5</td>
<td>NA</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Esophageal</td>
<td>NA</td>
<td>&lt;5</td>
<td>NA</td>
<td>&lt;5</td>
<td>NA</td>
</tr>
<tr>
<td>Head and neck</td>
<td>100</td>
<td>7</td>
<td>NA</td>
<td>&lt;5</td>
<td>NA</td>
</tr>
<tr>
<td>Liver</td>
<td>100</td>
<td>11</td>
<td>NA</td>
<td>&lt;5</td>
<td>NA</td>
</tr>
<tr>
<td>Lung</td>
<td>100</td>
<td>23</td>
<td>NA</td>
<td>&lt;5</td>
<td>100</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</td>
<td>&lt;5</td>
<td>NA</td>
<td>&lt;5</td>
<td>20</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.

There is 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.

**p16**

**p16** is a cell cycle-inhibitory protein, which has been found to be commonly mutated in a variety of human cancers (134–140). Epigenetic control of p16 expression by DNA methylation may also inactivate this cancer susceptibility gene (141, 142). Germ-line mutations in p16 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 has been considered a melanoma susceptibility gene; however, half of the chromosome 9p-linked melanoma families do not show p16 germline mutations. **In vitro** studies...
have shown that germline melanoma-predisposing mutants are unable to inhibit the catalytic activity of cyclin D1/CDK4 and cyclin D1/CDK6 complexes (144). The absence of \( p16^{INK4} \) alteration in a large portion of the melanoma-prone families linked to chromosome 9p21 supports the possible existence of additional melanoma susceptibility genes. There are some interesting distinctions between the germline and somatic \( p16^{INK4} \) mutational spectra (Fig. 6). Deletions and insertions, which constitute 33% of the somatic \( p16^{INK4} \) mutation, are represented by only 5% of germline mutations. CC to TT tandem double mutations, which are characteristically induced by UV radiation, constitute 4% of somatic \( p16^{INK4} \) mutations but are never found as a germline mutation. Although several germline polymorphisms also have been reported for \( p16^{INK4} \), their functional and prognostic significance is not yet known.

**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 5382insC, 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
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.

A recent study involving 5318 Ashkenazi Jews, the 185delAG and 5382insC mutations in BRCA1 and the 6174delI 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, 6174delI and 997del5. The 6174delI 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 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 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
MOLECULAR EPIDEMIOLOGY OF HUMAN CANCER

Table 3 Assessment of causation by the Bradford-Hill criteria

| • Strength of association |
| • Consistency |
| Positive dose-response correlation between estimated dietary aflatoxin B<sub>1</sub> exposure and the frequency of 249<sup>AGG</sup> &gt; AGT p53 mutations in three different ethnic populations on three continents (179, 180, 206) |
| • Specificity |
| 249<sup>AGG</sup> p53 mutant cells are observed in nontumorous liver in high HCC incidence geographic areas (177) |
| • Biological plausibility |
| • Aflatoxin B<sub>1</sub> is a potent mutagen and carcinogen in laboratory studies (207–209) |
| • Aflatoxin B<sub>1</sub> is enzymatically activated by human hepatocytes (210, 211), and the 8,9-AFB<sub>1</sub> oxide binds to the third base (G) in codon 249 (212) |
| • AFB<sub>1</sub> exposure to human liver cells (178, 184) in vitro produces codon 249<sup>AGG</sup> p53 mutations |
| • 249<sup>AGG</sup> p53 expression inhibits apoptosis (213) and p53-mediated transcription (118) and enhances liver cell growth in vitro (214) |

* AFB<sub>1</sub>, aflatoxin B<sub>1</sub>.
* HCC, hepatocellular carcinoma.

Molecular Linkage between Carcinogen Exposure and Cancer

A number of specific p53 mutational hotspots have been recognized in different types of human cancer (Fig. 8). 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 (174–176). 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 nontumorous 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<sup>−8</sup> to 10<sup>−10</sup>, and these frequencies increase only 10- to 1000-fold on exposure to a mutagen. Therefore, methods that allow the detection of few altered DNA sequences from 10<sup>3</sup> to 10<sup>10</sup> copies of the corresponding wild-type sequences in the presence of large quantities of cellular DNA are required. The development of a highly sensitive genotypic assay is allowing the detection of low-frequency mutations in normal-appearing human tissues (177, 178). The detection of a particular mutation in normal-appearing tissue provides further support for the involvement of a specific carcinogen in a particular human cancer and may help identify individuals at increased cancer risk.

In liver tumors from persons living in geographic areas in which aflatoxin B<sub>1</sub> and hepatitis B virus are cancer risk factors, most p53 mutations are at the third nucleotide pair of codon 249 (179–182). A dose-dependent relationship between dietary aflatoxin B<sub>1</sub> intake and codon 249<sup>AGG</sup> p53 mutations is observed in hepatocellular carcinoma from Asia, Africa, and North America (reviewed in Refs. 8 and 183). The mutation load of 249<sup>AGG</sup> mutant cells in nontumorous liver also is positively correlated with dietary aflatoxin B<sub>1</sub> exposure (177). Exposure of aflatoxin B<sub>1</sub> to human liver cells in vitro produces 249<sup>AGG</sup> p53 mutations (178, 184). These results indicate that expression of the 249<sup>AGG</sup> p53 protein provides a specific growth and/or survival advantage to liver cells (185) 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 (186), whereas they are rarely reported in other types of cancers (73). In vitro studies have shown the induction of the characteristic CC to TT mutations by UV exposure (187–191). Sunlight-exposed normal skin and precancerous skin contain CC to TT tandem mutations (192, 193). These results indicate that CC to TT mutations induced by sunlight 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. Furthermore, 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 (Fig. 9). 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 (73, 194). A dose-dependent increase in p53 G to T transversion mutations with cigarette smoking has been reported in lung cancer (127). 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 (119). Cigarette smoke condensate or BP can neoplastically transform human bronchial epithelial cells in vitro (195).

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
The difference in mutation spectra and cancer incidence within a population or race could be due to either the effect of a regional environmental agent or the polymorphic variation in genes responsible for carcinogen metabolism and DNA repair. Comprehensive databases of tumor suppressor genes with mutations can be used as an important tool in molecular epidemiology. Examples of databases available on the Internet include the following: (a) dietary aflatoxin B₁ exposure that can cause codon 249ser (AGG to AGT) p53 mutations during human liver carcinogenesis (Table 3); (b) the chemical carcinogen BP in tobacco smoke that can cause p53 hotspot mutations in human lung carcinogenesis (Table 4); and (c) sunlight exposure that can cause a characteristic CC to TT double mutation in p53 in human skin cancer (Table 5). Furthermore, the influence of geographic location, race, and gender needs to be taken into account when determining an individual's susceptibility or risk for cancer (13). The p53 mutation spectra in breast cancer from Japanese women show a different pattern when compared with that of European and American women (197, 198). Even among African-American and Caucasian-American women, a different p53 mutational pattern in breast cancer is observed (199). The difference in mutation spectra and cancer incidence within a population or race could be due to either the effect of a regional environmental agent or the polymorphic variation in genes responsible for carcinogen metabolism and DNA repair.

Comprehensive databases of tumor suppressor genes with mutations can be used as an important tool in molecular epidemiology. Examples of databases available on the Internet include the following: (a) for p53, http://www.iarc.fr/p53/home page.html; (b) for germline mutations in Li-Fraumeni syndrome, http://cuni.cz/win/projects/read me.htm (71, 200); (c) for BRCA1 and BRCA2, http://www.nhgri.nih.gov/Intramural research/ Lab transfer/Bic/; and (d) for APC, http://perso.curie.fr/Thierry.Soussi/p53database.html. We have previously discussed the limitations and advantages of the methodologies to detect mutations and the need for improved epidemiological design and quality control (73). Further improvement of these databases by including information on geographical location, race, ethnicity, mutation exposure, clinical course, genetic polymorphism, and so forth will enhance the analyses of the mutation spectra of these genes.

Acknowledgments

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

References


155. Klein-Szanto, A. J., Inzunza, J., Garcia-Palazco, I., Caamano, J., Metcalfe, R., Welsh, J., and Harris, C. C. A tobacco-specific N-nitrosamine or cigarette smoke


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.